UNIT – I

BUSINESS INTELLIGENCE

PART- A

1. Define Business Intelligence?(APR/MAY2017)

Business intelligence may be defined as a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for complex decision-making processes.

2. Define Churn?

The marketing manager of a mobile phone company realizes that a large number of customers are discontinuing their service, leaving her company in favor of some competing provider. As can be imagined, low customer loyalty, also known as customer attrition or churn , is a critical factor for many companies operating in service industries.

3 What is the main purpose of business Intelligence System?

The main purpose of business intelligence systems is to provide knowledge workers with tools and methodologies that allow them to make *effective* and *timely* decisions.

4 **Define Data?**

Data: Generally, data represent a structured codification of single primary entities, as well as of transactions involving two or more primary entities.

5 Define Knowledge?

Knowledge. Information is transformed into knowledge when it is used to make decisions and develop the corresponding actions. Therefore, we can think of knowledge as consisting of information put to work into a specific domain, enhanced by the experience and competence of decision makers in tackling and solving complex problems.

6 **Define Information?**

Information. Information is the outcome of extraction and processing Activities carried out on data, and it appears meaningful for those who receive it in a specific domain.

7 Define Knowledge Management?

The activity of providing support to knowledge workers through the integration of decision-making processes and enabling information technologies is usually referred to as *knowledge management*.

8 **Define Optimization.**

The optimization models that allow us to determine the best solution out of a set of alternative actions, which is usually fairly extensive and sometimes even infinite.

9 **Define Design**

The second phase includes two sub-phases and is aimed at deriving a provisional plan of the overall architecture, taking into account any development in the near future and the evolution of the system in the midterm.

10. What is Planning Stage?

The planning stage includes a sub-phase where the functions of the business intelligence system are defined and described in greater detail. Sub-sequent, existing data as well as other data that might be retrieved externally are assessed.

11. What do your meant by Data exploration?

At the third level of the pyramid we find the tools for per- forming a *passive* business intelligence analysis, which consist of query and reporting systems, as well as statistical methods. These are referred to as passive methodologies because decision makers are requested to generate prior hypotheses or define data extraction criteria, and then use the analysis tools to find answers and confirm their original insight.

13.What is Effective decisions?

The application of rigorous analytical methods allows decision makers to rely on information and knowledge which are more dependable. As a result, they are able to make better decisions and devise action plans that allow their objectives to be reached in a more effective way. Indeed, turning to formal analytical methods forces decision makers to explicitly describe both the criteria for evaluating alternative choices and the mechanisms regulating the problem under investigation. Furthermore, the ensuing in-depth examination and thought lead to a deeper awareness and comprehension of the underlying logic of the decision-making process.

14. Define Timely decisions.

Enterprises operate in economic environments characterized by growing levels of competition and high dynamism. As a consequence, the ability to rapidly react to the actions of competitors and to new market conditions is a critical factor in the success or even the survival of a company.

PART-B

1. Explain the role of mathematical m odels?

A business intelligence system provides decision makers with information and knowledge extracted from data, through the application of mathematical models and algorithms. In some instances, this activity may reduce to calculations of totals and percentages, graphically represented by simple histograms, whereas more elaborate analyses require the development of advanced optimization and learning models.

In general terms, the adoption of a business intelligence system tends to promote a scientific and rational approach to the management of enterprises and complex organizations. Even the use of a spreadsheet to estimate the effects on the budget of fluctuations in interest rates, despite its simplicity, forces decision makers to generate a mental representation of the financial flows process.

Classical scientific disciplines, such as physics, have always resorted to

mathematical models for the abstract representation of real systems. Other disciplines, such as operations research, have instead exploited the application of scientific methods and mathematical models to the study of artificial systems, for example public and private organizations. Part II of this book will describe the main mathematical models used in business intelligence architectures and decision support systems, as well as the corresponding solution methods, while Part III will illustrate several related applications.

The rational approach typical of a business intelligence analysis can be summarized schematically in the following main characteristics.

- First, the objectives of the analysis are identified and the performance indicators that will be used to evaluate alternative options are defined.
- Mathematical models are then developed by exploiting the relationships among system control variables, parameters and evaluation metrics.
- Finally, *what-if* analyses are carried out to evaluate the effects on the performance determined by variations in the control variables and changes in the parameters.

Although their primary objective is to enhance the effectiveness of the decision-making process, the adoption of mathematical models also affords other advantages, which can be appreciated particularly in the long term. First, the development of an abstract model forces decision makers to focus on the main features of the analyzed domain, thus inducing a deeper understanding of the phenomenon under investigation. Furthermore, the knowledge about the domain acquired when building a mathematical model can be more easily transferred in the long run to other individuals within the same organization, thus allowing a sharper preservation of knowledge in comparison to empirical decision-making processes. Finally, a mathematical model developed for a

specific decision-making task is so general and flexible that in most cases it can be applied to other ensuing situations to solve problems of similar type.

2.Explain in detail about Business intelligence architectures? (APR/MAY2017, 2019)

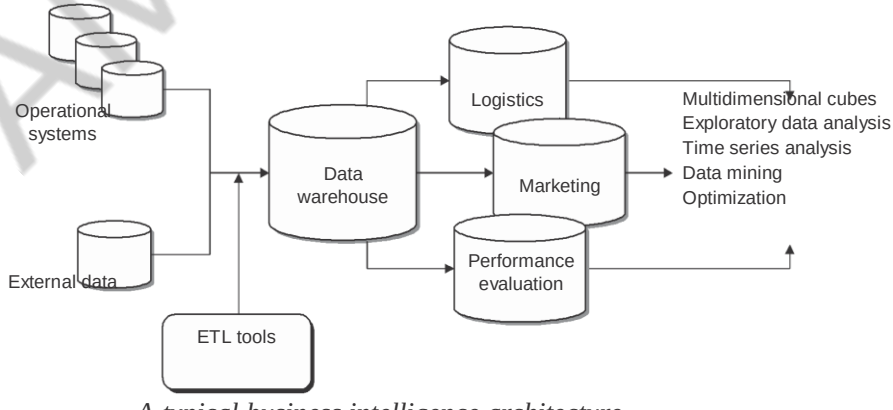
The architecture of a business intelligence system, depicted in includes three major components.

Data sources. In a first stage, it is necessary to gather and integrate the data stored in the various primary and secondary sources, which are heterogeneous in origin and type. The sources consist for the most part of data belonging to operational systems, but may also include unstructured documents, such as emails and data received from external providers. Generally speaking, a major effort is required to unify and integrate the different data sources, as shown in Chapter 3.

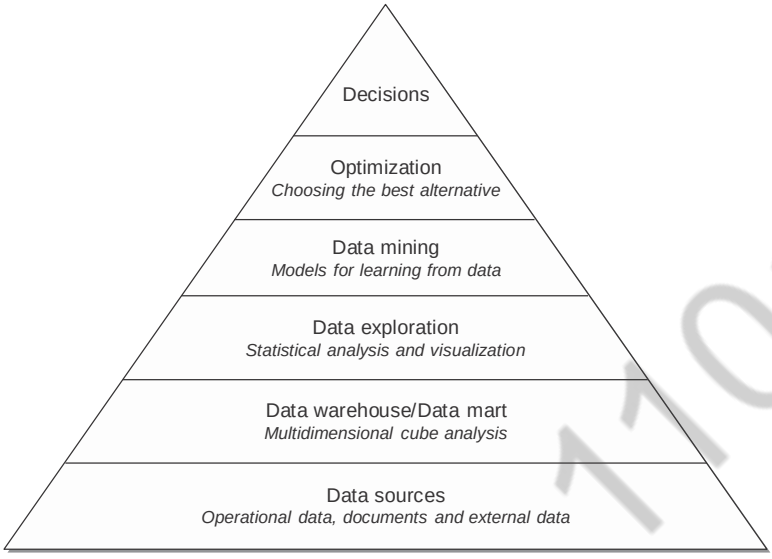
Data warehouses and data marts. Using extraction and transformation tools known as *extract, transform, load* (ETL), the data originating from the different sources are stored in databases intended to support business intelligence analyses. These databases are usually referred to as *data warehouses* and *data marts*, and they will be the subject of Chapter 3.

Business intelligence methodologies. Data are finally extracted and used to feed mathematical models and analysis methodologies intended to support decision makers. In a business intelligence system, several decision support applications may be implemented, most of which will be described in the following chapters:

- multidimensional cube analysis;
- exploratory data analysis;



A typical business intelligence architecture



The main components of a business intelligence system

- time series analysis;
- inductive learning models for data mining;
- optimization models.

The pyramid in Figure 1.3 shows the building blocks of a business intelligence system. So far, we have seen the components of the first two levels We now turn to the description of the upper tiers.

Data exploration. At the third level of the pyramid we find the tools for performing a *passive* business intelligence analysis, which consist of query and reporting systems, as well as statistical methods. These are referred to as passive methodologies because decision makers are requested to generate prior hypotheses or define data extraction criteria, and then use the analysis tools to find answers and confirm their original insight. For instance, consider the sales manager of a company who notices that revenues in a given geographic area have dropped for a specific group of customers. Hence, she might want to bear out her hypothesis by using extraction and visualization tools, and then apply a statistical test to verify that her conclusions are adequately supported by data.

Data mining. The fourth level includes *active* business intelligence methodologies, whose purpose is the extraction of information and knowledge from data.

These include mathematical models for pattern recognition, machine learning and data mining techniques Unlike the tools described at the previous level of the pyramid, the models of an active kind do not require decision makers to formulate any prior hypothesis to be later verified. Their purpose is instead to expand the decision makers' knowledge.

Optimization. By moving up one level in the pyramid we find optimization models that allow us to determine the best solution out of a set of alternative actions, which is usually fairly extensive and sometimes even infinite.

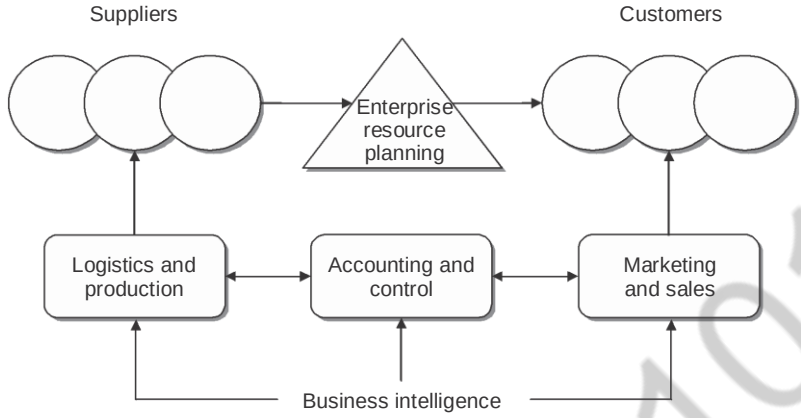
Decisions. Finally, the top of the pyramid corresponds to the choice and the actual adoption of a specific decision, and in some way represents the natural conclusion of the decision-making process. Even when business intelligence methodologies are available and successfully adopted, the choice of a decision pertains to the decision makers, who may also take advantage of informal and unstructured information available to adapt and modify the recommendations and the conclusions achieved through the use of mathematical models.

As we progress from the bottom to the top of the pyramid, business intelligence systems offer increasingly more advanced support tools of an active type. Even roles and competencies change. At the bottom, the required competencies are provided for the most part by the information systems specialists within the organization, usually referred to as *database administrators*. Analysts and experts in mathematical and statistical models are responsible for the intermediate phases. Finally, the activities of decision makers responsible for the application domain appear dominant at the top.

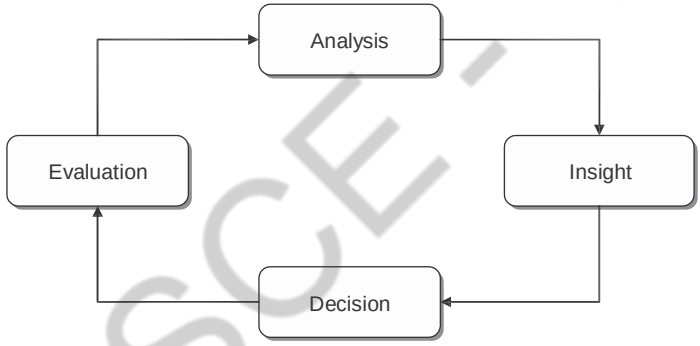
As described above, business intelligence systems address the needs of different types of complex organizations, including agencies of public administration and associations. However, if we restrict our attention to enterprises, business intelligence methodologies can be found mainly within three departments of a company.

3. What is the Cycle of a business intelligence analysis?

Each business intelligence analysis follows its own path according to the application domain, the personal attitude of the decision makers and the available analytical methodologies. However, it is possible to identify an ideal cyclical



Departments of an enterprise concerned with business intelligence systems



Cycle of a business intelligence analysis

path characterizing the evolution of a typical business intelligence analysis, as shown in Figure 1.5, even though differences still exist based upon the peculiarity of each specific context.

Analysis. During the analysis phase, it is necessary to recognize and accurately spell out the problem at hand. Decision makers must then create a mental representation of the phenomenon being analyzed, by identifying the critical factors that are perceived as the most relevant. The availability of business intelligence methodologies may help already in this stage, by permitting decision makers to rapidly develop various paths of investigation. For instance, the exploration of data cubes in a multidimensional analysis, according to different logical views as described in Chapter 3, allows decision makers to modify their

hypotheses flexibly and rapidly, until they reach an interpretation scheme that they deem satisfactory. Thus, the first phase in the business intelligence cycle leads decision makers to ask several questions and to obtain quick responses in an interactive way.

Insight. The second phase allows decision makers to better and more deeply understand the problem at hand, often at a causal level. For instance, if the analysis carried out in the first phase shows that a large number of customers are discontinuing an insurance policy upon yearly expiration, in the second phase it will be necessary to identify the profile and characteristics shared by such customers. The information obtained through the analysis phase is then transformed into knowledge during the insight phase. On the one hand, the extraction of knowledge may occur due to the intuition of the decision makers and therefore be based on their experience and possibly on unstructured information available to them. On the other hand, inductive learning models may also prove very useful during this stage of analysis, particularly when applied to structured data.

Decision. During the third phase, knowledge obtained as a result of the insight phase is converted into decisions and subsequently into actions. The availability of business intelligence methodologies allows the analysis and insight phases to be executed more rapidly so that more effective and timely decisions can be made that better suit the strategic priorities of a given organization. This leads to an overall reduction in the execution time of the *analysis–decision–action–revision* cycle, and thus to a decision-making process of better quality.

Evaluation. Finally, the fourth phase of the business intelligence cycle involves performance measurement and evaluation. Extensive metrics should then be devised that are not exclusively limited to the financial aspects but also take into account the major performance indicators defined for the different company departments. Chapter 15 will describe powerful analytical methodologies for performance evaluation.

4. Explain the Enabling factors in business intelligence projects

Some factors are more critical than others to the success of a business intelligence project: *technologies*, *analytics* and *human resources*.

Technologies. Hardware and software technologies are significant enabling factors that have facilitated the development of business intelligence systems within enterprises and complex organizations. On the one hand, the computing capabilities of microprocessors have increased on average by 100% every 18 months during the last two decades, and prices have fallen. This trend has

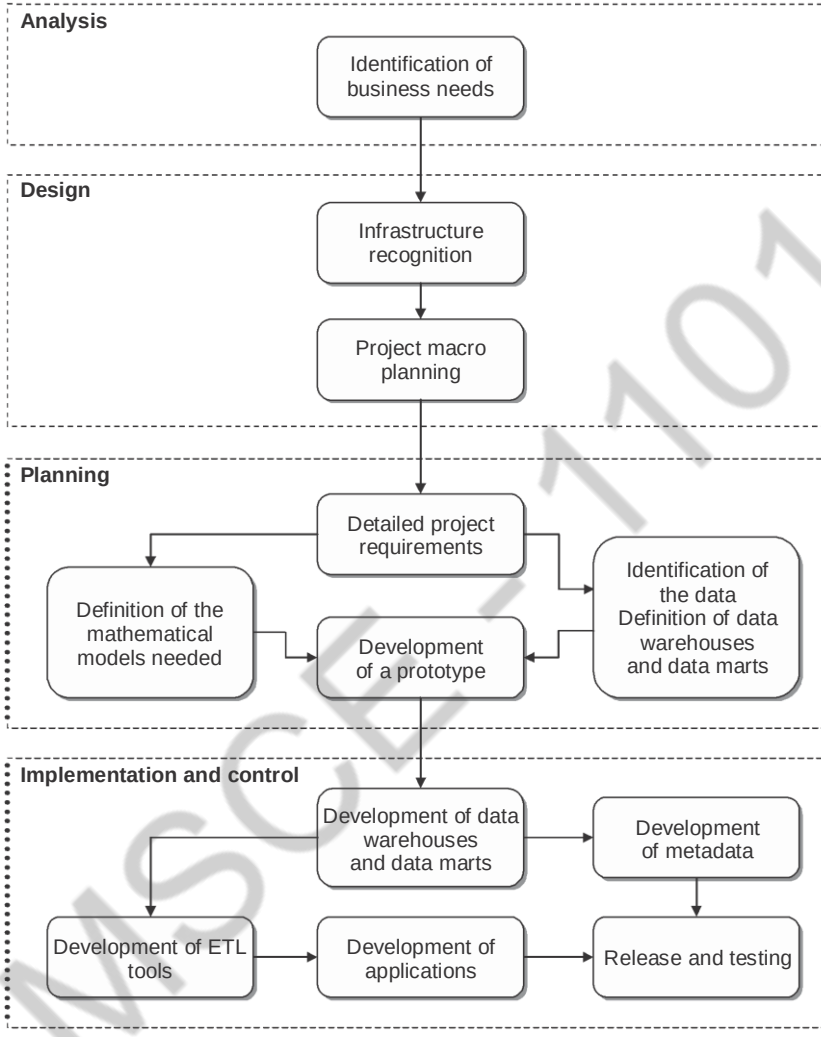
enabled the use of advanced algorithms which are required to employ inductive learning methods and optimization models, keeping the processing times within a reasonable range. Moreover, it permits the adoption of state-of-the-art graphical visualization techniques, featuring real-time animations. A further relevant enabling factor derives from the exponential increase in the capacity of mass storage devices, again at decreasing costs, enabling any organization to store terabytes of data for business intelligence systems. And network connectivity, in the form of *Extranets* or *Intranets*, has played a primary role in the diffusion within organizations of information and knowledge extracted from business intelligence systems. Finally, the easy integration of hardware and software purchased by different suppliers, or developed internally by an organization, is a further relevant factor affecting the diffusion of data analysis tools.

Analytics. As stated above, mathematical models and analytical methodologies play a key role in information enhancement and knowledge extraction from the data available inside most organizations. The mere visualization of the data according to timely and flexible logical views, as described in Chapter 3, plays a relevant role in facilitating the decision-making process, but still represents a passive form of support. Therefore, it is necessary to apply more advanced models of inductive learning and optimization in order to achieve active forms of support for the decision-making process.

Human resources. The human assets of an organization are built up by the competencies of those who operate within its boundaries, whether as individuals or collectively. The overall knowledge possessed and shared by these individuals constitutes the *organizational culture*. The ability of knowledge workers to acquire information and then translate it into practical actions is one of the major assets of any organization, and has a major impact on the quality of the decision-making process. If a given enterprise has implemented an advanced business intelligence system, there still remains much scope to emphasize the personal skills of its knowledge workers, who are required to perform the analyses and to interpret the results, to work out creative solutions and to devise effective action plans. All the available analytical tools being equal, a company employing human resources endowed with a greater mental agility and willing to accept changes in the decision-making style will be at an advantage over its competitors.

5. Explain in detail about the Development of a business intelligence system.(APR/MAY 2017)

The development of a business intelligence system can be assimilated to a project, with a specific final objective, expected development times and costs, and the usage and coordination of the resources needed to perform planned



Phases in the development of a business intelligence system

activities. Obviously, the specific path followed by each organization might differ from that outlined in the figure. For instance, if the basic information structures, including the data warehouse and the data marts, are already in place, the corresponding phases indicated.

Analysis. During the first phase, the needs of the organization relative to the development of a business intelligence system should be carefully identified. This preliminary phase is generally conducted through a series of interviews of

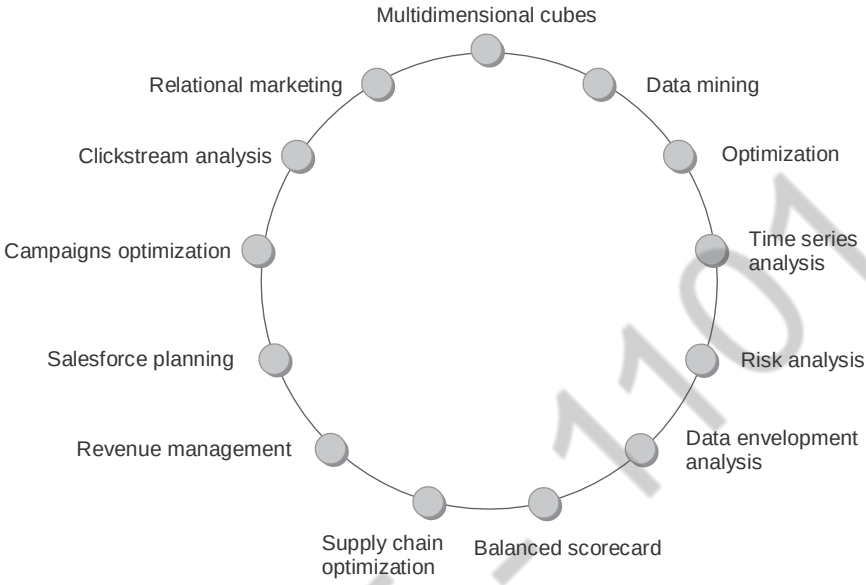
knowledge workers performing different roles and activities within the organization. It is necessary to clearly describe the general objectives and priorities of the project, as well as to set out the costs and benefits deriving from the development of the business intelligence system.

Design. The second phase includes two sub-phases and is aimed at deriving a provisional plan of the overall architecture, taking into account any development in the near future and the evolution of the system in the mid term. First, it is necessary to make an assessment of the existing information infrastructures. Moreover, the main decision-making processes that are to be supported by the business intelligence system should be examined, in order to adequately determine the information requirements. Later on, using classical project management methodologies, the project plan will be laid down, identifying development phases, priorities, expected execution times and costs, together with the required roles and resources.

Planning. The planning stage includes a sub-phase where the functions of the business intelligence system are defined and described in greater detail. Subsequently, existing data as well as other data that might be retrieved externally are assessed. This allows the information structures of the business intelligence architecture, which consist of a central data warehouse and possibly some satellite data marts, to be designed. Simultaneously with the recognition of the available data, the mathematical models to be adopted should be defined, ensuring the availability of the data required to feed each model and verifying that the efficiency of the algorithms to be utilized will be adequate for the magnitude of the resulting problems. Finally, it is appropriate to create a system prototype, at low cost and with limited capabilities, in order to uncover beforehand any discrepancy between actual needs and project specifications.

Implementation and control. The last phase consists of five main sub-phases. First, the data warehouse and each specific data mart are developed. These represent the information infrastructures that will feed the business intelligence system. In order to explain the meaning of the data contained in the data warehouse and the transformations applied in advance to the primary data, a *metadata* archive should be created, as described in Chapter 3. Moreover, ETL procedures are set out to extract and transform the data existing in the primary sources, loading them into the data warehouse and the data marts. The next step is aimed at developing the core business intelligence applications that allow the planned analyses to be carried out. Finally, the system is released for test and usage.

It provides an overview of the main methodologies that may be included in a business intelligence system, most of which will be described


Multidimensional cubes
Data mining
Optimization
Time series analysis
Risk analysis
Data envelopment analysis
Balanced scorecard
Supply chain optimization
Revenue management
Salesforce planning
Campaigns optimization
Clickstream analysis
Relational marketing

UNIT - 2

KNOWLEDGE DELIVERY

PART – A

1. **Define Knowledge delivery?**
   The attraction of delivery and presentation of actionable knowledge must be balanced with the effort necessary for preparing the data so that there is something.

   That can actually be delivered. That being said, once the key components of the data integration, data warehousing, and BI framework have been put in place, it is time to consider what types of equation the different BI users are looking to ask have answered, and the best way of presenting results.

2. **What are the different types of consumers of the results from the and analytics environment?**
   - Power users
   - Business users
   - Casual users
   - Data aggregators or Information Providers
   - Operational analytical users
   - Extend enterprise users
   - IT users

3. **Define Business users?**
   Business users, who rely on domain – specific reporting and analyses prepared by power users, but who also rely on their own ad hoc queries and desire access to raw data for drilling down, direct interaction with analytical servers, extraction, and then further manipulation, perhaps using desktop utility tools.

4. What do you mean by data aggregators?

Data aggregators or information Providers, which are business that collect industry or society wide data and enhance and reorganize that data as a way of providing value added services to customers and subscribers. Some examples include database marketing services, financial and credit information services real estate business information services, audience measurement services, market research providers, and national statistical agencies, among others.

5. Defines Standard Reports?

Reports which are derived from user specification to provide a consistent view of particular aspects of the business, generated in batch and typically delivered on a scheduled basis through a standard (web) interface.

6. Define Interactive Analysis?(APR/MAY2017)

The pre cursor to dimensional analysis provides some level of analysis and is often manifested in pivot table which is called interface Analysis.

These pivot table enabled broader flexibility in grouping data within ordered hierarchies, development of static graphs and charts or just persuing the data from different angles.

7. What are the caveats to allow users to formulate and execute adhoc queries?

There are some caveats to allow users to formulate and execute adhoc queries, they are

Performance

Semantic consistency

Repeatability

8. Define Semantic Consistency?

Semantic Consistency: Allowing users to write their own queries implies they know and understand the meanings of the data elements they have selected to include in their result sets. However, without comprehensive, standardized business term glossaries and metadata repositories, users may see data element names and impulse their

definitions, potentially assigning meanings that are different than what was intended by the data creators. These discrepancies may impact believability of the results.

9. Define Repeatability?

Repeatability: The adhoc process involves a sequences consisting of multiple iterations of the two – phased query and review of the result set process. The operational process allows the analyst to effectively follow a thread or a train o thought, but without a means for capturing the thought process driving the sequence, it is difficult to capture the intuition that derives the ultimate result. In other words, the sequence may yield some results, but it may bed difficult to replicate that process a second or third time.

10. What do you meant by Parameterized reports?

Parameterized reports provide one approach to self – services business intelligence, or "self – service BI". In a self – services BI framework, tools simplify the different aspects of generating results and reports, including simplifying:

- The data discovery process by presenting a palette of data sets that the user can access and use;
- The data access methods by masking or virtualizing access to the data to be queried;
- The documentation of the "make – up" of the report via collaborative means so that the results, and more importantly, the process for generating the results, can be shared with other analysis; and
- The development of the presentation layer, whether that is simple row/column reports.

11. Define Multi dimensional Analysis?

The multi dimensional analysis provided by OLAP tools helps analysis :slice and dice" relationships between variables within different levels of their own hierarchies. Some examples include reviewing "times sales by time period by region"(in which there are different types of times, different time period durations, and different.

12. What are the different methods of Visualizing?

The various methods for Visualizing are:

1. Chart
2. Graphs
3. Widgets

13. Define gauge?

Gauge: A gauge is an indicator of magnitude in the context of critical value ranges. A gauge is good for conveying relative status of critical variables and points that should trigger some action. A traditional examples is an automobile's fuel gauge, which indicates the relative fullness of the tank, as well as an area close to the "empty" measure marked as red to indicates the need for refuelling.

14. Define Spider Chart?

Spider or radar Chart: A spider chart displays a series of variables values across a collections of dimensions. Each dimensions is representation as an axis emanating as from the centre with specific gradations. A set of observations can be mapped as points (and connected with lines). Different observations can be graphed using different colours. An example using a spider chart looks at a number of different characteristics of products (price, height, width, weight, mean time between failure), and relative success, allowing the analyst to quickly compare different products and look for correlations of the variable values.

15. Define Spark Line?

Spark Line: Spark lines are small line graphs without axes or coordinates. Many spark line can be used in relative comparison regarding trends. As an example, the trends of different stock price histories for similar companies can be compared to determine of there are industry relating to stock price.

16. Define Score Cards?

A Scorecards usally presents the values of Key performance indicators as well as indicators reflecting whether those KPI values are acceptable or not. The scorecard presentations may also be enhanced with historical trends and indications if the KPIs have been improving or

not over time. Scorecards are often updated on a periodic basis (e.g., daily or hourly).

17. What is Dashboards?(APR/MAY2019)

Dashboards provides some degree of flexibility to the user in crafting the presentations that is most relevant to the way he/she operates, Given an inventory of presentations graphics (such as those described in the previous sections), an analyst and business user can work together in selecting the most appropriate methods of presentations. Dashboards can connect to real – time sources, and allow the business data consumer to customize an up to date presentation of summarized performance metrics, allowing continuous monitoring throughout the day.

18. What are the Characteristics of business process that suits to integrate analytics?

Some characteristics of business process that are nicely suited to integrated analytics include:

- The business process has distinct performance objectives
- The business process involves decision points by one or more actors.
- The process's performances can be impaired by absence of information
- The process's performance can be improved with well – informed decision making.
- Participants do not need to be "tech – savvy" to be informed.

19. Define Directional Indicators?

Directional Indicators (arrow up or down): These are also indicators that are used for comparison to prior values. Often these are represented using three images – one to indicate improvement, one to indicate no change, and one to indicate degradation of the value. For example, directional indicators can be used as part of time series presentations of stock prices to indicate whether the end of day price is higher, the same, or lower than the previous day's price.

20. What do you meant by Heat Map?

Heat map: This is a graph that tiles a two – dimensional space using tiles of different sizes and colors. A heat map is good for displaying many simultaneous value yet highlighting specific one based on their values. As an examples, a heat map can display the number of times each particular link on web page was clicked, and can highlight the areas of greatest activity.

21. What is alerts or notifications?

Alerts or Notifications: When you think about the ways that individuals usally review the data in the standard report's layout, you will recognize that in many cases, the individual's attention is only focused on one or two key pieces of information. In these situations, the individual's goal is examining some specific variable's value, and either verifying that the value is within an expected range, or determining that the value is outside the expected range and then taking some actions.

PART – B

1. Explain in detail about the seven types of consumers of the results fro the BI and analytics environment?

Power users, who constitute of experienced sophisticated analysts who want to use complex tools and techniques to analyze data and whose results will inform decision – making processes.

Business users, who rely on domain – specific reporting and analyses prepared by power users, but who also rely on their own ad hoc queries and desire access to raw data for drilling down, direct interaction with analytics servers, extractions, and then further manipulation, perhaps using desktop utility tools.

Casual users, who may represents more than one area of the business, and rely on rolled – up entrics from across functions or operational areas summarized from predesigned reports presented via scorecards or dashboards.

Data aggregators or information providers, which are business that collect industry – or societywide data and enhance and reorganize that data

as a way of providing value – added services to customers and subscribers. Some example include database marketing services, financial and credit information services, real estate business information services audience measurement services, market research providers, and national statistical agencies, among others.

Operational analytics users, who indirectly rely on the results of analyrics embedded within operational applications. Examples include call center representatives whose scripts are adjusted interactively in relation to customer protiles, predicted behavioural predispositions and ad placement, or users of retail shelf management systems that adjust stock levels based on demand across multiple regions.

Extended enterprise, users comprising external parties, customers, regulators, external business analysts, partners, suppliers, or anyone with a need for reported information for tactional decision making;

IT users, mostly involved in the development aspects of BI, and whose use of BI is more for supporting the needs of other information consumers.

Each of these user types has different expectations for his or her interactive experience, and each is framed within the context of the driving factors for delivery of actionable knowledge and the types of actions each would take based on the presented results for non – IT and non – aggregator users.

The upshot is that there are many different modes of presentations that are relevant user types. The next sections provide an overview of the typical methods employed for delivery and presentation.

2. Explain in detail about Standard Reports?

Standards Reports

The most "generic" approach to presentation of information reflects a relatively basic two – dimensional alignment of information, characterized within a grid of rows and columns. Standard, static reports derived from user specifications provide a consistent view of particular aspects of the business, generated in batch and typically delivered on a scheduled basis through a standard (web) interface.

The columns typically articulate the item or characteristics being measurement while the rows will generally correspond to the division and hierarchies for which those measures are provided. The intersection of each row and column provides the specific measure for the column's characteristics

for the row's item. For example, let's example a sample report provided by the US Census Bureau.

In this example, there are two measures (in the columns) the estimated of owner occupied housing units with a mortgage, and the margin of error associated with the measure. There are four groups of items being measured;

- The number of owner – occupied hosing units with a mortgage
- Value of the houses
- Mortgage status
- Household income for the previous 12 months

Within some of these groups, these are further hierarchical breakdowns, such as the dollar groupings for value, or the categories for mortgage status. These are relatively generic categories/ hierarchies, and this is reflected in the fact that these are indeed "canned" (or static) reports that have been already prepared for presentation. The presumption is that the static nature of standard reports will drive the need for alternative methods for additional insight. In order words, standard reports present analytical results unless any of the reported numbers are perceived to be beyond the bounds of expectations. And in either case, the standard report only provides a view into what was intended to be shared, but is limited in providing answers to specific business questions.

Users Type	Example users	Actionable knowledge	Type of Action	Example usage Expectations
Power users	Data scientist research analytics "data mines"	Behaviour patterns statistical anomalies analytical results, predictive modes	Recommendations to corporate strategy develop and deliver prdedictive models, adjustments to comporate tactics	Unonncumbered access to data ability to slice and dice dimensional data, integrated data mining models, ad hoc queries, accepts to analytical patterns and programming models.
Business users	Operations managers, division managers	Operational reports, behaviour that derivatives from expectations within business function	Further drill down communication with front line employees, recommendations to senior management	Periodically updated standard reports notifications and alerts, presentations of currents status within monitored business function ad hoc queries
Casual users	Senior managers	Deviation from expected behaviour order activity, perssmization across business function	Alert business managers and users for further investigation and resolution	Monitoring of continuously updated reports and manners across different business functions notifications and alerts.
Operational analytics users	Line or production managers staff workers	Onegoing operational statuses anomaloes behaviour	Mitigation of someduration of rises or emergent issues	Notifications and alerts integrations within existing applications
Extended enterprise users	Your business's customers	Current status of integrations with the business	Purchase or sale of assets revising porticilo payment	Secure presentation via the web

Example Standard Report for US housing Characteristics, Taken from American fact Finder

Subject	United States Owner – occupied housing

3. Explain in detail about Interactive analysis and Adhoc Querying? (APR/MAY 2017)

Interactive Analysis and Ad hoc Querying

BI users looking for additional details regarding information delivered in standard reports may not opt to drill into the data, either with broader visibility into the existing data or with a finer level of granularity. Both are intended to go beyond the relatively strict format of the standard report, even if they open up different views into the data.

The first option involves taking data formatted into a standard report and downloading it into a framework that allows you to slice and dice the existing data more freely. One example involves data from the report into a desktop spreadsheet tool that provides organizations around hierarchies. This precursor to dimensional analysis provides some level of interactive analysis, and is often manifested as a pivot table. These pivot tables enable broader flexibility in grouping data within ordered hierarchies, development of static graphs and charts, or just perusing the data from different angles.

The second option is more powerful in enabling finer granularity by allowing more sophisticated users to execute their own queries into the analytical data platform. Users with an understanding of the data warehouse's data inventory and who have some skill at understanding their help users describe the data they'd like to review. These tools reformulate those requests in SQL queries that are executed directly.

The result sets are also suitable for loading into desktop tools for further organization and analysis, as well as forming the basis for static charts and graphs. However, there are some caveats when allowing users to formulate and execute ad hoc queries here are some.

Performance. Writing efficient queries is a skill, and many queries involve joints across multiple tables that can bring a system's performances to knees. The users would be expected to be highly before letting many implies they their own queries.

Semantic consistency. Allowing users to write their own queries implies they know and understand the meanings of the data elements they have selected to include in their result sets. However, without comprehensive, standardized business term glossaries and metadata repositories, users may see data element names and impulse their definitions, potentially assigning meanings that are different than what was intended by the data creators. These discrepancies may impact believability of the results.

Repeatability. The ad hoc process involves a sequence consisting of multiple iterations of the two – phased query and review of the result set process. The operational process allows the analyst to effectively follow a thread or train of thought, but without a means for capturing the thought process driving the sequences, it is difficult to capture the intuition that drives the ultimate result. In other words, the sequence may yield some results, but it may be difficult to replicate that process a second or third time.

Standard reports can provide knowledge to a broad spectrum of consumers, even if those consumers must have contextual knowledge to identify the key indicators and take action. Ad hoc queries enable grater drill – down and potential for insight. However, given the growth of data into the petabytes coupled with the complexity and performances impacts of ad hoc queries, standard reporting is rapidly yielding to more organized methods for delivering results, through parameterized reporting, dimensional analysis, and notification, alerts exception reporting.

4. What do you meant by Parameterized reports and Self – service reporting?(APR/MAY2017)

Parameterized reports and Self – service reporting

After monitoring the types of ad hoc queries performed, it because apparent in many scenarios that users within similar categories were executing very similar queries. The problem was that despite the similarity of the queries, each was being executed in isolation, with each contributing to degradation of overall performance. However, knowledge of the similarly of query patterns allows the system managers to optimize the environment to help reduce system load but reorganizing the data to make it more ameanable to the similar queries, preprocessing some aspects of those queries, or caching parts of the data to reduce memory access and network latency.

Precomputed of even via form – based drop downs. In other words, the queries are generally static, and differ only by defined set of parameter values. Satisfying these parameterized reports bridges. The parameterized approach is particularly beneficial in operational scenarios in which similar queries and drill – downs are done over and performances by location the queries are always the same, they just differ by the location parameter.

Parameterized reports provide one approach to self- service business intelligence, or "self – service BI". In a self – service BI always the same, they just differ by the location parameter.

Parameterized reports provide one approach to self – service business intelligence, or "self – service BI". In a self – service BI framework, tools simplify the different aspects of generating results and reports, including simplifying:

- The data discovery process by presenting a palette of data sets that the user can access and use
- The data access methods by masking or virtualizing access to the data to be queried
- The documentation of the "make – up" of the report via collaborative means so that the results, and more importantly, the process for generating the results, can be shared with other analysts
- The development of the presentation layer, whether that is simple row/ column reports, or using more sophisticated visualization techniques (as we will explore in the next few sections)

Another benefit of self – services BI is that it is intended to reduce or eliminate the IT bottleneck. In many environments, the IT department is responsible for developing reports and as the BI program gains more acceptance, there will be greater demand for IT resources for report development. This becomes a bottleneck when the time for responding to a request exceeds the window of opportunity for exploiting the actionable performs within its first week of release so that adjustment and tweaks can be made; if it takes three weeks for the report to be readied, it is already too late to take action.

5. Explain in detail about DIMENSIONAL ANALYSIS?

Multidimensional Analysis and online analytical processing (OLAP) add a layer on top of the pivot table approaches used within desktop spreadsheet tools. The multidimensional analysis provided by OLAP tools helps analysis "slice and dice" relationships between different variables within different levels of their own hierarchies. Some examples include reviewing " items sales by time period by region" (in which there are different types of items, different time period durations, and different.

For example, a national call center manager might review average hold times by regional call center. As long as the average hold time is between 30 and 60 seconds, the averages remain within the acceptable level of service. However, once an average hold time for any regions exceeds 60 seconds, the call center manager will need to reach out to the regional call center manager to investigate why the hold times are longer than expected.

Of course, you can envision many similar scenarios in which the action needs to triggered only when certain variables hit specific values. And in each of these cases, reviewing the entire reports is overkill – the business user only need to know the specific variables's value, and only when that value would need to trigger an action; otherwise the variable's value can be ignored. This realization means that instead of presenting an entire report, alerts or notifications can be an alternative method for delivering actionable knowledge.

This method is nicely suited to operational environments in which notifications can be delivered via different methods. Some examples include email, instant messages, direct messages delivered through (potentially internal) social networking sites, smart phones, other mobile devices, radio transmissions or even visual cues (such as scrolling, message boards, light

banks or visual consoles). In these situations, the notifications method can embody the context; for example , a flashing amber lights provides the medium for notification as well as the message. This approach not only simplifies the delivery of the critical peace of information, it reduces the effort for inspecting the critical values and there by enable actions.

Levels of granularity for regions or "product availability by product classification by supplier by location " (in which there are multiple layers of product classifications and different levels of granularity for locations)

The use of the word "by" suggests a pivot around which the data can be viewed, allowing us to look at items sold, grouped by item classification and then grouped by time periods, then by regions or the other way around, grouped by regions then by time periods. OLAP lets the analyst drill up and down along the hierarchies in the different dimensions to uncover dependent relationships that are hidden within the hierarchies.

Since OLAP queries are generally organized around partial aggregations along the different dimensions, the data can be organized along the different dimension in queries that "slice" or "dice" the data. "Slicing" fixes the region (Northeast) and reviewing items sales grouped by classification and or more dimensions, such as selecting a collection of item classifications and then presenting those selected items by time period and locations.

Any of the dimensions can be drilled through (also referred to a drill – down) by navigating along the different levels of a dimension's hierarchy. For example, once a region is selected (Northeast), sales by item by time period can be reviewed at the next level down in the regional hierarchy (such as by each of the states in the Northeast region)

OLAP environments presents the data aligned along selected dimensions. The presentations layer often provides a palette from which dimensions can be selected for visualizations and those dimensions can be pivoted around each other. The data can be presented in the same grid format as the standard report, or can be visualized using graphical components. The slicing, dicing, and drill through provided by the OLAP presentations provides much greater flexibility for the power user performing data to identify anomalous behaviour or to look for potential patterns may also benefit through the use OLAP environment.

6. What is visualization? Explain different types of visualizations modes of data?(APR/MAY2019)

The methods discussed have largely focused on the means of delivery of the analytical results, but less on the presentation. Yet our conclusion at the end of the previous section suggested that presentations methods for specific pieces of information might be better in conveying a message or triggering the and then comparing analytical results.

There are many different types of visualizations modes for data, and while this is not intended to provide a comprehensive overview of visualization techniques, it is meant to provide an overview of a handful of ways to present actionable knowledge.

Line chart: A line chart maps point on a grid connected by line segments. A line chart can be used to show a series of connected values, such as a time series. An example would be mapping the rice and fall of gas prices per gallon using the price of a gallon of gas on the first day of each month for the previous 36 months.

Bar chart: A bar chart maps values using rectangles whose lengths corresponds to the charted values. Bar charts are good for comparing different values of the same variable across different contexts. An example would be chart of the average life expectancy in years across different countries.

Pie chart: A pie chart is conveyed as a circle that is broken out into sectors representing some percentage of a whole. A pie chart is good for showing distributions of values across a single domain. An example is showing the relative percentage of owner occupied homes by ethnicity within a Zip code area. The total of all the components always will add up to 100% and each slice of the pie represents a percentage of the whole.

Scatter plot. A scatter plot graphs points showing relationship between two variables. Typically one variable is fixed (the dependent variable) and the other is not (the independent variable). In a two – dimensional scatter plot, the x axis represents the independent variable value and the y axis represents the dependent variable. A scatter plot is used to look for correlation between the dependent and independent variable. An example graphs an individual's age (the dependent variable) and the individual's observed weight (the independent variable).

Bubble chart. A bubble chart is variation on a scatter plot in which a third variable can be represented using the size of the item in the chart. An example would graph the dollar sales volume by the number of items sold, and the bubbles could represent the percentage of the overall market share.

Gauge. A gauge is an indicator of magnitude in the context of critical value ranges. A gauge is good for conveying relative status of critical variable and points that should trigger some action. A traditional example is an automobile's fuel gauge, which indicates the relative fullness of the tank, as well as an area close to the "empty" measure marked as red to indicate the need for refuelling.

Directional indicators (arrow up or down). These are also indicators that are used for comparison to prior values. Often these are represented using three images one to indicate improvement, on to indicate no changes, and one to indicate degradation of the value. For example, directional indicators can be used as part of a time series presentation of stock prices to indicate whether the end of day price is higher, the same, or lower than the previous day's price.

Heat map. This is graph that tiles a two – dimensional space using tiles of different sizes and colors. A heat map is good for displaying many simultaneous values yet highlighting specific ones based on their values. As an example, a heat map can display the number of times each particular link on a web page was clicked, and can highlight the times of greatest activity.

Spider or radar chart. A spider chart displays a series of variable values across a collection of dimensions. Each dimension is represented as an axis emanating from the centre with specific gradations. A set of observations can be mapped as points (and connected with lines). Different observations can be graphed using different colors. An example using a spider chart looks at a number of different characteristics of products (price, height, width, weight, mean time between failure) and relative success allowing the analyst to quickly compare different products and look for correlation of the variable values.

Spark line. Spark line are small line graphs without axes or coordinates. Many spark line can be used in relative comparison regarding trends. As an example, the trends of different stock price histories for similar companies can be compared to determine of there are industry trends relating to stock price.

There are many other types of visual "widgets" that can be used for presentation of information. A good resource for understanding graphic visualization is the classic.

7. Explain in detail about Score cards and dash boards? (APR/MAY2017)

Beginning with notification and alerts, we can extend the concept of carefully crafting the visual presentation of broader collections of relevant analytical results. In other words, if a trained eye is required to scan key performances metrics form standard reports, simplifying the presentation of the key performances metrics may better enable the knowledge worker to transition from seeing what has already happened to understanding the changes necessary to improve the business process.

Scorecards and dashboards are two different approaches for consolidating the presentation of reported results to a particular user type. A scorecard usally presents the value of key performances indicators as well ass indicators reflecting whether those KPI values are acceptable or not. The scorecard presentation may also be enhanced with historical trends and indications if the KPIs have been improving or not over time. Scorecards are often updated on a periodic basis (e.g., daily or hourly)

Dashboards provide some degree of flexibility to the user in crafting the presentation that is most relevant to the way he/she operates. Given an inventory of presentation graphics (such as those described in the previous section), an analyst and business user can work together in selecting the most appropriate methods of data consumer to customize an up – to – date presentation of summarized performance metrics, allowing continuous monitoring throughout the day.

Pervasive delivery mechanism can push dashboards to a large variety of channels, ranging from the traditional browser – based format to handheld mobile devices. Through the interactive nature of the dashboards, the knowledge worker can drill down through the key indicators regarding any emerging opportunities, as well as take action through integrated process – flow and communication engines.

Another approach to dashboards is the concept of a mashup, which allows the knowledge consumers themselves the ability their own combination of analytics and reports with external data streams, news feeds, social networks, and other web2.0 resources in a visualization

framework that specifically suits their own business needs and objectives. The mashup framework provides the "glue" for integrating data stream and BI with interactive business applications.

8. What do you meant by Geographic visualization?

Geographic Visualization

In aspects of location intelligence and spatial analytics, and the results of that type of analysis can be presented within the context of a map. Instead of using the standard graphical widgets described in a previous section, aggregate values and totals can be attributed to a visual representation of a map. For example, population statistics for each country in the European Union can be superimposed on top of a map of Europe.

These maps can satisfy the desire to drill down; interactive selection or clicking on one segment of the map can be zoom in from a geographic standpoint. In addition, spatial analysis results can be layered within the mapping interface. For example, in an insurance management application, hazard zones can be superimposed on top of regions potentially affected by weather events to help guide determination of heightened risk areas the company's customer base. Another example employs the heat map concept to geographic regions using and colors to present a collection of variable values.

Often dashboards will link more than one visualization component to others, and this can be easily applied to geographic visualization. For example a dimensional analysis presentation (such as pivot table) for a geographic hierarchy can be presented in one frame while the aggregated values are displayed within a map. Realigning the dimensions in the grid will automatically update the map, and drilling through regions on the map will trigger a reload of the drilled through data within the grid.

9. Explain about Integrated Analytics?

Integrated Analytics

As specific analytic values trigger specific actions within well – defined business process, employing a combination of alerts and notifications with visualization tools reduces the need for the end users to have deep training in the use of the BI tool set. In other words, when the analytical results are

fed directly into operational activities, the end – user may be a consumer of BI and not even be aware of it.

Some characteristics of business process that are nicely suited to integrated analytics include.

- The business process has distinct performances objectives
- The business process involves decision points by one or more actors
- The process's performances can be impaired by absence of information
- The process's performances can be impaired by ill – informed decisions.
- The process's can be improved with well – informed decision making
- Participants do not need to be "tech – savvy" to be informed

Yet in order to make integrated analytics work, the implements must make sure that all the information necessary can be delivered to the appropriate person with the right time frame to facilitate the best decisions. That might imply the need for real time integration of data from multiple sources of both analytics results and operational data. In turn, the delivery of the actionable knowledge must be seamlessly presented in a way that is best suited to business operations and is seamlessly integrated with the common suites of desktop productivity tools. This type of event driven notification allows analytics to be directly embedded within operational process and supporting applications across multiple channels. As this approach gains popularity more widespread adoption of BI services coupled with lowered barriers to deployment will open new opportunities for integrating BI results.

10. What are considerations in Optimizing the Presentation for the Right Message?

Considerations: Optimizing the Presentation for the Right Message

We have covered the range of presentation from the traditional grid style report all the way to fancy visualization, all intended to help convey the results of analyses that should trigger actions to benefit the business. However, the wide palette of available graphs, charts, indicators, dials,

knobs, and such can, at times, detract from the content when the presentation overwhelms the values that are being presented. Here are somne quick guidelines to keep in mind when laying out a BI dashboard.

Choose the right visualization graphic: Don't let the shiny graphics fool you into using a visual component that does not properly convey the indent result. For example, line charts are good for depicting historical trends of the same variable over time but bar charts may not be as good a choice.

Manage your "real estate". The available screen space limits what can be displayed at one time, and this is what is referred to as screen "real estate".

When optimizing the presentation for the right message, consider the following considerations; the main context, consistency, simplicity of the message, and engagement.

Maintain context: You must recognize that the presentation of value is subjected to variant interpretations when there is no external context defining its meaning. For example, presenting a value on a dial – gauge conveys the variable's magnitude, but not whether that value is good, bad, or indifferent. Adjusting the dial gauge with a red zone (to indicate a bad value) and a green zone (to indicate a good value) provides the context of the displayed magnitude.

Be Consistent. When the power of self – service dashboard development is placed in the hands of many data consumers, their own biases will lead to an explosion of variant ways of representing the same or similar ideas. The result is that what makes sense to one grows with wider dissemination. Consistent representations and presentations (and corresponding selection of standard visualization graphics) will help to ensure consistent interpretations.

Keep it simple. Don't inundate presentation with fancy – looking graphics that don't add to the decision making process. Often the simpler the presentation, the more easily the content is conveyed.

Engage. Engage the user community and agree on standards practices, and a guide book for developing visualization parameters for delivery and presentation.

Unit – III

EFFICIENCY
PART-A

1. What is the purpose of DEA?

The purpose of *data envelopment analysis* (DEA) is to compare the operating performance of a set of units such as companies, university departments, hospitals, bank branch offices, production plants, or transportation systems. In order for the comparison to be meaningful, the units being investigated must be homogeneous.

2. Define inputs and outputs in DEA?

Data envelopment analysis relies on a productivity indicator that provides a measure of the efficiency that characterizes the operating activity of the units being compared. This measure is based on the results obtained by each unit, which will be referred to as *outputs*, and on the resources utilized to achieve these results, which will be generically designated as *inputs* or *production factors*.

3. Define Decision Making Units?

In data envelopment analysis the units being compared are called *decision-making units* (DMUs), since they enjoy a certain decisional autonomy.

4. How to measure the efficiency of Decision making Units?

Assuming that we wish to evaluate the efficiency of n units, let $N = \{1, 2, \ldots, n\}$ denote the set of units being compared.

If the units produce a single output using a single input only, the *efficiency* of the jth decision-making unit DMU_j, $j \in N$, is defined as

$$\theta_j = \frac{y_j}{x_j}, \qquad \theta_j = \frac{v}{x_j}$$

in which y_j is the output value produced by DMU_j and x_j the input value used.

the efficiency of DMU_j is defined as

$$\theta_j = \frac{u_1 y_{1j} + u_2 y_{2j} + \ldots + u_m y_{mj}}{v_1 x_{1j} + v_2 x_{2j} + \ldots + v_s x_{sj}} = \frac{\sum_{r \in K} u_r y_{rj}}{\sum_{i \in H} v_i x_{ij}},$$

$$\theta_j = \frac{u_1 y_{1j} + u_2 y_{2j} + \ldots + u_m y_{mj}}{v_1 x_{1j} + v_2 x_{2j} + \ldots + v_s x_{sj}} = \frac{\sum_{r \in K} u_r y_{rj}}{\sum_{i \in H} v_i x_{ij}}$$

$$\theta_j^o = \frac{y_j}{y_j^*}$$

$$\theta_j = \frac{y}{x_j}$$

for weights $u_1, u_2, \ldots, u_m$ associated with the outputs and $v_1, v_2, \ldots, v_s$ assigned to the inputs.

5. Define Efficient Frontier?

The *efficient frontier*, also known as *production function*, expresses the relationship between the inputs utilized and the outputs produced. It indicates the maximum quantity of outputs that can be obtained from a given combination of inputs. At the same time, it also expresses the minimum quantity of inputs that must be used to achieve a given output level. Hence, the efficient frontier corresponds to *technically efficient* operating methods.

6. Define efficient and in-efficient units?

The line with the maximum slope, represented in Figure 15.1 by a solid line, is the effi- cient frontier for all branches being analyzed. The branches that are on this line correspond to efficient units, while the branches that are below the efficient frontier are inefficient units.

7. Define Production Possibility Set?

The area between the efficient frontier and the positive horizontal semi-axis is called the Production Possibility Set.

8. Define input oriented and output oriented efficiency?

In particular, for each DMU_j, $j \in N$, the *input-oriented* efficiency θ_j^I can be defined as the ratio between the ideal input quantity x^* that should be used by the unit if it were efficient and the actually used quantity x

Similarly, the *output-oriented* efficiency θ_j^O is defined as the ratio between the quantity of output y_j actually produced by the unit and the ideal quantity y^* that it should produce in conditions of efficiency: $\theta_j^O = \frac{y_j}{y^*}$

$$\theta^O{}_j \frac{y_j}{y^*}. \qquad 4)$$

9. Define CCR model?

Using data envelopment analysis, the choice of the optimal system of weights for a generic DMU_j involves solving a mathematical optimization model whose decision variables are represented by the weights u_r, $r \in K$, and v_i, $i \in H$, associated with each output and input. Various formulations have been proposed, the best-known of which is probably the Charnes–Cooper–Rhodes (CCR) model.

10.Define Peer group?

Data envelopment analysis identifies for each inefficient unit a set of excellent units, called a *peer group*, which includes those units that are efficient if evaluated with the optimal system of weights of an inefficient unit. The peer group, made up of DMUs which are characterized by operating methods similar to the inefficient unit being examined, is a realistic term of comparison which the unit should aim to imitate in order to improve its performance.

11.What is the need to identify efficient units?

The need to identify the efficient units, for the purpose of defining the best operating practices, stems from the principle itself on which data envelopment analysis is grounded, since it allows each unit to evaluate its own degree of efficiency by choosing the most advantageous structure of weights for inputs and outputs.

What are the methods used differentiate efficient units?

To differentiate these units, we may resort to a combination of differ- ent methods: *cross-efficiency analysis* , evaluation of *virtual inputs* and *virtual outputs*, and *weight restrictions*.

12.Define cross-efficiency analysis? (APR/MAY 2017)

The analysis of *cross-efficiency* is based on the definition of the *efficiency matrix*, which provides information on the nature of the weights system adopted

by the units for their own efficiency evaluation.

13.Define square efficiency matrix?

The square efficiency matrix contains as many rows and columns as there are units being compared. The generic element θ_{ij} of the matrix represents the efficiency of DMU_j evaluated through the optimal weights structure for DMU_i, while the element θ_{jj} provides the efficiency of DMU_j calculated using its own optimal weights.

14.What are the quantities of interest derived from the efficiency matrix?

Two quantities of interest can be derived from the efficiency matrix.

The first represents the average efficiency of a unit with respect to the optimal weights systems for the different units, obtained as the average of the values in the jth column. The second is the average efficiency of a unit measured applying its optimal system of weights to the other units. The latter is obtained by averaging the values in the row associated with the unit being examined.

15.What is the purpose of Virtual inputs and virtual outputs?

Virtual inputs and virtual outputs provide information on the relative importance that each unit attributes to each individual input and output, for the purpose of maximizing its own efficiency score. Thus, they allow the specific competencies of each unit to be identified, highlighting at the same time its weaknesses.

The *virtual inputs* of a DMU are defined as the product of the inputs used by the unit and the corresponding optimal weights. Similarly, *virtual outputs* are given by the product of the outputs of the unit and the associated optimal weights.

.

16.Define linear regression.(APR/MAY 2017)

Linear regression refers to a model that can show relationship between two variables and how one can impact the other. In essence, it involves showing how the variation in the "dependent variable" can be captured by change in the "independentvariables".

17.Write a short notes on outliers (APR/MAY 2019)

An outlier, in mathematics, statistics and information technology, is a specific data point that falls outside the range of probability for a data set. In other words, the outlier is distinct from other surrounding data points in a particular way. Outlier analysis is extremely useful in various kinds of analytics and research, some of it related to technologies and IT systems.

PART- B

1. Explain about Date envelopment analysis?

The purpose of *data envelopment analysis* (DEA) is to compare the operating performance of a set of units such as companies, university departments, hospitals, bank branch offices, production plants, or transportation systems. In order for the comparison to be meaningful, the units being investigated must be homogeneous.

The performance of a unit can be measured on several dimensions. For example, to evaluate the activity of a production plant one may use quality indicators, which estimate the rate of rejects resulting from manufacturing a set of products, and also flexibility indicators, which measure the ability of a system to react to changes in the requirements with quick response times and low costs.

Data envelopment analysis relies on a productivity indicator that provides a measure of the efficiency that characterizes the operating activity of the units being compared. This measure is based on the results obtained by each unit, which will be referred to as *outputs*, and on the resources utilized to achieve these results, which will be generically designated as *inputs* or *production factors*. If the units represent bank branches, the outputs may consist of the number of active bank accounts, checks cashed or loans raised; the inputs may be the number of cashiers, managers or rooms used at each branch. If the units are university departments, it is possible to consider as outputs the number of active teaching courses and scientific publications produced by the members of each department; the inputs may include the amount of financing received by each department, the cost of teaching, the administrative staff and the availability of offices and laboratories.

2. What are the Efficiency measures for Data Envelopment analysis?

In data envelopment analysis the units being compared are called *decision-making units* (DMUs), since they enjoy a certain decisional autonomy. Assuming that we wish to evaluate the efficiency of n units, let $N = \{1, 2, \ldots, n\}$ denote the set of units being compared.

If the units produce a single output using a single input only, the *efficiency* of the jth decision-making unit DMU_j, $j \in N$, is defined as

$$\theta_j = \frac{y_j}{x_j},$$

in which y_j is the output value produced by DMU_j and x_j the input value used.

If the units produce multiple outputs using various input factors, the efficiency of DMU_j is defined as the ratio between a weighted sum of the outputs and a weighted sum of the inputs. Denote by $H = \{1, 2, \ldots, s\}$ the set of production factors and by $K = \{1, 2, \ldots, m\}$ the corresponding set of outputs. If x_{ij}, $i \in H$, denotes the quantity of input i used by DMU_j and y_{rj}, $r \in K$, the quantity of output r obtained, the efficiency of DMU_j is defined as

$$\theta_j = \frac{u_1 y_{1j} + u_2 y_{2j} + \ldots + u_m y_{mj}}{v_1 x_{1j} + v_2 x_{2j} + \ldots + v_s x_{sj}} = \frac{\sum_{r \in K} u_r y_{rj}}{\sum_{i \in H} v_i x_{ij}},$$

for weights $u_1, u_2, \ldots, u_m$ associated with the outputs and $v_1, v_2, \ldots, v_s$ assigned to the inputs.

In this second case, the efficiency of DMU_j depends strongly on the system of weights introduced. At different weights, the efficiency value may undergo relevant variations and it becomes difficult to fix a single structure of weights that might be shared and accepted by all the evaluated units. In order to avoid possible objections raised by the units to a preset system of weights, which may privilege certain DMUs rather than others, data envelopment analysis evaluates the efficiency of each unit through the weights system that is best for the DMU itself – that is, the system that allows its efficiency value to be maximized. Subsequently, by means of additional analyses, the purpose of data envelopment analysis is to identify the units that are efficient in absolute terms and those whose efficiency value depends largely on the system of weights adopted.

3. What is Efficient Fronier? Explain in detail?

The *efficient frontier*, also known as *production function*, expresses the relationship between the inputs utilized and the outputs produced. It indicates the maximum quantity of outputs that can be obtained from a given combination of inputs. At the same time, it also expresses the minimum quantity of inputs that must be used to achieve a given output level. Hence, the efficient frontier corresponds to *technically efficient* operating methods.

The efficient frontier may be empirically obtained based on a set of observations that express the output level obtained by applying a specific combination of input production factors. In the context of data envelopment analysis, the observations correspond to the units being evaluated. Most statistical methods of parametric nature, which are based for instance on the calculation of a regression curve, formulate some prior hypotheses on the shape of the production function. Data envelopment analysis, on the other hand, forgoes any assumptions on the functional form of the efficient frontier, and is therefore nonparametric in character. It only requires that the units being compared are not placed above the production function, depending on their efficiency value. To further clarify the notion of efficient frontier.

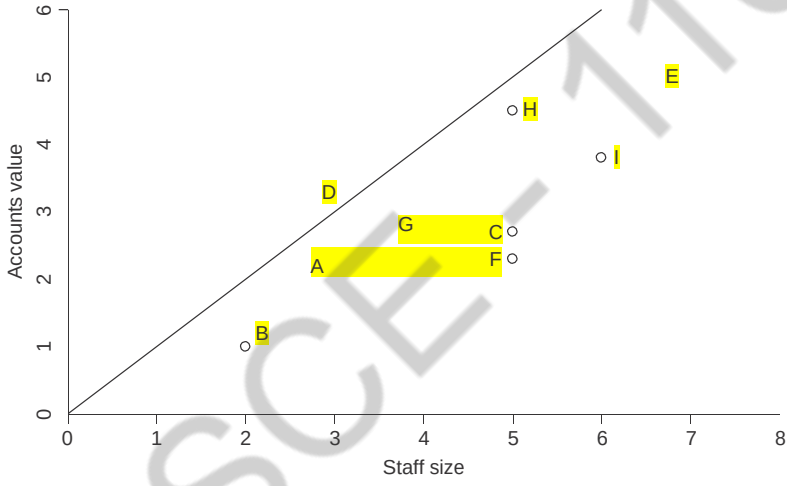
Evaluation of the efficiency of bank branches. A bank wishes to compare the operational efficiency of its nine branches, in terms of staff size and total value of savings in active accounts. It shows for each branch the total value of accounts, expressed in hundreds of thousands of euros, and the number of staff employed, with the cor- responding efficiency values calculated based on definition .The graph shows for each branch the number of employ- ees on the horizontal axis and the value of accounts on the vertical axis. The slope of the line connecting each point

to the origin represents the efficiency value associated with the corresponding branch. The line with the maximum slope, represented by a solid line, is the effi- cient frontier for all branches being analyzed. The branches that are on this line correspond to efficient units, while the branches that are below the efficient frontier are inefficient units. The area between the efficient frontier and the positive horizontal semi-axis is called the *production possibility set*.

A possible alternative to the efficient frontier is the regression line that can be obtained based on the available observations, indicated in Figure 15.1 by a dashed line. In this case, the units that fall above the regression line may be deemed excellent, and the degree of excellence of each unit could be expressed by its distance from the line. However, it is appropriate to underline the difference that exists between the prediction line obtained using a regression model and the efficient frontier obtained using data envelopment analysis. The

Input and output values for the bank branches

bank branch	staff size	accounts value	efficiency
A	3	2.5	0.733
B	2	1.0	0.500
C	5	2.7	0.540
D	3	3.0	1.000
E	7	5.0	0.714
F	5	2.3	0.460
G	4	3.2	0.700
H	5	4.5	0.900
I	6	4.5	0.633



Evaluation of efficiency of bank branches

regression line reflects the average behavior of the units being compared, while the efficient frontier identifies the best behavior, and measures the inefficiency of a unit based on the distance from the frontier itself.

Notice also that the efficient frontier provides some indications for improving the performance of inefficient units. Indeed, it identifies for each input level the output level that can be achieved in conditions of efficiency. By the same token, it identifies for each output level the minimum level of input that should be used in conditions of efficiency. In particular, for each DMU_j, $j \in N$, the *input-oriented* efficiency θ_j^I can be defined as the ratio between the ideal input quantity x^* that should be used by the unit if it were efficient and the actually used quantity x_j:

$$\theta_j^I = \frac{x^*}{x_j}. \qquad (15.3)$$

Similarly, the *output-oriented* efficiency θ_j^O is defined as the ratio between the quantity of output y_j actually produced by the unit and the ideal quantity y^* that it should produce in conditions of efficiency:

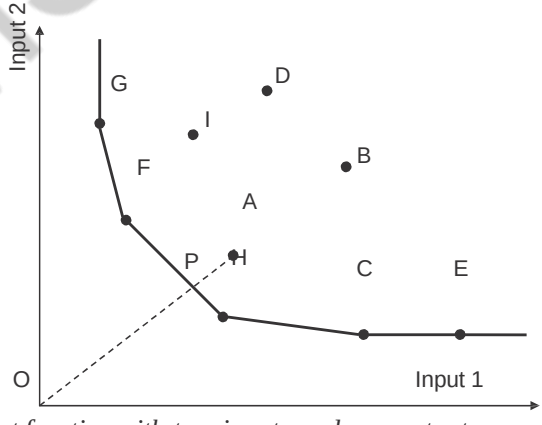
$$\theta^O \quad \underline{y_j} \qquad\qquad j = y^* .$$

The problem of making an inefficient unit efficient is then turned into one of devising a way by which the inefficient unit can be brought close to the efficient frontier.

If the unit produces a single output only by using two inputs, the efficient frontier assumes the shape. In this case, the inefficiency of a given unit is evaluated by the length of the segment connecting the unit to the efficient frontier along the line passing through the origin of the axes. For the example illustrated in Figure 15.2, the efficiency value of DMU_A is given by

$$\overline{OP} \quad \theta_A = \quad OA , \qquad)$$

where $\overline{OP}$ and $\overline{OA}$ represent the lengths of segments OP and OA, respectively. The inefficient unit may be made efficient by a displacement along segment OA that moves it onto the efficient frontier. Such displacement is tantamount to progressively decreasing the quantity of both inputs while keeping unchanged the quantity of output. In this case, the production possibility set is defined as the region delimited by the efficient frontier where the observed units being compared are found.



Efficient frontier with two inputs and one output

4.Explain in detail about CCR model? (APR/MAY 2017)

Using data envelopment analysis, the choice of the optimal system of weights for a generic DMU_j involves solving a mathematical optimization model whose decision variables are represented by the weights u_r, $r \in K$, and v_i, $i \in H$, associated with each output and input. Various formulations have been proposed, the best-known of which is probably the Charnes–Cooper–Rhodes (CCR) model. The CCR model formulated for DMU_j takes the form

$$\begin{aligned} \max \quad & \vartheta = \frac{\sum_{r \in K} u_r y_{rj}}{\sum_{i \in H} v_i x_{ij}}, \\ \text{s.to} \quad & \frac{\sum_{r \in K} u_r y_{rj}}{\sum_{i \in H} v_i x_{ij}} \leq 1, \qquad j \in N, \\ & u_r, v_i \geq 0, \qquad r \in K, i \in H. \end{aligned} \tag{}$$

The objective function involves the maximization of the efficiency measure for DMU_j. Constraints require that the efficiency values of all the units, calculated by means of the weights system for the unit being examined, be lower than one. Finally, conditions guarantee that the weights associated with the inputs and the outputs are non-negative. In place of these conditions, sometimes the constraints $u_r, v_i \geq \delta$, $r \in K$, $i \in H$ may be applied, where $\delta > 0$,
preventing the unit from assigning a null weight to an input or output.

Model can be linearized by requiring the weighted sum of the inputs to take a constant value, for example 1. This condition leads to an alterna- tive optimization problem, the *input-oriented* CCR model, where the objective function consists of the maximization of the weighted sum of the outputs

$$\begin{aligned} \max \quad & \vartheta = \sum_{r \in K} u_r y_{rj}, \\ \text{s.to} \quad & \sum_{i \in H} v_i x_{ij} = 1, \\ & \sum_{r \in K} u_r y_{rj} - \sum_{i \in H} v_i x_{ij} \leq 0, \qquad j \in N, \\ & u_r, v_i \geq 0, \qquad r \in K, i \in H. \end{aligned}$$

Let ϑ^* be the optimum value of the objective function corresponding to the optimal solution $(\mathbf{v}^*, \mathbf{u}^*)$ of problem. DMU_j is said to be *efficient* if $\vartheta^* = 1$ and if there exists at least one optimal solution $(\mathbf{v}^*, \mathbf{u}^*)$ such that $\mathbf{v}^* > \mathbf{0}$ and $\mathbf{u}^* > \mathbf{0}$.

By solving a similar optimization model for each of the n units being compared, one obtains n systems of weights. The flexibility enjoyed by the

units in choosing the weights represents an undisputed advantage, in that if a unit turns out to be inefficient based on the most favorable system of weights, its inefficiency cannot be traced back to an inappropriate evaluation process. However, given a unit that scores $\vartheta^* = 1$, it is important to determine whether its efficiency value should be attributed to an actual high-level performance or simply to an optimal selection of the weights structure.

Dual of the CCR model

For the input-oriented CCR model, the following dual problem, which lends itself to an interesting interpretation, can be formulated:

$$\begin{aligned} \min \quad & \vartheta, \\ \text{s.to} \quad & \sum_{j \in N} \lambda_j x_{ij} - \vartheta x_{ij} \leq 0, \quad i \in H, \\ & \sum_{j \in N} \lambda_j y_{rj} - y_{rj} \geq 0, \quad r \in K, \\ & \lambda_j \geq 0, \quad j \in N. \end{aligned}$$

Based on the optimum value of the variables $\lambda_j^*, j \in N$, the aim of model
is to identify an ideal unit that lies on the efficient frontier and represents
a term of comparison for DMU_j. Constraints of the model require that this unit produces an output at least equal to the output produced by DMU_j, and uses a quantity of inputs equal to a fraction of the quantity used by the unit examined. The ratio between the input used by the ideal unit and
the input absorbed by DMU_j is defined as the optimum value ϑ^* of the dual variable ϑ. If $\vartheta^* < 1$, DMU_j lies below the efficient frontier. In order to be efficient, this unit should employ $\vartheta^* x_{ij}$, $i \in H$, of each input.

The quantity of inputs utilized by the ideal unit and the level of outputs to be produced are expressed as a linear combination of the inputs and outputs associated with the n units being evaluated:

$$\begin{aligned} x_i^{\text{ideal}} &= \sum_{j \in N} \lambda_j^* x_{ij}, \quad i \in H, \\ y_r^{\text{ideal}} &= \sum_{j \in N} \lambda_j^* y_{rj}, \quad r \in K. \end{aligned}$$

For each feasible solution (ϑ, λ) to problem , the slack variables s_i^-, $i \in H$, and s_r^+, $r \in K$, can be defined, which represent respectively the quantity of input i used in excess by DMU_j and the quantity of output r produced in

shortage by the DMU_j with respect to the ideal unit:

$$s_i^- = \vartheta x_{ij} - \sum_{j \in N} \lambda_j x_{ij}, \quad i \in H,$$

$$s_r^+ = \sum_{j \in N} \lambda_j y_{rj} - y_{rj}, \quad r \in K.$$

As with the primal problem, it is possible also for the dual problem to provide a definition of efficiency. DMU_j is efficient if $\vartheta^* = 1$ and if the optimum value of the slack variables is equal to zero: $s_i^{-*} = 0$, $i \in H$, and $s_r^{+*} = 0$, $r \in K$. In other words, DMU_j is efficient if it is not possible to improve the level of an input used or the level of an output produced without a deterioration in the level of another input or of another output. If $\vartheta^* < 1$, DMU_j is said to be *technically inefficient*, in the sense that, in order to obtain the same output, the input quantities used could be simultaneously reduced in the same proportion. The maximum reduction allowed by the efficient frontier is defined by the value $1 - \vartheta^*$. If $\vartheta^* = 1$, but some slack variables are different from zero, DMU_j presents a *mix inefficiency* since, keeping the same output level, it could reduce the use of a few inputs without causing an increase in the quantity of other production factors used.

5.What are the target objectives of CCR model?

In real-world applications it is often desirable to set improvement objectives for inefficient units, in terms of both outputs produced and inputs utilized. Data envelopment analysis provides important recommendations in this respect, since it identifies the output and input levels at which a given inefficient unit may become efficient. The efficiency score of a unit expresses the maximum proportion of the actually utilized inputs that the unit should use in conditions of efficiency, in order to guarantee its current output levels. Alternatively, the inverse of the efficiency score indicates the factor by which the current output levels of a unit should be multiplied for the unit to be efficient, holding constant the level of the productive inputs used. Based on the efficiency values, data envelopment analysis therefore gives a measure for each unit being compared of the savings in inputs or the increases in outputs required for the unit to become efficient.

To determine the target values, it is possible to follow an input- or output-oriented strategy. In the first case, the improvement objectives primarily concern the resources used, and the target values for inputs and outputs are given by

$$x_{ij}^{\text{target}} = \vartheta^* x_{ij} - s_i^{-*}, \quad i \in H,$$

$$y_{rj}^{\text{target}} = y_{rj} + s_r^{+*}, \quad r \in K.$$

In the second case, target values for inputs and outputs are given by

$$x_{ij}^{\text{target}} = x_{ij} - \frac{s_i^{-*}}{\vartheta^*}, \quad i \in I,$$

$$y_{rj}^{\text{target}} = \frac{y_{rj} + s_r^{+*}}{\vartheta^*}, \quad r \in K.$$

Other performance improvement strategies may be preferred over the proportional reduction in the quantities of inputs used or the proportional increase in the output quantities produced:

- priority order for the production factors – the target values for the inputs are set in such a way as to minimize the quantity used of the resources to which the highest priority has been assigned, without allowing variations in the level of other inputs or in the outputs produced;
- priority order for the outputs – the target values for the outputs are set in such a way as to maximize the quantity produced of the outputs to which highest priority has been assigned, without allowing variations in the level of other outputs or inputs used;
- preferences expressed by the decision makers with respect to a decrease in some inputs or an increase in specific outputs.

6.Explain about Peer groups in Data Envelopment Analysis?

Data envelopment analysis identifies for each inefficient unit a set of excellent units, called a *peer group*, which includes those units that are efficient if evaluated with the optimal system of weights of an inefficient unit. The peer group, made up of DMUs which are characterized by operating methods similar to the inefficient unit being examined, is a realistic term of comparison which the unit should aim to imitate in order to improve its performance.

The units included in the peer group of a given unit DMU_j may be identified by the solution to model (15.9). Indeed, these correspond to the DMUs for which the first and the second member of constraints (15.11) are equal:

$$E_j = \{ j : \sum_{r \in K} u_r^* y_{rj} = \sum_{i \in H} v_i^* x_{ij} \}.$$

Alternatively, with respect to formulation , the peer group consists of those units whose variable λ_j in the optimal solution is strictly positive:

$$E_j = \{ j : \lambda_j^* > 0 \}.$$

Notice that within a peer group a few excellent units more than others may represent a reasonable term of comparison. The relative importance of a unit belonging to a peer group depends on the value of the corresponding variable λ_j in the optimal solution of the dual model.

The analysis of peer groups allows one to differentiate between really efficient units and apparently efficient units for which the choice of an optimal system of weights conceals some abnormal behavior. In order to draw this distinction, it is necessary to consider the efficient units and to evaluate how often each belongs to a peer group. One may reasonably expect that an efficient unit often included in the peer groups uses for the evaluation of its own efficiency a robust weights structure. Conversely, if an efficient unit rarely represents a term of comparison, its own system of optimal weights may appear distorted, in the sense that it may implicitly reflect the specialization of the unit along a particular dimension of analysis.

7. What are the Identification of good operating practices? (APR/MAY 2017)

By identifying and sharing *good operating practices*, one may hope to achieve an improvement in the performance of all units being compared. The units that appear efficient according to data envelopment analysis certainly represent terms of comparison and examples to be imitated for the other units. However, among efficient units some more than others may represent a target to be reached in improving the efficiency.

The need to identify the efficient units, for the purpose of defining the best operating practices, stems from the principle itself on which data envelopment analysis is grounded, since it allows each unit to evaluate its own degree of efficiency by choosing the most advantageous structure of weights for inputs and outputs. In this way, a unit might appear efficient by purposely attributing a non-negligible weight only to a limited subset of inputs and outputs. Furthermore, those inputs and outputs that receive greater weights may be less critical than other factors more intimately connected to the primary activity performed by the units being analyzed. In order to identify good operating practices, it is therefore expedient to detect the units that are really efficient, that is, those units whose efficiency score does not primarily depend on the system of weights selected. To differentiate these units, we may resort to a combination of different methods: *cross-efficiency analysis*, evaluation of *virtual inputs* and *virtual outputs*, and *weight restrictions*.

Cross-efficiency analysis

The analysis of *cross-efficiency* is based on the definition of the *efficiency matrix*, which provides information on the nature of the weights system adopted

by the units for their own efficiency evaluation. The square efficiency matrix contains as many rows and columns as there are units being compared. The generic element θ_{ij} of the matrix represents the efficiency of DMU_j evaluated through the optimal weights structure for DMU_i, while the element θ_{jj} provides the efficiency of DMU_j calculated using its own optimal weights. If DMU_j is efficient (i.e. if $\theta_{jj} = 1$), although it exhibits a behavior specialized along a given dimension with respect to the other units, the efficiency values in the column corresponding to DMU_j will be less than 1.

Two quantities of interest can be derived from the efficiency matrix. The first represents the average efficiency of a unit with respect to the optimal weights systems for the different units, obtained as the average of the values in the jth column. The second is the average efficiency of a unit measured applying its optimal system of weights to the other units. The latter is obtained by averaging the values in the row associated with the unit being examined. The difference between the efficiency score θ_{jj} of DMU_j and the efficiency obtained as the average of the values in the jth column provides an indication of how much the unit relies on a system of weights conforming with the one used by the other units in the evaluation process. If the difference between the two terms is significant, DMU_j may have chosen a structure of weights that is not shared by the other DMUs in order to privilege the dimensions of analysis on which it appears particularly efficient.

Virtual inputs and virtual outputs

Virtual inputs and virtual outputs provide information on the relative importance that each unit attributes to each individual input and output, for the purpose of maximizing its own efficiency score. Thus, they allow the specific competencies of each unit to be identified, highlighting at the same time its weaknesses.

The *virtual inputs* of a DMU are defined as the product of the inputs used by the unit and the corresponding optimal weights. Similarly, *virtual outputs* are given by the product of the outputs of the unit and the associated optimal weights.

Inputs and outputs for which the unit shows high virtual scores provide an indication of the activities in which the unit being analyzed appears particularly efficient. Notice that model admits in general multiple optimal solutions, corresponding to which it is possible to obtain different combinations of virtual inputs and virtual outputs.

Two efficient units may yield high virtual values corresponding to different combinations of inputs and outputs, showing good operating practices in different contexts. In this case, it might be convenient for each unit to follow the principles and operating methods shown by the other, aiming at improving its own efficiency on a specific dimension.

Weight restrictions

To separate the units that are really efficient from those whose efficiency score largely depends on the selected weights system, we may impose some restrictions on the value of the weights to be associated with inputs and outputs. In general, these restrictions translate into the definition of maximum thresholds for the weight of specific outputs or minimum thresholds for the weight of specific inputs. Notice that, despite possible restrictions on the weights, the units still enjoy a certain flexibility in the choice of multiplicative factors for inputs and outputs. For this reason it may be useful to resort to the evaluation of virtual inputs and virtual outputs in order to identify the units with the most efficient operating practices with respect to the usage of a specific input resource or to the production of a given output.

UNIT – IV
BUSINESS INTELLIGENCE APPLICATIONS
PART –A

1. Define Marketing Intelligence?

The first theme is particularly broad and concerns the application of predictive models to support relational marketing strategies, whose purpose is to customize and strengthen the relationship between a company and its customers. After a brief introduction to relational marketing, we will describe the main streams of analysis that can be dealt with in this domain of application, indicating for each of them the classes of predictive models that are best suited to dealing with the problems considered. The subjects discussed in this context can be partly extended to the relationship between citizens and the public administration.

2. Define Salesforce Management?

The second theme concerns *salesforce management*. First, we will provide an overview of the major decision-making processes emerging in the organization of a sales staff, highlighting also the role played by response functions. Then, we will illustrate some optimization models which aim to allocate a set of geographical territories to sales agents as well as planning the activities of sales agents.

3. What are the aim of relational marketing?

The aim of a *relational marketing* strategy is to initiate, strengthen, intensify and preserve over time the relationships between a company and its stakeholders, represented primarily by its customers, and involves the analysis, planning, execution and evaluation of the activities carried out to pursue these objectives.

4. What are the Motivations and objectives of Marketing?

The reasons for the spread of relational marketing strategies are complex and interconnected. Some of them are listed below:

- The increasing concentration of companies in large enterprises and the resulting growth in the number of customers have led to greater complexity in the markets.
- The increased flow of information and the introduction of e-commerce have enabled global comparisons. Customers can use the Internet to compare features, prices and opinions on products and services offered by the various competitors.
- Customer loyalty has become more uncertain, primarily in the service industries, where often filling out an on-line form is all one has to do to change service provider.

5. Define Customer relationship Management?

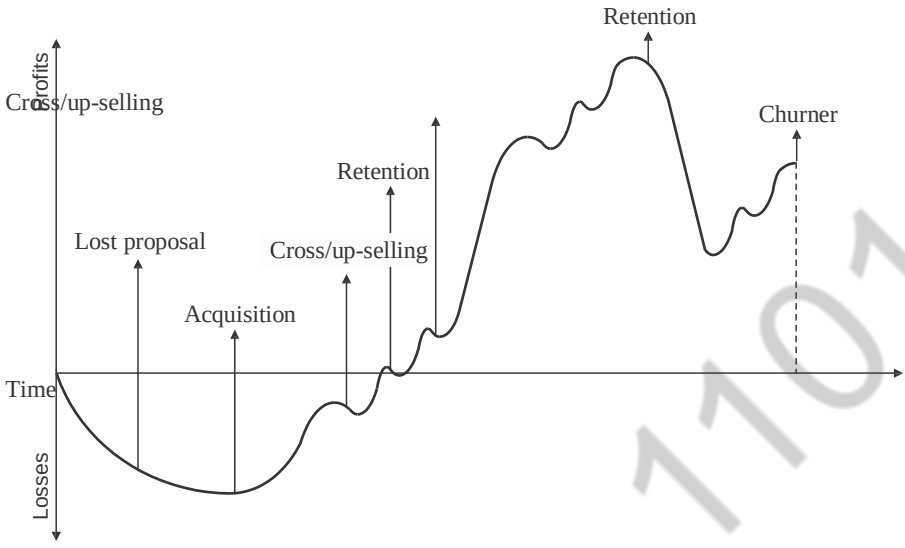
To obtain the desired advantages, a company should turn to relational marketing strategies by following a correct and careful approach. In particular, it is advisable to stress the distinction between a relational marketing vision and the software tools usually referred to as *customer relationship management* (CRM).

6. Define Relational Marketing?

Relational marketing is not merely a collection of software applications, but rather a coherent project where the various company departments are called upon to cooperate and integrate the managerial culture and human resources, with a high impact on the organizational structures.

7. What are the main stages of a customer lifetime?

The main stages during the customer *lifetime*, shows the cumulative value of a customer over time..
prospect,
acquisition
lost cross-selling
up-selling and
voluntary
forced
unintentional
cross-selling up-selling
Retention
Churner



Lifetime of a customer

8. What is Acquisition ?

Although retention plays a prominent role in relational marketing strategies, for many companies the *acquisition* of new customers also represents a crit- ical factor for growth. The acquisition process requires the identification of new prospects, as they are potential customers who may be totally or partially unaware of the products and services offered by the company, or did not possess in the past the characteristics to become customers, or were customers of competitors.

9. What is Retention?

The maturity stage reached by most products and services, and the subsequent saturation of their markets, have caused more severe competitive conditions.

There are also economic reasons for devoting substantial efforts to customer *retention*: indeed, it has been empirically observed that the cost of acquiring a new customer, or winning back a lost customer, is usually much higher – of the order of 5 to 9 times higher – than the cost of the marketing actions aimed at retaining customers considered at risk of churning. Furthermore, an action to win back a lost customer runs the risk of being too late and not achieving the desired result.

10.Define Cross-selling ?

The term *cross-selling* refers to the attempt to sell an additional product or service to an active customer, already involved in a long-lasting commercial relationship with the enterprise. By means of classification models, it is possible to identify the customers characterized by a high probability of accepting a cross-selling offer, starting from the information contained in the available attributes.

In other instances, it is possible to develop an *up-selling* initiative, by persuading a customer to purchase an higher-level product or service, richer in functions for the user and more profitable for the company, and therefore able to increase the lifetime value curve of a customer.

11.What is the purpose of relational marketing model?

The purpose of the model is to identify the most interesting segments, corresponding to customers who have taken up the gold service in the past, and who appear therefore more appreciative of the additional services offered by the gold card. The segments identified in this way represent the target of up-selling actions.

12. What is the need of Market basket analysis?

The purpose of *market basket analysis* is to gain insight from the purchases made by customers in order to extract useful knowledge to plan marketing actions. It is mostly used to analyze purchases in the retail industry and in e-commerce activities, and is generally amenable to unsupervised learning problems. It may also be applied in other domains to analyze the purchases made using credit cards, the complementary services activated by mobile or fixed telephone customers, the policies or the checking accounts acquired by a same household.

13. Define Basket?

. Each transaction consists of a list of purchased items. This list is called a *basket*, just like the baskets available at retail points of sale.

14. Define web mining?

The web is a critical channel for the communication and promotion of a company's image. Moreover, e-commerce sites are important sales channels. Hence, it is natural to use *web mining* methods in order to analyze data on the activities carried out by the visitors to a website.

15. What are the main purposes of web mining methods?

Web mining methods are mostly used for three main purposes, as shown in *content mining*, *structure mining* and *usage mining*.

16. Define Content mining.

Content mining involves the analysis of the content of web pages to extract useful information. Search engines primarily perform content mining activities to provide the links deemed interesting in relation to keywords supplied by users. Content mining methods can be traced back to data mining problems for the analysis of texts, both in free format or HTML and XML formats, images and multimedia content.

17. Define Structure mining.

The aim of this type of analysis is to explore and understand the topological structure of the web. Using the links presented in the various pages, it is possible to create graphs where the nodes correspond to the web pages and the oriented arcs are associated with links to other pages.

18. Define Usage mining.

Analyses aimed at *usage mining* are certainly the most relevant from a relational marketing standpoint, since they explore the paths followed by navigators and their behaviors during a visit to a company website. Methods for the extraction of association rules are useful in obtaining correlations between the different pages visited during a session.

19. Define Salesforce m anagement?

Most companies have a sales network and therefore rely on a substantial number of people employed in sales activities, who play a critical role in the profitability of the enterprise and in the implementation of a relational marketing strategy. The term *salesforce* is generally taken to mean the whole set of people and roles that are involved, with different tasks and responsibilities, in the sales process.

20. What are the advantages of Decision processes in salesforce management?

The design and management of a salesforce raise several decision-making problems,. When successfully solved, they confer multiple advantages: maximization of profitability, increased effectiveness of sales actions, increased efficiency in the use of resources, and greater profes- sional rewards for sales agents.

21. Write the various categories of sales force management?

In particular, the decision-making processes relative to salesforce manage- ment can be grouped into three categories: *design*, *planning* and *assessment*.

22. What is the purpose of Assessment in salesforce management?

The purpose of *assessment* and *control* activities is to measure the effectiveness and efficiency of individuals employed in the sales network, in order to design appropriate remuneration and incentive schemes. To measure the efficiency of the sales agents it is necessary to define adequate evaluation criteria that take into account the actual personal contribution of each agent, having removed effects due to the characteristics that may make an area or product more or less advantageous than others.

23. What are Response functions?

Response functions play a key role in the formulation of models for designing and planning a sales network. In general terms, a response function describes the elasticity of sales in terms of the intensity of the sales actions, and is a formal method to describe the complex relationship existing between sales actions and market reactions.

24. What is Sales territory design ?

Sales territory design involves allocating sales coverage units to individual agents so as to minimize a weighted sum of two terms, representing respectively the total distance between areas belonging to the same territory and the imbalance of sales opportunities for the agents.

25.What are the categories of Analysis and results?

The data available for the acquisition analysis are subdivided into three categories.

Prospects and customers.
Vehicles.

Works.

26.What is an effective Supply chain Management?

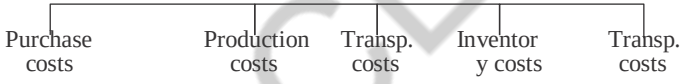
An effective *supply chain* management, understood as the logistic and production processes of a single enterprise as well as the network of companies composing the production chain of a given industry.

27.Define Supply Chain?

A *supply chain* may be defined as a network of connected and interdependent organizational units that operate in a coordinated way to manage, control and improve the flow of materials and information originating from the suppliers and reaching the end customers, after going through the procurement, processing and distribution subsystems of a company.

The aim of the integrated planning and operations of the supply chain is to combine and evaluate from a systemic perspective the decisions made and the actions undertaken within the various sub processes that compose the logistic system of a company.

The major purpose of an integrated logistic process is to minimize a function expressing the total cost, which comprises processing costs, transportation costs for procurement and distribution, inventory costs and equipment costs.



28.Define backlogging?

The feature that needs to be modeled in some logistic systems is *backlog- ging*. The term *backlog* refers to the possibility that a portion of the demand due in a given period may be satisfied in a subsequent period, incurring an addi- tional penalty cost. Backlogs are a feature of production systems more likely to occur in B2B or make-to-order manufacturing contexts.

29. Define Revenue Management System?

Revenue management is a managerial policy whose purpose is to maximize profits through an optimal balance between demand and supply. It is mainly intended for marketing as well as logistic activities and has found growing interest in the service industry, particularly in the air transportation, tourism and hotel sectors.

PART - B

1. Explain Relational Marketing in detail?

In order to fully understand the reasons why enterprises develop relational marketing initiatives, consider the following three examples: an insurance company that wishes to select the most promising market segment to target for a new type of policy; a mobile phone provider that wishes to identify those customers with the highest probability of churning, that is, of discontinuing their service and taking out a new contract with a competitor, in order to develop targeted retention initiatives; a bank issuing credit cards that needs to identify a group of customers to whom a new savings management service should be offered. These situations share some common features: a company owning a massive database which describes the purchasing behavior of its customers and the way they make use of services, wishes to extract from these data useful and accurate knowledge so as to develop targeted and effective marketing campaigns.

The aim of a *relational marketing* strategy is to initiate, strengthen, intensify and preserve over time the relationships between a company and its stakeholders, represented primarily by its customers, and involves the analysis, planning, execution and evaluation of the activities carried out to pursue these objectives.

Relational marketing became popular during the late 1990s as an approach to increasing customer satisfaction in order to achieve a sustainable competitive advantage. So far, most enterprises have taken at least the first steps in this direction, through a process of cultural change which directs greater attention toward customers, considering them as a formidable asset and one of the main sources of competitive advantage. A relational marketing approach has been followed in a first stage by service companies in the financial and telecommunications industries, and has later influenced industries such as consumer goods, finally reaching also manufacturing companies, from automotive and commercial vehicles to agricultural equipments, traditionally more prone to a vision characterized by the centrality of products with respect to customers.

Motivations and objectives

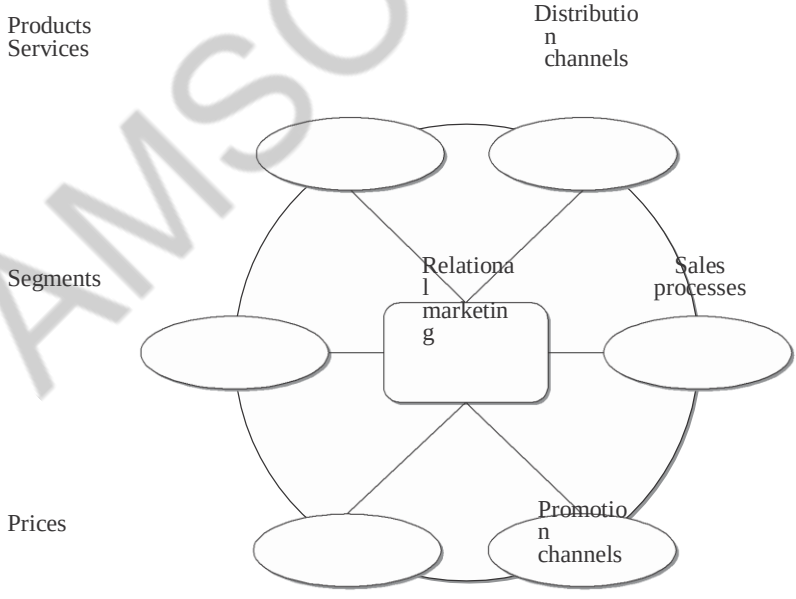
The reasons for the spread of relational marketing strategies are complex and interconnected. Some of them are listed below:

- The increasing concentration of companies in large enterprises and the resulting growth in the number of customers have led to greater complexity in the markets.
- Since the 1980s, the innovation – production – obsolescence cycle has progressively shortened, causing a growth in the number of customized

options on the part of customers, and an acceleration of marketing activities by enterprises.

- The increased flow of information and the introduction of e-commerce have enabled global comparisons. Customers can use the Internet to compare features, prices and opinions on products and services offered by the various competitors.
- Customer loyalty has become more uncertain, primarily in the service industries, where often filling out an on-line form is all one has to do to change service provider.
- In many industries a progressive commoditization of products and ser- vices is taking place, since their quality is perceived by consumers as equivalent, so that differentiation is mainly due to levels of service.
- The systematic gathering of sales transactions, largely automated in most businesses, has made available large amounts of data that can be transformed into knowledge and then into effective and targeted marketing actions.
- The number of competitors using advanced techniques for the analysis of marketing data has increased.

Relational marketing strategies revolve around the choices which can be effectively summarized as formulating for each segment, ideally



Decision-making options for a relational marketing strategy

Components of a relational marketing strategy

for each customer, the appropriate offer through the most suitable channel, at the right time and at the best price.

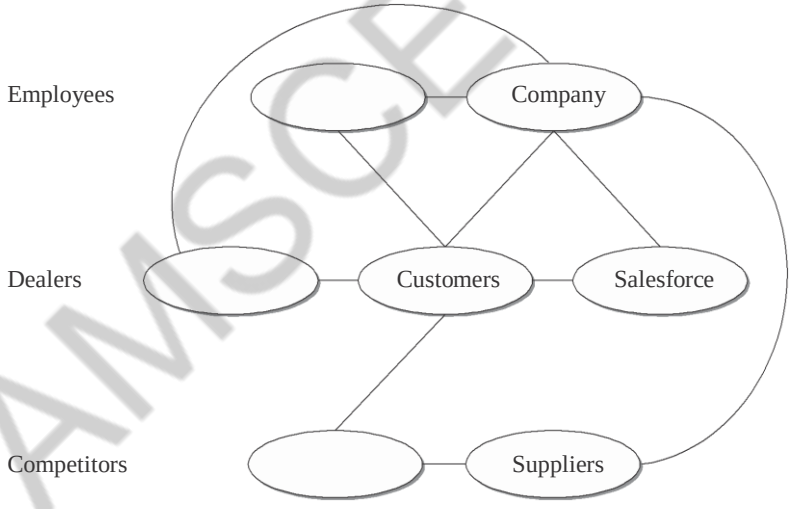
The ability to effectively exploit the information gathered on customers' behavior represents today a powerful competitive weapon for an enterprise. A company capable of gathering, storing, analyzing and understanding the huge amount of data on its customers can base its marketing actions on the knowledge extracted and achieve sustainable competitive advantages. Enterprises may profitably adopt relational marketing strategies to transform occasional contacts with their customers into highly customized long-term relationships. In this way, it is possible to achieve increased customer satisfaction and at the same time increased profits for the company, attaining a win–win relationship.

To obtain the desired advantages, a company should turn to relational marketing strategies by following a correct and careful approach. In particular, it is advisable to stress the distinction between a relational marketing vision and the software tools usually referred to as *customer relationship management* (CRM). Relational marketing is not merely a collection of software applications, but rather a coherent project where the various company departments are called upon to cooperate and integrate the managerial culture and human resources, with a high impact on the organizational structures. It is then necessary to create within a company a true *data culture*, with the awareness that customer-related information should be enhanced through the adoption of business intelligence and data mining analytical tools.

Based on the investigation of cases of excellence, it can be said that a successful relational marketing strategy can be achieved through the development of a company-wide vision that puts customers at the center of the whole organization. Of course, this goal cannot be attained by exclusively relying on innovative computer technologies, which at most can be considered a relevant enabling factor.

The overlap between relational marketing strategies and CRM software led to a misunderstanding with several negative consequences. On one hand, the notion that substantial investments in CRM software applications were in themselves sufficient to generate a relational marketing strategy represents a dangerous simplification, which caused many project failures. On the other hand, the high cost of software applications has led many to believe that a viable approach to relational marketing was only possible for large companies in the service industries. This is a deceitful misconception: as a matter of fact, the essential components of relational marketing are a well-designed and correctly fed marketing data mart, a collection of business intelligence and data mining analytical tools, and, most of all, and the cultural education of the decision makers. These tools will enable companies to carry out the required analyses and translate the knowledge acquired into targeted marketing actions.

The relationship system of an enterprise is not limited to the dyadic relation- ship with its customers, represented by individuals and companies that purchase the products and services offered, but also includes other actors, such as the employees, the suppliers and the sales network.



Network of relationships involved in a relational marketing strategy

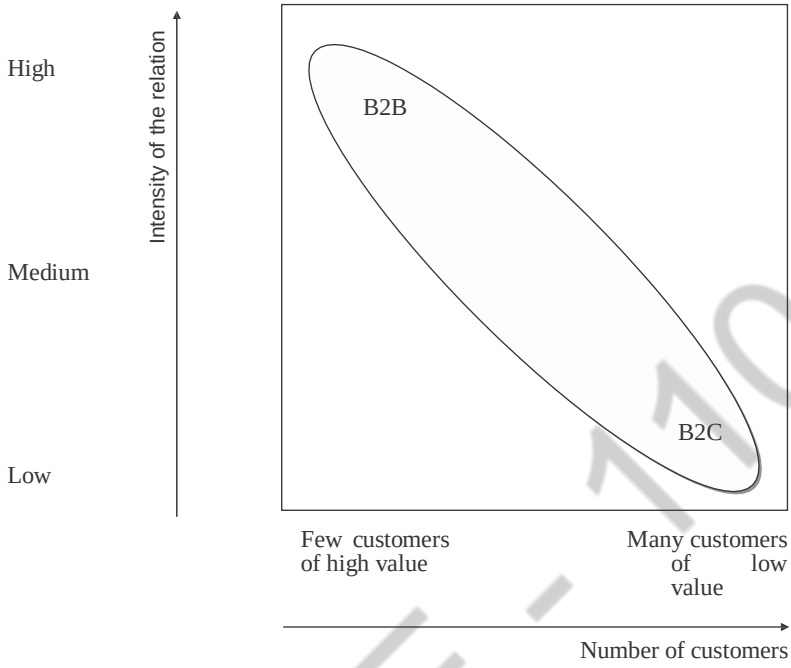
a mutually beneficial exchange occurs between the different subjects involved. More generally, we can widen the boundaries of relational marketing systems to include the stakeholders of an enterprise. The relationship between an enterprise and its customers is sometimes mediated by the sales network, which in some instances can partially obstruct the visibility of the end customers.

Let us take a look at a few examples to better understand the implications of this issue. The manufacturers of consumer goods, available at the points of sale of large and small retailers, do not have direct information on the consumers purchasing their products. The manufacturers of goods covered by guarantees, such as electrical appliances or motor vehicles, have access to personal information on purchasers, even if they rarely also have access to information on the contacts of and promotional actions carried out by the network of dealers. Likewise, a savings management company usually places shares in its invest ment funds through a network of intermediaries, such as banks or agents, and often knows only the personal data of the subscribers. A pharmaceutical enterprise producing prescription drugs usually ignores the identity of the patients that use its drugs and medicinal products, even though promotional activities to influence consumers are carried out in some countries where the law permits.

It is not always easy for a company to obtain information on its end customers from dealers in the sales network and even from their agents. These may be reluctant to share the wealth of information for fear, rightly or wrongly, of compromising their role. In a relational marketing project specific initiatives should be devised to overcome these cultural and organizational barriers, usually through incentives and training courses.

The number of customers and their characteristics strongly influence the nature and intensity of the relationship with an enterprise, The relationships that might actually be established in a specific economic domain tend to lie on the diagonal shown in the figure. At one extreme, there are highly intense relationships existing between the company and a small number of customers of high individual value. Relationships of this type occur more frequently in *business-to-business* (B2B) activities, although they can also be found in other domains, such as private banking. The high value of each customer justifies the use of dedicated resources, usually consisting of sales agents and key account managers, so as to maintain and strengthen these more intense relationships. In situations of this kind, careful organization and planning of the activities of sales agents is critical. Therefore, optimization models for *sales- force automation* (SFA).

At the opposite extreme of the diagonal are the relationships typical of consumer goods and *business-to-consumer* (B2C) activities, for which a high number of low-value customers get in contact with the company in an impersonal

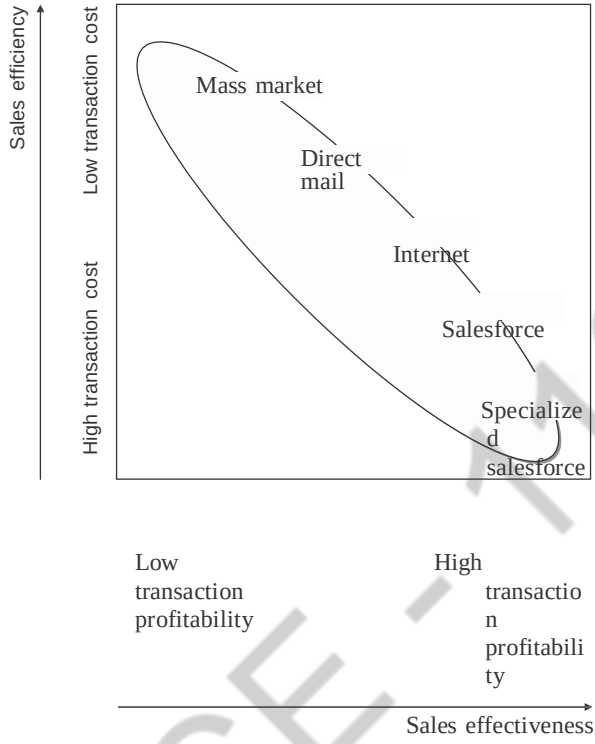


Intensity of customer relationships as a function of number of customers

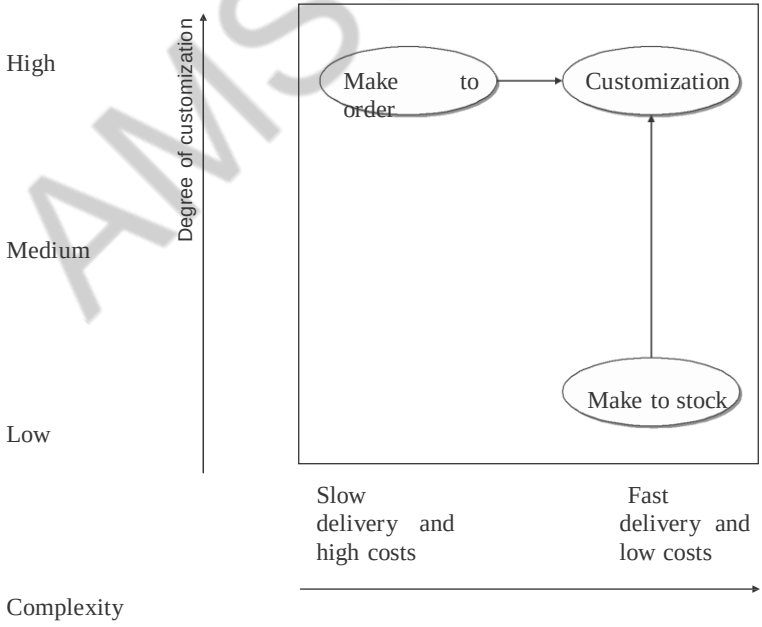
way, through websites, call centers and points of sale. Data mining analyses for segmentation and profiling are particularly valuable especially in this con- text, characterized by a large number of fragmented contacts and transactions. Relational marketing strategies, which are based on the knowledge extracted through data mining models, enable companies develop a targeted customization and differentiation of their products and/or services, including companies more prone toward a mass-market approach.

It contrasts the cost of sales actions and the corresponding revenues. Where transactions earn a low revenue per unit, it is necessary to implement low-cost actions, as in the case of mass-marketing activities. Moving down along the diagonal in the figure, more evolved and intense relationships with the customers can be found. The relationships at the end of the diagonal presuppose the action of a direct sales network and for the most part are typical of B2B relational contexts.

It shows the ideal path that a company should follow so as to be able to offer customized products and services at low cost and in a short time. On the one hand, companies operating in a mass market, well acquainted with fast delivery at low costs, must evolve in the direction of increased customization, by introducing more options and variants of products and services offered



Efficiency of sales actions as a function of their effectiveness



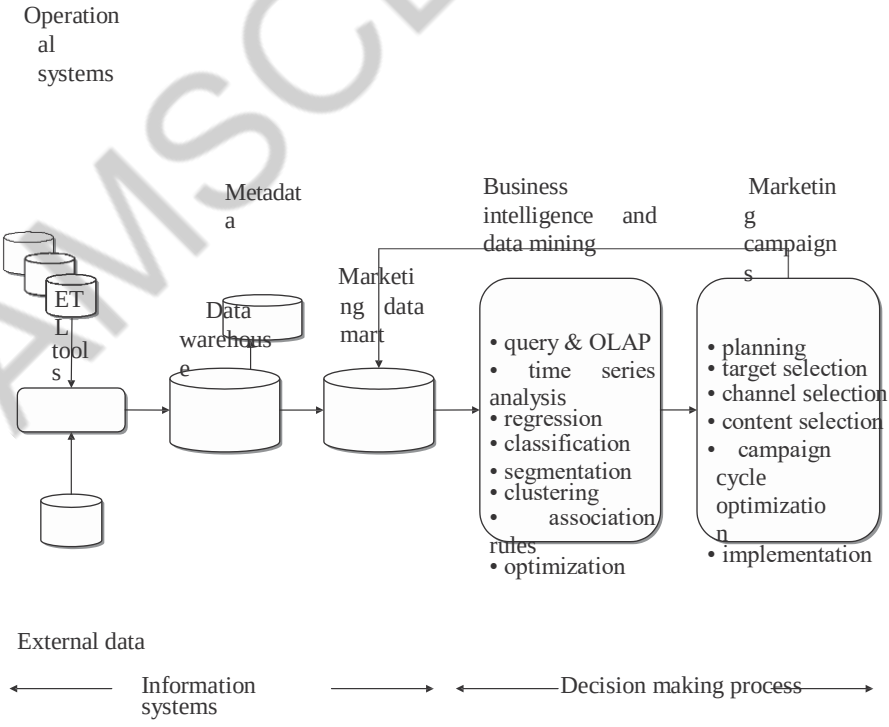
Level of customization as a function of complexity of products and services

to the various market segments. Data mining analyses for relational market- ing purposes are a powerful tool for identifying the segments to be targeted with customized products. On the other hand, the companies oriented toward make-to-order production must evolve in a direction that fosters reductions in both costs and delivery times, but without reducing the variety and the range of their products.

2.Explain about the environment for relational marketing analysis

shows the main elements that make up an environment for relational marketing analysis. Information infrastructures include the company's data warehouse, obtained from the integration of the various internal and external data sources, and a marketing data mart that feeds business intelligence and data mining analyses for profiling potential and actual customers. Using pattern recognition and machine learning models as described in previous chapters, it is possible to derive different segmentations of the customer base, which are then used to design targeted and optimized marketing actions. A classification model can be used, for example, to generate a *scoring* system for customers according to their propensity to buy a service offered by a company, and to direct a cross-selling offer only toward those customers for whom a high probability of acceptance is predicted by the model, thus maximizing the overall *redemption* of the marketing actions.

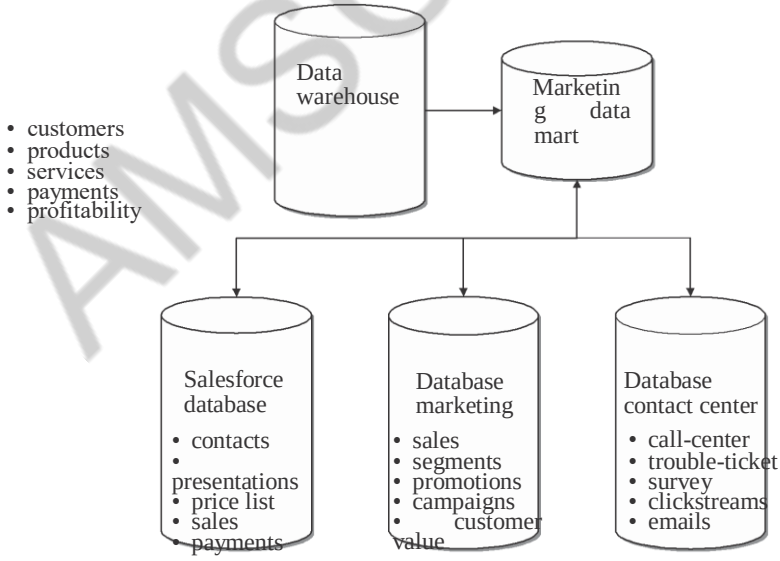
Effective management of frequent marketing campaign cycles is certainly a complex task that requires planning, for each segment of customers, the content of the actions and the communication channels, using the available



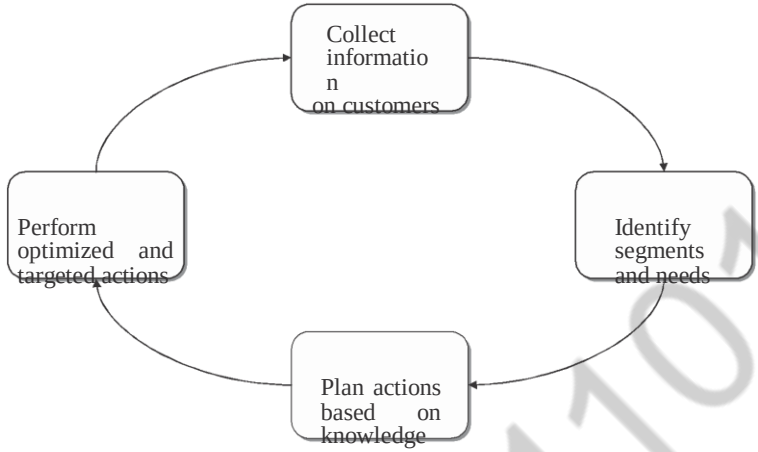
Components of an environment for relational marketing analysis

human and financial resources. The corresponding decision-making process can be formally expressed by appropriate optimization models. The cycle of marketing activities terminates with the execution of the planned campaign, with the subsequent gathering of information on the results and the redemption among the recipients. The data collected are then fed into the marketing data mart for use in future data mining analyses. During the execution of each campaign, it is important to set up procedures for controlling and analyzing the results obtained. In order to assess the overall effectiveness of a campaign, it would be advisable to select a *control group* of customers, with characteristics similar to those of the campaign recipients, toward whom no action should be undertaken.

The main types of data stored are discussed in a data mart for relational marketing analyses. A company data warehouse provides demographic and administrative information on each customer and the transactions carried out for purchasing products and using services. The marketing database contains data on initiatives carried out in the past, including previous campaigns and their results, promotions and advertising, and analyses of customer value. A further possible data source is the salesforce database, which provides information on established contacts, calls and applicable sales conditions. Finally, the *contact center* database provides access to data on customers' contacts with the call center, problems reported, sometimes called *trouble tickets*, and



Types of data feeding a data mart for relational marketing analysis



Cycle of relational marketing analysis

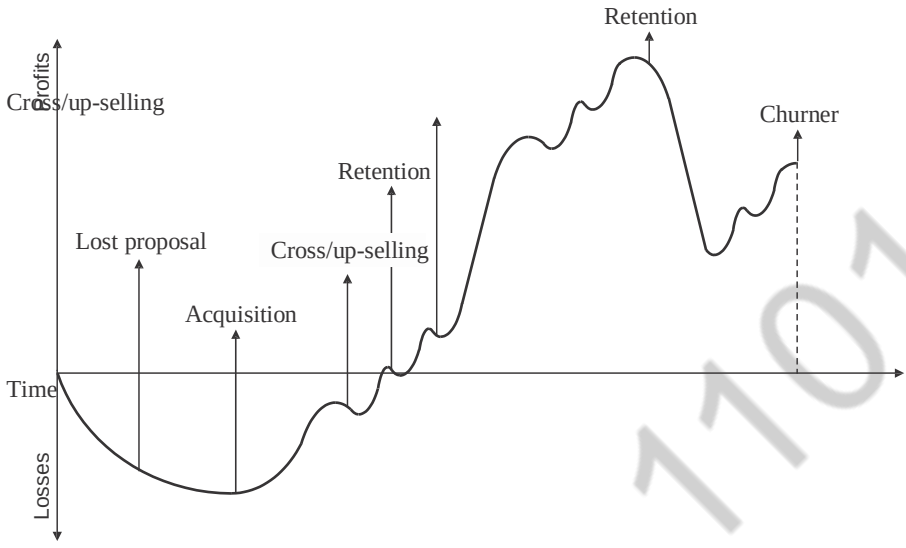
related outcomes, website navigation paths and forms filled out on-line, and emails exchanged between customers and the support center.

The available data are plentiful, providing an accurate representation of the behaviors and needs of the different customers, through the use of inductive learning models.

The first step is the exploration of the data available for each customer. At a later time, by using inductive learning models, it is possible to extract from those data the insights and the rules that allow market segments characterized by similar behaviors to be identified. Knowledge of customer profiles is used to design marketing actions which are then translated into promotional campaigns and generate in turn new information to be used in the course of subsequent analyses.

Lifetime value

The main stages during the customer *lifetime*, showing the cumulative value of a customer over time. The figure also illustrates the different actions that can be undertaken toward a customer by an enterprise. In the initial phase, an individual is a *prospect* , or potential customer, who has not yet begun to purchase the products or to use the services of the enterprise. Toward potential customers, *acquisition* actions are carried out, both directly (telephone contacts, emails, talks with sales agents) and indirectly (advertising, notices on the enterprise website). These actions incur a cost that can be assigned to each customer and determine an accumulated loss that lasts until a critical event



Lifetime of a customer

in the relationship with a customer occurs: a prospect becomes a customer. This event may take various forms in different situations: it may consist of a service subscription, the opening of a bank account, the first purchase at a retailer point of sale with the activation of a loyalty card. Before becoming a new customer, a prospect may receive from the enterprise repeated proposals aiming acquiring her custom, shown in the figure as *lost proposals* , which have a negative outcome. From the time of acquisition, each customer generates revenue, which produces a progressive rise along the curve of losses and cumulated profits. This phase, which corresponds to the maturity of the relationship with the enterprise, usually entails alternating *cross-selling* , *up-selling* and *retention* actions, in an effort to extend the duration and the profitability of the relationship so as to maximize the lifetime value of each customer. The last event in a customer lifetime is the interruption of the relationship. This may be *voluntary* , when a customer discontinues the services of an enterprise and switches to those of a competitor, *forced* , when for instance a customer does not comply with payment terms, or *unintentional* , when for example a customer changes her place of residence.

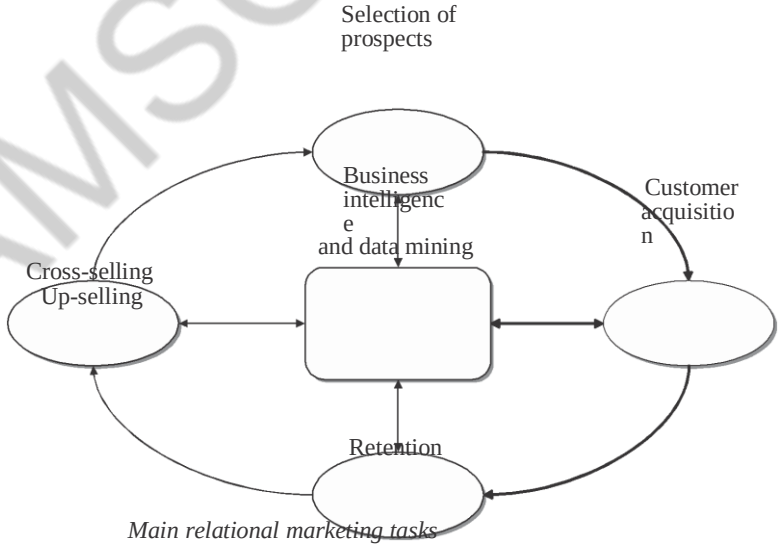
The progress of a customer lifetime highlights the main tasks of relational marketing. First, the purpose is to increase the ability to acquire new customers. Through the analysis of the available information for those customers who in the past have purchased products or services, such as personal socio-demographic characteristics, purchased products, usage of services,

previous contacts, and the comparison with the characteristics of those who have not taken up the offers of the enterprise, it is possible to identify the segments with the highest potential. This in turn allows the enterprise to optimize marketing campaigns, to increase the effectiveness of acquisition initiatives and to reduce the waste of resources due to offers addressed to unpromising market segments.

Furthermore, relational marketing strategies can improve the loyalty of customers, extending the duration of their relationship with the enterprise, and thus increasing the profitability. In this case, too, the comparative analysis of the characteristics of those who have remained loyal over time with respect to those who have switched to a competitor leads to predictions of the likelihood of churning for each customer. Retention actions can therefore be directed to the most relevant segments, represented by high-value customers with the highest risk of churning.

Finally, relational marketing analyses can be used to identify customers who are more likely to take up the offer of additional services and products (*cross-selling*), or of alternative services and products of a higher level and with a greater profitability for the enterprise (*up-selling*).

The tasks of acquisition, retention, cross-selling and up-selling, shown in Figure 13.11, are at the heart of relational marketing strategies and their aim is to maximize the profitability of customers during their lifetime. These analysis tasks, which will be described in the next sections, are clearly amenable to classification problems with a binary target class. Notice that attribute selection plays a critical role in this context, since the number of available explanatory variables is usually quite large and it is advisable for learning models to use a



Main relational marketing tasks

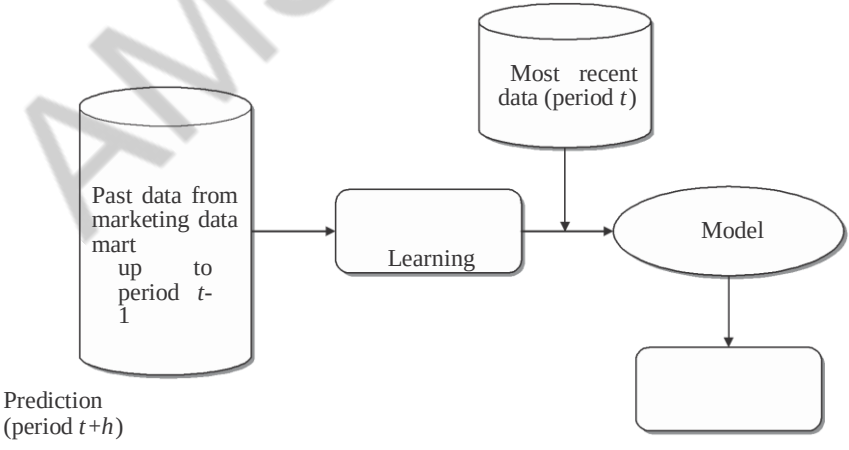
limited subset of predictive features, in order to generate meaningful and useful classification rules for the accurate segmentation of customers.

The effect of latency in predictive models

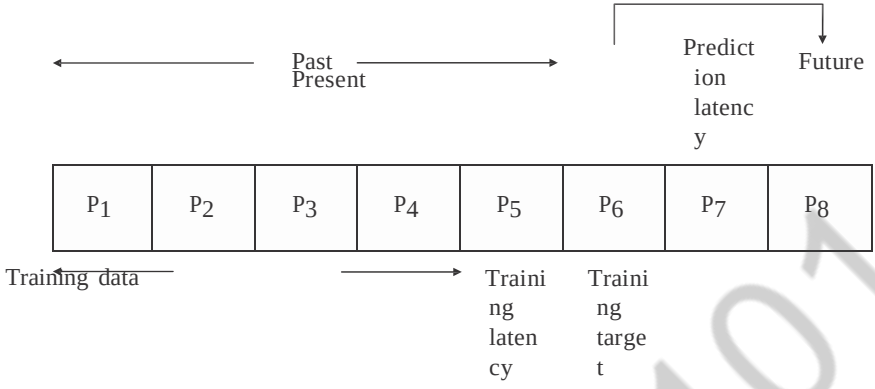
The logic of development of a classification model for a relational marketing analysis, also taking into account the temporal dimension. Assume that t is the current time period, and that we wish to derive an inductive learning model for a classification problem. For example, at the beginning of October a mobile telephone provider might want to develop a classification model to predict the probability of churning for its customers. The data mart contains the data for past periods, updated as far as period $t-1$. In our example, it contains data up to and including September.

Furthermore, suppose that the company wishes to predict the probability of churning h months in advance, since in this way any retention action has a better chance of success. In our example, we wish to predict at the beginning of October the probability of churning in November, using data up to September. Notice that the data for period t cannot be used for the prediction, since they are clearly not available at the beginning of period t.

To develop a classification model we use the value of the target variable for the last known period $t-1$, corresponding to the customers who churned in the month of September. It should be clear that for training and testing the model the explanatory variables for period $t-2$ should not be used, since in the training phase it is necessary to reproduce the same situation as will be faced when using the model in the prediction stage. Actually, the target



Development and application flowchart for a predictive model



Latency of a predictive model

variable must be predicted $h = 2$ periods in advance, and therefore there is an intermediate period of future data that are still unknown at time t (the month of October in our example). To reflect these dynamics, the training phase should be carried out without using August data. In general, the $h-1$ periods corresponding to data still unknown during the prediction phase, and not used during the training phase, are referred to as the model's *latency*

Acquisition

Although retention plays a prominent role in relational marketing strategies, for many companies the *acquisition* of new customers also represents a critical factor for growth. The acquisition process requires the identification of new prospects, as they are potential customers who may be totally or partially unaware of the products and services offered by the company, or did not possess in the past the characteristics to become customers, or were customers of competitors. It may also happen that some of the prospects were former customers who switched their custom to competitors, in which case much more information is usually available on them.

Once prospects have been identified, the enterprise should address acquisition campaigns to segments with a high potential profitability and a high probability of acquisition, in order to optimize the marketing resources. Traditional marketing techniques identify interesting segments using predefined pro- filing criteria, based on market polls and socio-demographic analyses, according to a top-down perspective. This approach can be successfully integrated or even replaced by a top-down segmentation logic which analyzes the data available in the data mart, as shown in (demographic information, contacts with prospects, use of products and services of competitors), and derives classification rules that characterize the most promising profiles for acquisition purposes. Also in this case, we are faced with a binary classification problem, which can be analyzed with the techniques are described.

Retention

The maturity stage reached by most products and services, and the subsequent saturation of their markets, have caused more severe competitive conditions. As a consequence, the expansion of the customer base of an enterprise consists more and more of switch mechanisms – the acquisition of customers at the expense of other companies. This phenomenon is particularly apparent in service industries, such as telecommunications, banking, savings management and insurance, although it also occurs in manufacturing, both for consumer goods and industrial products. For this reason, many companies invest significant amounts of resources in analyzing and characterizing the phenomenon of *attrition*, whereby customers switch from their company to a competitor.

There are also economic reasons for devoting substantial efforts to customer *retention*: indeed, it has been empirically observed that the cost of acquiring a new customer, or winning back a lost customer, is usually much higher – of the order of 5 to 9 times higher – than the cost of the marketing actions aimed at retaining customers considered at risk of churning. Furthermore, an action to win back a lost customer runs the risk of being too late and not achieving the desired result. In many instances, winning back a customer requires investments that do not generate a return.

One of the main difficulties in loyalty analysis is actually recognizing a churn. For subscription services there are unmistakable signals, such as a formal notice of withdrawal, while in other cases it is necessary to define adequate indicators that are correlated, a few periods in advance, with the actual churning. A customer who reduces by more than a given percentage her purchases at a selected point of sale using a loyalty card, or a customer who reduces below a given threshold the amount held in her checking account and the number of transactions, represent two examples of disaffection indicators. They also highlight the difficulties involved in correctly defining the appropriate threshold values.

To optimize the marketing resources addressed to retention, it is therefore necessary to target efforts only toward high-value customers considered at risk of churning. To obtain a scoring system corresponding to the probability of churning for each customer, it is necessary to derive a segmentation based on the data on past instances of churning. Predicting the risk of churning requires analysis of records of transactions for each customer and identifying the attributes that are most relevant to accurately explaining the target variable. Again, we are faced with a binary classification problem. Once the customers with the highest risk of churning have been identified, a retention action can be directed toward them. The more accurately such action is targeted, the cheaper it is likely to be.

Cross-selling and up-selling

Data mining models can also be used to support a relational marketing analysis aimed at identifying market segments with a higher propensity to purchase additional services or other products of a company. For example, a bank also offering insurance services may identify among its customers segments interested in purchasing a life insurance policy. In this case, demographic information on customers and their past transactions recorded in a data mart can be used as explanatory attributes to derive a classification model for predicting the target class, consisting in this example of a binary variable that indicates whether the customer accepted the offer or not.

The term *cross-selling* refers to the attempt to sell an additional product or service to an active customer, already involved in a long-lasting commercial relationship with the enterprise. By means of classification models, it is possible to identify the customers characterized by a high probability of accepting a cross-selling offer, starting from the information contained in the available attributes.

In other instances, it is possible to develop an *up-selling* initiative, by persuading a customer to purchase an higher-level product or service, richer in functions for the user and more profitable for the company, and therefore able to increase the lifetime value curve of a customer. For example, a bank issuing credit cards may offer customers holding a standard card an upgrade to a gold card, which is more profitable for the company, but also able to offer a series of complementary services and advantages to interested customers. In this case too, we are dealing with a binary classification problem, which requires construction of a model based on the training data of customers' demographic and operational attributes. The purpose of the model is to identify the most interesting segments, corresponding to customers who have taken up the gold service in the past, and who appear therefore more appreciative of the additional services offered by the gold card. The segments identified in this way represent the target of up-selling actions.

Market basket analysis

The purpose of *market basket analysis* is to gain insight from the purchases made by customers in order to extract useful knowledge to plan marketing actions. It is mostly used to analyze purchases in the retail industry and in e-commerce activities, and is generally amenable to unsupervised learning problems. It may also be applied in other domains to analyze the purchases made using credit cards, the complementary services activated by mobile or fixed telephone customers, the policies or the checking accounts acquired by a same household.

The data used for this purpose mostly refer to purchase transactions, and can be associated with the time dimension if the purchaser can be tracked through a loyalty card or the issue of an invoice. Each transaction consists of a list of purchased items. This list is called a *basket*, just like the baskets available at retail points of sale.

If transactions cannot be connected to one another, say because the purchaser is unknown, one may then apply association rules,

to extract interesting correlations between the purchases of groups of items. The rules extracted in this way can then be used to support different decision-making processes, such as assigning the location of the items on the shelves, determining the layout of a point of sale, identifying which items should be included in promotional flyers, advertisements or coupons distributed to customers.

Clustering models are also useful in determining homogeneous groups of items, once an incidence matrix X has been created for the representation of the dataset, where the rows correspond to the transactions and the columns to the items.

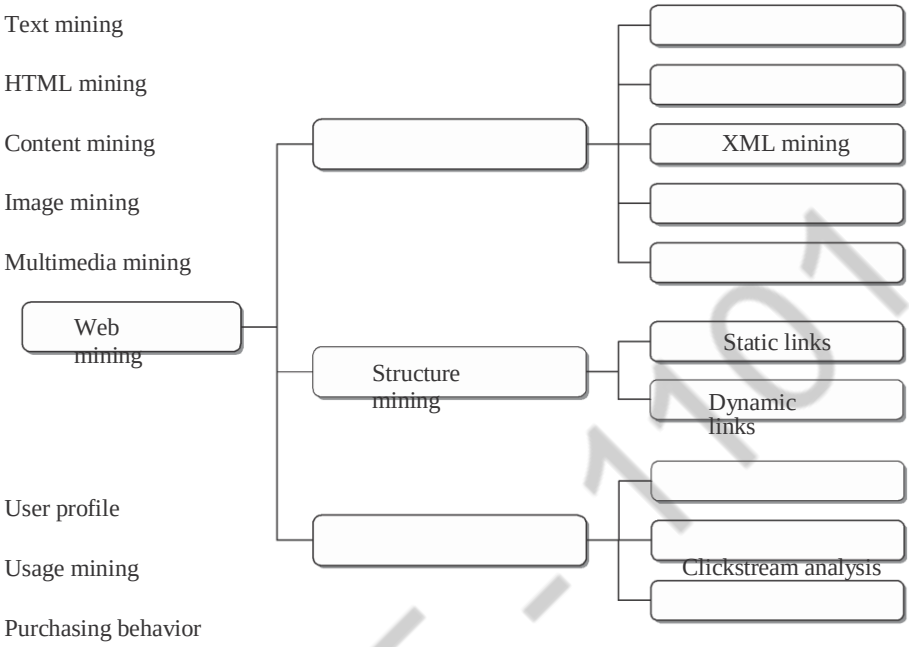
If customers are individually identified and traced, besides the above techniques it is also possible to develop further analyses that take into account the time dimension of the purchases.

3.What is Web mining? Explain in detail about various web mining methods?

The web is a critical channel for the communication and promotion of a com pany's image. Moreover, e-commerce sites are important sales channels. Hence, it is natural to use *web mining* methods in order to analyze data on the activities carried out by the visitors to a website.

Web mining methods are mostly used for three main purposes are
: *content mining*, *structure mining* and *usage mining*.

Content mining. Content mining involves the analysis of the content of web pages to extract useful information. Search engines primarily perform content mining activities to provide the links deemed interesting in relation to keywords supplied by users. Content mining methods can be traced back to data mining problems for the analysis of texts, both in free format or HTML and XML formats, images and multimedia content. Each of these problems is in turn dealt with using the learning models described in previous chapters. For example, text mining analyses are usually handled as multicategory classification problems, where the target variable is the subject category to which the text refers, while explanatory variables correspond to the meaningful words contained in the text.



Taxonomy of web mining analyses

Once it has been converted into a classification problem, text mining can be approached using the methods described. Text mining techniques are also useful for analyzing the emails received by a support center. Notice that the input data for content mining analyses are easily retrievable, at least in principle, since they consist of all the pages that can be visited on the Internet.

Structure mining. The aim of this type of analysis is to explore and understand the topological structure of the web. Using the links presented in the various pages, it is possible to create graphs where the nodes correspond to the web pages and the oriented arcs are associated with links to other pages. Results and algorithms from graph theory are used to characterize the structure of the web, that is, to identify areas with a higher density of connections, areas disconnected from others and maximal *cliques*, which are groups of pages with reciprocal links. In this way, it is possible to pinpoint the most popular sites, or to measure the distance between two sites, expressed in terms of the lowest number of arcs along the paths that connect them in the links graph. Besides analyses aimed at exploring the *global* structure of the web, it is also possible to carry out *local* investigations to study how a single website is articulated. In some investigations, the local structure of websites is associated with the

time spent by the users on each page, to verify if the organization of the site suffers from inconsistencies that jeopardize its effectiveness. For example, a page whose purpose is to direct navigation on the site should be viewed by each user only briefly. Should this not be the case, the page has a problem due to a possible ambiguity in the articulation of the links offered.

Usage mining. Analyses aimed at *usage mining* are certainly the most relevant from a relational marketing standpoint, since they explore the paths followed by navigators and their behaviors during a visit to a company website. Methods for the extraction of association rules are useful in obtaining correlations between the different pages visited during a session. In some instances, it is possible to identify a visitor and recognize her during subsequent sessions. This happens if an identification key is required to access a web page, or if a cookie-enabling mechanism is used to keep track of the sequence of visits. Sequential association rules or time series models can be used to analyze the data on the use of a site according to a temporal dynamic. Usage mining analysis is mostly concerned with *clickstreams* – the sequences of pages visited during a given session. For e-commerce sites, information on the purchase behavior of a visitor is also available.

4.Explain Salesforce management in detail?

Most companies have a sales network and therefore rely on a substantial number of people employed in sales activities, who play a critical role in the profitability of the enterprise and in the implementation of a relational marketing strategy. The term *salesforce* is generally taken to mean the whole set of people and roles that are involved, with different tasks and responsibilities, in the sales process. A preliminary taxonomy of salesforces is based on the type of activity carried out.

Residential. *Residential* sales activities take place at one or more sites managed by a company supplying some products or services, where customers go to make their purchases. This category includes sales at retail outlets as well as wholesale trading centers and *cash-and-carry* shops.

Mobile. In *mobile* sales, agents of the supplying company go to the customers' homes or offices to promote their products and services and collect orders. Sales in this category occur mostly within B2B relationships, even though they can also be found in B2C contexts.

Telephone. *Telephone* sales are carried out through a series of contacts by tele- phone with prospective customers.

There are various problems connected with managing a mobile sales force management, which will be the main focus of this section. They can be subdivided into a few main categories:

- designing the sales network;
- planning the agents' activities;
- contact management;
- sales opportunity management;
- customer management;
- activity management;
- order management;
- area and territory management;
- support for the configuration of products and services;
- knowledge management with regard to products and services. Designing the sales

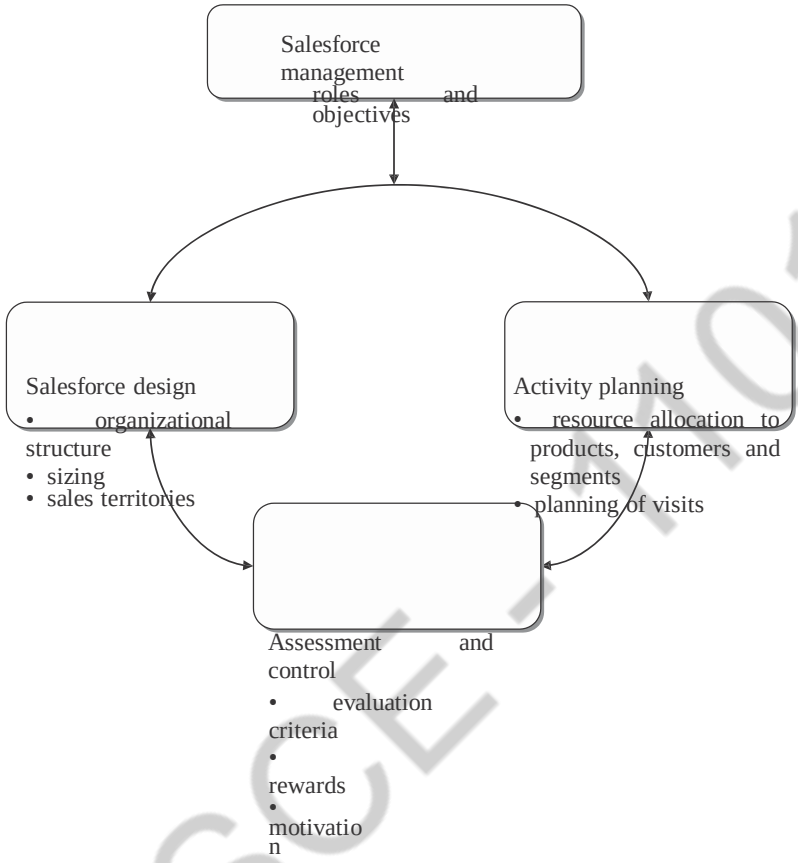
network and planning the agents' activities involve decision-
making tasks that may take advantage of the use of optimization models. The remaining activities are operational in nature and may benefit from the use of software tools for *salesforce automation* (SFA), today widely implemented.

5.Explain about Decision processes in salesforce management?

The design and management of a salesforce raise several decision-making problems. When successfully solved, they confer multiple advantages: maximization of profitability, increased effectiveness of sales actions, increased efficiency in the use of resources, and greater professional rewards for sales agents.

The decision processes should take into account the strategic objectives of the company, with respect to other components of the marketing mix, and conform to the role assigned to the salesforce within the broader framework of a relational marketing strategy. The two-way connections indicated in the figure suggest that the different components of the decision-making process interact with each other and with the general objectives of the marketing department.

In particular, the decision-making processes relative to salesforce manage- ment can be grouped into three categories: *design*, *planning* and *assessment*.



Decision processes in salesforce management

6.Discuss about salesforce design?

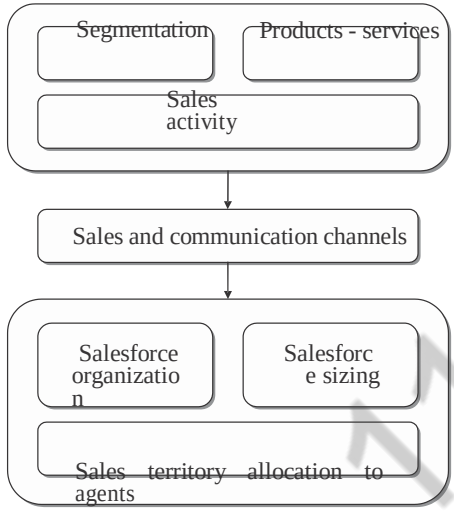
Salesforce design is dealt with during the start-up phase of a commercial activity or during subsequent restructuring phases, for example following the merger or acquisition of a group of companies.

The design phase is usually preceded by the creation of market segments through the application of data mining methods and by the articulation of the offer of products and services, which are in turn subdivided into homogeneous classes. Salesforce design includes three types of decisions.

Organizational structure. The *organizational structure* may take different forms, corresponding to hierarchical agglomerations of the agents by group of prod- ucts, brand or geographical area. In some situations the structure may also be differentiated by markets. In order to determine the organizational structure, it is necessary to analyze the complexity of customers, products and sales activities, and to decide whether and to what extent the agents should be specialized.

Sizing. Sales network *sizing* is a matter of working out the optimal number of agents that

should operate within the selected structure, and depends on several factors, such as the number of customers and prospects, the desired level of



Salesforce design process

sales area coverage, the estimated time for each call and the agents' traveling time. One should bear in mind that a reduction in costs due to a decrease in the sales force size is often followed by a reduction in sales and revenues. A better allocation of the existing sales force, devised during the planning phase by means of optimization models, is usually more effective than a variation in size.

Sales territories. Designing a *sales territory* means grouping together the geo- graphical areas into which a given region has been divided and assigning each territory to an agent. The design and assignment of sales territories should take into account several factors, such as the sales potential of each geographical area, the time required to travel from one area to another and the total time each agent has available. The purpose of the assignment consists of determining a balanced situation between sales opportunities embedded in each territory, in order to avoid disparities among agents. The assignment of the geographical areas should be periodically reviewed since the sales potential balance in the various territories tends to vary over time.

Decisions concerning the design of the sales force should take into account decisions about sales force planning, and this explains the two-way link between the two corresponding blocks.

Planning

Decision-making processes for *planning* purposes involve the assignment of sales resources, structured and sized during the design phase, to market entities.

Resources may correspond to the work time of agents or to the budget, while market entities consist of products, market segments, distribution channels and customers.

Allocation takes into account the time spent pitching the sale to each customer, the travel time and cost, and the effectiveness of the action for each product, service or market segment. It is also possible to consider further ancillary activities carried out at the customers' sites, such as making suggestions that are conducive to future sales or explaining the technical and functional features of products and services.

Sales force planning can greatly benefit from the use of optimization models, as explained below.

Assessment

The purpose of *assessment* and *control* activities is to measure the effectiveness and efficiency of individuals employed in the sales network, in order to design appropriate remuneration and incentive schemes. To measure the efficiency of the sales agents it is necessary to define adequate evaluation criteria that take into account the actual personal contribution of each agent, having removed effects due to the characteristics that may make an area or product more or less advantageous than others. Data envelopment analysis, described in Chapter 15, provides useful models that can be applied to assess agents' performance.

7.What are the Models for sales force management?

In what follows we will describe some classes of optimization models for designing and planning the sales force. These models are primarily intended for educational purposes, to familiarize readers with the reasoning behind specific aspects of a sales network, through the formulation of optimization models. For the sake of clarity and conciseness, for each model we have limited the extensions to a single feature. Sales networks simultaneously possess more than one of the distinctive features previously described, and therefore the models developed in real-world applications, just like those described in the last section of the chapter, are more complex and result from a combination of different characteristics.

Before proceeding, it is useful to introduce some notions common to the different models that will be described. Assume that a region is divided into J geographical sales areas, also called *sales coverage units*, and let $J = \{1, 2, \ldots, J\}$. Areas must be aggregated into disjoint clusters, called *territories*, so that each area belongs to one single territory and is also connected to all the areas belonging to the same territory. The connection property implies that from each area it is possible to reach any other area of the same territory. The

time span is divided into T intervals of equal length, which usually correspond to weeks or months, indicated by the index $t \in T = \{1, 2, \ldots, T\}$.

Each territory is associated with a sales agent, located in one of the areas belonging to the territory, henceforth considered as her area of residence. The choice of the area of residence determines the time and cost of traveling to any other area in the same territory. Let I be the number of territories and therefore

the number of agents that form the sales network, and let $\boldsymbol{I} = \{1, 2, \ldots, \boldsymbol{I}\}$.

In each area there are customers or prospects who can be visited by the agents as part of their promotions and sales activities. In some of the models that will be presented, customers or prospects are aggregated into segments, which are considered homogeneous with respect to the area of residence and possibly to other characteristics, such as value, potential for development and purchasing behaviors. Let H be the number of market entities, which in different models

may represent either single customers or segments, and let $H = \{1, 2, \ldots, H\}$.

Let $D_{\boldsymbol{j}}$ be the set of customers, or segments of customers where necessary, located in area $\boldsymbol{j}$.

Finally, assume that a given agent can promote and sell K products and services during the calls she makes on customers or prospects, and let $K =$

$\{1, 2, \ldots, K\}$.

8.Explain about Response functions?

Response functions play a key role in the formulation of models for designing and planning a sales network. In general terms, a response function describes the elasticity of sales in terms of the intensity of the sales actions, and is a formal method to describe the complex relationship existing between sales actions and market reactions.

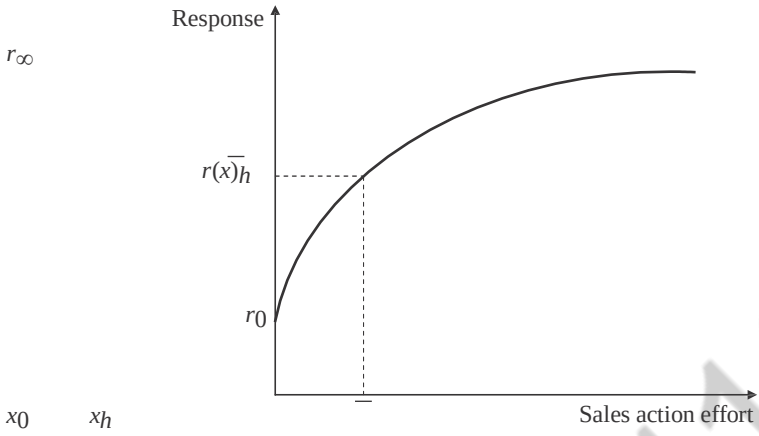
Sales to which the response function refers are expressed in product units

or monetary units, such as revenues or margins. For the sake of uniformity, in the next sections response functions are assumed to be expressed as sales revenues. The intensity of a sales action can be related to different variables, such as the number of calls to a customer in each period, the number of mentions of a product in each period, and the time dedicated to each customer in each period.

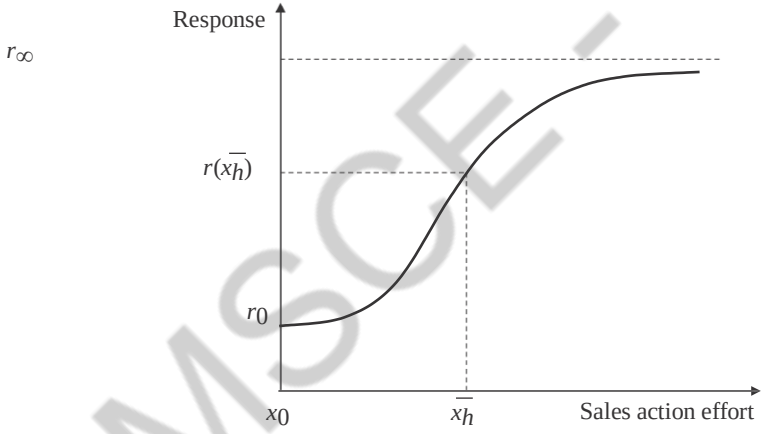
In principle, it is possible to consider a response function in relation to each factor that is deemed critical to sales: the characteristics of customers and sales territories; the experience, education and personal skills of the agents; promotions, prices, markdown policies operated by the company and the corresponding features for one or several competitors.

The two possible shapes of the response function,

obtained by placing the sales of a product or service on the vertical axis and the intensity of the sales action of interest on the horizontal axis. To fix ideas, we



A concave response function



A sigmoidal response function

will assume that the number x_h of calls that a specific agent makes to customer h in each period of the planning horizon is placed on the horizontal axis.

The concave response function shown can be interpreted in the following way: as the number of calls increases, revenues grow at a decreasing rate approaching 0, according to the principle of decreasing marginal revenues. In general, a lack of sales actions toward a given customer does not imply a lack of sales, at least for a certain number of periods. This is an effect of the actions executed in previous periods that lasts over time. For this reason the response function is greater than 0 at $x_h = 0$.

The sigmoidal response function reflects a different hypothesis of sales growth as a function of the actions carried out. The assumption made in this case is that the central interval of values on the horizontal axis corresponds to a higher rate of sales growth, while outside that area the growth rate is lower.

It is worth noting that each decision concerning the allocation of sales resources is based on a response function hypothesis, which is implicit and unaware in intuitive decision-making processes, while it is explicit and rigorous in mathematical models such as those presented below.

Response functions can be estimated by considering two types of informa- tion. On the one hand, one can use past available data regarding the intensity of the actions carried out and the corresponding sales, to develop a parametric regression model through variants of regression methods. On the other hand, interviews are carried out with agents and sales managers to obtain subjective information which is then incorporated into the procedure for calculating the response function.

We will now show by means of an example how the procedure for estimat- ing the response function works. Let $r_h(x_h)$ be the sales value for customer h associated with a number x_h of calls during a given period. More generally, the variable that determines the response function r expresses the intensity of the sales action that has been carried out. A parametric form should first be selected in order to express the functional dependence. The following function, which may assume both concave and sigmoidal shapes by varying the parameters, can be used:

$$r_h(x_h) = r_0 + (r_\infty - r_0)\,\gamma\,\frac{x_h^\sigma}{\mu^\sigma + x_h^\sigma}.$$

The parameters in the expression $r_h(x_h)$ have the following meaning: r_0 rep- resents the sales level that would be obtained at a sales action intensity equal to 0, as a prolonged effect of previous actions; r_∞ represents the maximum sales level, irrespective of the intensity of the sales action; γ and σ are two parameters to be estimated.

To obtain an estimate of the four parameters appearing in the expression for $r_h(x_h)$ it is possible to proceed in two complementary ways. Past sales data can be used to set up a regression model and determine the values through the least squares method. In order to increase the value of the opinions of the sales agents, it is also possible to ask agents and sales managers to estimate the value of the parameters r_0 and r_∞, as well as the values of the expected sales at other three critical points of the response function: $r(\bar{x}_h)$, corresponding to

associated respectively with increasing and decreasing the number of calls by 50% with respect to the current value. Based on a subjective evaluation of

the five response values derived through the procedure described above, an estimate by interpolation of the scale parameters γ and σ is then obtained.

9.Explain about Sales territory design?

Sales territory design involves allocating sales coverage units to individual agents so as to minimize a weighted sum of two terms, representing respectively the total distance between areas belonging to the same territory and the imbalance of sales opportunities for the agents.

Each region is subdivided into J geographical areas, which should then be clustered into I territories, whose total number has been determined beforehand. A sales agent will be associated with each territory, and she should be located in one of the sales coverage units, to be considered as her area of residence. It is further assumed that travel times within each area are negligible with respect to the corresponding travel times between a pair of distinct areas.

Each area will be identified by the geographical coordinates (e_j, f_j) of one of its points, considered as representative of the entire sales coverage unit. One might, for instance, choose the point whose coordinates are obtained as the average of the coordinates of all points belonging to that area. For each territory, let (e_i, f_i) denote the coordinates of the area where the agent associated with the territory resides. This area will be called *centroid* of territory i.

The parameters in the model are as follows: d_{ij} is the distance between centroid i and area j, given by

$$d_{ij} = \sqrt{(e_i - e_j)^2 + (f_i - f_j)^2}; \qquad (13.2)$$

a_j is the opportunity for sales in area j; and β is a relative weight factor between total distance and sales imbalance.

Consider a set of binary decision variables Y_{ij} defined as

$$Y_{ij} = \begin{cases} 1 & \text{if area } j \text{ is assigned to territory } i \\ 0 & \text{otherwise.} \end{cases}$$

Define I additional continuous variables that express the deviations from the average sales opportunity value for each territory:

S_i deviation from the average opportunity value $\left| \sum_{j \in J} \frac{1}{I} a_j \right|$ for territory i.

Hence, the corresponding optimization problem can be formulated as

$$\min \sum_{i \in I} \sum_{j \in J} a_j d_{ij}^2 Y_{ij} + \beta \sum_{i \in I} S_i ,$$

$$\text{s.to} \quad \sum_{j \in J} a_j Y_{ij} - \frac{1}{I} \sum_{j \in J} a_j \leq S_i , \qquad i \in I,$$

$$\sum_{j \in J} a_j Y_{ij} - \frac{1}{I} \sum_{j \in J} a_j \geq -S_i , \qquad i \in I,$$

$$\sum_{i \in I} Y_{ij} = 1, \qquad j \in J,$$

$$S_i \geq 0, \; Y_{ij} \in \{0, 1\}, \qquad i \in I, j \in J.$$

The purpose of constraints is to bound by means of variable S_i the absolute deviation between each territory sales opportunity and the average sales opportunity, to make the assignment to territories more uniform with respect to sales opportunities, hence balancing the sales chances across the agents. constraints represent a multiple choice condition imposed to guarantee that each sales coverage unit is exclusively assigned to one territory, and hence to one and only one agent.

Model is a mixed binary optimization problem, which can be solved by a branch-and-bound method, possibly truncated to limit the computing time and to achieve suboptimal solutions. Alternatively, an approximation algorithm can be devised for its *ad hoc* solution.

10.What is Calls and product presentations planning?

Optimization models for calls and product presentations planning are intended to derive for each agent the optimal sales activity plan.

Calls planning

The aim of the first model described is to identify the optimal number of calls to each customer or prospect (taken together as *market entities* in what follows) located in the territory assigned to a specific agent. The objective function expresses the difference between revenues and transfer costs.

The decision variables are defined as

X_h = number of calls to market entity h, W_j = number of trips to market area j,

while the parameters have the following meanings:

a_h = strategic relevance of market entity h, c_j = transfer cost to area j,
v_j = transfer time to area j,
t_h = time spent with market entity h in each call, l_h = minimum number of calls to market entity h, u_h = maximum number of calls to market entity h, b = total time available to the sales agent.

The corresponding optimization problem can be formulated as

$$\max_{h \in H} \quad a_h r(X_h) - \sum_{j \in J} c_j W_j,$$

$$\text{s.to} \quad t_h X_h + \sum_{j \in J} v_j W_j \leq b,$$

$$X_h \leq u_h, \quad X_h \geq l_h, \quad h \in H,$$

$$W_j \geq X_h, \quad j, h \in D,$$

$$X_h, W_j \geq 0 \text{ and integer}, \quad h \in H, j \in J.$$

Constraint expresses a bound on the total time available to the sales agent within the planning horizon. Constraints impose a lower and an upper bound, respectively, on the number of calls to each market entity. Finally, constraints establish a logical consistency condition between the decision variables X_h and W_j.

To obtain a solution, one may proceed as follows. First, the response function is approximated with a piecewise linear function, deriving a set of linear mixed integer optimization problems. These are then solved by using a branch-and-bound method, possibly truncated to limit the computing time and to achieve subopti- mal solutions. Alternatively, again, an approximation algorithm can be devised for *ad hoc* solution.

Product presentations planning

The aim of this model is to determine for each period in the planning horizon the optimal number of mentions for each product belonging to the sales portfolio of a given agent. Through an index called *relative exposure* the model also

incorporates the dynamic effects determined by the mentions of each product made in past periods.

The decision variables of the model are consequently defined as

X_{kt} = number of calls for product k in period t,

Z_{kt} = cumulated exposure level for product k in period t.

The parameters are

d_{kt} = number of units of product k available in period t, p = maximum number of mentions for each product,

λ = memoryless parameter.

The quantity $\sigma(X_{kt})$ expresses the relative exposure of product k as a function of the number of times k has been mentioned in period t. The relative exposure formalizes the relationship between the level of cumulative exposure and the number of mentions made in period t through constraints (13.15) in the subsequent optimization model. The response function then depends on the level of cumulated exposure.

The resulting optimization model is formulated as

$$\max \sum_{t \in T} \sum_{k \in K} d_{kt} r_k(Z_{kt}),$$

$$\text{s.to} \sum_{k \in K} X_{kt} \leq Kp, \qquad t \in T,$$

$$Z_{kt} = \lambda \sigma(X_{kt}) + (1-\lambda)\sigma(X_{kt}-1), \quad k \in K, t \in T,$$

$$X_{kt}, Z_{kt} \geq 0 \text{ and integer}, \qquad k \in K, t \in T.$$

Constraints impose a limitation on the maximum number of mentions that can be made in each period. Constraints express, through a recursive formula, the relationship between the cumulative exposure level and the number of mentions made in each period.

Model is also a nonlinear mixed integer optimization problem, whose solution can be obtained analogously to model

Calls and product presentations planning

The aim of the model described in this section is to determine the optimal number of calls to each market entity belonging to a given segment and, for each call, the number of mentions for each product in the sales portfolio. The aggregation of market entities into segments has the purpose of simplifying

the estimation of the response function, by limiting its evaluation only to the segments identified within the scope of the analysis.

The decision variables are defined as

X_{kh} = number of mentions of product k to a customer in segment h, W_h = number of calls to a customer in segment h.

The parameters are

p_h = maximum number of products mentioned in each call to a customer in segment h,

s_h = number of customers in segment h,

b = maximum number of calls that can be made by each agent.

The resulting optimization problem is formulated as

$$\max \sum_{k \in K} \sum_{h \in H} r_{kh}(X_{kh}),$$

$$\text{s.to} \quad \sum_{k \in K} X_{kh} \leq p_h W_h, \qquad h \in H,$$

$$X_{kh} \leq W_h, \qquad k \in K, h \in H,$$

$$\sum_{h \in H} s_h W_h \leq bI,$$

$$X_{kh}, W_h \geq 0 \text{ and integer}, \qquad k \in K, h \in H.$$

Constraints express the limitation on the total number of mentions made to each segment. Constraints represent a logical consistency condition between decision variables. Finally, constraint establishes an overall upper bound on the number of calls that sales agents can make.

Model is a nonlinear mixed integer optimization problem, to which remarks similar to those made regarding apply, in particular for its approximate solution.

A general model for sales resources planning

It is possible to provide a somewhat general formulation for salesforce planning problems by adopting a representation framework that involves listing, at least ideally, all tasks that can be assigned to each agent. The resulting model described in this section, like the one discussed in the previous section, derives the optimal plan for the sales agents across multiple time periods, taking into account different shared resources.

For each agent i and for each period t, the set of all possible sales actions, which represent the plan of calls and product presentations to different customers, is identified in advance and denoted by S_{it}. The required resources, denoted by the index $g \in G = \{1, 2, \ldots, G\}$, represent the overall budget avail- able to implement the sales actions, or other technical factors needed to adopt the different actions.

The required binary decision variables Y_{iut} are defined as

$$Y_{iut} = \begin{cases} 1 & \text{if action } u \in S_{it} \text{ is selected for agent } i \text{ in period } t, \\ 0 & \text{otherwise.} \end{cases}$$

The parameters are

S_{it} = set of feasible actions for agent i in period t,

w_{giut} = quantity of resource g required to implement action $u \in S_{it}$ by agent i in period t,

V_{gt} = quantity of resource g available in period t, v_{iut} = profit value associated with action $u \in S_{it}$.

The resulting optimization problem is formulated as

$$\max \sum_{t \in T} \sum_{i \in I} \sum_{u \in S_{it}} v_{iut} Y_{iut},$$

$$\text{s.to} \sum_{i \in I} \sum_{u \in S_{it}} w_{giut} Y_{iut} \leq V_{gt}, \qquad g \in G, t \in T,$$

$$\sum_{u \in S_{it}} Y_{iut} = 1, \qquad i \in I, t \in T,$$

$$Y_{iut} \in \{0, 1\}, \qquad i \in I, u \in S_{it}, t \in T.$$

Constraints express the upper limit on the amount available for each resource in each period. Constraints represent a multiple choice condition imposed to guarantee that each agent in each period will perform exactly one action.

Model is a binary optimization problem belonging to the class of generalized multiple choice knapsack problems, for whose solution remarks similar to those made for model apply. Notice that it is usually advisable to reduce in advance the number of available actions for each agent, by means of a preprocessing phase aimed at discarding those actions that are regarded as less convenient or less profitable.

11.Explain about Business case studies in marketing models?

In this section we will briefly describe some business case studies that illustrate the application to real-world problems of the methods for marketing analysis presented above. For confidentiality reasons, numerical data for the examples presented will not be given. The purpose of these case studies is to offer readers some ideas on the possible fields of application of business intelligence systems in marketing-related decision-making processes.

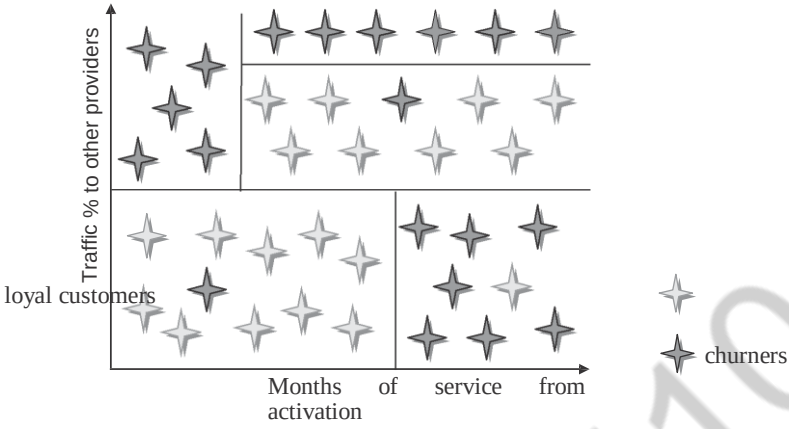
Retention in telecommunications:

Companies operating in the mobile telephone industry were among the first to use learning models and data mining methods to support relational marketing strategies. One of the main objectives has been customer retention, also known as churn analysis. The effect of market saturation and strong competition have combined to cause instability and disaffection among consumers, who can choose a company based on the rates, services and access methods that they deem most convenient. This phenomenon is particularly critical with regard to prepaid telephone cards, very popular in the mobile phone industry today, as they make changing a telephone service provider quite easy and of little cost. Due to the very nature of the services offered, telephone providers possess a vast array of data on their customers and are in the best position to achieve the maximum benefit from data mining in order to target marketing actions and to optimize the use of resources.

Company and objectives:

A mobile phone company wishes to model its customers' *propensity* to churn, that is, a predictive model able to associate each customer with a numerical value (or *score*) that indicates their probability of discontinuing service, based on the value of the available explanatory variables. The model should be able to identify, based on customer characteristics, homogeneous segments relative to the probability of churning, in order to later concentrate on these groups the marketing actions to be carried out for retention, thus reducing attrition and increasing the overall effectiveness. Figure 13.19 shows the possible segments derived using a classification model, using only two predictive attributes in order to create a two-dimensional chart for illustration purposes. The segments with the highest density of churners allow to identify the recipients of the marketing actions.

After an initial exploratory data analysis, the decision is made to develop more than one predictive model, determining a priori some market macro- segments that appear heterogeneous. In this way it is possible to obtain several



An example of segmentation for retention analysis in a mobile telephone company

accurate models instead of a single model related to the entire customer base. The analysis carried out using clustering methods confirms the appropriateness of the segments considered, and leads to the subdivision of customers into groups based on the following dimensions:

- customer type (business or private);
- telephone card type (subscription or prepaid);
- years of service provision, whether above or below a given threshold;
- area of residence.

The marketing data mart provides for each customer a large amount of data:

- personal information (socio-demographic);
- administrative and accounting information;
- incoming and outgoing telephone traffic, subdivided by period (weeks or months) and traffic direction;
- access to additional services, such as fax, voice mail, and special service numbers;
- calls to customer assistance centers;
- notifications of malfunctioning and disservice;
- emails and requests through web pages.

There are approximately 100 explanatory variables available for constructing predictive models.

Analysis and results:

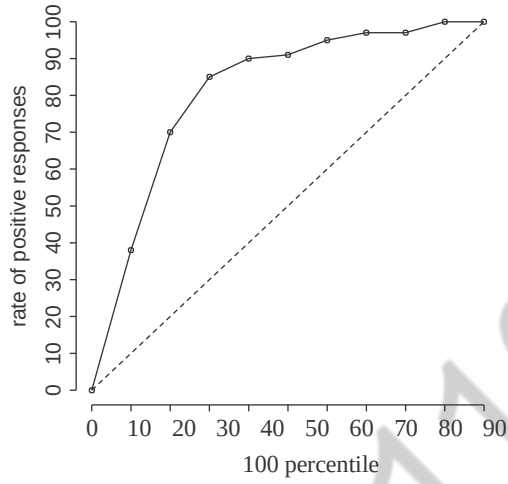
Once the dataset for developing the models has been extracted from the data mart, a detailed exploratory analysis can be carried out. On the one hand, it shows a certain number of anomalies, in the form of outliers and missing values, whose removal improves the quality of data. On the other hand, additional variables can be generated through appropriate transformations in order to highlight relevant trends and correlations identified by exploratory data analysis. After applying feature reduction and extraction, the new dataset contains about 150 predictive variables, after the addition of derived variables and removal of some original variables deemed uninfluential on the target. An indicator variable is defined to denote churning by a customer in cases where official notification of service discontinuation is not required, as is the case for prepaid cards. For the different macro-segments, the related *churning signal* is thus defined. For example, if a private customer makes fewer outgoing calls than a preset threshold and receives a number of incoming calls that is below a second threshold value, she is believed to be at a churning stage.

The retention analysis is therefore brought back to a binary classification problem, and models for each macro-segment are then constructed. Different methodologies and different parameterizations are used to obtain several alter- native models, the most effective of which can be chosen later based on a comparison that takes into account the indicators of accuracy and the interpretability of the rules generated.

At the end of the development phase, two classes of predictive models are identified. The models based on support vector machines achieve a significantly higher accuracy than other methods, but the corresponding rules they derive are more cumbersome. Classification trees based on axis-parallel splitting rules lead to interpretable rules which are simple and intuitive, but achieve a lower accuracy. The former are preferable for generating the lists of optimal recipients to target marketing campaigns aimed at retention, while the latter are better for investigating loyalty and highlighting relevant market niches. Figure 13.20 shows the cumulative gain curve associated with a classifier based on discrete variants of support vector machines.

Acquisition in the automotive industry:

Companies in the automotive industry are striving to develop initiatives aimed at strengthening competitiveness and increasing market share. In this scenario,



Cumulative gains chart for retention analysis in a mobile telephone company

relational marketing projects have been started with the purpose of optimizing the marketing actions and offering products in line with customers' needs, anticipating the evolution of markets and demand. A further element that leads to relational marketing initiatives targeting the sales network is the European directive called the *block exception rule* which introduced new scenarios in the relationships between manufacturing companies and partners in the sales channels, allowing dealers to carry out multi-brand sales activities, and promotion and sales actions across the entire EU territory.

In particular, the use of business intelligence methods is a key factor in strengthening the knowledge of prospects and customers and hence improving new customer acquisition processes and the loyalty of the customer base. It is worth observing that, for some markets in the automotive industry, it is possible to integrate the internal data contained in the marketing data mart with external data sources, which provide a thorough description of the purchases of automobiles and industrial vehicles made by private customers and by companies, easily found at motor vehicles registries. Such an opportunity can be used to enhance segmentation analyses.

Company and objectives:

A company manufacturing industrial vehicles wishes to develop a predictive model that can assign to each prospect a score that indicates his propensity to positively respond to a marketing action aimed at acquisition. The main purpose of the model is to provide guidance for promotion actions carried out by the

network of dealers. But also, in order to stay one step ahead of competitors, it is required to better understand the trends of the future demand, refining the knowledge of the customer base and of the market scenarios, and identifying the distinctive features that characterize current customers in order to design initiatives directed to stimulate new acquisitions.

Analysis and results:

The data available for the acquisition analysis are subdivided into three categories.

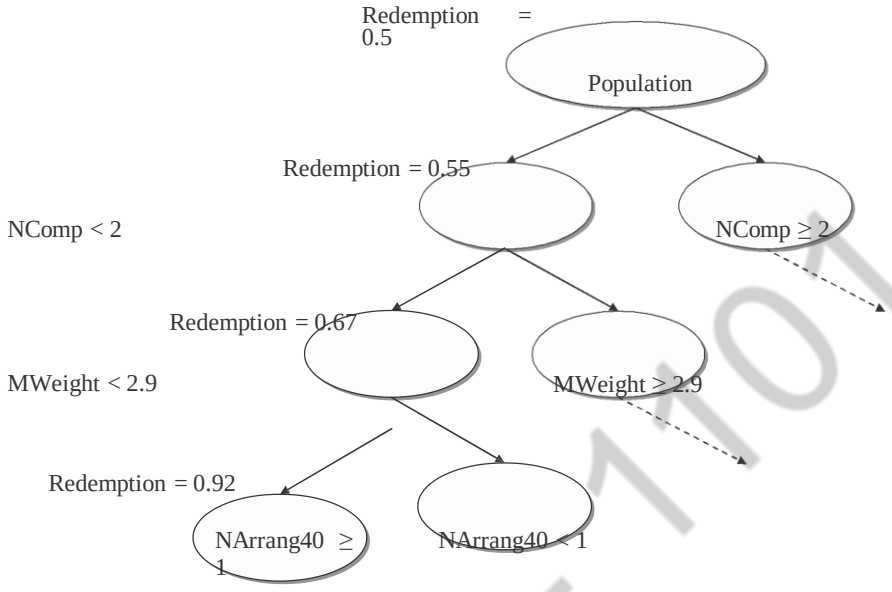
Prospects and customers. The first group of data concerns current or potential customers of the company, including customers owning vehicles not necessarily produced by the company. Besides demographic information, the data include 15 explanatory variables that gather meaningful information, such as the top-rated type of vehicle owned by each prospect, the number of new and used vehicles produced by competitors and owned by each prospect, and the number of new and used vehicles bought in the past by each prospect.

Vehicles. The second group of data includes 16 attributes that enable the vehicles owned by each prospect to be identified, among them model, weight class, optional features, fuel type, first and last registration date and status (new or used vehicle). These attributes define for each vehicle the timing of sales transactions, or transfers of ownership.

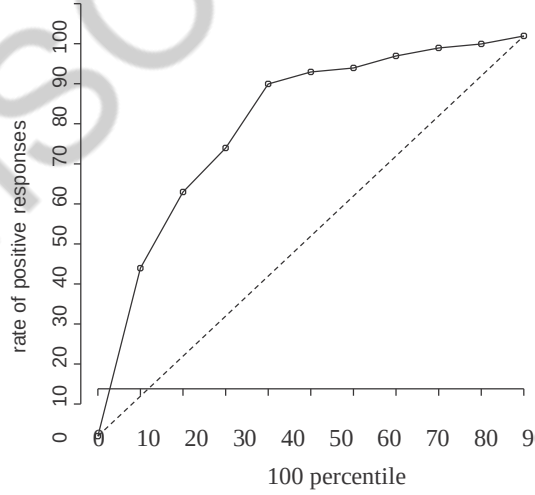
Works. The third group of data refers to maintenance and repair work under- gone by the vehicles during the warranty period and beyond. In particular, these data indicate the type of work carried out, a description of the problem, the vehicle mileage at the time of service, the car shop where the repair was carried out and the date of admission and release of the vehicle to and from the shop.

After exploratory data analysis, which led to the removal of a number of anomalies and outliers, selection of variables was carried out. The analysis led to the addition of new explanatory variables for the time interval that normally elapses between two subsequent purchases, the number of vehicles owned by each prospect, subdivided by weight class, the proportion of vehicles of other brands owned by each potential customer, and the proportion of vehicles of each class included in the vehicle portfolio of each prospect.

Classification models were then developed using different methods. Similar considerations to those in the previous section about retention analysis apply also in this case. Discrete support vector machines turned out to be the most effective method, while classification trees generate rules that can be more easily interpreted.



Classification tree for acquisition analysis in an industrial vehi- cle company



Cumulative gains chart for the acquisition analysis in an industrial vehicle company

Cross-selling in the retail industry:

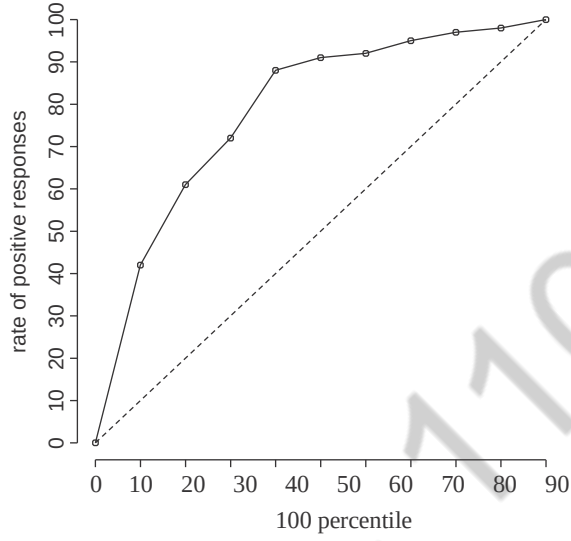
The company considered in this section operates in retail consumer electronics, and wishes to segment its customer base in order to optimize marketing actions aimed at promoting a specific product or group of products. The goal is therefore to develop a predictive model able to assign to each customer a score that indicates her propensity to respond positively to a cross-selling offer. Besides prediction purposes, the model should be used also to interpret explanatory factors that have a greater effect on the purchase of the product promoted. Finally, the model is to be used for assessing the existence of causal and temporal correlations between the purchase of the product promoted and the purchase of other items.

Analysis and results:

The data available for cross-selling analysis are mainly transactional, referring to customers who have signed up for a loyalty card at one of the company's retail stores. These data include the following information:

- personal information (socio-demographic);
- date of signing up for the loyalty card, which can be regarded as the starting date for the relationship between customer and company;
- dates of first and last purchase, marking the boundary of the time interval within which purchases have been made by each individual customer;
- cash slips, indicating which items, and in what quantities, have been purchased by each customer;
- purchases of sale items made by each customer;
- participation in point-earning programs and related prizes won;
- consumer financing requested to make purchases.

Hence, a binary classification problem can be formulated, where the target variable corresponds to the purchase of the specific product to be promoted. Since the prediction should be available a month in advance, classification models should take into account the corresponding latency. Exploratory data analysis enabled the detection and removal of anomalies and missing data. Then a data preparation stage took place at which those variables were removed that showed a low correlation with the target. Finally, some new explanatory variables were generated through transformations of the original attributes, with the purpose of highlighting trends in the temporal sequence of the expenditure amounts.



Cumulative gains chart for the cross-selling analysis in a retail company

At the end of the exploratory analysis, the dataset used to develop the classification models included approximately 120 explanatory attributes. Although the methods based on variants of support vector machines turned out to be more accurate, a model based on classification trees was deemed more appropriate, in consideration of the interpretability that was indicated among the primary objectives of the analysis. The cumulative gains curve corresponds to a model based on support vector machines, able to reach a lift of 8 at a quantile of 0.05, corresponding to 5% of the customers considered by the model to have the highest propensity to buy the promoted product.

Attention then turned to the hypothesis of a causal correlation between the purchase of the product promoted and some other items, identified by marketing analysts on a subjective basis. We proceeded by evaluating the association rules which relate the product promoted to each of the other items, in a series of dyadic relations. The analysis showed that the presence of the product promoted in the head or body of the rule provides similar support and confidence values. This led to rejection of the hypothesis of a significant correlation between purchases. To further confirm this conclusion, sequential rules were analyzed to assess whether the purchase of the item promoted was preceded by the purchase of one of the other items considered, or whether the former was preceded by the latter. Also in this case, the analysis allowed the existence of causal relationships between purchases to be ruled out.

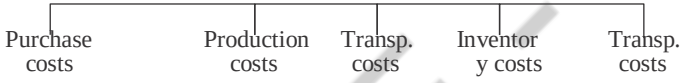
12.Explain about Supply chain optimization?

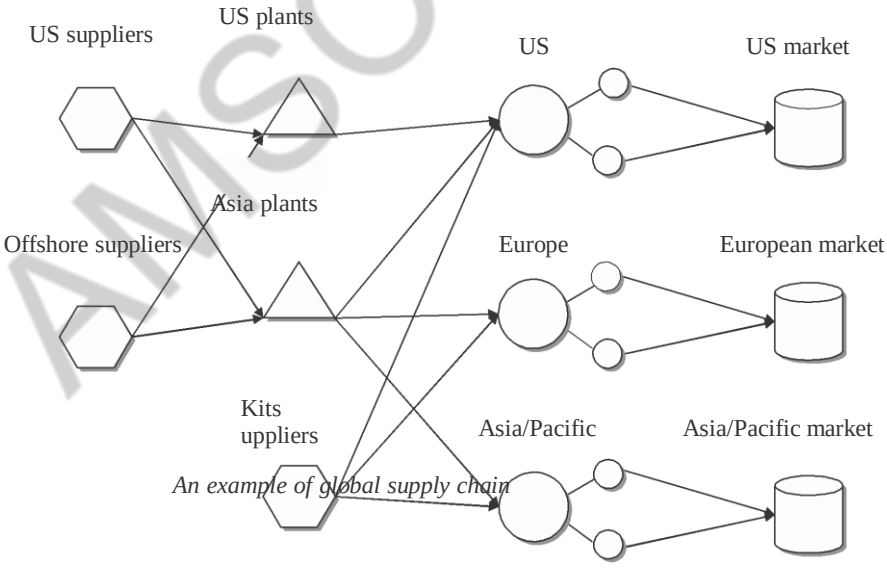
In a broad sense, a *supply chain* may be defined as a network of connected and interdependent organizational units that operate in a coordinated way to manage, control and improve the flow of materials and information originating from the suppliers and reaching the end customers, after going through the procurement, processing and distribution subsystems of a company.

The aim of the integrated planning and operations of the supply chain is to combine and evaluate from a systemic perspective the decisions made and the actions undertaken within the various sub processes that compose the logistic system of a company.

Many manufacturing companies, such as those operating in the consumer goods industry, have concentrated their efforts on the integrated operations of the supply chain, even to the point of incorporating parts of the logistic chain that are outside the company, both upstream and downstream.

The major purpose of an integrated logistic process is to minimize a function expressing the total cost, which comprises processing costs, transportation costs for procurement and distribution, inventory costs and equipment costs. Note that





An example of global supply chain

the optimization of the costs for each single phase does not generally imply that the minimum total cost of the entire logistic process has been achieved, so that a holistic perspective is required to attain a really optimized supply chain.

The need to optimize the logistic chain, and therefore to have models and computerized tools for medium-term planning and for capacity analysis, is particularly critical in the face of the high complexity of current logistic systems, which operate in a dynamic and truly competitive environment. We are referring here to manufacturing companies that produce a vast array of products and that usually rely on a multicentric logistic system, distributed over several plants and markets, characterized by large investments in highly auto- mated technology, by an intensive usage of the available production capacity and by short-order processing cycles. The features of the logistic system we have described reflect the profile of many enterprises operating in the consumer goods industry.

In the perspective outlined above, the aim of a medium-term planning process is therefore to devise an optimal logistic production plan, that is, a plan that is able to minimize the total cost, understood as the sum of procurement, processing, storage, distribution costs and the penalty costs associated with the failure to achieve the predefined service level. However, to be implemented in practice, an optimal logistic production plan should also be feasible, that is, it should be able to meet the physical and logical constraints imposed by limits on the available production capacity, specific technological conditions, the structure of the bill of materials, the configuration of the logistic network, min- imum production lots, as well as any other condition imposed by the decision makers in charge of the planning process.

Optimization models represent a powerful and versatile conceptual paradigm for analyzing and solving problems arising within integrated supply chain planning, and for developing the necessary software. Due to the complex interactions occurring between the different components of a logistic production system, other methods and tools intended to support the planning activity seem today inadequate, such as electronic spreadsheets, simulation systems and planning modules at infinite capacity included in *enterprise resource planning* software. Conversely, optimization models enable the development of realistic mathematical representations of a logistic production system, able to describe with reasonable accuracy the complex relationships among critical components of the logistic system, such as capacity, resources, plans, inventory, batch sizes, lead times and logistic flows, taking into account the various costs. Moreover, the evolution of information technologies and the latest developments in optimization algorithms mean that decision support systems based on optimization models for logistics planning can be efficiently developed.

13.What are the various Optimization models for logistics planning?

In this section we will describe some optimization models that may be used to represent the most relevant features of logistic production systems. when introducing salesforce planning models, for the sake of simplicity we have chosen to illustrate for each model a single feature of a logistic system. Readers should keep in mind that real-world logistic production systems feature simultaneously more than one of the elements considered, so that the models developed in applications, will be substantially more complex as they result from the combination of the different features.

Before proceeding with the description of specific models, it is useful to introduce some notation common to most models presented in this section. The logistic system includes I products, which will be denoted by the index $i \in \boldsymbol{I} = \{1, 2, \ldots, I\}$. The planning horizon is subdivided into T time intervals $t \in T = \{1, 2, \ldots, T\}$, generally of equal length and usually corresponding to weeks or months.

The manufacturing process has at its disposal a set of critical resources shared among the different products and available in limited quantities. These resources may consist of production and assembly lines, to manpower, to spe- cific fixtures and tools required by manufacturing. The R critical resources considered in the logistic production system will be denoted by the index $r \in \boldsymbol{R} = \{1, 2, \ldots, R\}$. Whenever a single resource is relevant to the manufacturing process, the index r will be omitted for sake of simplicity.

Tactical planning:

In its simplest form, the aim of tactical planning is to determine the production volumes for each product over the T periods included in the medium-term planning horizon in such a way as to satisfy the given demand and capacity limits for a single resource, and also to minimize the total cost, defined as the sum of manufacturing production costs and inventory costs.

We therefore consider the decision variables

P_{it} = units of product i to be manufactured in period t,

I_{it} = units of product i in inventory at the end of period t,

and the parameters

d_{it} = demand for product i in period t,

c_{it} = unit manufacturing cost for product i in period t,

h_{it} = unit inventory cost for product i in period t,

e_i = capacity absorption to manufacture a unit of product i, b_t = capacity available in period t.

The resulting optimization problem is formulated as follows:

$$\min \sum_{t \in T} \sum_{i \in \boldsymbol{I}} (c_{it} P_{it} + h_{it} I_{it})$$

$$\text{s.to} \quad P_{it} + I_{i,t-1} - I_{it} = d_{it}, \qquad i \in \boldsymbol{I}, t \in T,$$

$$\sum_{i \in \boldsymbol{I}} e_i P_{it} \le b_t, \qquad t \in T,$$

$$P_{it}, I_{it} \ge 0, \qquad i \in \boldsymbol{I}, t \in T.$$

Constraints express the balance conditions among production, inventory and demand, by establishing a connection between successive periods along the planning horizon. Inequalities constrain the absorbed capacity not to exceed the available capacity for each period.

Model is a linear optimization problem which can be therefore solved efficiently even with a very large number of variables and constraints, of the order of a few million, by means of current state-of-art algorithms and computer technologies.

Extra capacity:

A first extension of the basic model deals with the possibility of resorting to *extra capacity*, perhaps in the form of overtime, part-time or third-party capacity. In addition to the decision variables already included in model ,we define the variables

O_t = extra capacity used in period t,

and the parameters

q_t = unit cost of extra capacity in period t.

The optimization problem now becomes

$$\min \sum_{t \in T} \sum_{i \in \boldsymbol{I}} (c_{it} P_{it} + h_{it} I_{it}) + \sum_{t \in T} q_t O_t$$

$$\text{s.to} \quad P_{it} + I_{i,t-1} - I_{it} = d_{it}, \qquad i \in \boldsymbol{I}, t \in T,$$

$$\sum_{i \in \boldsymbol{I}} e_i P_{it} \le b_t + O_t, \qquad t \in T,$$

$$P_{it}, I_{it}, O_t \ge 0, \qquad i \in \boldsymbol{I}, t \in T.$$

Constraints have been modified to include the available extra capacity. The extended model is still a linear optimization problem which can be therefore solved efficiently.

Multiple resources:

If the manufacturing system requires R critical resources, a further extension of model can be devised by considering multiple capacity constraints. The decision variables already included in model remain unchanged, though it is necessary to consider the additional parameters

b_{rt} = quantity of resource r available in period t,

e_{ir} = quantity of resource r absorbed to manufacture one unit of product i.

The resulting optimization problem is given by

$$\min \sum_{t \in T} \sum_{i \in \boldsymbol{I}} (c_{it} P_{it} + h_{it} I_{it})$$

$$\text{s.to} \quad P_{it} + I_{i,t-1} - I_{it} = d_{it}, \qquad i \in \boldsymbol{I}, t \in T,$$

$$\sum_{i \in \boldsymbol{I}} e_{ir} P_{it} \leq b_{rt}, \qquad r \in \boldsymbol{R}, t \in T,$$

$$P_{it}, I_{it} \geq 0, \qquad i \in \boldsymbol{I}, t \in T.$$

Constraints have been modified to take into account the upper limits on the capacity of the R resources in the system. Model remains a linear optimization problem which can be solved efficiently.

Backlogging:

Another feature that needs to be modeled in some logistic systems is *backlogging*. The term *backlog* refers to the possibility that a portion of the demand due in a given period may be satisfied in a subsequent period, incurring an additional penalty cost. Backlogs are a feature of production systems more likely to occur in B2B or make-to-order manufacturing contexts. In B2C industries, such as mass production consumer goods, on the other hand, one is more likely to find a variant of the backlog, known as *lost sales*, in which unfulfilled demand in a period cannot be transferred to a subsequent period and is lost.

To model backlogging, it is necessary to introduce new decision variables

B_{it} = units of demand for product i delayed in period t,

and the parameters

g_{it} = unit cost of delaying the demand for product i in period t.

The resulting optimization problem is

$$\min \sum_{t \in T} \sum_{i \in \boldsymbol{I}} (c_{it} P_{it} + h_{it} I_{it} + g_{it} B_{it})$$

$$\text{s.to} \quad P_{it} + I_{i,t-1} - I_{it} + B_{it} - B_{i,t-1} = d_{it}, \qquad i \in \boldsymbol{I}, t \in T,$$

$$\sum_{i \in \boldsymbol{I}} e_i P_{it} \leq b_t, \qquad t \in T,$$

$$P_{it}, I_{it}, B_{it} \geq 0, \qquad i \in \boldsymbol{I}, t \in T.$$

The balance constraints (14.14) have been modified to take backlog variables into account. Specifically, in each period t one is allowed to delay a portion of the demand d_{it}, given precisely by the backlog variable B_{it}, whereas the demand d_{it} itself is increased by the units held as backlog in the previous period. Model (14.13) is again a linear optimization problem which can be therefore solved efficiently.

An alternative way to model backlog is to suppose that the demand d_{it} is made up of separate orders k, each characterized by a request for w_{ik} units of product i and a due delivery date t_{ik}:

$$d_{it} = \sum_{k \in K_{it}} w_{ik}, \qquad K_{it} = \{k | t_{ik} = t\}.$$

We further assume that each order should be completed within at most two periods from its due delivery date.

To model this second form of backlogging, define the following binary decision variables:

$$\beta_{ik0} = \begin{cases} 1 & \text{if order } k \text{ for product } i \text{ is delivered on time,} \\ 0 & \text{otherwise;} \end{cases}$$

$$\beta_{ik1} = \begin{cases} 1 & \text{if order } k \text{ for product } i \text{ is delayed by one period,} \\ 0 & \text{otherwise;} \end{cases}$$

$$\beta_{ik2} = \begin{cases} 1 & \text{if order } k \text{ for product } i \text{ is delayed by two periods,} \\ 0 & \text{otherwise.} \end{cases}$$

Consider also the parameters

g_{ik1} = cost of delivering order k for product i delayed by one period, g_{ik2} = cost of delivering order k for product i delayed by two periods.

The resulting optimization problem is formulated as

$$\min \sum_{t\in T}\sum_{i\in \boldsymbol{I}}(c_{it}P_{it}+h_{it}I_{it})+\sum_{t\in T}\sum_{i\in \boldsymbol{I}}\Big(\sum_{k\in K_{i,t-1}} g_{ik1}\beta_{ik1}+\sum_{k\in K_{i,t-2}} g_{ik2}\beta_{ik2}\Big)$$

$$\text{s.to } P_{it}+I_{i,t-1}-I_{it}=\sum_{k\in K_{it}} w_{ik}\beta_{ik0}+\sum_{k\in K_{i,t-1}} w_{ik}\beta_{ik1}+\sum_{k\in K_{i,t-2}} w_{ik}\beta_{ik2}, \quad i\in \boldsymbol{I}, t\in T,$$

$$\sum_{i\in \boldsymbol{I}} e_i P_{it}\le b_t, \quad t\in T,$$

$$\beta_{ik0}+\beta_{ik1}+\beta_{ik2}=1, \quad i\in \boldsymbol{I}, k\in K,$$

$$P_{it}, I_{it}\ge 0,\ \beta_{ik0}, \beta_{ik1}, \beta_{ik2}\in\{0,1\}, \quad i\in \boldsymbol{I}, t\in T.$$

The balance constraints (14.19) have been modified to take the binary backlog variables into account. Specifically, in each period t are fulfilled the orders k due for period t for which there is no delayed delivery (i.e. such that $\beta_{ik0}=1$), the orders due for the previous period $t-1$ for which the delivery is delayed by one period (i.e. such that $\beta_{ik1}=1$), and finally the orders due for period $t-2$ for which the delivery is delayed by two periods (i.e. such that $\beta_{ik2}=1$). The multiple choice constraints (14.21) establish that each order be fulfilled in exactly one of the three alternative ways corresponding to variables $\beta_{ik0}, \beta_{ik1}, \beta_{ik2}$.

Compared to previous model, formulation allows us to attach a different penalty cost to the K orders, therefore assigning preferences to some customers, for example those considered of strategic importance. In B2B customer–supplier relationships this possibility can be of considerable value.

Unlike the models previously considered, model (14.18) is a mixed binary linear optimization problem, whose solution requires computation times that grow exponentially fast with the number of variables and constraints. How- ever, when the problem size is too large to yield an optimal solution in a reasonable time, for instance through a general purpose exact algorithm

such as branch-and-bound, it is usually possible to devise an approximation algorithm achieving suboptimal solutions. This can be done by truncating a branch-and-bound procedure, or by designing *ad hoc* approximation algorithms.

Finally, notice that a multi-objective optimization model can also be formulated by requiring, for instance, that a proportion of at least n_1 of the orders be delivered on time and that a proportion of at most $1-n_2$ be delayed by two periods, by introducing into model (14.18) the additional constraints

$$\sum_{k=1}^{K_i} \beta_{ik0} \geq n_1 K_i, \qquad \sum_{k=1}^{K_i} \beta_{ik0} + \beta_{ik1} \geq n_2 K_i,$$

where K_i is the total number of orders referring to product i.

Minimum lots and fixed costs:

A further feature often appearing in manufacturing systems is represented by *minimum lot* conditions: for technical or scale economy reasons, it is sometimes necessary that the production volume for one or more products be either equal to 0 (i.e. the product is not manufactured in a specific period) or not less than a given threshold value, the minimum lot.

To incorporate minimum lot conditions into the model, we define the binary decision variables

$$Y_{it} = \begin{cases} 1 & \text{if } P_{it} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

and the parameters

l_i = minimum lot for product i,

γ = constant value larger than any producible volume for i.

The optimization problem is now

$$\min \sum_{t \in T} \sum_{i \in \boldsymbol{I}} (c_{it} P_{it} + h_{it} I_{it})$$

$$\text{s.to} \quad P_{it} + I_{i,t-1} - I_{it} = d_{it}, \qquad i \in \boldsymbol{I}, t \in T,$$

$$\sum_{i \in \boldsymbol{I}} e_i P_{it} \leq b_t, \qquad t \in T,$$

$$P_{it} \geq l_i Y_{it}, \qquad i \in \boldsymbol{I}, t \in T,$$

$$P_{it} \leq \gamma Y_{it}, \qquad i \in \boldsymbol{I}, t \in T,$$

$$P_{it}, I_{it} \geq 0, Y_{it} \in \{0, 1\}, \qquad i \in \boldsymbol{I}, t \in T.$$

Constraints express the minimum lot conditions. Constraints are logical consistency conditions between the variables Y_{it} and P_{it} needed to force the binary variable Y_{it} to take the value 1 whenever the corresponding production volume P_{it} is greater than 0. The constant γ in must be chosen sufficiently large that the condition does not constitute an actual upper bound on the producible volume for product i. Indeed, it should be bounded above only by the available capacity and by the assigned demand.

In all previous model formulations we have implicitly assumed that production costs are proportional to production volumes. For some logistic systems, however, in order to manufacture a product it may be necessary to set up a machine and incur a *setup cost*. However, such costs are required only if the production volume is strictly greater than 0, that is, if production of the product concerned is actually accomplished. A further parameter,

$$f_{it} = \text{unit setup cost for product } i \text{ in period } t,$$

is then assigned and the optimization problem becomes

$$\begin{aligned}
\min \quad & \sum_{t \in T} \sum_{i \in I} (c_{it} P_{it} + h_{it} I_{it} + f_{it} Y_{it}) \\
\text{s.to} \quad & P_{it} + I_{i,t-1} - I_{it} = d_{it}, && i \in I, t \in T, \\
& \sum_{i \in I} e_i P_{it} \le b_t, && t \in T, \\
& P_{it} \le \gamma Y_{it}, && i \in I, t \in T, \\
& P_{it}, I_{it} \ge 0, Y_{it} \in \{0, 1\}, && i \in I, t \in T.
\end{aligned}$$

Constraints represent the logical consistency conditions between variables Y_{it} and P_{it}, as already observed for model

Models and mixed binary linear optimization problems,
for whose solution the same remarks as for model apply.

Bill of materials:

A further extension of the basic planning model deals with the representation of products with a complex structure, described via the so-called *bill of materials*, where end-items are made by components that in turn may include other components.

Formally, the following parameters are defined to describe the structure of the bill of materials:

$$a_{ij} = \text{units of product } i \text{ directly required by one unit of product } j,$$

where the term *product* refers here to both end-items and components at various levels of the bill of materials. For each product i we assign an *external* demand d_{it} and an *internal* demand, the latter induced by the requirements of product i needed to manufacture the components or the end-items for which i represents a direct component. The external demand for components may originate from other plants of the same manufacturing company or from outside customers that also buy components.

The resulting optimization problem is formulated as

$$\min \sum_{t \in T} \sum_{i \in \boldsymbol{I}} (c_{it} P_{it} + h_{it} I_{it})$$

$$\text{s.to} \quad P_{it} + I_{i,t-1} - I_{it} = d_{it} + \sum_{j \in \boldsymbol{I}, j = i} a_{ij} P_{jt}, \qquad i \in \boldsymbol{I}, t \in T,$$

$$\sum_{i \in \boldsymbol{I}} e_i P_{it} \leq b_t, \qquad t \in T,$$

$$P_{it}, I_{it} \geq 0, \qquad i \in \boldsymbol{I}, t \in T.$$

The balance constraints have been modified to take into account the demand internally generated. Model is a linear optimization problem which can be therefore solved efficiently.

Multiple plants:

In this section it is assumed that a manufacturing company has a network of M production plants, located in geographically distinct sites, that manufacture a single product. The logistic system is responsible for supplying N peripheral depots, located in turn at distinct sites. Each production plant $m \in \boldsymbol{M} = \{1, 2, \ldots, M\}$ is characterized by a maximum availability of product, denoted by s_m, while each plant $n \in N = \{1, 2, \ldots, N\}$ has a demand d_n. We further assume that a transportation cost c_{mn} is incurred by sending a unit of product from plant m to depot n, for each pair (m, n) of origins and destinations in the logistic network. The objective of the company is to determine an optimal logistic plan that satisfies at minimum cost the requests of the depots, without violating the maximum availability at the plants. It should be clear that the problem described arises frequently in logistic systems, at different levels in the logistic network (e.g. from suppliers to plants, from plants to warehouses or from warehouses to customers).

The decision variables needed to model the problem described represent the quantity to be transported for each plant–depot pair,

x_{mn} = unit of product to be transported from m to n.

The resulting optimization problem is

$$\min \sum_{m \in M} \sum_{n \in N} c_{mn} x_{mn}$$

$$\text{s.to} \quad \sum_{n \in N} x_{mn} \leq s_m, \qquad m \in M,$$

$$\sum_{m \in M} x_{mn} \geq d_n, \qquad n \in N,$$

$$x_{mn} \geq 0, \qquad m \in M, n \in N.$$

Constraints ensure that the availability of each plant is not exceeded, whereas constraints establish that the demand of each depot be satisfied.

Model is a linear optimization problem, and can be therefore solved efficiently.

14.Explain in detail about Revenue management systems?

Revenue management is a managerial policy whose purpose is to maximize profits through an optimal balance between demand and supply. It is mainly intended for marketing as well as logistic activities and has found growing interest in the service industry, particularly in the air transportation, tourism and hotel sectors. More recently these methods have also begun to spread within the manufacturing and distribution industries.

The strong interest shown by such enterprises in the themes considered by revenue management should come as no surprise, if we consider the complexity and strategic relevance of decision-making processes concerning demand management, which are addressed by marketing and logistics managers. Consider, for example, the complex interactions among decisions on pricing, sales pro- motions, markdowns, mix definition and allocation to points of sale, in a highly dynamic and competitive context characterized by multiple sales channels and several alternative ways of contacting customers.

Despite the potential advantages that revenue management initiatives may offer for enterprises, there are certain difficulties that hamper the actual implementation of practical projects and actions aimed at adopting revenue management methodologies and tools. We can identify several explanations for the gap between intentions and initiatives actually undertaken. Certainly the fear of implementation costs and uncertainty over the results that can be achieved play an important role, as happens for many innovation projects. Empirical investigations show, however, that the primary reason for prudence in adopting revenue management should be sought in the prerequisite conditions necessary to successfully start a revenue management project. There is a high level

of interaction between revenue management and two other themes that we described earlier optimization of the supply chain and relational marketing. On the one hand, in order to apply revenue management methods and tools it is necessary to have an integrated and optimized logistic chain that guarantees the efficiency and responsiveness of the logistic flows. On the other hand, it is also necessary to possess a deep knowledge of the customers and an accurate micro-segmentation of the market, achieved through data mining analytical models and therefore based on the analysis of the actual purchasing behaviors regularly recorded in the marketing data mart. Hence, to profitably adopt revenue management a company should be able to enhance and trans- form into knowledge, through the use of business intelligence methodologies, the huge amount of information collected by means of automatic data gathering technologies.

Decision processes in revenue management

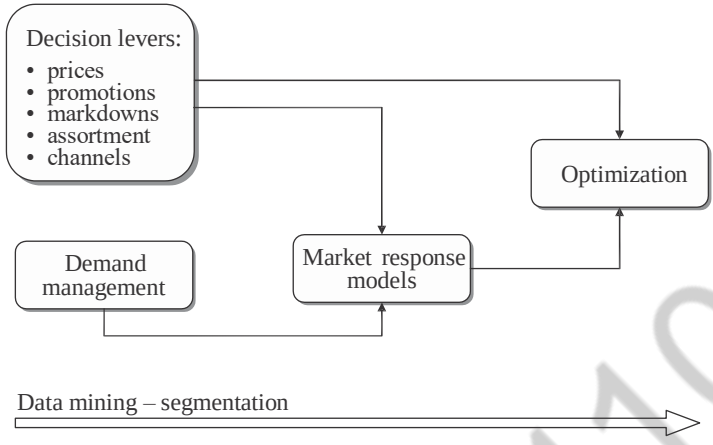
Revenue management involves the application of mathematical models to predict the behavior of customers at a micro-segmentation level and to optimize the availability and price of products in order to maximize profits. In this respect, to summarize relational marketing objectives: to formulate for each segment, ideally for each customer, the appropriate offer through the most suitable channel, at the right time and at the best price.

The purpose of revenue management is therefore to maximize profits, aligning the offer of products and services to the expected demand, using both the major levers of the marketing mix (e.g. prices, promotions, assortment) and the levers of logistics (e.g. efficiency and timeliness). Specific and innovative features of revenue management strategies are a closer focus on demand than supply and a greater emphasis on costs than revenues; such features are often absent from the managerial policies adopted by most enterprises.

As already observed, in recent years revenue management has been applied with more and more success by many companies operating in the service industry. Among the pioneers in this field are airlines, hotel chains, automobile rental companies, theme parks, theaters and other entertainment-related enterprises. The common characteristics of these fields are well apparent: a highly perish- able product, a fairly low marginal sales cost and the possibility of applying dynamic pricing policies and exploiting multiple sales channels.

Revenue management affects some highly complex decision-making processes of strategic relevance,

- market segmentation, by product, distribution channel, consumer type and geographic area, performed using data mining models;



Decision processes in revenue management

- prediction of future demand, using time series and regression models;
- identification of the optimal assortment, i.e. the mix of products to be allocated to each point of sale;
- definition of the market response function, obtained by identifying models and rules that explain the demand based on company actions, the initiatives of competitors and other exogenous contextual events;
- management of activities aimed at determining the price of each product (*pricing*) as well as the timing and the amount of markdowns;
- planning, management and monitoring of sales promotions, and assessment of their effectiveness;
- sales analysis and control, and use of the information gathered to evaluate market trends;
- material procurement and stock management policies, such as control policy, frequency of issued orders, reorder quantities;
- integrated management of the different sales and distribution channels. Revenue management relies on the following basic principles:

- To address sales to micro-segments: segmentation carried out by means of business intelligence and data mining models is critical to achieve an adequate knowledge of the market.

- To exploit the product value cycle: to generate the highest revenues, it is required to grasp the value cycle of products and services, in order to optimally synchronize their availability over time and to determine the price for each market micro-segment. Notice that the value cycle also depends on the sensitivity of micro-segments to price variations.
- To have a price-oriented rather than cost-oriented approach in balancing supply and demand: when supply and demand are out of balance, most enterprises tend to react by increasing or decreasing capacity. In many instances it might, however, be more convenient to adopt price variations, avoiding repeated variations in capacity.
- To make informed and knowledge-based decisions: a consistent use of prediction models tends to mean that decisions rest on a more robust knowledge basis. In particular, a correct prediction of consumer purchasing behaviors is essential to evaluate elasticity and reactions to price variations.
- To regularly examine new opportunities to increase revenues and profits: the possibility of timely access to the available information, combined with the possibility of considering alternative scenarios, strengthens the competencies of marketing analysts and increases the effectiveness of their activity.

The adoption of revenue management methods and tools requires a few pre- requisite conditions to be satisfied within a company, since without them the expected results are unlikely to be achieved. As with any innovation project, it is the people and the organization that constitute a key success factor rather than the use of specific software tools. In this case too, the culture and the structure of the processes within an organization must be prepared to adopt powerful tools that may turn out to be unsafe and disrupting if improperly used. It is therefore necessary to develop within the enterprise an information culture, particularly among those knowledge workers who operate in the marketing and logistics departments, more directly involved with the application of revenue management strategies. This means that all marketing data must be systematically gathered, controlled, normalized, integrated and stored in a data mart. To segment the market and to create micro-segments, business intelligence methods and analytical models should be used. It is therefore advisable for an enterprise turning to revenue management to have already developed relational marketing initiatives or at least to be able to carry out data mining analyses.

On the other hand, the decisions involved in revenue management strategies share many aspects with the logistics department, and in particular with the

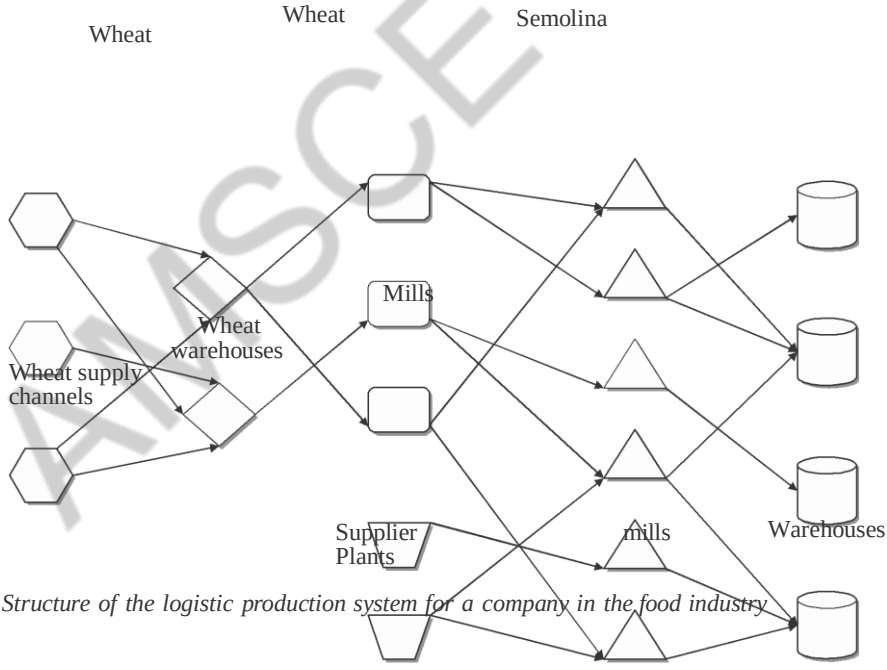
management of flows in the supply chain. In this case too, particularly for manufacturing companies, it is advisable for an enterprise considering revenue management to have previously embarked on supply chain integration and rationalization projects, in order to guarantee an adequate cost reduction that, combined with the increased revenues obtained through revenue management, may lead to a significant increase in profits. Moreover, effective supply chain management is also required to guarantee timely restocking.

15.Explain Business case studies for Logistics?

This section describes two examples of real-world applications of optimization models for logistic and production planning. The first is concerned with an enterprise operating in the food industry, while the second refers to a company that manufactures integrated solutions for liquid food product packaging.

Logistics planning in the food industry

The logistic system of the food manufacturing company consists of a network whose nodes represent suppliers of raw materials, production plants and central and peripheral warehouses. Retail and wholesale distribution to the points of sale, placed downstream of the warehouses, is regarded as external to the logistic subsystem considered here.



Structure of the logistic production system for a company in the food industry

Logistics planning is driven by the demand plan, which in turn depends on the sales forecasts for individual items, disaggregated by warehouse and by period. Supply, production and primary distribution plans are therefore meant to feed end items into central and peripheral warehouses.

Industrial processing manufactures products of low complexity on *transfer lines*, that is, highly automated production lines designed to achieve high out- put rates even with a limited plant flexibility. The processing cycle basically includes two major phases: during the *production* phase, raw materials are transformed into semi-finished goods, which during the *packaging* phase are then turned into end items. Both production and packaging lines operate in a batch processing mode. Generally, each item can be produced at several plants and on several lines within a given plant. Each plant includes both production and packaging lines, while the transfer of semi-finished goods from one plant to another is not allowed. Based on the demand profile, and on the related production capacity requirements, each of the two phases of the production cycle may represent a bottleneck of the manufacturing system during the periods included in the planning horizon. Since the enterprise targets consumer goods markets, if the demand for an item for a specific warehouse cannot be satisfied in a given period, it is assumed that this translates into lost sales. Therefore, the company does not allow any demand backlog to be met during following periods.

The logistic system is made up of dozens of plants housing hundreds of production and packaging lines. More than a thousand end items are produced and later stored in central and peripheral warehouses. Each production or pack- aging line may produce up to dozens of semi-finished goods or end products. There are plants that manufacture a small number of items (two or three) and plants that are able to produce a considerable portion of the product range.

Several decisions depend on the company's logistic plan, and would benefit from a production and distribution plan that is more efficient and stable. Some of these decisions are:

- the optimal sizing of the stocks, so as to guarantee the required service level;
- the plan of allocation to the plants of the demand originating at each market warehouse, in order to achieve an optimal balance between production, distribution and supply costs;
- the plan of allocation of the production capacity, with the possible addition of extra capacity, by using extra work shifts and part-time work or by subcontracting some of the production to external partners (*co-packers*); moreover, the company needs to set out medium-term contractual obligations with co-packers, which are significantly affected by the stability and reliability of the logistic plan devised;

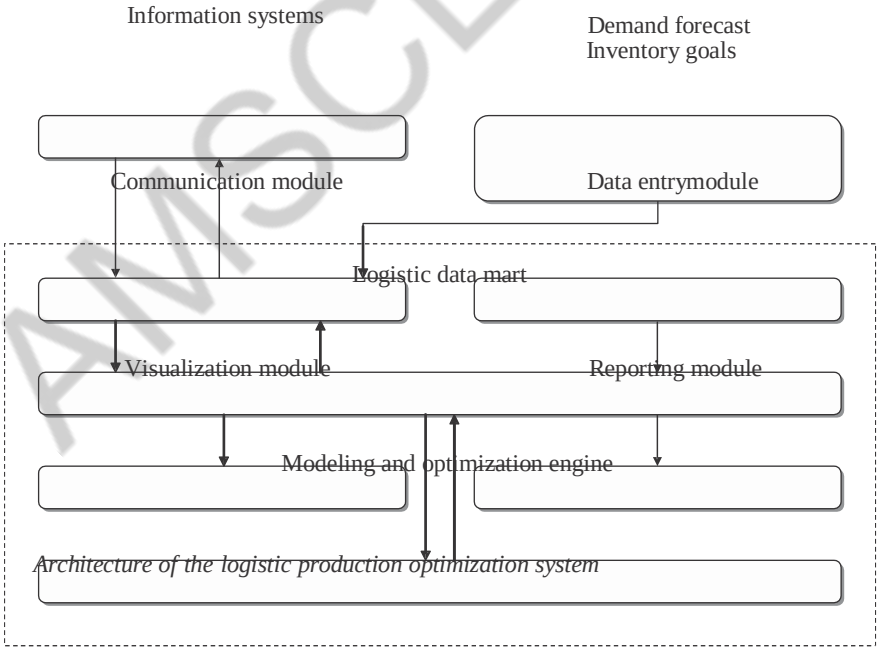
- the supply plan, and the corresponding medium-term contracts with suppliers for the procurement of raw materials and packaging.

Before the new planning system based on an optimization model was introduced, it was customary to work out the logistic plan in an aggregate way, by families of items and with infinite capacity, without duly considering costs and therefore with no attempt at optimization. Subsequent simulations were used to determine the corresponding level of production capacity engagement and the portions of the logistic plan that turned out not to be feasible. In some instances, such preliminary processing activities were followed by more focused analyses developed by the planners using simple spreadsheets. This development process proved largely unable to effectively manage such a highly complex supply chain; furthermore, it was also inefficient, since operators were required to continuously apply corrections and assessments.

The optimization system for logistics planning

The reengineering of the supply chain management process has led to the development of a logistic decision support system whose *intelligence* is represented by an integrated optimization model of the entire logistic production system. shows a sketch of the system architecture.

The system uses a logistic data mart that constitutes a local database, also to increase the efficiency in the generation and solution of the optimization model by means of the algorithmic engine. The information contained in the data



Architecture of the logistic production optimization system

mart is entered and updated in two different ways: on the one hand, through the automatic acquisition of data from the company information system; and on the other hand, through direct data entry by the users for those data and parameters required to devise the optimal logistic plan and not contained in the company information system.

Primary input data acquired from the information system are represented by the demand plan, formulated by item, warehouse and period. The demand plan is determined based on sales forecasts, which are available at the highest level of accuracy. Sales forecasts for each point of sale are therefore aggregated, allocating the demand to each central and peripheral warehouse. Finally, to obtain the demand plan, the aggregated sales forecasts are corrected to take into account the desired level of stocks, defined for the purpose of guaranteeing a preset level of service to customers.

The logistic plan, updated every week on a *rolling* basis, is usually developed over a time span of a year, divided into weeks. Downstream, the planning system is integrated with other modules that allow raw and packaging material requirements to be determined via *manufacturing resource planning*.

At the heart of the planning decision support system is the modeling and algorithmic module, which in turn is based on a finite-capacity optimization model of the entire logistic system. In particular, this is a discrete-time deterministic model that uses a network representation of the logistic system. The nodes of the network represent the supply centers, plants and warehouses where the demand is placed, as indicated in Figure 14.3. For each plant, the model considers the two phases of the manufacturing process – both production and packaging lines. The two phases are planned in a coordinated way, by properly adjusting the volume of products processed at each phase during each period of the planning horizon.

Each production and packaging line is assigned the available capacity, expressed in hours, for each period of the planning horizon. The available manpower capacity at the packaging lines is also assigned, plant by plant, since the human resources employed in packaging are multi-skilled workers and can be shared among different lines of the same plant. Finally, one assigns the potential manpower availability in terms of overtime, holidays or seasonal work, which can be used when needed at additional cost.

Regarding stock management, the model input data specify both a safe level of stock, defined on the basis of the demand profile and the desired service level, and an overstock value not to be exceeded. The logistic plan must guarantee at each period that stocks are above the safe level and below the overstock threshold. For each violation of these conditions, a penalty cost assigned as input is applied.

The size of the minimum lots is also specified, both for production and packaging lines, depending on the specific semi-finished goods and end items. Finally, the model deals with secondary conditions, to be assigned at the discretion of the planners, such as enforcing minimum and maximum volumes for selected end items, or specifying a set of semi-finished goods and products allowed to be manufactured at the beginning or end of each period.

The model also determines the supply of semolina to the plants. This can be procured through two channels: it can be purchased from external suppliers; or it can be obtained from production plants (mills) owned by the company, which in turn have to obtain stocks of wheat from supply markets. In brief, the input data to the system are:

- the demand plan, specified for each warehouse, item and period;
- the initial inventory for each warehouse;
- the bill of materials;
- the technological maps, which list all possible combinations and the corresponding yields between semi-finished goods and production lines, as well as between end items and packaging lines;
- the capacity of production and packaging lines over the periods of the planning horizon;
- the time availability of labor, in terms of regular working hours, extra shifts, holidays and seasonal working time;
- the minimum lots and the forced excess of production and packaging;
- the availability of the different types of wheat on the supply markets;
- the capacity at the mills to transform wheat into semolina;
- the availability of semolina at third-party mills;
- the blending quality requirements for the transformation of wheat into semolina.

The objective function of the model takes the following cost factors into account:

- the transportation costs of transferring end items from each plant to each warehouse;
- the production costs for production lines, net of labor costs;

- the production costs for packaging lines, net of labor cost;
- the penalty cost of failing to reach the desired stock level;
- the penalty cost of overstocking;
- the cost determined by lost sales due to the required product being unavailable in the appropriate period;
- the unit cost of labor, regular working hours, extra shifts, holidays and seasonal working;
- the cost of different wheat types available on the supply markets;
- the milling cost;
- the cost of semolina bought at third-party mills.

The representation of the logistic system described leads to a large-scale mixed binary optimization model, due to the presence of binary decision variables. However, despite its high level of complexity, the model requires no more than a few minutes' computation time, due to the existence of an *ad hoc* algorithm which yields an approximate solution.

At the end of the computation, the model provides an optimal integrated logistic plan, providing the following information:

- the allocation of the demand to the plants;
- the distribution plan, expressing the volumes of end products shipped weekly from each plant to the central and peripheral warehouses;
- the production and packaging plans for each plant, indicating the volumes of semi-finished goods and end products processed weekly by each production and packaging line, with the possibility of processing ahead of time with respect to the due dates, whenever needed or advantageous;
- the inventory plan, specifying the optimal levels of weekly stock for the end products at each warehouse;
- the possibly unmet demand for end items for each warehouse;
- the production capacity engagement plan, for both the lines and the labor;
- the supply of raw materials and packaging;

- the employment of additional labor, in the form of extra shifts, holiday work or seasonal work;
- the purchase from the supply markets and storage cost of wheat, based on the foreseen availability;
- the transportation of wheat from supply markets to the company's own mills;
- the blending and processing of wheat into semolina at the company's own mills;
- the purchase of semolina from the mills of third-party suppliers;
- the transportation of semolina from internal mills and from supplier mills to plants.

16.What are the advantages of using decision support system for logistics planning?

The use of the decision support system for logistics planning affords several advantages, both in terms of efficiency in devising the plan and effectiveness of the plans generated.

It requires less effort on the part of planners, while at the same time contributing to a substantial job enrichment, since they are required to play a decision-making role that is more gratifying on a personal level and of greater value to the company. Indeed, the system enables users to perform and manage different scenario analyses by experimenting with the input data and the parameters of the model. In this way, planners may simulate and evaluate the effect of different conditions and assumptions, carrying out what-if analyses in order to achieve the most effective logistic plan. For example, they can modify some cost parameters, or the level of the required stocks, or the labor availability, in order to assess the consequences of such changes. Particularly noteworthy and of great practical advantage is the possibility of assigning predetermined quantities for the production volumes, by single item or group of items, as well as by single period or group of periods.

The planners responsible for wheat procurement can also use the system to easily carry out what-if analyses of different alternative scenarios, formulated by varying model parameters such as transportation costs, wheat purchase costs, currency exchange rates, and the processing capacity of the company's own mills.

In addition, the system encourages closer integration between the sales planning department and the supply chain management department, by reducing interdepartmental conflicts and improving the quality of decisions regarding

the marketing mix, intended to maximize the overall economic benefits for the company.

Other advantages that are worth mentioning concern the vast array of managerial decisions that benefit from the introduction of an optimized logistic plan, among which are:

- the possibility of assessing the feasibility of the restocking plan devised by the sales department;
- the definition of contracts with the suppliers of raw materials and pack- aging, as well as with third-party co-packers;
- the definition of the budget, formulated on a rolling basis and with a planning horizon of 18 months;
- the hiring and training plans for seasonal labor;
- the possibility of assessing the impact of expansion plans for the production capacity, obtainable through the expansion of production and packaging lines within existing plants, the construction of new manufacturing plants and the activation of relationships with new co-packers;
- the optimal allocation of the demand to the plants consistent with the plan of distribution to the markets.

17.Explain the use of Logistics planning in the packaging industry

The second case study considered here refers to an enterprise that produces integrated solutions for processing, packaging and handling liquid food products. The highly complex logistic and production network includes approximately
60 plants, 1000 different end products and 200 market areas assigned to more than 100 sales divisions, called *market companies*.

The organizational structure entails considerable independence in decision making on the part of local market companies, which leads them to assign production to their preferred plant, without taking into account global optimization goals at the enterprise level. The policy of independence in decision making also means a lack of homogeneity of costs and service performance among the various production sites. Delivery costs and times depend on the number of distinct products manufactured at a given plant, and the lack of homogeneity in the allocations is considerable, since the number of products assigned to the various production plants may vary from 10 to 100.

A decision support system for medium-term logistics planning was developed, with functionalities and features not much dissimilar from those described

in the previous section. Its primary objective was to allocate the demand to the plants so as to minimize the overall logistic and production costs.

The optimization model representing the *intelligence* component of the sys- tem considers for each plant the two stages of production required by the processing technology of the company. For each stage, the model takes into account:

- the production costs by line and by plant;
- the cost of procurement of raw materials from suppliers;
- the transportation costs for raw materials;
- the transportation costs from plants to markets for end products;
- the limitations imposed by the capacity of the production lines;
- the limited availability of some technological components required by the production process;
- the technological maps that describe all possible combinations of lines and products, along with the relative processing times.

The optimization model also determines transfers from one plant to another for some technological process components, whose total number is limited by cost considerations. The solution of the resulting model leads to an optimal logistic and production plan, which includes the following choices:

- the supply plan of raw materials from suppliers, for each plant and for each period;
- the production plan for each line, plant, product and period, and there- fore the optimal allocation of the demand to plants;
- the distribution plan for each product, plant, market area and period;
- the allocation and possible transfer between plants of critical techno- logical process components in each period.

UNIT – V

FUTURE OF BUSINESS INTELLIGENCE

PART – A

1. What is the future of Business Intelligence?

The future of business intelligence centers on making BI relevant for everyone, not only for information workers and internal employees, but also beyond corporate boundaries, to extend the reach of BI to customers and suppliers.

2. Define web based business intelligence?

Web-based business intelligence and dashboards were rated the highest, with predictive analytics and alerting also at the top. Surprising to me, Microsoft Office Integration, BI Search, and Mobile BI were selected by *only* a small percentage of survey respondents.

3. What are the factors to be considered before to proceed an innovation with BI?

- Embrace Items in the upper-right quadrant

- Adopt Where Appropriate Items in the lower-right quadrant
- Evaluate and Test Items in the upper-left quadrant are relatively new but will have'a profound impact on user adoption.
- Monitor and Understand Items in the lower-left quadrant are so new that they may be riskier investments.

4. What do you mean by predictive analysis?

Traditionally, predictive analytics has been a backroom task performed by a Limited few statisticians who would take a snapshot of the data (either from a data warehouse or from a purpose-built extract from the source system), build a model, test a model, finalize it, and then somehow disseminate the results.

5.What is the use of Market Basket Analysis?

Market basket analysis helps retailers understand which products sell together and provide product recommendations. In the past,

Corporate Express provided these recommendations by logical product pairings. So if a customer ordered a stapler, the online store would rec-omrnend a staple remover as the marketing team had marked this as a complementary product.

5. What are the benefits of a BI Search?

BI Search offers a number of promising benefits to business intelligence:

- Simple user interface.
- A more complete set of information to support decision making, with
  . the Integration of structured (quantitative) and unstructured content (textual). Structured data refers to the numerical values typically cap-tured in the operational systems and subsequently stored in a data warehouse. Unstructured content refers to information stored in tex-tual comment Fields, documents, annual reports, websites, and so on.

- Users can find what they need through search, rather than through navigating a long list of reports.

6. Define Text Analytics?

Text analytics is closely related to search in that unstructured information or text can be transformed into quantitative data. For example, *it* allows for searching of information in a comment field to see how many times a customer praised a particular product. Text analytics is the numerical analysis of textual information.

7. What is the use of BI Search interface?

A BI Search interface promises to change the way users access information.

Picture a Google interface to BI.

The added benefit is that in addition to displaying reports coming from the £1 server, the search engine will also list textual information that may be relevant a customer letter, sales can notes, headline news.

8. What is BCBS?

BlueCross Blue Shield (BCBS) of Tennessee (TN) is an early adopter of these capabilities. BCBS of TN is a not-for-profit provider of health insurance. In 2006, it paid $7 billion in benefits for its 2 million commercial members. Managing claims and negotiating rates with providers is critical in ensuring BCBS can meet its obligations to the members it insures.

9. What is Report – based Interactivity?

Report-based interactivity is a term that warrants a better name.
It is used to create report as well as to format and navigate the report. It also don't fully capture the value of this capability. I suspect poor terminology and lack of awareness also explains why survey respondents ranked this capability on the low end of importance for emerging technologies . So after much thought and brainstorming with some colleagues, I will refer to this capability as "rich reportlets."

10. Define Rich reportlets?

Rich reportlets are powered by Web 2.0 technologies to create rich Internet applications (RIA). When Bl suites were first re-architected, report consumers could only view a static page. Given how static a display this was, more sophisticated users would export the data to Excel for analysis.

With rich reportlets, someone accesses a report over the Web but in a much more interactive

11. Define Technical innovation?

Technical innovation is only one aspect that will help increase Bl's prevalence. In discussing future plans with many of the case study com- panies, much of their concern was not about technology, but rather, in Finding new ways to use BI to address common business problems.

PART - B

1. Expalin about the Future of Business Intelligence?

The future of business intelligence centers on making BI relevant for everyone,not only for information workers and internal employees, but also beyond corporate boundaries, to extend the reach of BI to customers and suppliers.

As the Successful BI case studies have demonstrated, when best practices are applied, BI usage can explode beyond the paltry 25% of employees today to a much more prevalent business tool. It will take cultural shifts, new ways of thinking, and continued techni cal innovation.

Business intelligence has the power to change people's way of working, to enable businesses to compete more effectively and Efficiently, and to help nonprofits stretch their dollars further, All of this is possible based on insights available at the click of a mouse, push of a button, or touch of a screen.

As discussed , much of the key to successful business intelligence has to do with the people, processes, and culture. Don't rely on technical innovation alone to solve the biggest barriers to Bf success, but by all means, do get excited about the innovations that will make BI easier and more prevalent. 'BI as a technology has changed dramatically since its inception in the early 1990s.

2. What are the emerging technologies of business Intelligence?

As part of the Successful BI Survey, respondents were asked to choose items from a list of emerging technologies that they believe will help their companies achieve greater success.

The majority of survey respondents believe web-based dishoards, alerting, and predictive analytics will allow greater success. These are considered most important in helping companies achieve greater success. Web-based business intelligence and dashboards were rated the highest, with predictive analytics and alerting also at the top. Surprising to me, Microsoft Office Integration, BI Search, and Mobile BI were selected by only a small percentage of survey respondents.

The view according to business users, however, is slightly different. Business users account for *only* 10% of the survey respondents. Those who describe themselves as hybrid business-IT personnel account for 23% of respondents. I have specifi*cally* excluded IT personnel and hybrids , to show the gap in perceived importance of certain technologies. When viewing responses only for business users, the importance of Microsoft Office integration moves to the top of the list, while alerting moves down.

Some of these differences can be explained by gaps in understanding of the feature benefits, but also by a respondent's point of view. For example, IT professionals have been burned in the past by the thousands of disconnected spreadsheets and the ensuing data chaos. As Microsoft Office integration with BI has improved dramatically in 2007, IT professionals may not realize that spreadsheet-based analysis Can now be "safely" enabled and can be something to be embraced for

knowledge workers familiar not just with Microsoft Excel, but also with Word, Outlook, and PowerPoint. In a similar fashion, if you are a BlackBerry user, you may rate Mobile BI high. While web-based business intelligence may have been introduced in the late 1990s, these solutions only reached the rich functionality of their desktop predecessors in the 2005 time frame. A number of companies are not yet on the latest releases, though, and still use client/server BI deployments.

Depending upon Where a survey respondent is in their web-based BI deployment will
influence how this capability was rated.

At The Data Warehousing Institute's (TDWI) Executive Summit in February 2007, I participated in a panel on the role of emerging technologies in extending the reach and impact of business intelligence. Attendance was restricted to BI directors and executive sponsors who influence their company's Bl strategy. Attendees could vote on a limited number of items that they thought would have the biggest impact in the next fe\v years. The most highly ranked item: performance management and predictive analytics, The things that got few to no votes were BI search, dashboards, and rich Internet applications, contrary to what I believe will have the biggest impact. As we delved into what these technologies mean, and in some cases, demonstrated them, the perceptions changed Significantly. In this way we sometimes don't know the impact any of these capabilities will have until the technology has become more mature and the industry understands it better. If you think about the way breakthroughs like the iPod and YouTube have revolutionized their markets, when they first were introduced, they were met with a mixture of fascination and confusion, without a clear understanding of where they would lead. Recent BI innovations must go through a similar process of the industry first understanding their potential.

The Y axis, then, indicates the degree to which an enabling technology will take BI's reach closer to 100% of employees. Business impact and BI prevalence are not linearly correlated, however, One enabling technology, such as predictive analytics, may yield a big value for a single decision, say, a $4 million savings by better marketing campaign management. Another enabling technology such as BI embedded in operational processes may affect thousands of users, each of whom makes dozens of decisions on a daily basis; the monetary value of these individual decisions It may be small when measured in isolation, but enormous when taken in aggregate. The size and shading of the bubbles give an indication of which items have a bigger single value. The bigger the bubble and darker the shading, the bigger the impact on a single decision or person.

3.How to proceed an innovation?

For each innovation, consider both the technical maturity and the business impact to decide how to proceed:

- Embrace Items in the upper-right quadrant show innovations that are mature and that should be embraced as they will help speed user adoption across multiple user segments.
- Adopt Where Appropriate Items in the lower-right quadrant show innovations that are mature but that may serve only specific segments of users. Mobile BI is an example of this; the technology is more mature than BI search, for example, but benefits only those users who have smart phones such as a BlackBerry.
- Evaluate and Test Items in the upper-left quadrant are relatively new but will have'a profound impact on user adoption. BI Search is a good example of this. The technology is very new and not well under stood. A number of usability and performance issues still need to be worked out, but the potential impact on user adoption is enormous.
- Monitor and Understand Items in the lower-left quadrant are so new that they may be riskier investments. Items here are less proven and have less market adoption.

It portrays broad industry maturity of these capabilities and the degree to which most vendors offer the capabilities. For clarity, I have selected only certain innovations; it is not meant to be an exhaustive list of all things going on in the industry . All items are in the context of business intelligence as a technology. So while performance management is certainly a mature concept and technology, the integration of performance management with business intelligence is still a work in progress, leaving this item positioned slightly behind web-based BI and Microsoft Office integration on the maturity spectrum.

4. Explain in detail about Predictive Analytics? (APR/MAY 2019)

Data mining, statistical analysis, and predictive analytics are nothing new. These technologies are well established and are used in a number of different applications such as fraud detection, customer scoring, risk analysis, and campaign management. What's changed is how they have become integrated in the Bl platform. Traditionally, predictive analytics has been a backroom task performed by a Limited few statisticians who would take a snapshot of the data (either from a data warehouse or from a purpose-built extract from the source system), build a model, test a model, finalize it, and then somehow disseminate the results. While the expertise to build such models remains a unique skill set, the indus try recognizes that the results of the analysis should be more broadly shared, not as a stand-alone application, but rather, as an integral part

of the BI solution. This does not mean that predictive analytics software will become "mainstream," but rather that the results of such analyses can be readily incorporated into everyday reports and decision making. The analysis, then, is what needs to become mainstream.

Predictive analytic tools from different vendors do continue to differ significantly in how they work and in what information is stored in the database versus calculated and presented in a report or incorporated into an operational process.

At Corporate Express, for example, predictive analytics are being used to improve customers' online shopping experience.

Market basket analysis helps retailers understand which products sell together and provide product recommendations. In the past, Corporate Express provided these recommendations by logical product pairings. So if a customer ordered a stapler, the online store would rec omrnend a staple remover as the marketing team had marked this as a complementary product.

ln analyzing the data, though, it turned· out that what was most often purchased with a stapler was not a staple remover, but rather a ruler, tape dispenser, and a wastepaper basket-s-items that indicate a purchase for a new employee. With the manually associated product recommendations, there was no significant impact on sales. Leveraging microStrategy and SPSS, Corporate Express tested a new market basket option. They analyzed past shopping carts and produces

recommendations to ensure the greatest lift. As a result, the average order size for market basket pairings doubled (versus those orders with no pairings), and the market basket application is expected to generate an incremental gross profit of more than $2 million in 2007.

Dow Chemical also has begun extending the reach of predictive analytics with SAS's JMP product (pronounced "jump"), a solution that combines visual analysis with statistics. Dow uses Business Objects and Cognos Powerplay as enterprise reporting and analysis standards. Through these tools and the data in the data warehouse, Dow began looking at the high cost of railroad shipments:
$400 million annually across North America."

A team of statistical experts studied the variables that most affected these costs and pulled data from the data warehouse and external data sources into SAS JMP. By benchmarking current payments versus industry norms, the analysis showed Dow was overpaying by 20%, of $80 million. In entering new contracts, the purchasing department now uses the software to predict appropriate rates, enabling them to negoti ate more aggressively,

For both Corporate Express and Dow Chemical, the move to predic tive analytics has been evolutionary. The underlying information archi tecture and a culture of fact-based decision making had to first reach a level of maturity and data quality before predictive analytics could be

embraced. While both companies have been doing statistical analysis for decades, the degree to which predictive analytics has now been incorporated into daily processes (online store at Corporate Express and purchasing negotiations at Dow Chemical) reflects the degree to which predictive analytics has shifted from the backroom to the front line, with the most casual of users deriving value from such analytics.

5. Explain in detail about BI Search? (APR/MAY 2017)

BI Search offers a number of promising benefits to business intelligence:

- Simple user interface.
- A more complete set of information to support decision making, with the Integration of structured (quantitative) and unstructured content (textual). Structured data refers to the numerical values typically cap tured in the operational systems and subsequently stored in a data warehouse. Unstructured content refers to information stored in tex tual comment Fields, documents, annual reports, websites, and so on.

- Users can find what they need through search, rather than through navigating a long list of reports.

A BI Search interface promises to change the way users access information. Picture a Google interface to BI. Without any training in a BI tool, users can enter a phrase such as "Recent sales for customer A" and then be presented with either a list of predefined reports or, in some cases, a newly generated query. The added benefit is that in addi tion to displaying reports coming from the £1 server, the search engine will also list textual information that may be relevant-a customer letter, sales can notes, headline news. When search capabilities are combined with text analytics, a report may include numerical data that scans the comment field to indicate number of complaints with number of posi tive comments. Never before has such unstructured data been so nicely accessible with structured or quantitative data.

If the integration of search and BI is successful, it is yet another innovation that will make BI accessible and usable by every employee in an organization. According to Tony Byrne, founder/president of EMS Watch, a technology evaluation firm focusing on enterprise search and content management systems, search as a technology has existed for more than 50 years. *Consumer* search (Google and Yahoo, for exam- ple) as a technology emerged with the Internet in the mid-1990s. In many respects, the success of consumer search has helped spur hype around *enterprise* search, ill which companies deploy search technology internally to search myriad document repositories.

and thus has helped business users to understand the possibilities. To illustrate the point, note that BI search was selected by only 27% of the Successful BI Survey respondents as a capability that would help foster greater success. Yet in discussing these technologies with individual executives who don't currently use business intelligence, a Google-like interface to BI generated the most enthusiasm.

6. Explain in detail about Text Analytics ? (APR/MAY 2017)

Text analytics is closely related to search in that unstructured infor mation or text can be transformed into quantitative data. For example, *it* allows for searching of information in a comment field to see how many times a customer praised a particular product. Text analytics is the numerical analysis of textual information.

Text analytics has existed for 25 years but with usage in limited sectors, particularly, the government. The convergence of search with business intelligence first emerged in 2006. Google is not the only enterprise search solution that BI vendors support but it is one that has the most consumer recognition

Despite all the improvements in data warehousing and BI front-end tools, users continue to feel overwhelmed with reports yet undersatisfied with meaningful information. They don't know what's available or where. Similar reports are created over and over because users don't know which reports already exist or how, for example, the report "Product Sales" differs from "Product Sales TID." Some of the most valuable information is hidden in textual data.

The incorporation of text analytics with traditional business intel ligence is still in its infancy. I place both Bi search and text analytics close to the Monitor and Understand quadrant but in the Evaluate and Test quadrant . Again, both technologies independent of Bl have existed for decades; it is that convergence with BI that is new. While the convergence is still relatively immature, the promise it brings for BI to reach more users and in the value of incorporating textual data is enormous.

The number of customers taking advantage of the BI Search and text analytics integration is only a handful. BlueCross Blue Shield (BCBS) of Tennessee (TN) is an early adopter of these capabilities. I BCBS of TN is a not-for-profit provider of health insurance. In 2006, it paid $ 7 billion in benefits for its 2 million commercial members. Managing claims and negotiating rates with providers is critical in ensuring BCBS can meet its obligations to the members it insures. While the insurer has had a mature business intelligence deployment for ten years," Frank Brooks, the senior manager of data resource management and chief data architect, recognized that there was value

in bringing the text data stored in comment fields from call center notes together with information in the data warehouse if Given how new the technology *is*, Brooks asked their Bl vendor, Cognos, along with IBM (who produces the search solution Omni Find) and SAS (who offers text analytics solution Text Miner) to work together to develop several prototypes and show the business users the concept of bringing BI, enterprise search, and text analytics together. WitI1 this capability, a business user can enter the key word "diabetes" in the Omni Find search box and be presented with a ranked list of things such as:

- Cognos reports and OLAP cubes that show claims paid for diabetic treatments
- Call center notes that involve diabetes
- New research on improving care for diabetes patients

The business was enthusiastic, There has been a high degree of collaboration between Be BS of TN and its information technology partners

category are on a steady decline, whereas consumer and small business segments show strength.

Creating this kind of display with standard BI software is theo retically possible, but one that would take many, many more steps. As well, if J am uncertain as to the best way to display the information, advanced visualization software can make suggestions. The capability to create easily such advanced visualizations is generally not available in BI suites. Users must rely on specialty products. In a theme similar to predictive analytics and search, visualization software has existed for years; the change is in its convergence with business intelligence such that advanced visualizations are appearing in dashboards and reports. In this regard, the emphasis for BI tools is changing from a focus of simply "getting to the data" to "what insights can I discover from the data and how can the most information be displayed in the smallest space."

7. Explain in detail about Rich Report lets?

Report-based interactivity is a term that warrants a better name.

It is an active report used for formatting and navigating reports. These are similar terms that also don't fully capture the value of this capability. I suspect poor terminology and lack of awareness also explains why survey respondents ranked this capability on the low end of importance for emerging technologies . So after much thought and brainstorming with some colleagues, I will refer to this capability as "rich reportlets." The difference in power and appeal with rich report- lets versus, say, green-bar paper reports and much of what is currently deployed over the Web, is comparable to the difference between a black Ford Model T and a red Mercedes sports coupe.

Rich reportlets are powered by Web 2.0 technologies to create rich Internet applications (RIA). When Bl suites were first re-architected for the web, report consumers could only view a static page. Given how static a display this was, more sophisticated users would export the data to Excel for analysis. Less sophisticated users would submit requests to IT or to the Bl team to modify the report design. The web in this case is only a delivery vehicle for data; it does not facilitate user adoption

and insight. With rich reportlets, someone accesses a report over the Web but in a much more interactive and appealing way. At a simple click, data can be re-sorted, filtered, or graphed, without having to launch a complicated report editor. With the use of either Adobe Flex or Macromedia Flash, these reports come to life in ways that make business intelligence fun. I have seen, for example, a bubble chart that displays bubbles dancing across the screen as the time axis marches onward. Such animation makes BI appealing as well as insightful as users see the trend in action. In this regard, the term "report" doesn't do justice to the capability that is more akin to a mini application.

This type of interactivity affects all BI users, whether casual or power users. The appeal makes BI more engaging, and while some technologists at: the importance of this, when other barriers to adoption exist, appeal matters. A lot the ability to interact with the data in a simple and intuitive way facilitates greater insight at the bands of the decision maker. The report consumer is not forced to delay this insight until a power user can modify the report. Lastly, the cost of ownership is lowered because a single reportlet can be "tweaked" to that decision maker's needs, without IT having to maintain thousands of individualized reports.

8. Explain in detail about the future beyond technology in BI?

Technical innovation is only one aspect that will help increase Bl's prevalence. In discussing future plans with many of the case study companies, much of their concern was not about technology, but rather, in Finding new ways to use BI to address common business problems. For the more large-scale deployments, some expressed concern about man- aging the risk of making any kind of major change to such a business critical, complex application.

With success, of course, comes greater demands on the systems and the people. Ensuring an effective way of prioritizing competing requests warrants constant attention.

One business leader expressed frustration at his department's inability to make wise investments, while witnessing other departments, working in more unison and getting more value from business intelligence. Yet he remains optimistic that his business will get there and that BI will be the first thing people look at, even before email.

"To have one screen I can get to with a single click, that shows sales, margin, price, opportunities in graphical form, with drill down-s-that

would be magic!" His comments remind me that the technology is sometimes the easy part.