POLITECNICO DI TORINO

Department of Engineering and Management

Master of Science in Engineering and Management

Master of Science Thesis

Implementation of a Business Intelligence Solution in a Manufacturing Company: A Predictive Analysis Approach

Tutors

Luca Settineri (Academic)
Fabio Ghi (Enterprise)

Candidate

Carlos Andrés Novoa Dita

October 2019

A veces me asustan mis sueños, porque suelen ser tan grandes que no sé por dónde empezar a hacerlos realidad.

Acknowledgments

I would first like to express my very profound gratitude to all my family, especially to my parents, Melbis Novoa and Merlys Dita, and to my little brothers, Andrea, Hector, and Luis Angel. I would also like to thank my girlfriend, Carmen Barbosa, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Secondly, thank my thesis advisor, Ph.D. Luca Settineri, who helped me whenever I ran into a trouble spot or had a question about my research and steered me in the right direction whenever he thought I needed it. In the same way, I thank the people of Bios Management, who led by Fabio Ghi, opened their doors and offered me the opportunity to grow professionally.

Finally, thanks to everyone who supported me and advised me when I needed it. In particular, thanks to my friends in Torino, Magangué, Cartagena and whatever place in the world because despite the distance and the time they always were there.

Contents

List of figures

List of tables

Abstract

In order to have an excellent performance in the future, companies must analyze in a meticulously way their historical data to understand the past and current behavior of their customers. The principal problem of this lies in the immeasurable amount of information that is usually generated every day (Big Data). With the aim of solving the problem mentioned above, most organizations are using Business Intelligence (BI) Tools for managing their data.

BI helps companies to convert raw data into valuable information. Currently, many BI Platforms are implementing Predictive Analytics Tools to improve the accuracy of their forecasting process. Thus allow an enhance the performance of their customer's decision-making process. The objective of this work is to implement BI solution in a manufacturing company, with a particular focus on the Predictive Analysis in order to which stand how it can improve the decision-making process and help in the decisions relating to the operational level.

CHAPTER 1:

INTRODUCTION

In the age of big data, a good understanding of the business helps the companies achieve success. The organizations know that the knowledge about the market (customers, suppliers, and competitors) is an essential and valuable asset. The decision-making process based on the knowledge obtained of gathering, managing, visualization, and interpretation of the data is crucial to companies can reach favorable results. It is not only to understand the behavior of the market, but also to be able to predict the possible changes that it may have. When a company has the ability to predict variations in the behavior of its customers and suppliers, it can adapt more quickly and be more competitive against its competitors. To achieve this, many companies are using predictive analysis; a branch of data mining allows companies to make predictions of future events through historical data.

This thesis presents the analysis and implementation of the activities carried out during the internship in the Italian company Bios Management, where it was developed and implemented a Business Intelligence solution for analyzes the processes performed in Carioca, an Italian company dedicated to the production of articles for writing, coloring, and drawing. The analysis, including also the implementation of the BI tool Board with a particular focus on the predictive analysis to facilitate the decision-making process.

Board Toolkit is an all-in-one Decision-Making Platform that provides all of the tools required to create and update databases, data presentations, analyses, and

process models in a single visual and interactive environment. A powerful tool including in the Board platform is BEAM, which (according to Board) "*seamlessly integrates advanced and predictive analytics with business intelligence and performance management, making it easier than ever for anyone to gain powerful business insights and take action quickly.*"

Carioca had no analytical model that allowed it to have an idea of his performance. Then, it was not possible to create an accurate prediction of the possible events in the future; all predictions about market behavior and growth were based on the intuition of managers. Through the Board, it was possible to create a predictive analysis that allowed estimating the optimum production levels during the different months of the year and at the same time, looking for possible solutions to problems such as seasonality.

Initially, this work presents an in-depth investigation of the concept of Business Intelligence and its essential contribution to the decision-making process. In the second part, an introduction is made to the BI tool, Board, where the functions of its most important components are explained. Once the necessary concepts have been clarified, a brief introduction to the target company is made. Subsequently, the implementation of Board and its BEAM tool is carried out, and finally, an analysis of the results obtained is performed.

CHAPTER 2:

LITERATURE REVIEW

2.1 DECISION MAKING

Decision-making is one of the most necessary tasks each day and every. Take an incorrect selection may want to lose a lot of quantity of cash the organization or even worse lead it goes bankrupt. As argued by way of Businska & Supulniece (2011), Decision making is the exercise which it uses the know-how obtained of reasoning about information saved to facilitate problem-solving. According to Bose (2009), it is feasible to classify the decisions in three types, strategic, tactical, and operational:

- **Strategic decisions** are taken questioning in the future; they are an emphasis on the long-term plans of a company.
- **Tactical decisions** center the attention on the high-quality way of implementing strategic decisions.
- **Operational decisions** guarantee that unique duties relate to the everyday operation hare carried out efficaciously and efficiently.

Figure 1 indicates the framework shape of the Decision-making method of Graves et al. (2013) "It describes a sequence of iterative steps in which risks and options a particular; near-term selections are made and implemented, and prerequisites are monitored to help refine those plans over time.". That is that selection type impact in the analytical techniques used for decision-making (Isik et al., 2013).

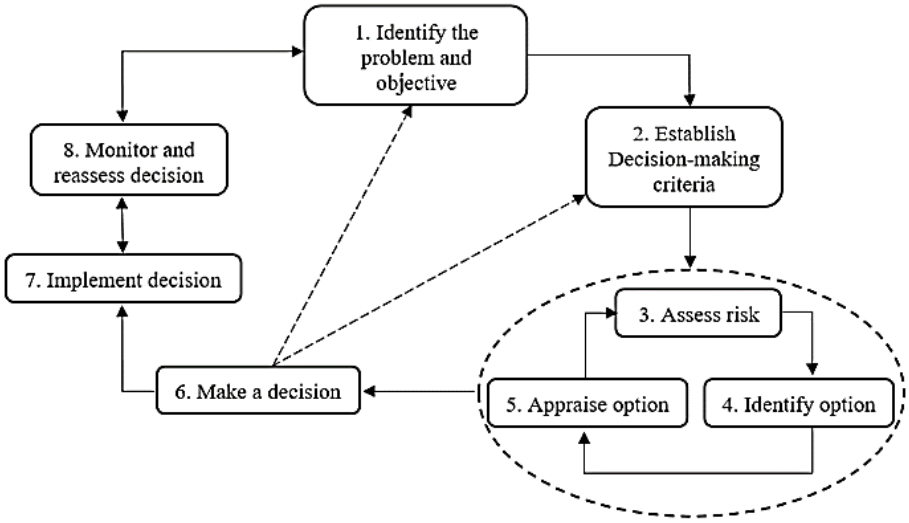


Figure 1. Decision-Making framework

2.1.1 Decision-making process and Business Intelligence

In the last section, it was possible to understand the concept and the importance of the decision-making process for an organization and realize bad decisions could do lose a lot amount of money the company or even worse lead it goes bankrupt. Due to the above, every organization should be careful at the moment of take their own and help themselves with the necessary tools to achieve their purposes. Here is where the Business Intelligence enters into action. The BI helps since it can extract crucial facts from a vast amount of data and transform them into actionable information that enables companies to make strategic decisions that can improve operational efficiency and business productivity.

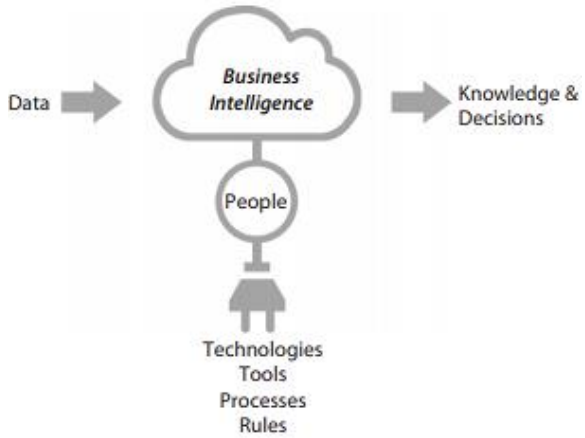


Figure 2. The relation between Business Intelligence and Decision-Making

Figure 2 describes the relationship between Business Intelligence and Decision-Making, where, according to Vizgaitytė & Rimvydas (2012), the data is transformed in knowledge and decisions through certain technologies, processes, tools, and rules (Business Intelligence) use for the people. In summary, it possible to say that Business intelligence can arm companies with rich data resources that can help them achieve their business goals and targets by guiding timely strategic decisions.

2.2 BUSINESS INTELLIGENCE

2.2.1 Definition of Business Intelligence

There are many explanations for Business intelligence (BI), but all of them conclude that it is how the data is used to make better decisions in an organization. As stated by Azma & Mostafapour (2012), the first definition of Business Intelligence was introduced in 1967 for the American professor Harold Wilensky[1]. He defines BI as the collection, processing, interpreting, communication, and data analysis of the companies that generate new knowledge and learning to make better decisions in the organization.

According to Chee, et al. (2009), over the years different authors gave to BI different concepts (Table 1) show a comparison between some authors and BI versions), but these definitions generally fall in three main groups: management (Process) aspect, technological aspect, and product aspect. Under the technological approach, the BI is considered as a group of software, solution, and all type of technological tools that help organizations in the decision-making process. That means that the BI importance is not in the process itself but on the technological tools that allow collecting, store, analyze, and mining the corporate data. The managerial aspect show BI as a process in which the most important things are in the coordination and management of the processes by which different information

[1] Organizational Intelligence: Knowledge and Policy in Government and Industry. New York: Basic Books, 1967.

sources (both inside and outside the company) can be integrated and analyzed to support the decision-making process. Moreover, from the product approach, BI is seen as a result that emerges from advanced and optimal processing of the data, information, and knowledge that support the decision-making and the assessing of the performance.

Table 1. Definition of BI: comparison between some authors and BI vendors

BI Vendor/Author	Definition of BI
Turban et al. (2007)	An umbrella term that encompasses tools, architectures, databases, data warehouses, performance management, methodologies, and so forth, all of which are integrated into a unified software suite.
Moss and Atre (2003)	It is an architecture and a collection of integrated operational as well as decision-support applications and databases that provide the business community easy access to business data.
Chang (2006)	The accurate, timely, critical data, information, and knowledge that supports strategic and operational decision making and risk assessment in uncertain and dynamic business environments. The source of the data, information, and knowledge are both internal organizationally collected as well as externally supplied by partners, customers, or third parties as a result of their own choice.
Gangadharan and Swami (2004)	The result of the in-depth analysis of detailed business data, including database and application technologies, as well as analysis practice.
Kulkarni and King (1997)	A product of analyzing business data using business intelligence tools. It emerges as a result of this analysis.
Moss and Hoberman (2004)	The processes, technologies, and tools needed to turn data into information, information into knowledge and knowledge into plans t hat drive profitable business action. BI encompasses data warehousing, business analytics tools, and content/knowledge management.
Adelman and Moss (2000)	A term encompasses a broad range of analytical software and solutions for gathering, consolidating, analyzing, and providing access to information in a way that is supposed to let an enterprise's users make better business decisions.

Gartner Research (Hostmann 2007)	An umbrella term that includes the analytic applications, the infrastructure, and platforms, as well as the best practices.
IBM (Whitehorn & Whitehorn 1999)	An umbrella term that broadly covers the processes involved in extracting valuable business information from the mass of data that exists within a typical enterprise.
Business Objects (2007)	The use of an organization's disparate data to provide meaningful information and analysis to employees, customers, suppliers, and partners for more effective decision making.
Cognos (2007)	Business intelligence brings people and data together, offering a variety of ways to see the information that backs fact-based decision-making.
SAS Institute (2007)	Delivering the right information to the right people at the right time to support better decision making and to gain competitive advantage.
Oracle (2007)	A portfolio of technology and applications that provides an integrated, end-to-end Enterprise Performance Management System, including financial performance management applications, operational BI applications, BI foundation and tools, and data warehousing.
Informatica, Teradata, MicroStrategy (Markarian, Brobst & Bedell 2007)	An interactive process for exploring and analyzing structured, domain-specific information (often stored in a data warehouse) to discern trends or patterns, thereby deriving insights and drawing conclusions.

Base on the different definitions of authors and considering the three aspects of BI, Chee et al. (2009) define the term Business Intelligence in the technological approach as a BI system, whereas in the process aspect is considering like the implementation of this BI system. Also, finally, in the product focus, BI is the result of the analysis of business data, which are gathered from different sources. Ponniah (2001) argued that BI from the point of view of the product has some particular characteristics, which are exposed in Table 2.

Table 2. BI characteristics: product approach

Characteristics	Descriptions
Integrated	Must have a single, enterprise-wide view.
Data integrity	The information must be accurate and must conform to business rules.
Accessible	Easily accessible with intuitive access paths and responsive for analysis.
Credible	Every business factor must have one and only one value.
Timely	The information must be available within the stipulated time frame.

2.2.2 Components of Business Intelligence

Olszak & Ziemba (2007) posit that a BI system is composed of the following essential elements:

- **ETL (Extraction-Transformation-Load):** These are the processes in which the data are extracted from the source systems are transformed through operations of "cleaning," uniformity, formatting, and finally, are loaded into the data warehouse. This is the most complex component of the BI system since the fill of the "data warehouse" depends on it.

- **Data warehouse (DW):** It is a collection of corporate information and data derived from operational systems and external data sources. A data warehouse is designed to support business decisions by allowing data consolidation, analysis, and reporting at different aggregate levels. Data is populated into the DW through the processes of extraction, transformation, and loading (ETL).

- **OLAP:** These are tools whose characterizing features are the aggregation of data on multiple dimensions and the ease of navigation of "natural" and parent-child hierarchies. This is combined with the ability to easily manipulate the form in which results are displayed, adding or eliminating

analysis dimensions, thus offering users the ability to create and manage their own data query sessions actively. The result of the queries is multidimensional, since it can arise from the intersection of several dimensions, but is represented in a table format, due to the two-dimensional nature of the visual supports.

- **Data mining[2]:** These are tools which use statistical techniques and machine learning to explore vast amounts of data; through them, patterns are extracted that allow to generate knowledge and, finally, apply these patterns to new data for predictive purposes, in other words, it offers a probabilistic indication of the occurrence of an event.

- **Reporting and ad hoc inquiry tools:** These tools allow and helpt in the process of creating and utilizing different synthetic reports.

- **Presentation layers** that include customized graphical and multimedia interfaces to provide users with information in an easy and accessible form.

[2] It will be deepened in the section 2.5.4

Figure 3 Show the way in which BI works throw an integration of all components described by Olszak & Ziemba (2007)

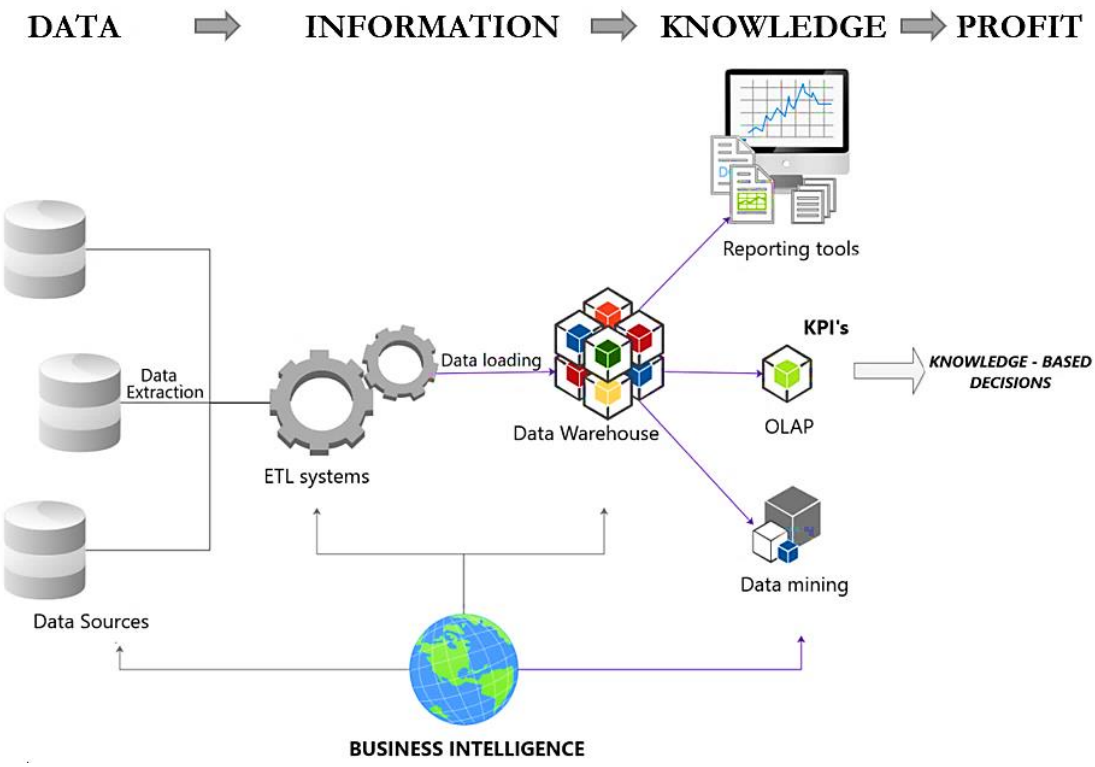


Figure 3. BI components integration

2.2.3 Contemporary applications of BI systems

Different industries have used the advantages of Business Intelligence as transportation, banking, pharmaceutics, etc. Table 3 describes some applications of BI in the areas mentioned above.

Table 3. Contemporary applications of BI systems

BI Applications	Benefits
Transportation Industry: Generally, transportation service providers utilize several tools and platforms provided by Business Intelligence (BI) vendors which	The direct benefits of the usage of a BI solution in the transportation industry are reduction in the turnaround time for preparation of reports, direct and faster access to the data needed to support

enables the delivery of information to decision-makers such as Query Tools Standard, Reporting Tools, Online Analytical Processing (OLAP) tools, Data Visualization Tools, and Data Mining Tools. Airline industry uses text mining to automatically extract useful information from different written resources such as incident reports.	decision-making, analyze the flow of businesses across services, regions, clients, pricing, currencies, and market factors in time, etc. The vast databases maintained by airlines have limited human interpretation, and the terminology appears different to a computer. Thus, the discovery of new, previously unknown knowledge can be found in a timely manner by using text mining. An incident report is prepared whenever an event occurs that might lead to a problem. Text mining of airline incident reports can identify potential dilemma. Text mining can be used with this large set of incident data reported to validate predetermined theories and to cull new patterns of knowledge.
Banking Industry Banking industry relies on the BI platform to make more effective decisions in a few areas such as Customer Analysis, Operations & Financial Analysis, Sales & Marketing Analysis.	BI applications help management to improve operational and strategic decisions based on better and timely information. Potential customers are identified through the analysis of purchasing data. Cross-selling opportunities will be recognized via the analysis of customer behavior.
Retail Industry BI is implemented for demand forecasting in the retail industry by generating reliable estimates for both short term and long-term demand based on the available customer data. Apart from demand forecasting, BI is also used to monitor customer loyalty by evaluating which customers are loyal and which are likely to leave.	When reliable estimates of customer demand are generated, service, and product distribution plans of a company would always be in place to meet its customer expectations. By monitoring customer loyalty, factors that influence their decisions to stay or go could be determined in order to devise better strategies to retain them.
Pharmaceuticals Industry BI systems help pharmaceuticals companies to identify which products are most profitable and monitor customer behavior in purchasing products.	By carefully tracking sales performance and consumer behavior, pharmaceuticals companies are able to set better marketing strategies and ensure proper allocation of marketing funds.

Health Care Industry The implementation of BI in the health care industry has enabled data to be delivered beyond administrative offices and directly to clinical staffs who can make the most use of it. In order to foster a broader adoption, interactive and user-friendly interfaces have been designed to provide users with simple and relevant data like the number of patients, treatments needed, and their hospitalization period.	Business decision making the process has become more effective where users can access any type of information with fast and consistent response time, independent of the data volumes analyzed or questions asked. The inter-operability application in BI reduces the operation cost in the health care industry by eliminating expensive custom-integration in the computing system.
Manufacturing Industry BI systems allow manufacturers to track their inventory usage across location and time by using alerts for instant notification of low inventory levels. Besides this, BI systems allow manufacturers to analyze data from multiple sources in order to set performance goals and create sophisticated profitability and financial models.	With the functionality of inventory monitoring, manufacturers can reduce over-capacity and ensure sufficient supplies for their production. Apart from this, BI systems also help manufacturers in financial management by identifying areas where they can increase profits and improve efficiency.

2.2.4 The market of the Business Intelligence

Nowadays, Business Intelligence has a critical acceptance in the market. According to the latest forecast from Gartner, Inc.[3] Global revenue in the business intelligence (BI) and analytics software market is expected to grow from $17.05 billion in 2017 to reach $29.48 billion by 2022. This growth is due to the increasing usage of data analytics and the rising dependency on data in the decision-making

[3] Gartner Inc. is a global leader in strategic consulting, research and analysis in the field of Information Technology. The main activity consists in supporting the investment decisions of its customers through research, consulting, benchmarking, events and news. Founded in 1979, Gartner is headquartered in Stamford, Connecticut, USA, and has 7,600 members, including more than 1,600 research analysts and consultants, and customers in 90 countries.

process. On the other hand, there is a tendency in small and medium-sized enterprises to adopt BI technologies (See Figure 4). The factors contributing to the high adoption rate for BI in small businesses include: currently these organizations requires more advanced analysis to function and to be able to scale in the markets, employees are more qualified and have the BI skills to help the organizations grow, the barriers and risks of adopting BI is lower in a small company compared with large firms, and finally BI culture is easy to implement in a small company as the employees tend to use analytics and BI apps daily (Columbus, 2018)

Gartner (2017), on the other hand, believes the rapidly evolving modern BI and analytics market is being influenced by seven dynamics:

- **Modern BI at scale will dominate new buying:** Modern BI tools support increased accessibility, agility, and analytical perception at the organization level, this lets in they can attain higher market share and dominate new purchases.

- **New innovative and established vendors will drive the next wave of market disruption:** The emergence of smart data discovery capabilities, machine learning and automation of the entire analytics workflow will drive a new flurry of buying because of its potential value to reduce time to insights from advanced analytics and deliver them to a broader set of people across the enterprise.

- **Need for complex datasets drives investments in data preparation:** Business users want to analyze a diverse, often large and more complex combinations of data sources and data models, faster than ever before. The ability to rapidly prepare, clean, enrich, and find trusted datasets in a more automated way becomes an essential enabler of expanded use.

- **Extensibility and embeddability will be key drivers of expanded use and value:** Both internal users and customers will either use more automated tools or will embed analytics in the applications they use in their context, or a combination of both. The ability to integrate and extend analytics content will be a crucial enabler of more pervasive adoption and value from analytics.

- **Support for real-time events and streaming data will expand use:** Organizations will increasingly leverage streaming data generated by devices, sensors, and people to make faster decisions. Vendors need to invest in similar capabilities to offer buyers a single platform that combines real-time events and streaming data with other types of source data.

- **Interest in cloud deployments will continue to grow**: Cloud deployments of BI and analytics platforms have the potential to reduce the cost of ownership and speed time to deployment. However, data gravity that still tilts to the majority of enterprise data residing on-premises continues to be a significant inhibitor to adoption. That reticence is abating, and Gartner expects the majority of new licensing buying likely to be for cloud deployments by 2020.

- **Marketplaces will create new opportunities for organizations to buy and sell analytic capabilities and speed time to insight**: The availability of an active market where buyers and sellers converge to exchange analytic applications, aggregated data sources, custom visualizations, and algorithms is likely to generate increased interest in the BI and analytics space and fuel its future growth.

Gartner (2017) stated that due to the dynamics mentioned above, the modern BI and analytics market is expected to decelerate; however, from 63.6 percent

growth in 2015 to a projected 19 percent by 2020. Gartner believes this reflects data and analytics are becoming mainstream. The market is growing in terms of seat expansion, but revenue will be dampened by pricing pressure.

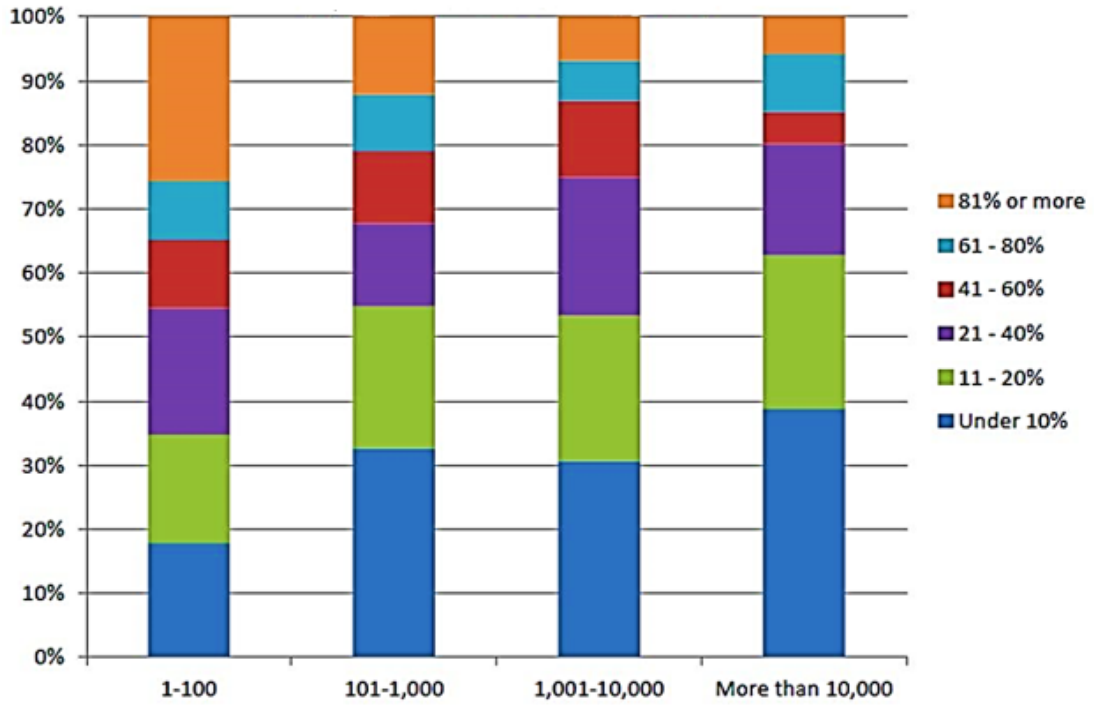


Figure 4. Penetration of BI Solutions today by organization size

The companies have figured out the usefulness of this tool, and every day more organizations adopt BI software for enhancing their process. In Figure 4, we can see that the power of BI had done that also small organizations with up to 100 employees have the highest rate of BI penetration or adoption in 2018.

According to Dresner Advisory Service's 2018 Wisdom of Crowds Business Intelligence Market Study[4], the principal roles that driving business to BI adoption are Executive Management, Operation, and Sales (See Figure 5). Moreover, a lot

[4] Dresner Advisory Services is an independent analyst firm specializing in Business Intelligence and related markets.

of companies that implemented BI software were found significant percentage changes in areas like Human resources (7.3%), Marketing (5.9%), and Sales (5%).

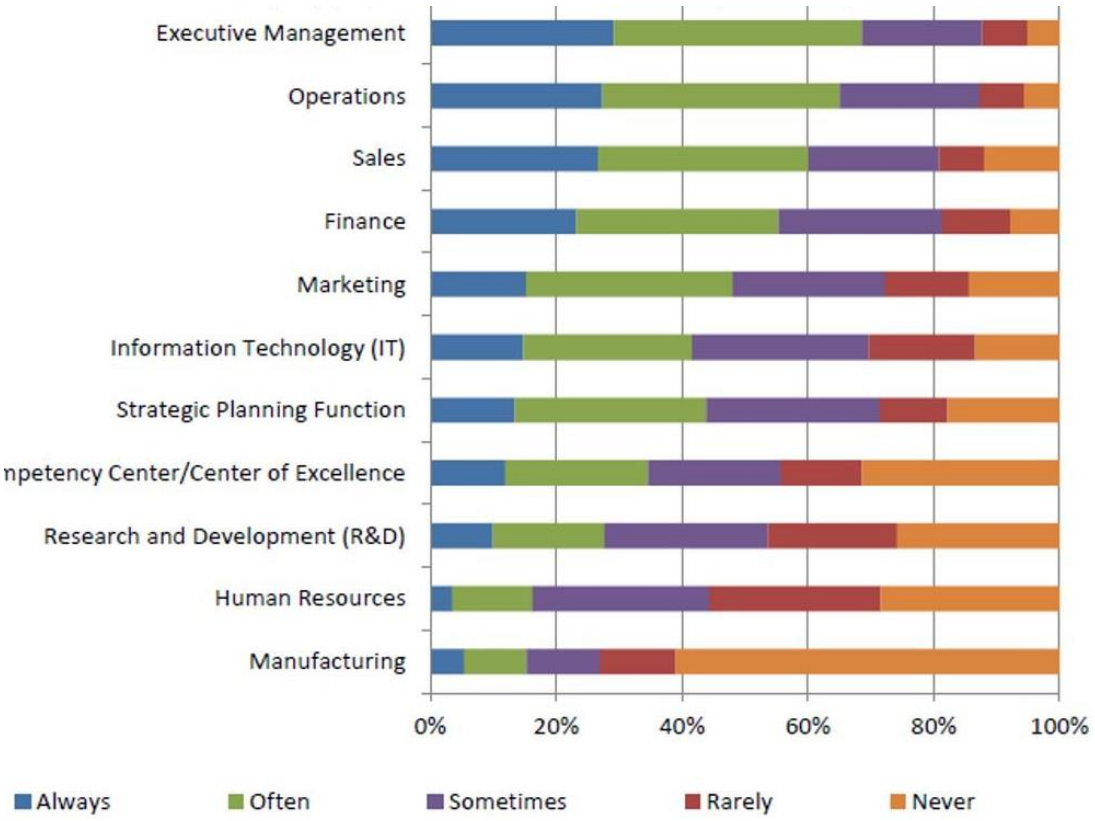


Figure 5. Functions Driving Business Intelligence

Making better decisions, improving operational efficiencies, growing revenues, increased competitive advantage, enhancing customer service, and attaining higher degrees of compliance and risk management are the main reasons that motivate organizations to implement the use of BI software. (See Figure 6 and Figure 7)

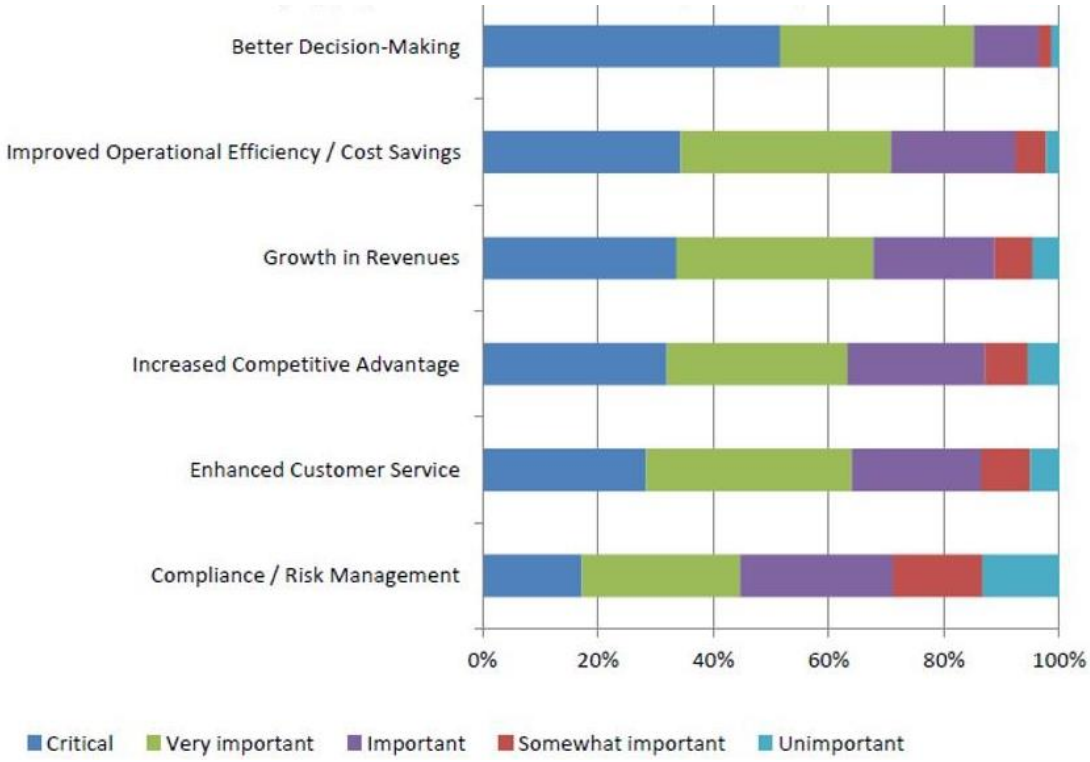


Figure 6. Business Intelligence Objectives

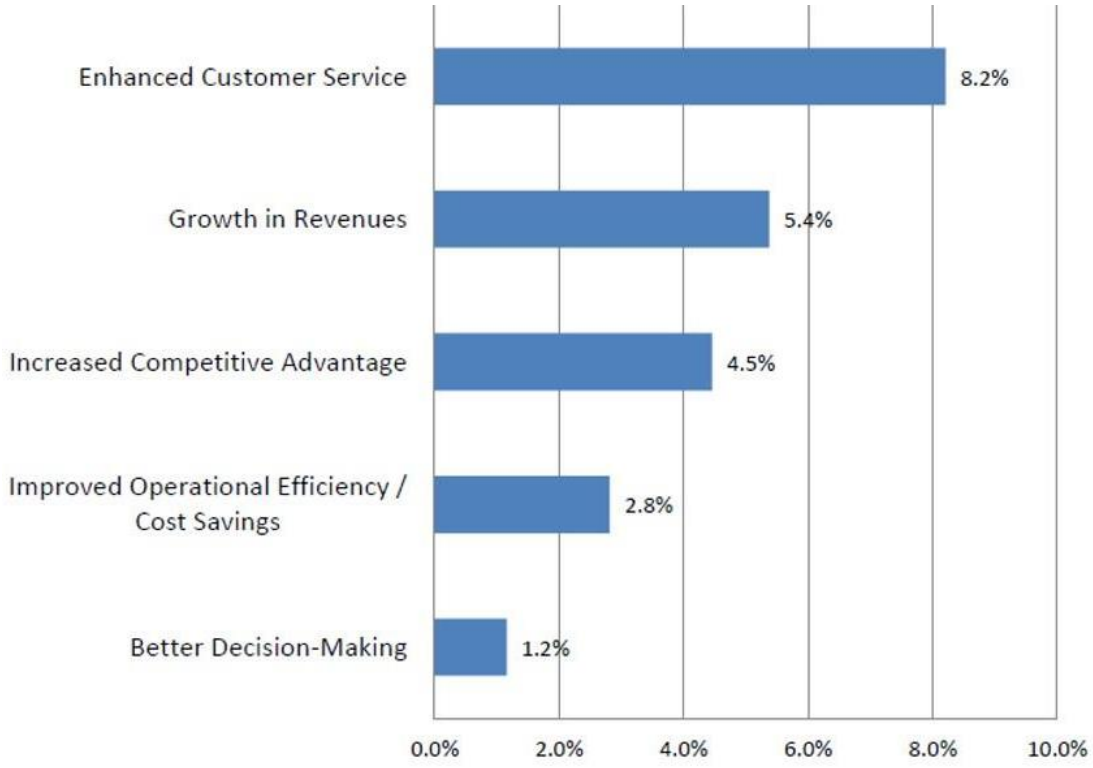


Figure 7. Percent change in BI Objectives 2017-2018

Insurance, Technology, and Business Services vertical industries have the highest rate of BI adoption today. The Insurance industry leads all others in BI adoption, followed by the Technology industry, with 40% of organizations having 41% or greater adoption or penetration. Industries whose BI adoption is above average include Business Services and Retail & Wholesale. The following graphic illustrates the penetration or adoption of Business Intelligence solutions today by industry.

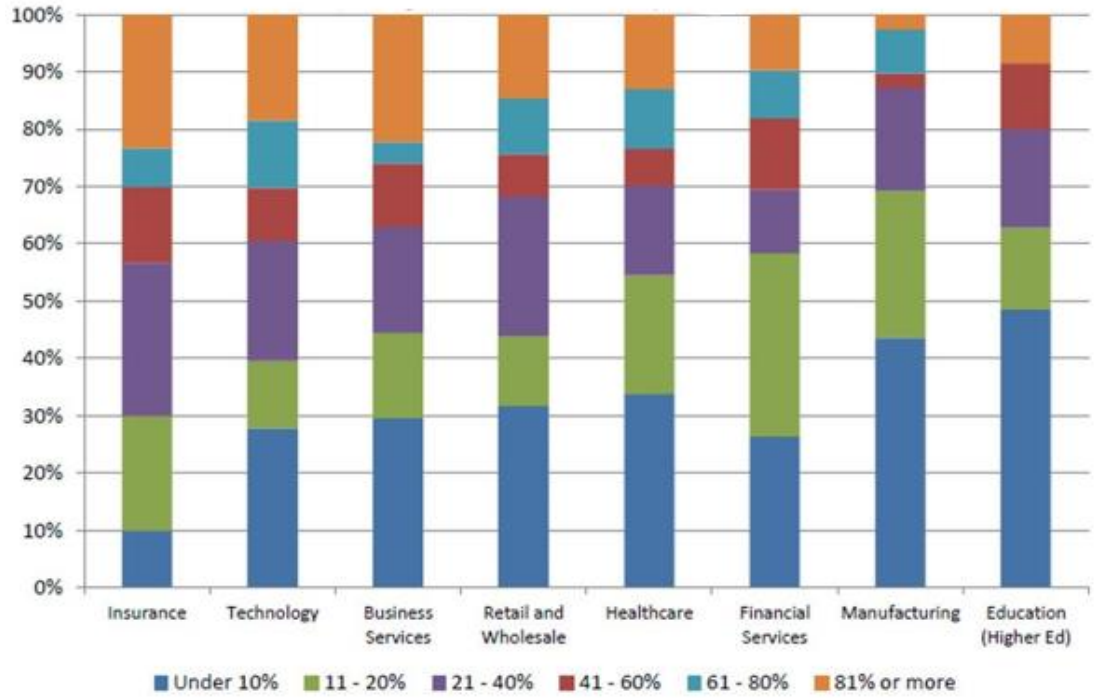


Figure 8. Penetration of Business Intelligence Solutions today by industry

2.3 BI PROJECTS

Unlike other projects, in a BI project, according to Moss & Atre (2003) (Cited in Kříž, 2015), the main feature is provided cross-organizational business analysis capabilities to all business people and all departments in the organization. A BI project can be defined as the ability of the project team to design, solve problems, understand ideas, learn, and to succeed (Rechenthin, 2013) In other words, as

argued by Gonçalves Teixeira (2016), "BI project can be defined as a set of activities or plans that after being concluded, can help in decision-making."

2.3.1 Managing BI Projects

Project Management (PM) "is a discipline that applies knowledge and tools in the elaboration of activities associated (projects) to reach an objective or a range of pre-defined objectives, like costs or deadlines" (Pelegrini et al., 2015. Cited by Gonçalves, 2016). According to Gonçalves (2016), "Varajão indicates that 'PM is essential for the development of successful projects, being transversal and having applications in many industries,' also the PM increases efficiency, effectiveness, and project success rate." On the other hands, Moss & Atre (2003) argued that it is possible to summarize in simplistic terms the activities of the project management answering four basic questions:

- What will be delivered? – Scope
- When will it be done? – Time
- How much will it cost? – Budget
- Who will do it? – Human resources

"These four questions have to be answered before the project manager starts creating a project plan. He or she must spend some time defining the project to understand requirements clearly, risks, constraints, and assumptions" (Moss & Atre, 2003 as quoted in Kříž, 2015). According to Moss (2003), there are certain activities that should be carried out to preparing well a project plan; these activities are exposed below:

- Create a work breakdown structure, listing activities, tasks, and subtasks.
- Estimate the effort hours for these activities, tasks, and subtasks.

- Assign resources to the activities, tasks, and subtasks.
- Determine the task dependencies.
- Determine the critical path based on the dependencies.
- Create a detailed project plan.

2.3.2 Evaluation of project success

The concept of success in a project is variable between authors. According to Gonçalves Teixeira (2016), authors like Ika (2009) and Serra & Kunc (2014) consider success equal as the capacity to combine efficiency (project management performance) and effectiveness. But, other authors like Khang & Moe (2008) (Cited by Gonçalves Teixeira, 2016) the success can evaluate with the performance of three dimensions: 'Efficiency of the implementation process, the perceived quality of the project and the client's satisfaction.

Gonçalves Teixeira (2016) argued that success in a BI project "can be measured through three levels: reports output level (enables the consult of information); cooperation level with other areas; and support level to the short time." In other words, a BI project will be successful when: allow access easily to crucial information through reports, dashboards, etc.; makes possible the collaboration among the different areas, helping the entire organization to be aligned in the same objective, and also, offer the necessary support that facilitates the decision-making process.

Figure 9 shows the principal steps of a traditional project management process: Design, Development, Analyze, and Evaluation, also adding the principal's characteristics of a BI project (according to the authors mentioned above) and include the three levels for assess the project BI success.

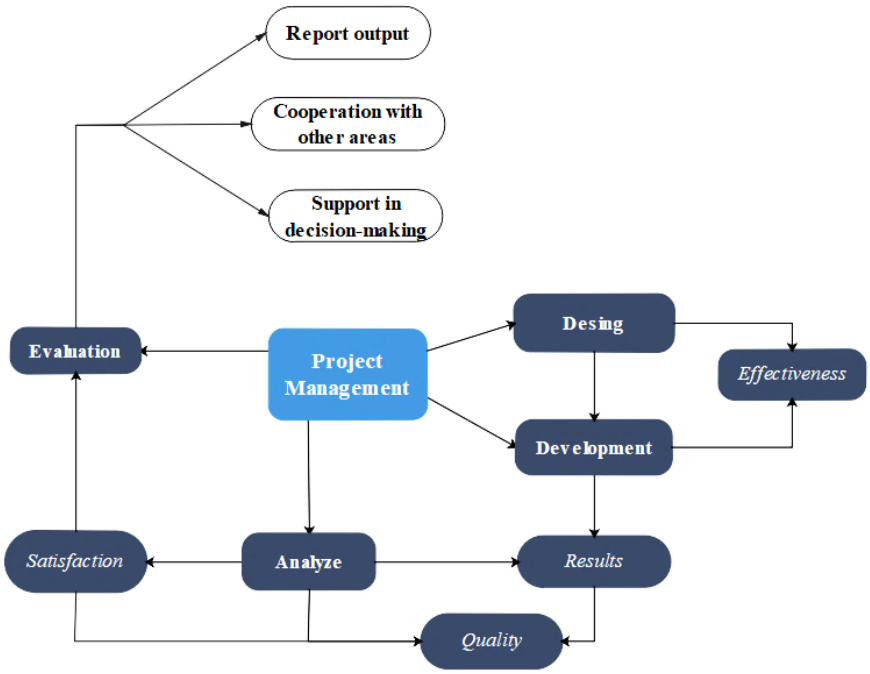


Figure 9. The classic PM and BI projects: Integration of the processes

Despite the increasing rate of acceptance of BI technologies and high implementation and development of BI projects nowadays, according to Gartner, Inc., the failure rate for new BI initiatives is roughly 70 percent. There are different reasons for the majority of BI projects never arrive at the desired finish. Based on the literature, it is possible to cluster the reasons why projects BI fails under three different approaches: user, data, and organization. Table 4 exposes various reasons according to each approach.

Table 4. Reasons why BI projects Fail

User	• Bad user experience. • Lack of user adoption. • Lack of proper human resources.
Data	• A lot of trivial data (No data cleansing activities) • A lot of trivial KPI. • No gathering requirements
Organization	• No support from above. • No investing in new technology. • No action plan.

- **User approach:** BI always will be about the user experiences; if the information gleaned through BI effort is not easy to read, understand, and present, the possibility of failure increases. In other words, users should be in the capacity to appropriate of the information. According to Wired,[5] a confusing or poorly designed dashboard can cause confusing things and making harder the use of the information. The problem, usually it is the lack of user experiences in the BI tools. If all appropriate parties are involved in BI and analytics, deployments are much more likely to succeed. Gathering data isn't all there is to BI, a lot less create dashboards or know how to work with the BI tools, an organization that lacks the staff to draw conclusions and implement changes based on the information it gathers, it can't learn from its analytic activities.

- **Data approach:** A common problem in BI projects is to gather and build data warehouses without a plan about the issues that should be solved because it is possible to run over budget just to hold it. It's vital to build efficient data warehouses with the necessary information for the project. An essential step for doing, it is establishing a good data source and an excellent cleaning program that can delete outliers and noise of the information in order to avoid going in the wrong direction. Furthermore, as more can be learned about what is useful data, It is possible to modify further and refine how to respond to changes.

- **Organization approach:** Develop a BI project is not easy because it is necessary to invest a lot of time and money. For this reason, executive participation is crucial. But a lack of administrative support usually is another significant boundary to BI success. Experts suggest that CIOs need

[5] Wired is a monthly American magazine, published in print and online editions, that focuses on how emerging technologies affect culture, the economy, and politics.

to step up and play key roles in selecting and implementing BI technology, having into account of both the CEO and their IT staff. On the other hand, the selection of the tools (software, programs, etc.) must be a shared decision between the CEO and the workforce. first of all, the selection of technology is fundamental because if the workers are not comfortable with the tools, there is a high probability of project failure. Furthermore, the companies need to be aware of the problems, aims, boundaries, and the other vital features for select the most appropriate technologies and plans, always under the guidance of the executive personal.

2.4 PREDICTIVE ANALYTICS

As said by S. Poornima and M. Pushpalatha (2018), the predictive analysis could be defined as "A progressive branch of data engineering that usually makes the prediction of any existence or probability of data. Predictive analytics makes use of data-mining methods for making predictions about the events in the future and then yields recommendations by these predictions".

According to Siegel (2010) (As quoted in Kříž, 2015), the predictive analytics play an essential role in the business world today, due to it enables management to forecast financial development or consolidation activities. Kříž (2015) also argued that the competitive advantage of the Predictive analytics lies in the possibility to anticipate the future better because if the companies know their customers better, they can plan their needs and increase its benefits. By analyzing the data, predictive analytics can provide actionable knowledge in a timely manner.

Siegel (2010) states (According to Brinkmann, 2015) that Predictive Analytics has meant a new phase in the enterprise evolution by applying organizational learning, which empowers the business to grow by the management of the new knowledge acquired through the historical data. Brinkmann (2015) argued that

"The value proposition of predictive analytics is to improve decision-making and operations within a company using applied organizational learning. This learning process provides a unique capability and reveals competitor weaknesses." Also, White (2010) (cited by Brinkmann, 2015) claimed one of the most important reasons why organizations adopt predictive analytics is to thrive in highly competitive environments.

The predictive analytics helps to make predictions of future events based on historical data assessment. This procedure uses different statistical and analytical methods that are employed to develop the models, which will make the prediction of future occurrences, events, or chances. An advantage of the predictive analytics, it is the possibility of dealing through continuous and discontinuous changes allowing betters models. In the end, the classification and possible predictions of the predictive analytics depend on the method and techniques used for the analysis but especially of the data. S. Poornima and M. Pushpalatha (2018)

The capacity to create models more accurate is the main difference between predictive analytics and traditional trends. The ability to extract the patterns and the buying trends of the customers is a powerful tool that helps the companies to identify the weakness in their competitors, thanks to the behavior trends encoded in the data. Also, the enterprise can then leverage this knowledge and act upon these opportunities by implementing precisely targeted marketing and sales activities. (Siegel 2010, as cited in Brinkmann, 2015)

2.4.1 Predictive Analytics and BI

BI looks for the trends at a macro or aggregate level of the business, trying to find the areas with over or under-performance. In contrast, the predictive analytics go one step further building analytic models at the lowest levels of the business at the individual customer, product, campaign, store, and device levels and looks for

predictable behaviors, propensities, and business rules (as can be expressed by an analytic or mathematical formula) Predictive analytics is about finding and quantifying hidden patterns in the data using complex mathematical models that can be used to predict future outcomes.

According to Schmarzo (2014), "Business intelligence is the world of descriptive analytics: retrospective analysis that provides a rearview mirror view on the business—reporting on what happened and what is currently happening. Predictive analytics is forward-looking analysis: providing future-looking insights on the business—predicting what is likely to happen (usually associated with a probability) and why it's likely to happen." And Figure 10 shows a difference between the two ways to view BI and predictive analytics.

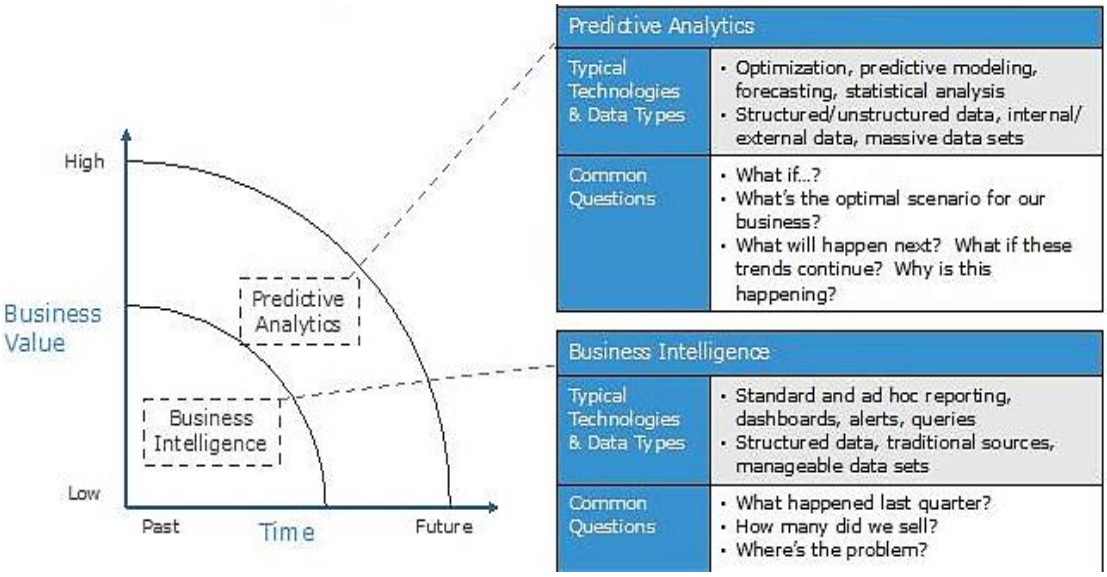


Figure 10. Business Intelligence versus Predictive Analytics

In summary, according to Schmarzo (2014) stated predictive analytics takes the questions that business intelligence is answering to the next level, moving from a retrospective set of answers to a set of responses focused on predicting performance and prescribing specific actions or recommendations. Therefore, Business Intelligence is necessary to know what really happened in the past, but

also Predictive Analytics is needed to optimize the resources as the decisions are made and action for the future.

2.5 DATA PROCESSING CHAIN

The data are embedded in the heart of business intelligence. Data have a hidden value that can be very beneficial for the companies if they are used in the correct way. Data can be modeled and stored in a database. The data from the warehouse can be combined with other sources of data and mined using data mining techniques to generate new insights. Relevant data can be extracted from the operational data stores according to specific reporting and analyzing purposes, and stored in a data warehouse. (Maheshwari, 2015) Figure 11 explains the progression of data processing activities.

Figure 11. Progression of the data processing activities.

2.5.1 Data

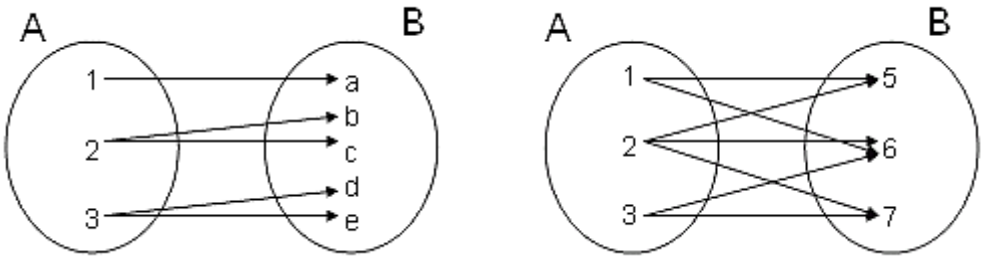
According to Maheshwari (2015), data is everything that can be recorded. Observation, facts, Anecdotes, and also opinions are different kinds of data. Data could come from different sources records compiled by the companies, government, or other agencies. Data could be numerical, like the temperature or alphanumeric as the name of employees in a company. There is also data of data called Metadata, for example, the format of a video file (whether it was a high-def file or lower resolution) is metadata.

The new trend of saving any kind of information is called ***datafication***. In other words, datafication means that almost every phenomenon is now being observed and stored. Every click on the web and every movement of the mobile devices is being recorded. More people are always connected to "the grid" by their phone network or the Internet, and so on. (Maheshwari, 2015)

2.5.2 Database

A database can be defined as a set, collection or deposit of data stored in a computer support, the data must be related and structured according to a model capable of collecting the semantic content of the stored data, due to the importance of data relationships in the real world, it is essential that the databases be able to save these interrelationships. (Llanos Ferraris, 2010) For example, the relationship between the customers and orders would be such that one customer can place many orders, but one order will be ordered by one and only one customer. It is called a one-to-many relationship. The relationship between orders and products is a little more complicated. One order may contain many products. And one product may be included in many different orders. This is called a many-to-many relationship. Different types of relationships can be modeled in a database. (Maheshwari, 2015)

Figure 12 shows the relations mentioned above.



[a]] One-to-many (Customers and Orders) [b] Many-to-many (Orders and Products)

Figure 12. Types of Relationship

According to Maheshwari (2015), the databases have overgrown over time. A decade ago, a terabyte-sized database was considered significant. Today the databases are in petabytes and exabytes. E-commerce and other web-based activities also generate vast amounts of data. The data generated through social networks have also created large databases. The files, included as email attachments in organizations, are of similar large size.

2.5.3 Data Warehouse

According to Frankenfield (2019), a Data Warehouse is a core component of the Business Intelligence and could be defined as electronic storage of a large amount of a business or organization information. This information helps makes management decisions (Maheshwari, 2015). Frankenfield (2019) also stated that "a data warehouse is designed to run query and analysis on historical data derived from transactional sources for business intelligence and data mining purposes."

It is essential to clarify that Data Warehouse and Database have not the same definition. A database is a transactional system that updates in real-time the data in order to have only the most recent data. On the other hand, a data warehouse is programmed to aggregate structured data over a period of time. (Frankenfield, 2019) For example, a database might only have the most recent acquisition of a customer, while a data warehouse might have all acquisitions that this customer has done for the past 10 years.

A data warehouse is created through specific steps. The first one is the extraction where the information is gathering from different sources points. After the data has been collected, it goes through ETL (Extract, Transform, and Load) process, where they are cleaned-up and converted from database format to a warehouse format. In this way, information is easier to use. (Frankenfield, 2019) Figure 13 summarized the creation of a data warehouse.

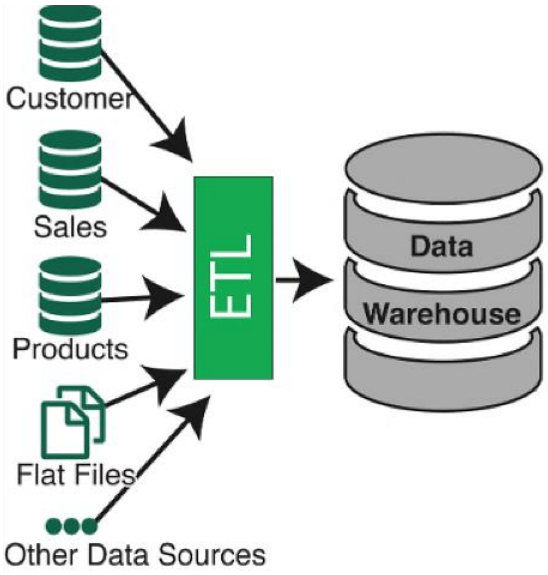


Figure 13. Data Warehouse Creation Process

2.5.4 Data Mining

Maheshwari (2015) defines Data Mining as "the art and science of discovering useful innovative patterns from data" this definition is complementing by Frankenfield (2019) that include that these patterns must have the capacity of improving the business processes.

Data Mining uses some techniques for the creation of the pattern models, below are exposed to the essential techniques.

- **Decision trees:** They are models that have tree-like structures that represent sets of decisions. These decisions generate rules for the classification of a set of data.

- **Regression:** This is a well-understood technique from the field of statistics. The goal is to find a best-fitting curve through the many data points.

- **Cluster analysis:** It groups a set of observations in a given number of clusters or groups; it is based on the idea of similarity of the groups.

- **Artificial neural network:** It lies in sequential learning, uses transformations of the original variables for prediction and nonlinearity of the model.

2.5.5 Data Visualization

According to Tableau[6] web site, it is possible to define data visualization like "the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data."

The visualization of the data plays a central role in the big data age. The companies need a way to analyze and understand a massive amount of information quickly. Through of differents dashboards, graphs, tables, etc. data visualization can show trends, outliers, and KPIs. The use of colors and forms is also crucial because, for the human eyes, it is very to easy different a red from green, a square from a circle. In this way, it is easy to maintain the focus in the patterns and identifies the most crucial information.

Data visualization is not merely to show whatever graph or table. Effective data visualization is a delicate balancing act between form and function. It is necessary to find the best type of data visualization that can fits perfectly with the information that will be exposed.

[6] Tableau Software is a software company with its headquarters in Seattle, United States, which develops interactive data visualization products that focus on business intelligence.

CHAPTER 3

BOARD

Board International is a Business Intelligence (BI), Corporate Performance Management (CPM)[7], and Advanced Analytics software vendor known for its Board Toolkit. The company is headquartered in Chiasso, Switzerland, where it was founded in 1994. Nowadays, Board International operates in 15 countries worldwide.

Board Toolkit is an all-in-one Decision-Making Platform that provides all of the tools required to create and update databases, data presentations, analyses, and process models in a single visual and interactive environment. Board Toolkit puts at the direct disposal of the organizations the tools to create from scratch, without specific programming languages, customized models and applications, perfectly responding to the needs of the company at that particular moment.

Unlike its main competitors, Board does not derive from a patchwork of different technologies and products obtained through acquisitions and forcibly integrated. Board is instead a unique and integrated product, which guarantees organizations the oversight of the entire area of Business Intelligence and Corporate Performance Management using a single interface, a single application construction environment, unique technology, and, therefore, unique know-how.

---

[7] Understanded as the combination of methodology, metrics, process and systems used for monitors and management of the performance of an organization.

Figure 14. Integrating Board tools

Figure 14 shows all integrating processes that offer the Board Toolkit. It is possible to note that Business Intelligence, Predictive Analytics, and Performance Management are unified in a single integrated environment.

3.1 BOARD POSITIONING

Figure 15 shows the important position of Board in the Gartner annual report (2016). Board stands out for being a platform easy of use for the users and also to bring substantial business benefits to the organizations that implemented its technology.

Board International has received differents recognition in the last years, positioning as one of the most important BI and CPM software. In the Dresner's 2018 Wisdom of Crowds Business Intelligence Market Study, Board was rated a "Technology Leader" in relation to Customer Experience and a "Trust Leader" in

relation to Vendor Credibility, whilst also being given a perfect score in terms of users' willingness to recommend the solution.

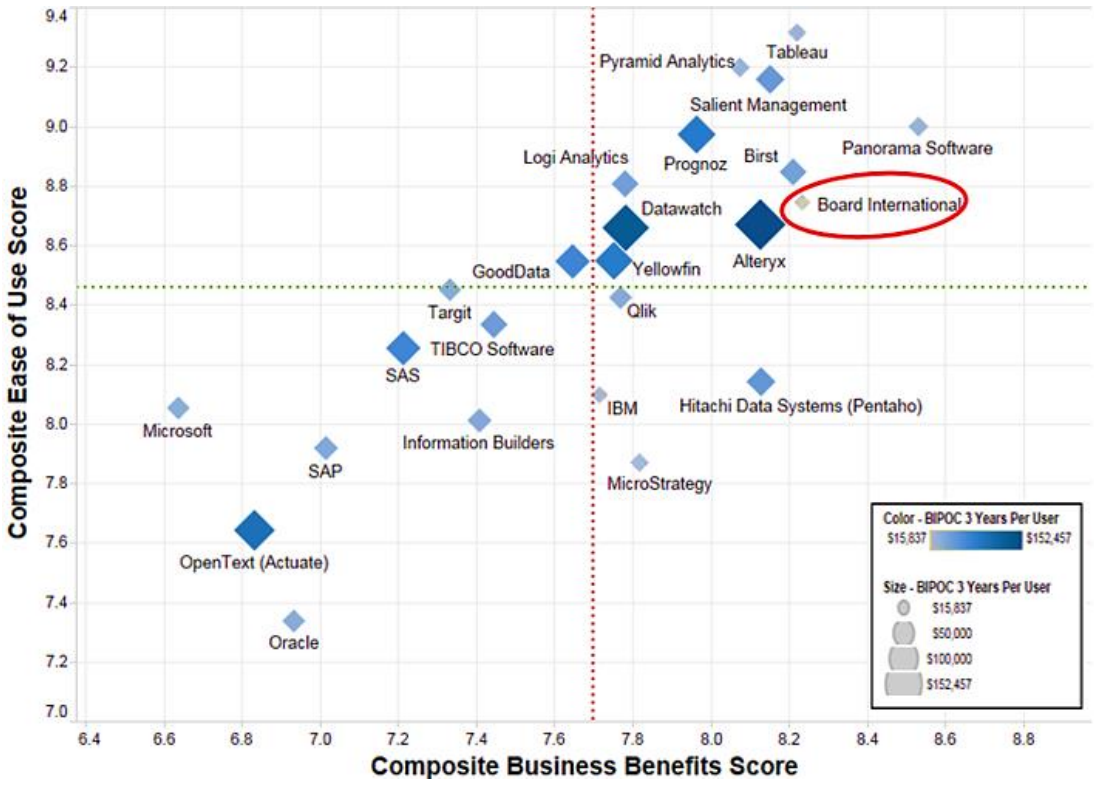


Figure 15. Platform cost-benefit by vendors

The study also highlights that "better decision-making" is still the top objective for organizations in the use of BI tools, alongside improving operational efficiency, growth in revenues, and the achievement of competitive advantage. This is particularly relevant given Board's unified analysis, simulation, and planning nature, which allows users to make better use of data for planning and forecasting within one solution.

On the other hand, Gartner has included Board in two important studies, the Magic Quadrant and the Frontrunner in the category analytics and Business Intelligence Platform. This indicates the leading role of Board in the BI market, and its quickly growing.

- **The Magic Quadrant for Analytics and Business Intelligence Platforms:** is a series of market research reports published by Gartner that rely on proprietary qualitative data analysis methods to demonstrate market trends, such as direction, maturity, and participants assess BI software and platforms performance across five Use Cases and for the vendors are used 15 Critical Capabilities. The Use Cases will define bellow, and the Critical Capabilities are exposed in Figure 16.

    1. **Agile, centralized BI provisioning:** Supports an agile IT-enabled workflow, from data to centrally delivered and managed analytic content, using the platform's self-contained data management capabilities.

    2. **Decentralized analytics:** Supports a workflow from data to self-service analytics, and includes analytics for individual business units and users.

    3. **Governed data discovery:** Supports a workflow from data to self-service analytics to system of record (SOR), IT-managed content with governance, reusability, and promotability of user-generated content to certified data and analytics content.

    4. **OEM or embedded analytics:** Supports a workflow from data to embedded BI content in a process or application.

    5. **Extranet deployment:** Supports a workflow similar to agile, centralized BI provisioning for the external customer or, in the public sector, citizen access to analytic content.

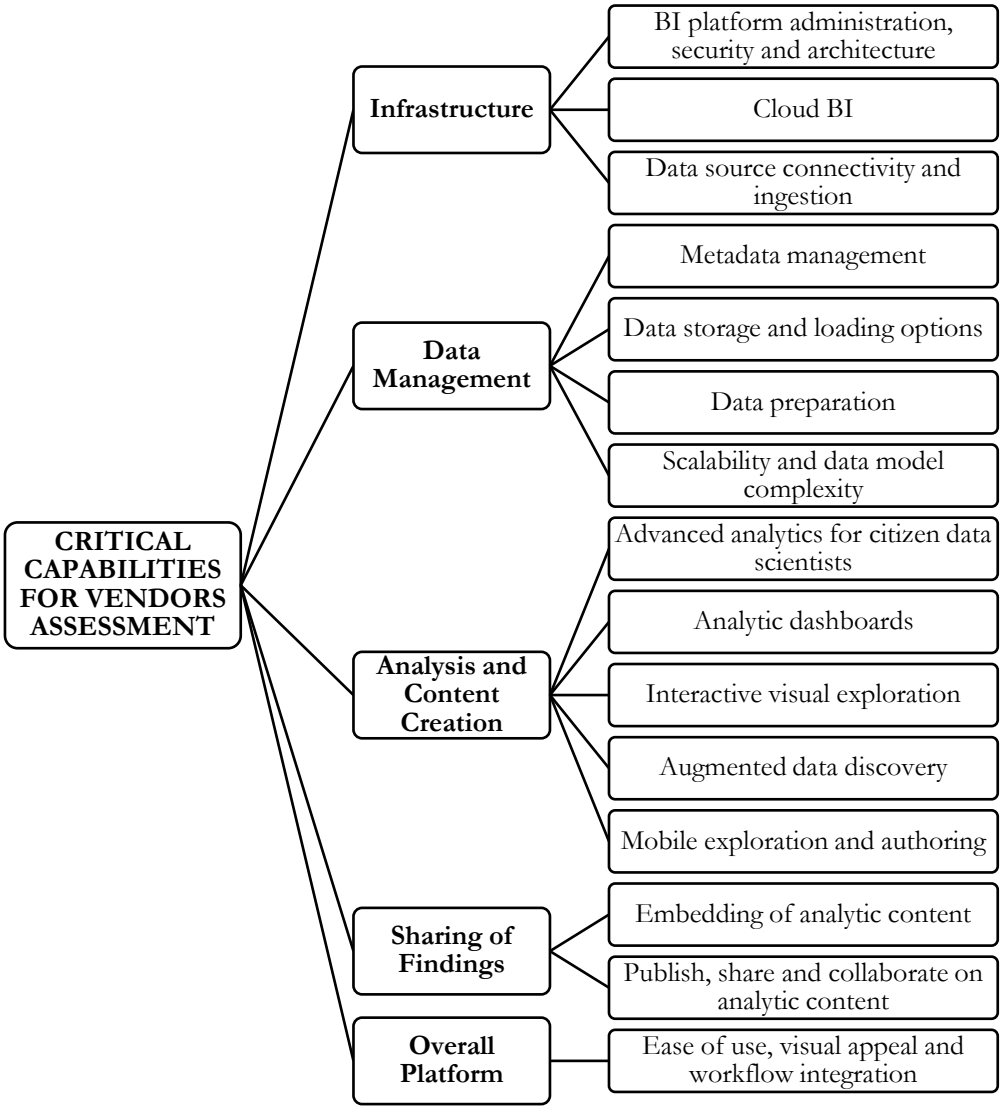


Figure 16. Critical Capabilities for Vendors Assessment

In the Magical Quadrant study Board has been included as a **Niche Player**[8] (see Figure 17) According to the Gartner analysis:

"Board successfully serves a submarket for centralized, single-instance analytics, BI and financial planning and analysis (FP&A) solutions. Board has a narrow focus, and the market's awareness of this vendor is limited, although its regional adoption is growing. Board has successfully transitioned to a cloud/subscription-based model. Its reference

[8] Its characteristics are deepened later.

customers reported that the most common use cases for Board's system are agile, centralized BI provisioning (61%), and decentralized analytics (57%)."

Figure 17. Magic Quadrant for Analytics and BI Platforms

Features of the companies according to the Gartner's Magic Quadrant:

- ***Leaders:*** *Vendors in the Leaders quadrant have the highest composite scores for their Completeness of Vision and Ability to Execute. A vendor in the Leaders quadrant has the market share, credibility, and marketing and sales capabilities needed to drive the acceptance of new technologies. These vendors demonstrate a clear understanding of market needs, they are innovators and thought leaders, and they have well-articulated plans that customers and prospects can use when designing their infrastructures and strategies. In addition, they have a presence in the five major geographical regions, consistent financial performance, and broad platform support.*

- ***Challengers:*** *A vendor in the Challengers quadrant participates in the market and executes well enough to be a serious threat to vendors in the Leaders quadrant. They have strong products, as well as sufficiently credible market position and resources to sustain continued growth. Financial viability is not an issue for vendors in the Challengers quadrant, but they lack the size and influence of vendors in the Leaders quadrant.*

- ***Visionaries****: A vendor in the Visionaries quadrant delivers innovative products that address operationally or financially significant end-user problems at a broad scale, but has not yet demonstrated the ability to capture market share or sustainable profitability. Visionary vendors have frequently privately held companies and acquisition targets for larger, established companies. The likelihood of acquisition often reduces the risks associated with installing their systems.*

- ***Niche Players:*** *Vendors in the Niche Players quadrant are often narrowly focused on a specific market or vertical segments. This quadrant may also include vendors that are adapting their existing products to enter the market under consideration, or more significant vendors having difficulty developing and executing on their vision.*

- **FrontRunners:** This is another important study that uses real reviews from real software users to highlight the top software products using in small businesses. This work is focused on the North American market. The aim of this analysis is to help small businesses to make a better decision about what is the better software for them. For the creation of the report were evaluated over 282 Business Intelligence (BI) products. Only those with the top ratings for Usability and User Recommended made rated as Front Runners.

  Figure 18 shows the position of the FrontRunners in the entire market, while Figure 19 exposes the principal's software business intelligence, according to the Gartner Study.

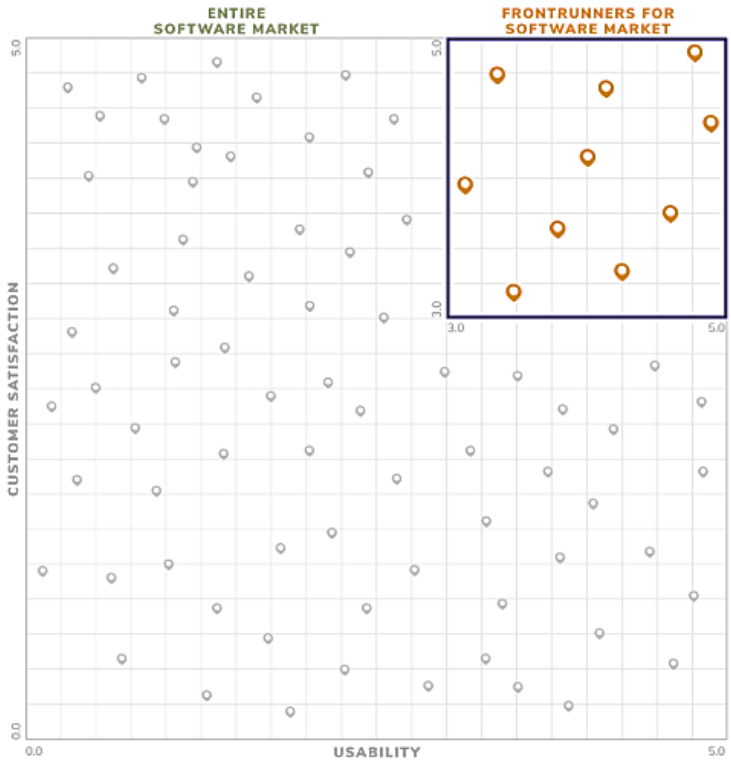


Figure 18. FrontRunners in the entire market

Figure 19. FrontRunner Business Companies

3.2 BOARD CAPSULES

A Capsule typically is a series of screens, each containing single or a couple of objects such as spreadsheets, charts, text notes, images, dashboards and indicators, and so on. A Capsule is an end-user environment that allows viewing and manipulating statistics from a Board database. It affords the usual multidimensional analysis features such as drill-down and slice and dice and countless more.

Capsules made of a number of screens typically include buttons allowing navigation from one display to another, presenting the consumer with a guided and logical route.

A Capsule does not always solely hold records presentation objects like spreadsheets and charts. It may additionally also consist of some database administration duties prepared in a Board Procedure. A Procedure is a sequence of actions such as a DataFlow, which it calculates the value of a Cubes from other existing Cubes (for instance calculating the Cost of good bought by using multiplying the Cubes Quantity instances the Product Standard Cost) or a DataReader, which is a system that loads statistics from a facts supply such as an ERP machine into a Board Database.

3.3 BOARD DATABASES

Board's database engine is designed to ensure the most efficient in managing vast volumes of information and to furnish top-level performance. The implementation of different multi-dimensional management techniques completely avoids the database explosion problem, usually related to multidimensional databases (also referred to as M-OLAP).

Board's databases are multidimensional and are, therefore, ultimate for online analytical processing (OLAP). Conceptually, a multidimensional database makes use of the concept of a data dice the place the dice cells contain values, and the dimensions signify the exclusive viable perspectives on data. For example, a "sales" cube, would include income values in its cells and may want to be considered by way of a number dimensions such as product (i.e., the sales figure per product), geography (i.e., income discern by way of metropolis or region), time.

A Board database is made up of:

3.3.1 Entities

These are information sets, generally text and codes like the set of Customers, Products, and Cities. Entities (and hierarchies) are the Cube's dimensions.

3.3.2 Hierarchies

When two or more entities have a many-to-one relationship, then a hierarchy can be defined. For example, the entities Customer, City, and State can be organized into the hierarchy Customer → City → State, since there is a many-to-one relationship existing between Customer and City and between City and State.

3.3.3 Cubes

The cubes are usually the most important part of the data (often numerical but not strictly) that can be analyzed and viewed by its different dimensions and hierarchy levels. Figure 20 shows the cube: Revenue, which has three main dimensions (Month, customer, and product). While Figure 21 exposes the detailed level of every dimension, which can be analyzed through the drill-down option.

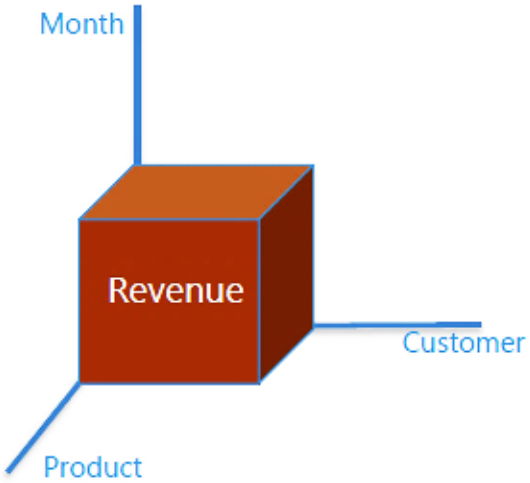


Figure 20. Cube Dimensions

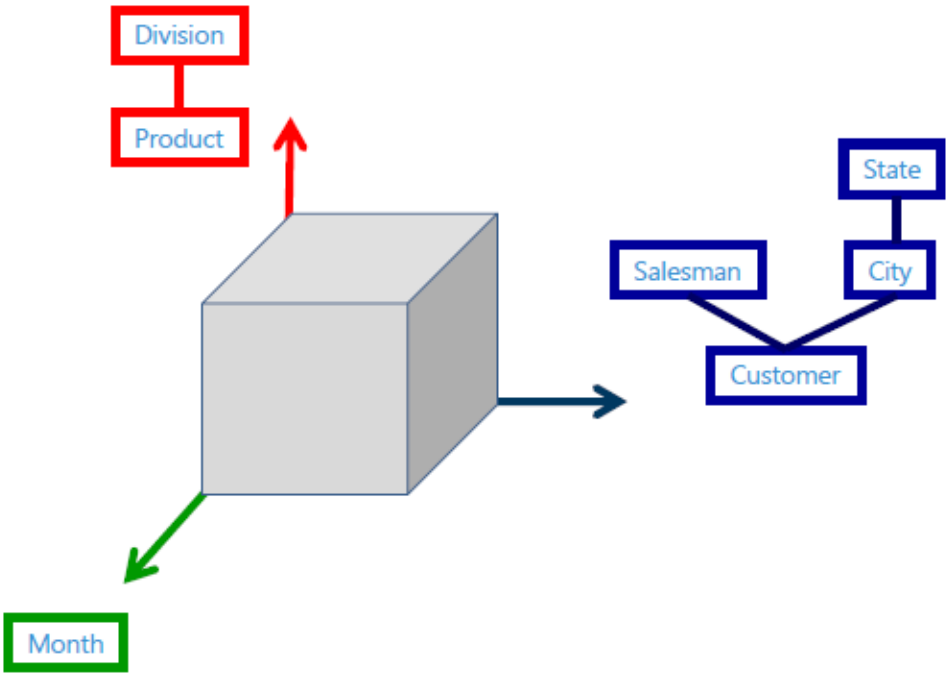


Figure 21. Cube Detailed Dimension

In Board, according to store needs, the Cubes could be MOLAP or ROLAP (See Figure 22). MOLAP method allows store the data in Board, every time a Data Reader or Protocol is run. On the other hand, in ROLAP method, the data is not stored in Board; it is only mapped from its original data sources.

Figure 22. Type of data storage in Board

There are two types of data structure in the cubes, Dense and Sparse. In the data structure dense, all combinations of the entities are created, and the Cube is full. The problem with this kind of data structure is the high use of the Ram and Hard Disk. Instead, with the data structure sparse, only the existing combination can be created.

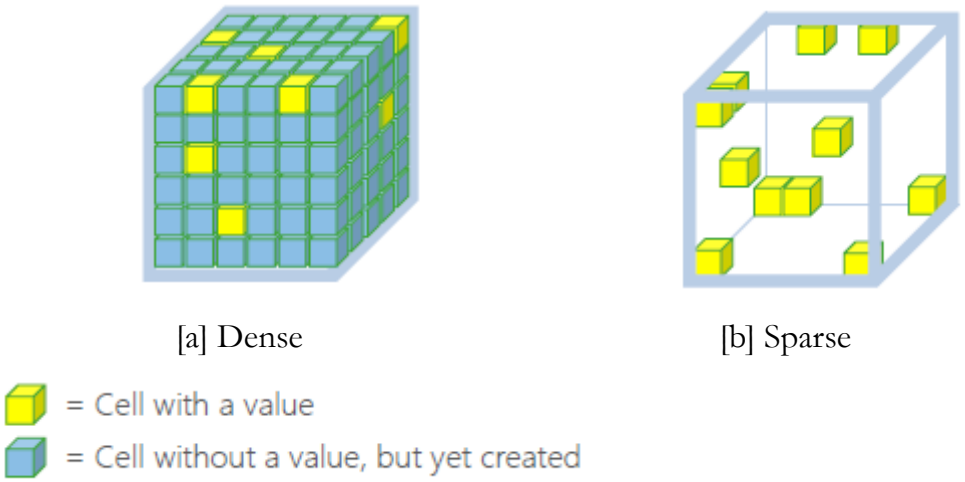


Figure 23. Cube: Data structure

3.3.4 Data Reader

The Data Reader allows import data from files source into a Board database. A Data Reader protocol defines how the external data should be introduced into Board database and mapped to entities and Cubes. A Data Reader protocol can

include a set of transformation formulas and validation rules to apply to incoming data. These rules are defined using the ETL function.

- **Append and Replace options:** A Data Reader protocol can include three types of fields, entity codes, entity descriptions, and Cubes. For each field type, the Append/Replace option regulates how and if external data is loaded or discarded.

  - **Entity code**

    Blank: The incoming record is only loaded if the value found in this field is the code of an existing member of the entity; otherwise, the entire record is discarded.

    Append: This option validates all incoming records; when the value from the data source is not the code of an existing member of the entity, a new member is created.

  - **Entity description**

    Blank: This option only loads the description if the current entity member description field is not populated; therefore, only descriptions for newly created items are loaded.

    Replace: This option overwrites the entity member description with the value from the source file. It is recommended to always enable this option in order to update entity member's descriptions with changes that might occur on the operational system master files.

- **Cube**

  Blank: This option adds incoming data to the Cube; if the Cube already contains values, then the new data is summed. The summing of data occurs only in the case where the Cube is numeric; for date and text Cubes, the current cell value is replaced with the source data.

  Replace: This option replaces the Cube cells belonging to all-time periods found in the incoming data, with the new values. When this option is enabled, the Data Reader performs the following actions scans the entire source file and memorizes the time periods, then clears the Cube cells belonging to the time periods found in the source file after that starts loading source data into the Cube.

3.4 BEAM

According to Board International, *"BEAM seamlessly integrates advanced and predictive analytics with business intelligence and performance management, making it easier than ever for anyone to gain powerful business insights and take action quickly."*

BEAM was developed in collaboration with IDSIA (Dalle Molle Institute for Artificial Intelligence Research), one of the most important world's institutions in the Artificial Intelligence area. According to Board, the creation of BEAM was due to two critical business needs. Firstly, the necessity of incorporating advanced analytics and predictive capability in the company's business processes. Secondly, it is the need to do the process in a simple way, without the necessity of spending a lot of money with experts at the moment of doing Business Intelligence (including the Data mining and Predictive Analytics) process.

The main advantage of BEAM is the capacity to complement the activities performed in Board without spoiling the regular performance of the processes carried out, being in this way a powerful tool for improving the decision-making process.

BEAM provides a solution that covers many analytical areas through three different modules: Predictive Analytics, Clustering, and Analytical Functions. (See Figure 24) BEAM offers advanced forecasting capabilities through the use of autoregressive linear models of the ARIMA[9] family. Board stated that "*BEAM can automatically evaluate the characteristics of each time series, consequently producing a suitable model and run forecasts thousands of times faster than the main competitors do. The foresight achieved can be refined by adding further information to the scenario to reflect the impact of external variables (covariates) on the forecast.*"

Figure 24. BEAM: Analytical Areas

BEAM predictive analytics section works following four main steps. They are described below:

[9] According to Chen (2019) "*An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends*"

3.4.1 Analysis and labeling of time series

Once the historical data have been loaded in Board Database, BEAM classifies the time series based on Average Demand Interval (ADI) [10].

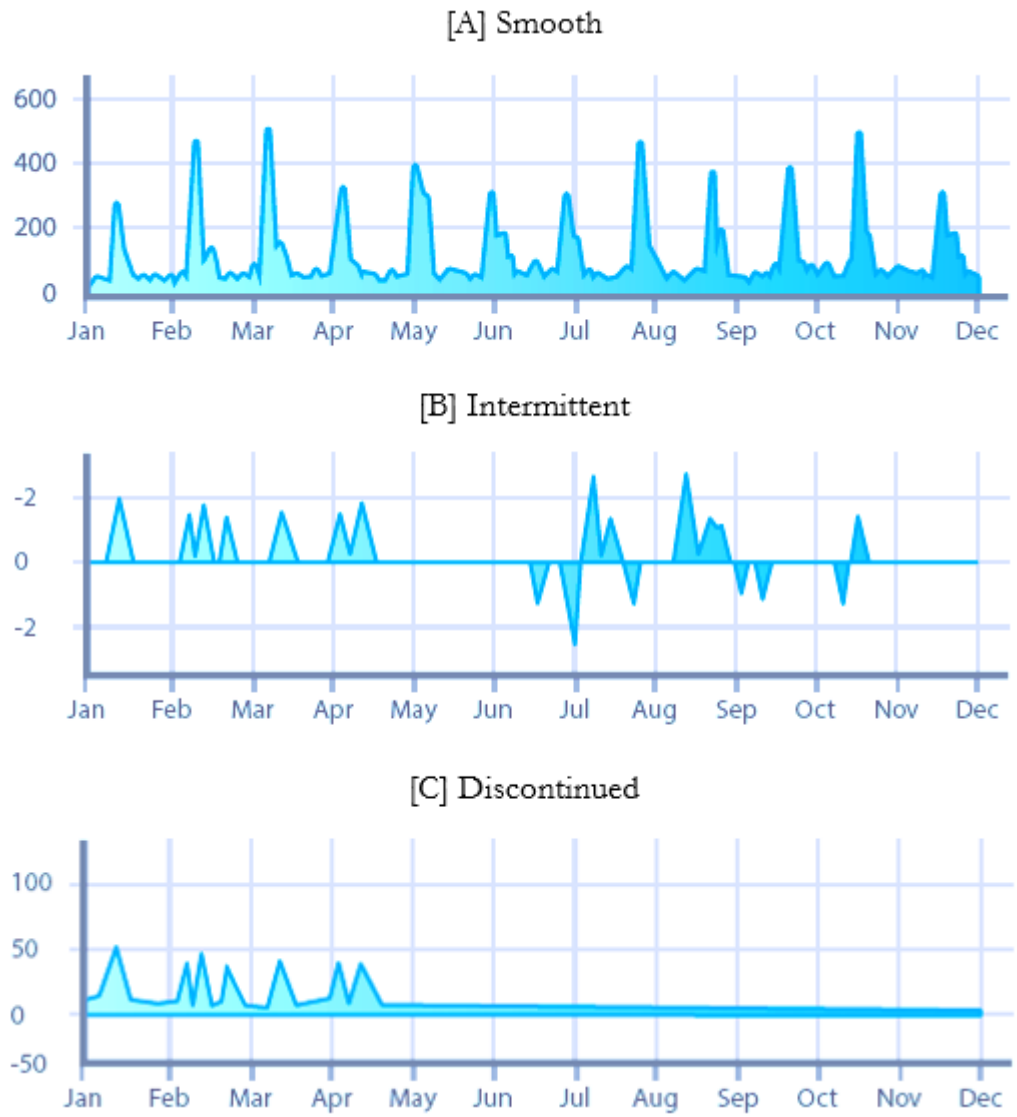


Figure 25. Types of Time Series

[10] It measures the demand regularity in time by computing the average interval between two demands.

The series could be Smooth, Intermittent, and Discontinued. (See Figure 25)

- **Smooth:** Observations are consistently different from 0. (ADI < 1,3)
- **Intermittent:** Observations are characterized by extended periods of no demands (ADI>1,3)
- **Discontinued:** The time series is constantly zero.

If the time series is discontinued, BEAM uses the Zero predictor for the Forecast. While if the time series is Intermittent, the Forecast is modeled according to Croston Model[11]. On the other hand, if the time series is labeled as Smooth, the ARIMA family model is used as described below.

3.4.2 Models Competition

Once the time is labeled as Smooth. Start the model competition process that allows a properly prediction of the trend, seasonality, and covariates. The model's competition follows the next steps:

- **Data series pre-processing:** In order to improve the predictive accuracy of the model in this phase, the system normalizes data (data trimming) and, based on a recursive statistical test, automatically de-trends and de-seasonalities any time series that needs it.

- **Training and Test set identification:** In this phase, any time series is split into two parts: a training set, on which candidate models are applied,

[11] According to Syntetos et al. (2014) *"Croston assumes that the distribution of nonzero demand sizes is normal, the distribution of inter-arrival times is geometric, and that demand sizes and inter-arrival times are mutually independent."*

and a test set, on which the accuracy of those models is measured. (Figure 26).

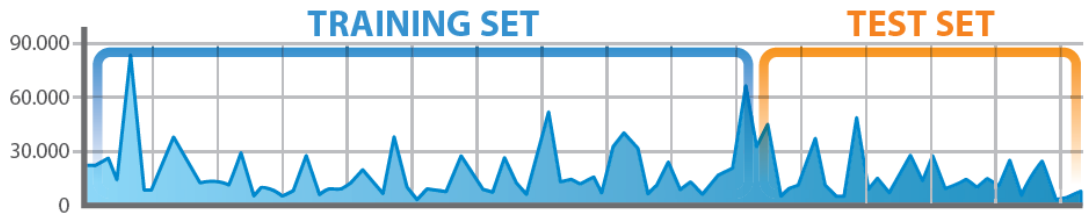


Figure 26. Training and test Identification

- **Model selection, the IDSI-ARX® (autoregressive exogenous) algorithm:** IDSI-ARX, is a linear autoregressive model. It assumes a linear relationship between the future values of a variable and past observations. The selection of the model structure is performed via competition among the candidate models (Persistent and Seasonal Naïve Predictors and linear autoregressive models belonging to the ARIMA Family). The model structure that shows the lowest error in the competition by minimizing the MASE (Mean Absolute Scaled Error) is eventually chosen.

- **Outliers detection and exclusion:** BEAM automatically identifies outstanding records in the time series, where the prediction of the model has a significant error. In specific, information perceived as strange is those in which the blunder is higher than 3.5 times the standard deviation. Once these exceptions of a time series are identified, the model brush aside them.

3.4.3 Covariates management

Covariates are external variables such as promotions, marketing campaigns, weather conditions, etc. that are useful to forecast a certain time series: for

example, usually, the sales of a product can be better predicted if its price history is available as external information (covariate).

BEAM ensures the management of covariates. The system is able to receive a set of covariates for a given time series and to automatically identify the significant ones, whilst discarding the irrelevant ones. Once fed with all the data, BEAM automatically creates alternative predictive model structures characterized by different sets of variables (covariates), runs a competition between them, and consequently selects the set of covariates that minimize the forecast error (MASE).

3.5 BIOS MANAGEMENT

Bios Management is a company that was born in 2004 with the objective of providing consultancy in the field of Direction, Management and Business Organization, with a vocation focused on the development of Business Intelligence, Performance & Knowledge Management models. Into this organization, this work was developed as part of an internship program that was focused on software development and applications of business intelligence and corporate performance management.

Figure 27. Bios Management Logo

Bios activity is synthesized in:

- Support for strategic planning and management control by facilitating the process of management decision making through Business Intelligence and Corporate Performance Management tools.

- Management consulting and training in areas like administration, finance, control, commercial area, marketing, human resources, production and logistics.

- Business Process Reengineering through activities of organizational analysis and review of business processes. Mapping of processes (primary and support), identification of critical points and areas of improvement, allowing to identify solutions and even redesign of these activities.

BIOS collaborates with leading companies nationally and internationally in various sectors: public services, health, pharmaceuticals, credit institutions, food and beverages, manufacturing, automotive and services. Operating in different project areas through Business Intelligence and Corporate Performance Management solutions.

CHAPTER 4:

OBJECTIVES AND METHODOLOGY

The thesis has been carried out within the consulting company: Bios Management. Who has proposed to use the Business Intelligence Software: Board, in order to improve the decision-making process of a customer that works in the manufacturing.

The objective of the thesis is developing a Business Intelligence environment that allows the users to create analysis capable of predicting with the significant accuracy the needed production level and the sales behavior.

Subsidiary aims are:

- Develop ETL, LDM, and Metrics.
- Create valuable KPIs'.
- Prepare Reports and Dashboards.
- Apply Predictive Analytics Methods.

In order to achieve the objectives mentioned above, the work has been articulated in six methodological phases:

1. **Target Company Review**: Firstly, it is necessary to explain the current situation of the company (Brief history, resources, structure, and production) in order to understand and facilitate this study.

2. **Data Source Identification:** In the second phase, the data provided by company Carioca will be analyzed in order to understand and classify it. The different entities, the hierarchy, and cubes based on the company data have been established

3. **Logical Data Model:** This phase explains in more detail the data initially received. Information related to the typology of this data and the implications related to working with each of those typologies of data is given.

4. **Data Loading Process:** Considering that all the data sources have different typologies, data cleaning, and data imputation to enforce data quality are made. Extract, Transform, and Load (ETL) processes are used to integrate data into a Board project.

5. **Building Metrics and Reports:** This phase provides the construction o with the predictive analytics models based on the company data. Also, the models are compared between them, and their level of accuracy is evaluated.

6. **Creating Dashboard:** In the last phase, the dashboards are created. They are composed of a series of tabs that can be used as canvasses to display critical data. Together, the dashboard composed a user-friendly environment where it is possible to analyze the differences between the predictive models.

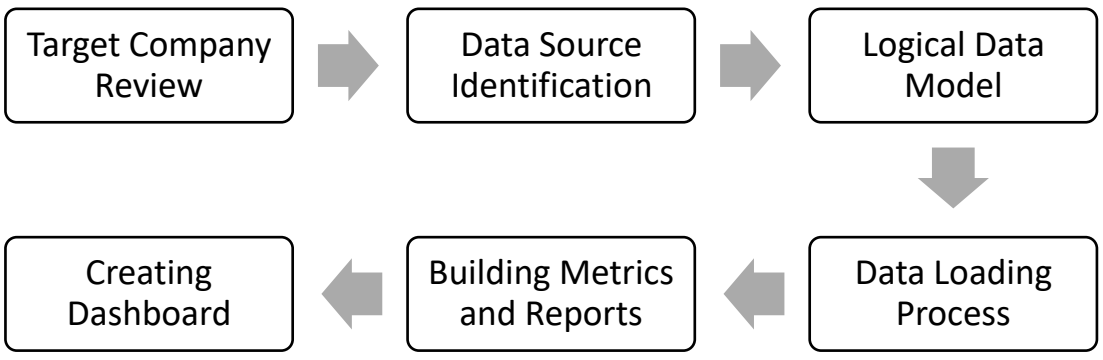


Figure 28. Methodology of the work

4.1 TARGET COMPANY REVIEW: CARIOCA

4.1.1 History

Carioca S.p.a came out of the legacy of one of the leading Italian companies in the production of tools for coloring, writing, and drawing: Universal S.p.a.

Carioca is an international love brand started in 1965 by the inspired entrepreneur Alessandro Frola who, in the Turin province in the mid-50s, founded the Universal Company, which gave rise to Carioca and Corvina and set the stage for the most critical writing area in Italy and the world. Thanks to the Carioca brand, Universal S.p.a soon became a global benchmark for felt tip pen production while Corvina becomes one of the best-selling ballpoint pens in the world.

After some years of crisis, in 2014 Universal S.p.a. initiated a procedure agreed in which Carioca S.p.a rent a part of the Company. Two years later, on April 2016, Carioca S.p.a acquires the Company definitively.

4.1.2 Structure and Resources

The team met in a major multinational company between the late '90s and early 2000s and today have come together to work on this new challenge. Today the company employs approximately 110 people, between its principal office, its plant in Settimo Torinese, and its Spanish subsidiary in Barcelona. The plant in Settimo Torinese is the main production center and exports over 85% of its products to over 60 countries

4.1.3 Production

Specializing in the production of coloring tools, Carioca is a symbol of the best of "made in Italy," with products designed to free the imaginative powers of adults and children in a shared play of creativity and dexterity. 85% of Carioca and Corvina products are proudly made in Italy and distributed around the world, with a global network active in over 60 countries and across 5 continents.

Carioca has a wide range of products, as felt tip pens, colored pencils, tempera, pastels, educational games, and many other drawing tools, all manufactured and tested with the utmost care, quality, and safety. The ink in the Carioca felt tip pens are produced entirely in the Settimo Torinese plant, using only food-grade dyes and ensuring unbeatable levels of washability and brilliance. The molding of the plastic material in granules is carried out in two highly automated departments with a production capacity of over 2,000,000 pens and felt tip pens daily. The complete robotization of item assembly is made possible by sophisticated electromechanical machines. Equipped with control devices, they enable a faster speed of production, maintaining a high standard of quality throughout the entire production process.

The molding of the plastic material in granules is made in two fully automated departments, able to produce more than 20,000,000 elements daily. The complete robotization of the articles assembly is carried out thanks to sophisticated electromechanical machines: equipped with control devices, they allow high production speed, maintaining a high standard of quality along the entire production process.

For Carioca, quality means attention to every detail, which is also the work philosophy of the company. Every day a team of designers, chemists, and craftsmen work together with passion in our factories to simple but technologically advanced writing tools, up to the best Italian design.

4.1.4 Brands

- **CARIOCA**

It is an international love brand created in 1965, specializing in the production of coloring tools. Its product range includes felt tip pens, colored pencils, tempera, pastels, educational games, and many other drawing items. The brand caught on in the early 60s with a historic advertising campaign starring the beloved character Carioca Jo, who is still featured on all Carioca packaging.

Carioca is an international brand born in 1965 and specialized in the production of drawing materials. The brand asserts itself at the beginning of the '60s with a famous advertising campaign starring Carioca Jò, the animated gunslinger, which is still a symbol of all Carioca packaging. Its product range includes fibre pens, colored pencils, tempera, crayons, educational games, and many other drawing materials.

Figure 29. Carioca Logo

- **CORVINA**

It has represented the company's passion for writing tools since 1965. Today, Corvina 51 is synonymous with "ballpoint pen" in many European countries. In 2014, the brand's image and product range were updated entirely in order to express the brand's particular identity and potential fully.

Corvina represents the passion of the group for writing tools since 1965. Even today, Corvina 51 is considered synonymous with "ballpoint pen" in many

European countries. In 2014 the brand image and the product range were redesigned to fully express the particular identity and the potential of the brand.

Figure 30. Corvina logo

- **CARIOCAPROMO**

It is the promotional division of Carioca and Corvina, producing fully personalizable items for promotional purposes.

Carioca offers vast customization possibilities for its product range. In addition to the complete customization of each item, thanks to a wide choice of colored plastics, the customer can customize clips and barrels through processes of screen printing, pad printing, digital printing, or punching. Cariocapromo offers full assistance in graphic and product development to help their clients better communicate their brand and message.

Figure 31. Carioca Promo Logo

4.2 DATA SOURCE IDENTIFICATION

Source data, provided by company Carioca, consists of some CSV[12] files. These files contain data about customers, products, sales, etc.

[12] A comma-separated values (CSV) file is a delimited text file that uses a comma to separate values. A CSV file stores tabular data (numbers and text) in plain text. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format.

4.2.1 Data Classification

Develop a BI project in Board required some logical steps. First of all, the data should be classified according to the type of data (entity or cube). The CSV analysis showed that the are 9 entities and 3 cubes. The entities are divided into 3 groups: Customer, Product, and Time Entities. (See Table 5)

Table 5. Entities and Cubes identified

Entity Group	Entity name
Customer	Country
	City
	Salesman
	Channel
	Customer
Product	Business Area
	Product Category
	Product
Time Entities	Month
Cubes	Sales Units
	Sales Amount
	Cost of Goods Sold

After classifying data between entities and cubes, the next step is to create a new Database in Board "Database Menu," as shown in Figure 32. The information needed to create a new database is shown in Figure 33. A name, default language, and date range are required. In this work, the date range goes from 2014 to 2018. In addition, the default time entity must be selected; if not, the month will be the default time entity.

Figure 32. Board Database Menu

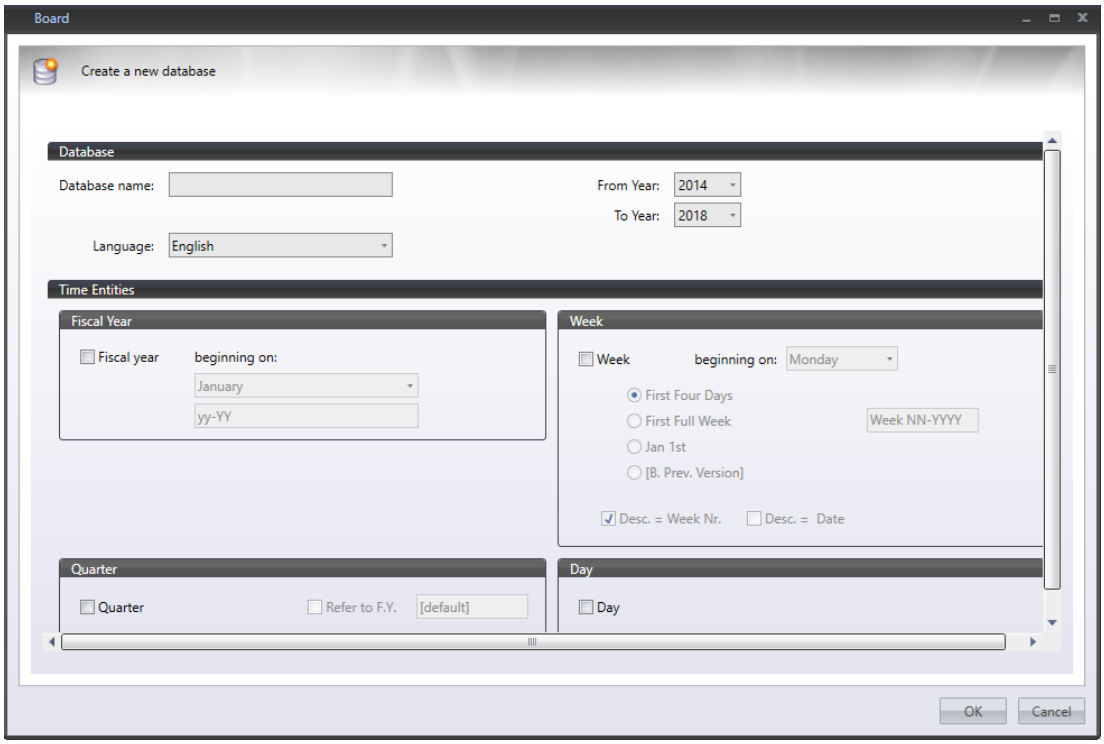


Figure 33. Window for creating a new DB

4.3 LOGICAL DATA MODEL

Logical data modeling is the process of representing the architecture and organization of data graphically without taking into account the physical implementation of the technology of the database management system involved in data storage. A logical data model provides all the information about the various entities and the relationships between the entities present in a database.

When the database is created, the entities, the relation between them, and the cubes should be created in the section Database Manager of the Board's "Database Menu." First, the entities are created. For every entity is needed a code width, description width, and the number of the maximum items. Figure 35 shows the configuration used for every item in this work. Note that the time entities are not defined due to they have another configuration section in the cases where it is necessary to create a customizable time entity. Otherwise, it is not necessary to do nothing because the software adds them by default.

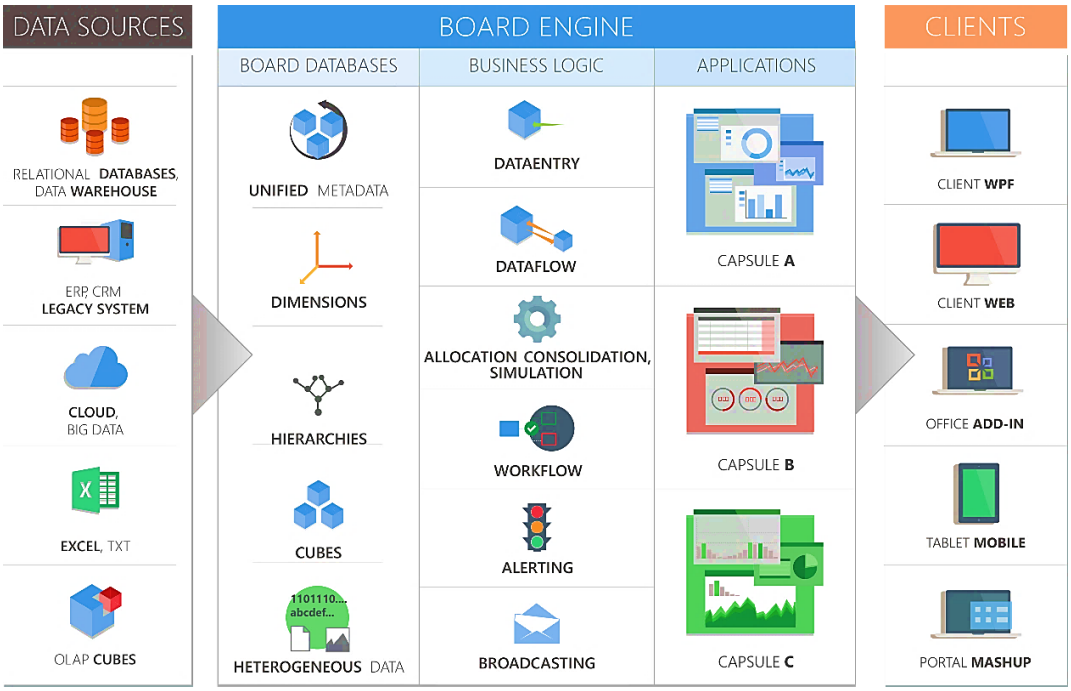


Figure 34. Board Logical Data Model

Entity	Code Width	Desc Width	Max Item Nr	Sort by	Display
Customer					
Customer	4	50	2000	Code	Desc
Salesman	10	50	50	Code	Desc
Channel	1	50	50	Code	Desc
City	10	50	800	Code	Desc
Country	10	50	70	Code	Desc
Product					
Product	20	50	5000	Code	Desc
Business Area	20	50	100	Code	Desc
Product Category	10	50	100	Code	Desc

Figure 35. Customer and Product entities in Board DB

Subsequently, the relations between the entities are set. Board only permits the relation one-to-many. For the two groups (Customer and Product), the relations are the way, as is shown in Figure 36. The customer, according to the hierarchy, the entity could be clustered into three groups: Channel, Salesman, and City. Moreover, City also could be grouped by country. In the case of the article, it only can be clustered in two groups, Product Category or Business Area.

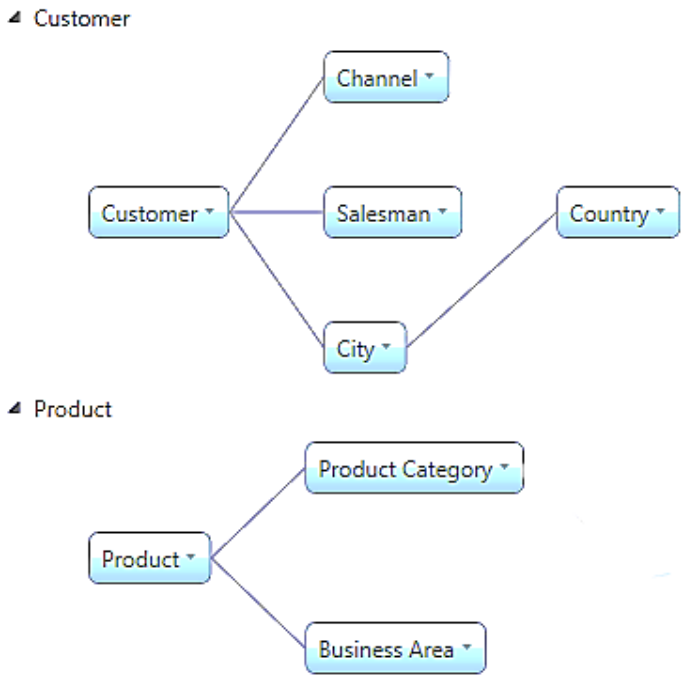
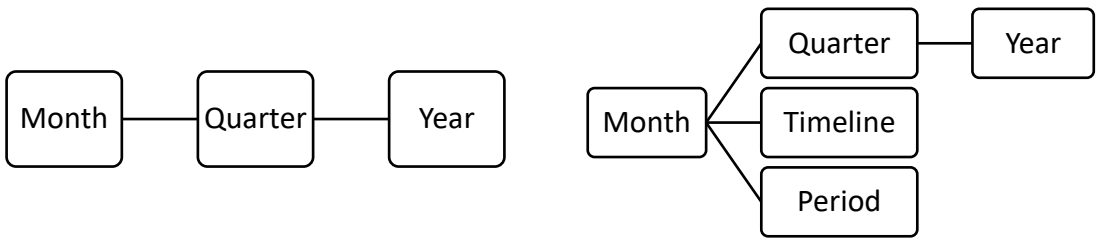


Figure 36. Customer and Product Hierarchy

For the purpose of this thesis, the time entities were configured. Figure 37 shows how the structure of the hierarchy of the time entities changed. Part [a] illustrates the structure before the configuration and the part [b] after that.

Timeline refers to what kind of period, the values could be Historical Time Series or Forecast Horizon. On the other hand, Period refers to the month but unlinked to year. Example: Into month entity, it is possible to find values like May 2016 or April 2015. Instead of the Period, the values are May or April. In other words, months could be clustered in periods.



[a] Time entities before configuration [b] Time entities after configuration

Figure 37. Time Entities Hierarchy

Once all entities (both regular entities and time entities) and hierarchies have been created, the next step is the creation of the cubes. Every cube must have one or more dimensions that could be density or sparsity[13].

According to the analysis of the CSV files' information, there are three central cubes, Sales Units, Sales Amounts, and Cost of Goods Sold. All of them have the same dimension: Month, Customer, and Product, which are the most detailed entities for this reason, and according to Figure 38, the values can be visualized through the different combinations and the drill-down option.

In order to understand better the Drill-Down option, Table 6 and Table 7 describe how works it in Board. In the first table, it is possible to see the cube: Sales unit, the columns are the first three months of the year while the rows contain the Cities per Year. On the other hand, the second table shows the same information but on a more detailed level. Now in the row are the Customer per Cities and the Cities per Year.

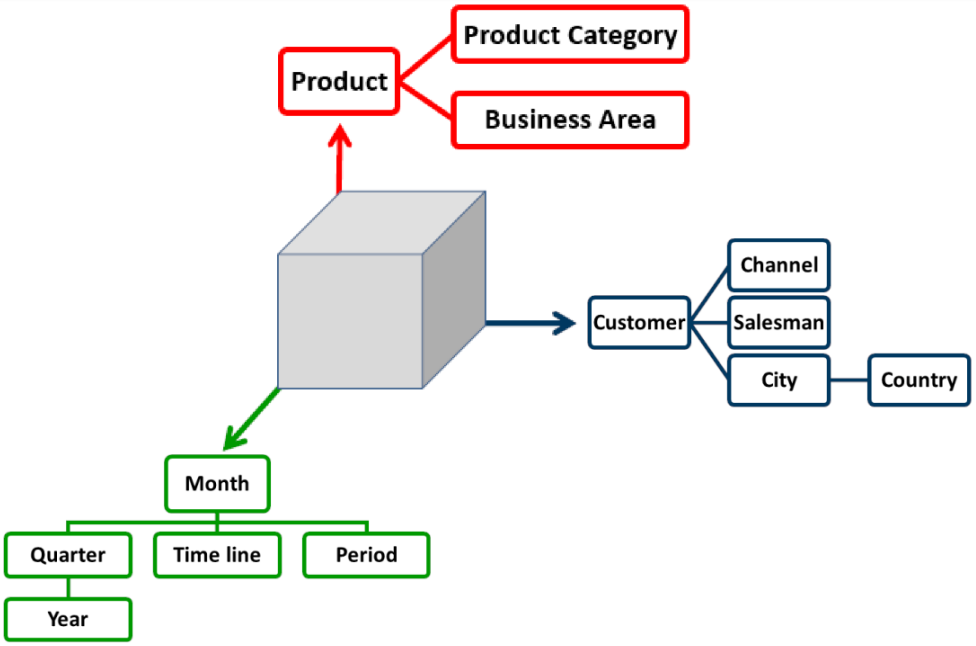


Figure 38. Maximum Detail Of The Work Cube

[13] This was explained in section 3.3.3

Table 6. Explanation of Drill-Down Option (part a)

Sales Units	Jan	Feb	Mar	Total
2014				
New York	10	30	40	80
London	20	90	100	210
Total 2014	30	120	140	290
2015				
New York	70	60	10	140
London	10	80	30	120
Total 2015	80	140	40	260
Gran Total	110	260	180	550

Table 7. Explanation of the Drill-Down Option (Part b)

Sales Units	Jan		Feb	Mar	Total
	2014				
	New York				
Customer A	5		15	19	39
Customer B	2		11	14	27
Customer C	3		4	7	14
Total New York	10		30	40	80
	London				
Customer D	10		44	64	118
Customer E	4		29	16	49
Customer F	6		17	20	43
Total London	20		90	100	210
Total 2014	30		120	140	290
	2015				
	New York				
Customer A	35		30	5	70
Customer B	14		22	3	39
Customer C	21		8	2	31
Total New York	70		60	10	140
	London				
Customer D	5		39	26	70
Customer E	2		24	6	32
Customer F	3		17	8	28
Total London	10		80	40	130
Total 2015	80		140	50	270
Grand Total	110		260	190	560

4.4 DATA LOADING PROCESS

When the logical data model is ready, the loading process can start. Entities and cubes contained in the different data sources (CSV files) provided by company Carioca can be loading in the Data reader option, located in the section Database Manager of the Board's "DataBase Menu."

On the other hand, It is imperative to say that Board allows the use of different data sources like SAP, SQL, Text files, etc. In this work, the Text File option for loading the data was used.

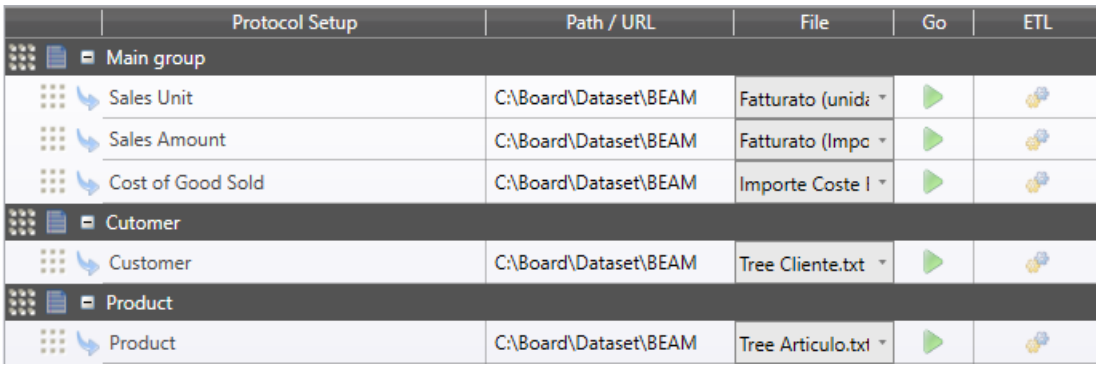


Figure 39. Data Reader Screen

The main parts of the Data Reader option are shown in Figure 39. Protocol Setup allows configured the loading process, which is selecting the cubes or entities that will be loaded and establish the type of separator in the text files. In the Path field, type the name of the directory where the text file is located. ; File, it is where the right document is selected for the data reader; Go, it is where the data readers are launch; and finally, ETL is the section where the data are modified according to the needs.

Figure 40 is exposed to the protocol setup layout, in the left part are the entities and cubes that will be loaded; they are selected in the right part where are all cubes and entities. In the bottom are the fields which contain the data of the CSV files.

(Note that in this case, the fields are empty because there is no selected any files) Position indicates the link between the entities and cubes created in Board with the CSV files, and finally, the actions depend on the needs and the type of data. i.e., the actions for entities code could be: Empty or Append and for the entity description and cubes: Empty or Replace. (See section 3.3.4)

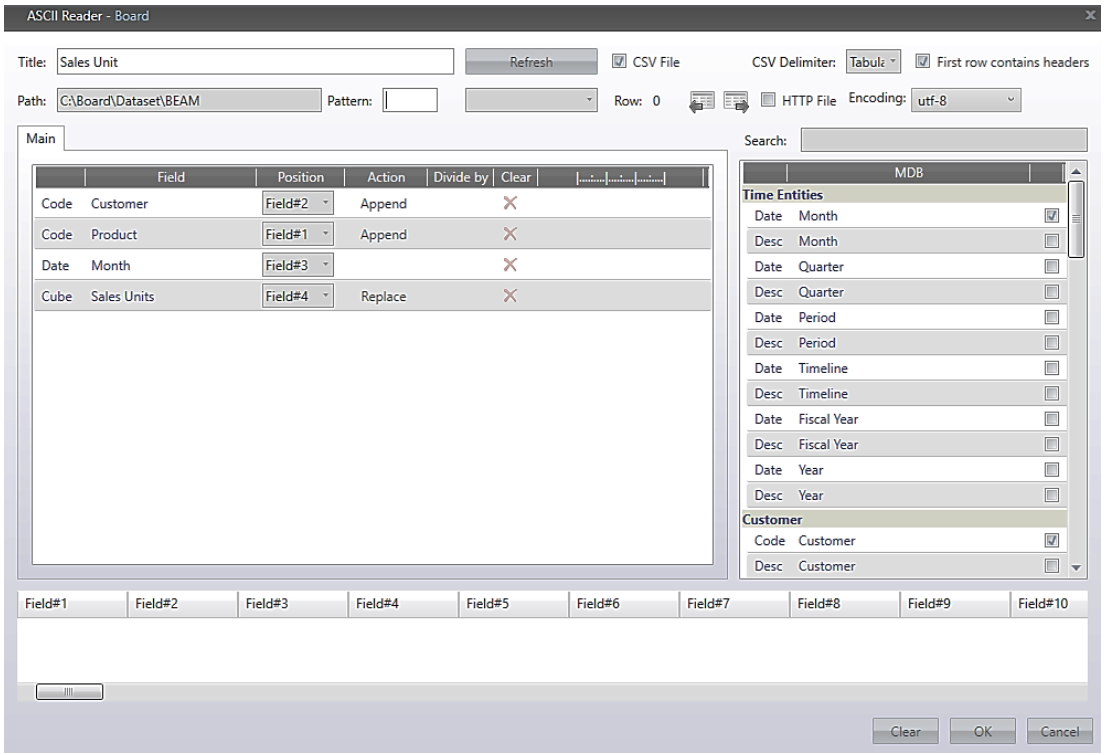
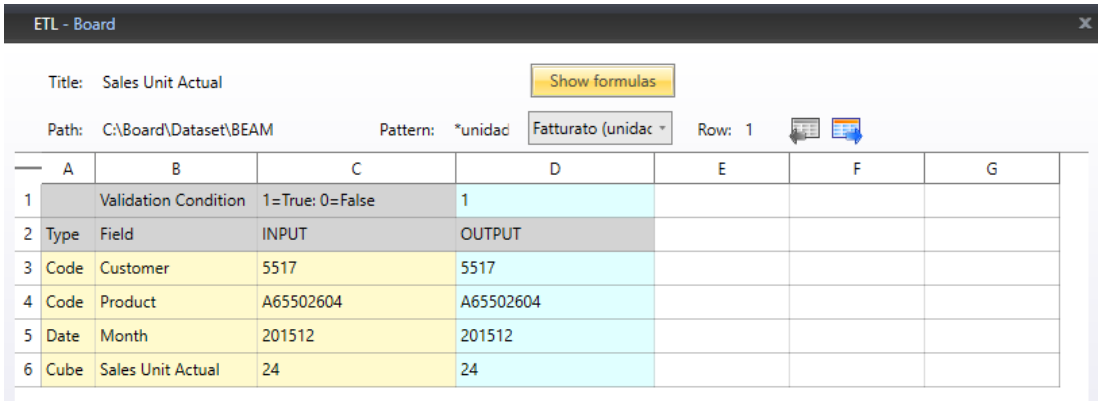


Figure 40. Data Reader: Protocol Setup

The ETL definition window () is a spreadsheet-like environment. The Cubes and Entities of the reader protocol are displayed in columns A and B (Type and Field). The ETL input (shown in column C) is the incoming external records from the source system. The ETL output, shown in column D, is the modified data (after applying the formulas) used to feed the Entities and Cubes.

In cell D1, it is feasible to define a record validation formula. If the end result of the expression is 1 (true), the record is validated and fed into Board database; if

the end result of the expression is zero (False), the document is discarded. The expression can refer to any cell of columns C to G.



	A	B	C	D	E	F	G
1		Validation Condition	1=True: 0=False	1			
2	Type	Field	INPUT	OUTPUT			
3	Code	Customer	5517	5517			
4	Code	Product	A65502604	A65502604			
5	Date	Month	201512	201512			
6	Cube	Sales Unit Actual	24	24			

Figure 41. ETL Definition Window

Table 8 is showing the summary of the data loaded after launch every data reader. Note that the column count only takes into consideration the unique data that means that, for example, if the files have three times the same cit, the column only considers this city like one value.

Table 8. Data Loaded Summary

Entity Group	Entity Level	Entity name	Count
Customer	3	Country	10
Customer	2	City	488
Customer	2	Salesman	26
Customer	2	Channel	8
Customer	1	Customer	1704
Product	2	Business Area	7
Product	2	Product Category	23
Product	1	Product	2094
Time Entities	3	Year	9
Time Entities	2	Period	12
Time Entities	2	Timeline	2
Time Entities	2	Quarter	36
Time Entities	1	Month	108

4.5 BUILDING METRICS AND REPORTS

When cubes, entities, and hierarchy have been created, and also, the data have been loaded. It is possible to start the building of metrics and procedures with the aim of creating reports. In order to build the necessary metrics, some cubes were created. These cubes will have the resultant information of the different processes and calculations that were carried out in Board as part of this work. Figure 42 shows all cubes created.

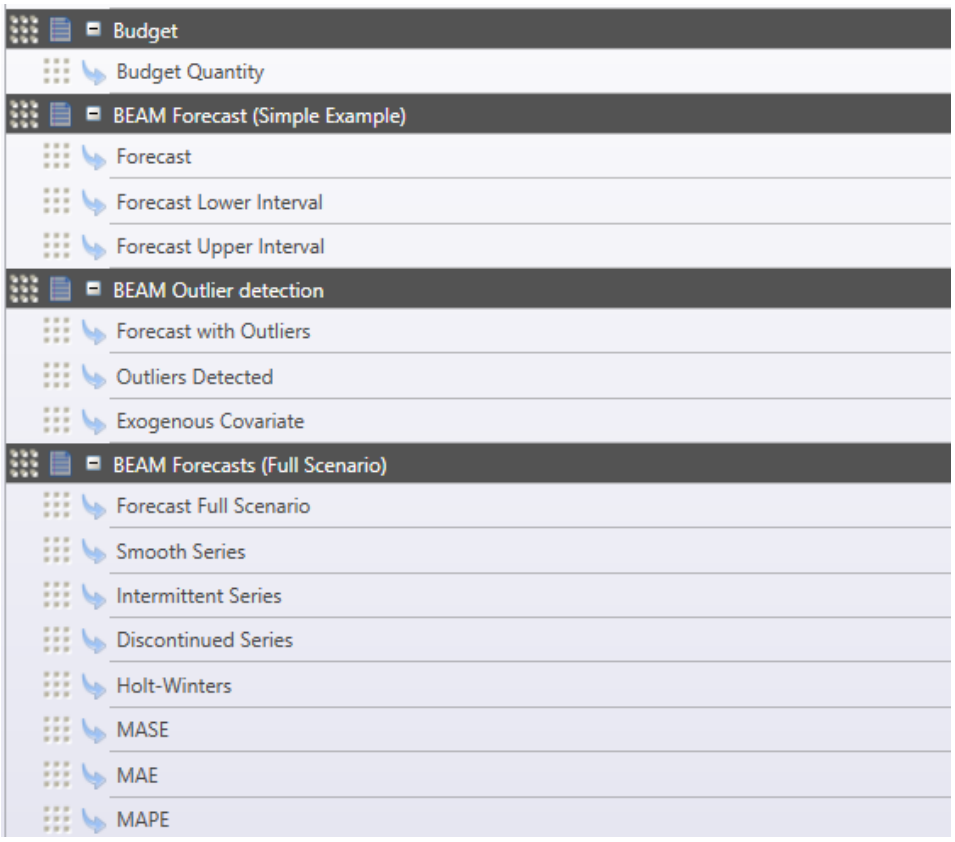


Figure 42. Principals cubes for the development of the model

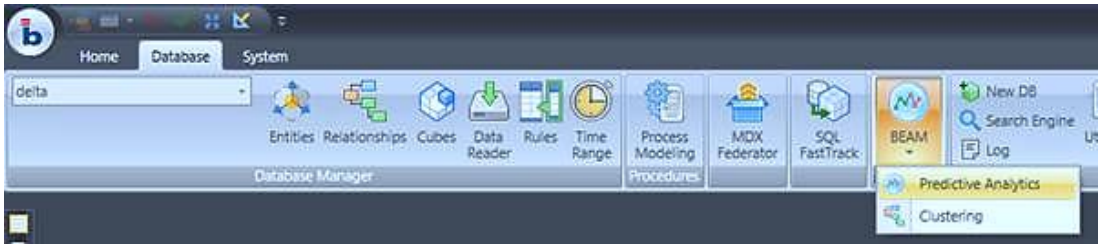


Figure 43. BEAM position in Board database

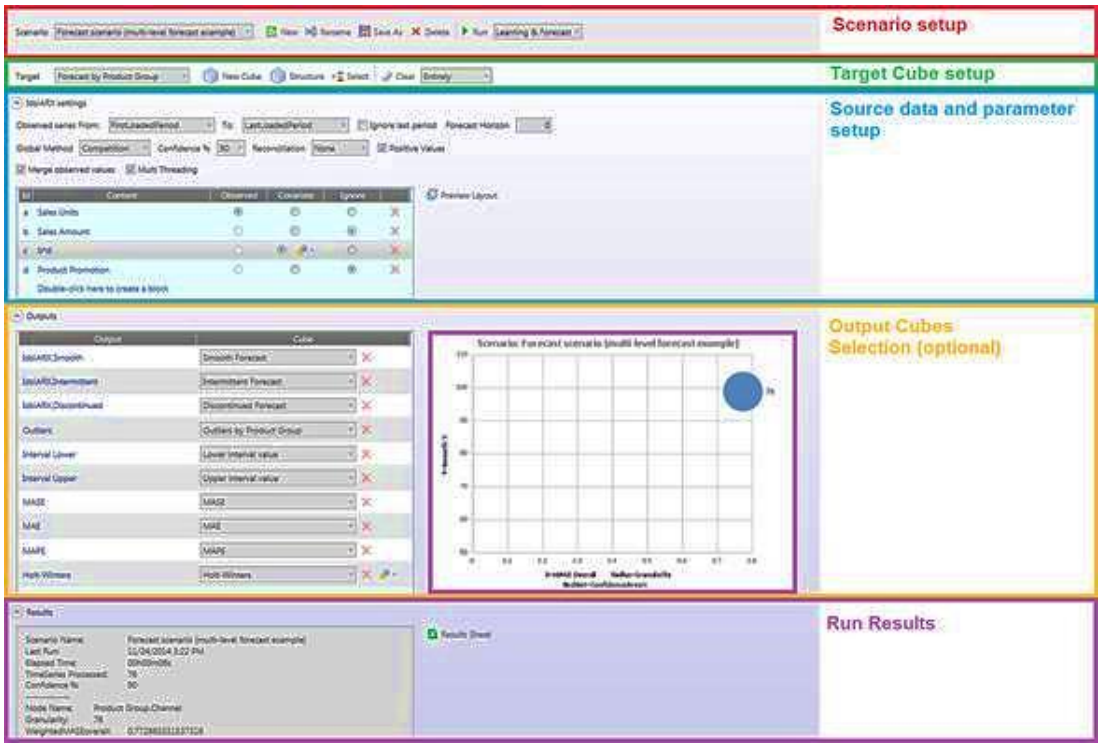


Figure 44. BEAM environment

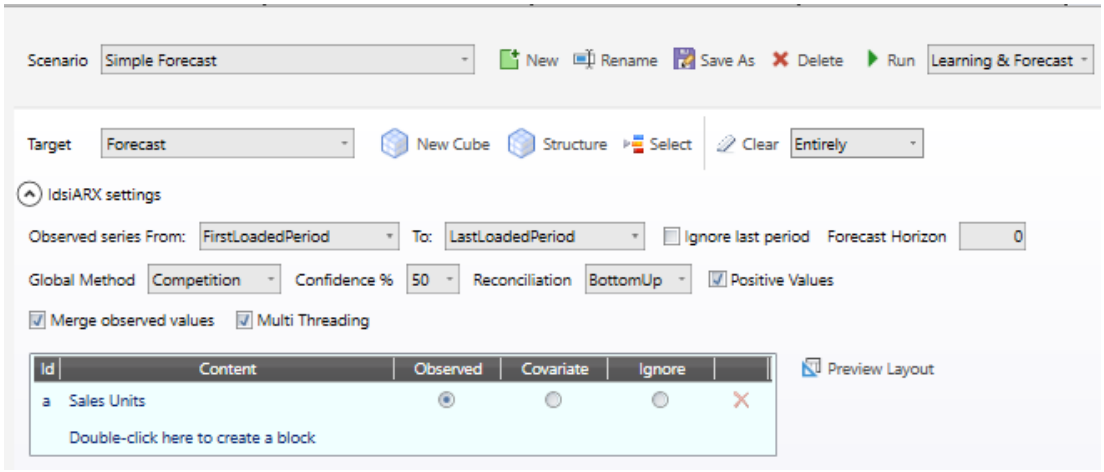


Figure 45. Source data and parameter setup

This section is the most important of the Scenario setup; the accuracy of the forecast mainly depends on the historical data set. In the two drop-down lists, we can select which periods will be observed, the default values are:

- **FirstLoadedPeriod:** The oldest value in the cube given the selection: please note that this value is unique and it is not evaluated for each time series;

- **LastLoadedPeriod:** The newest value in the cube given the selection: please note that this value is unique, and it is not evaluated for each time series.

If the last period data is not complete, we can exclude it in a dynamic way flagging the "ignore last period" checkbox.

- **Forecast Horizon:** here, it is possible to decide the amount of periods to forecast. The default 0 value will push the forecast until the end of the time range.

- **Confidence:** here, the level of confidence to calculate upper and lower intervals is decided.

- **Reconciliation:** Select one of the three reconciliation types; this is only needed with multi versioned target cubes.

- **Positive values:** The user can flag this option if he wants to automatically discard all the models that give some negative results in some period.

- **Merge observed values:** When this setting is on, the historical data will be copied to the cube along with the forecast.

Here we can decide the cubes and algorithms to use as source data. Cubes and algorithms can be set up as:

- **Observed:** there can be only one observed cube/algorithm. This will be the quantity the user is going to forecast.

- **Covariate:** there can be as much covariate as users want. Covariates are cubes containing time series with values on both past and future that have some sort of impact on the observed series (for example, it is possible to use the average temperature by month if the objective is to analyze Ice cream sales).

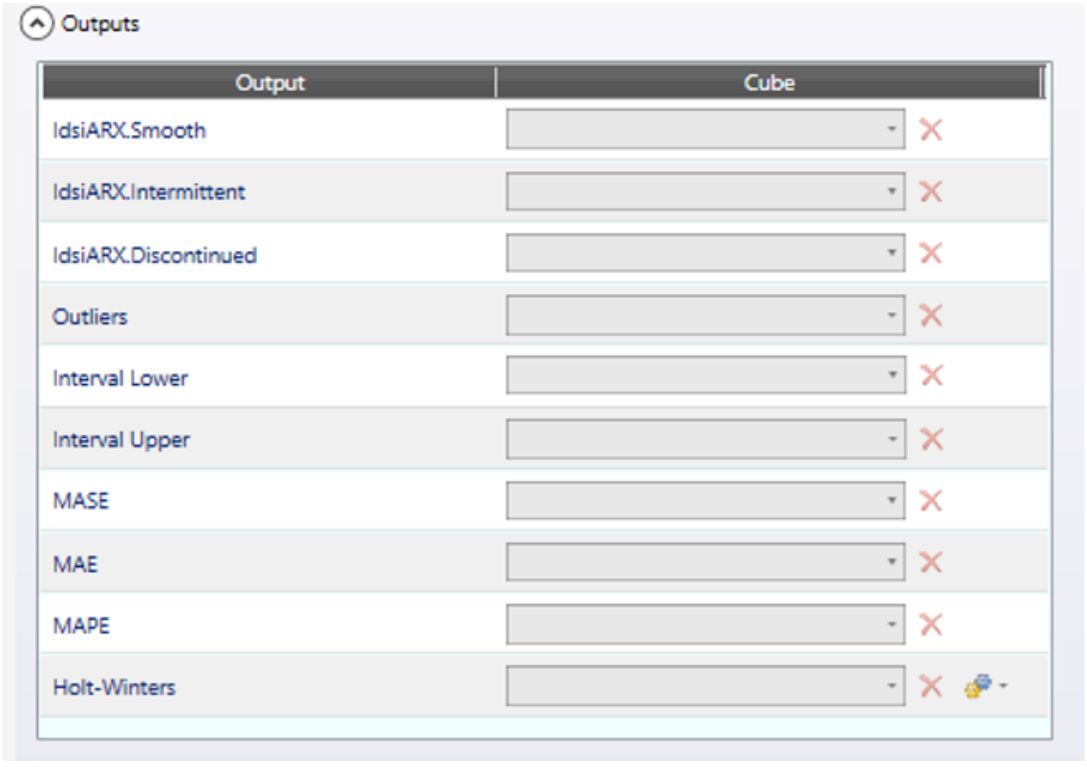


Figure 46. Output section

Other than the forecast in the target cube, Predictive Analytics outputs a lot of data. The user can decide to put this data into a series of cubes if he needs it.

The output cubes are not mandatory, as the target cube. Their structure will be the same as the target cube (if they have different structures, they will be automatically converted).

- **IdsiARX.Smooth:** This cube will be a slice of the target cube that will contain only the smooth time series.

- **IdsiARX.Intermittent:** This cube will be a slice of the target cube that will contain only the intermittent time series.

- **IdsiARX.Discontinued:** This cube will be a slice of the target cube that will contain only the discontinued time series.
- **Outliers:** this cube will only be populated on the past with the anomalous values of the various time series.

- **Interval lower:** this cube will contain the lower limit of the forecast; actual values will be higher than the interval lower and less than the interval upper with a probability equal to the confidence level.

- **Interval upper:** this cube will contain the upper limit of the forecast; actual values will be higher than the interval lower and less than the interval upper with a probability equal to the confidence level.

- **MASE:** Cube containing the MASE of each time series, the MASE of a period is the MASE of the model against the time series until that period.

- **MAE:** Cube containing the MAE of each time series, the MAE of a period is the MAE of the model against the time series until that period.

- **MAPE:** Cube containing the MAPE of each time series, the MAPE of a period is the MAPE of the model against the time series until that period.

- **Holt-Winters:** Also known as triple exponential smoothing, it will output the triple exponential smoothing of the series. The alpha, beta, and gamma parameters can be set up directly from the interface. Please note that Holt-Winters is the algorithm beneath the forecast time function in the block editor.

CHAPTER 5:

RESULTS

This chapter presents the results of the thesis. The content and function of every dashboard will be explained in order to understand its purpose in this work. Furthermore, the predictive analysis models created with the Board tools will be presented, the analysis of the trend and historical data will be created with the aim of gaining a superior comprehension of the actual and predictive data.

5.1 CREATING DASHBOARDS

A Dashboard is an information management tool that allows visualizing the most critical information of a set of data. The dashboards usually are used to simplify the information and only see the indicators (KPIs) that shows the information that could be added value to work or process. Dashboards can contain different visualization tools as tables, maps or graphs that help in the simplification and comprehension process of the information.

For this work, five main dashboards were designed. The Initial Dashboard, illustrated in Figure 47, allows selecting the two main functions of the project environment, the Analytical Function, or Predictive Analytics. If the user selects the Analytical Function he will be automatically redirected to the Analytic Dashboard (Figure 48) Where it is possible to obtain the statistics of the historical data, like the maximum and minimum value, standard deviation, average value and also let know when was the last or first no zero value.

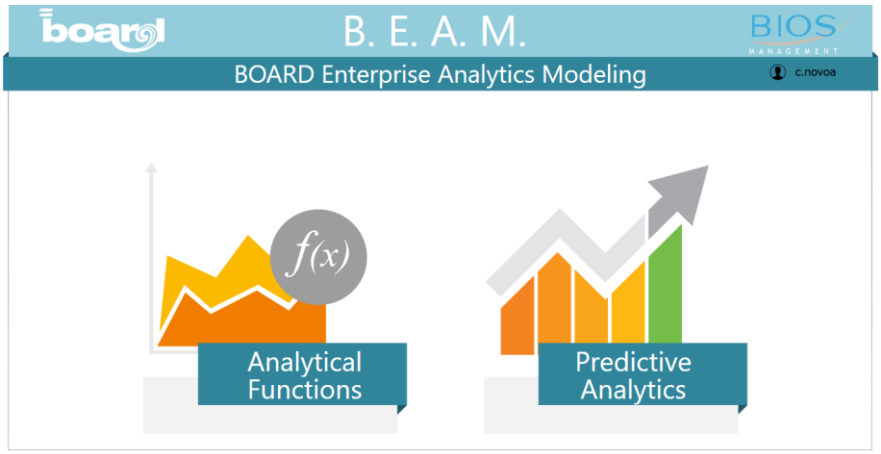


Figure 47. Initial Dashboard

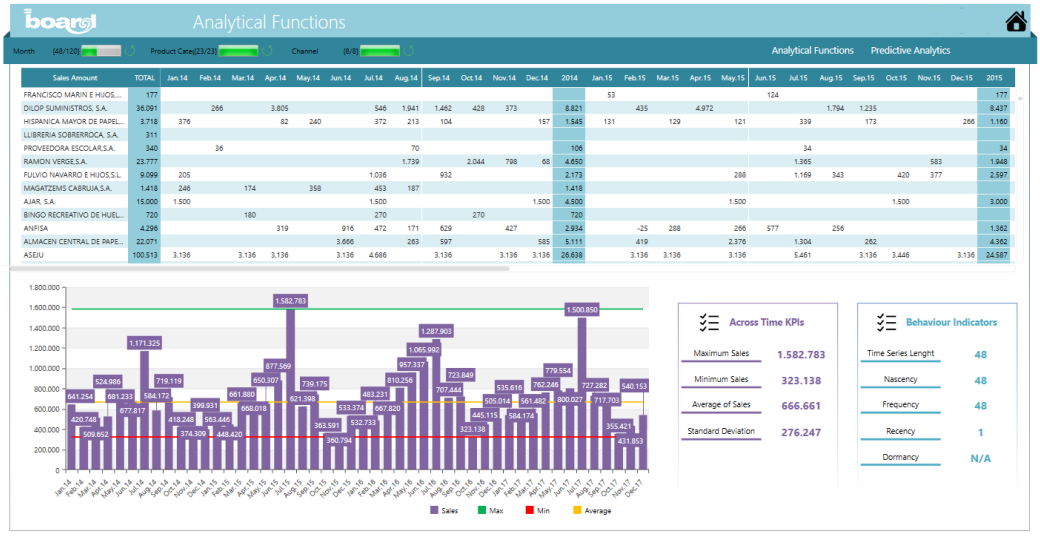


Figure 48. Analytic Dashboard

In addition to the above, the Analytical section has a dashboard that contains a comparison between the linear regression and the Forecast function of Board in order to understand the best accuracy for the sales in the company. Figure 49 and Figure 50 are a close view of the two tables present in the dashboard. The first one shows the linear regression analysis, where it is possible to appreciate that exists a high accuracy in the total value (0,35%) But, in the single month, the accuracy level exceeds 85%.

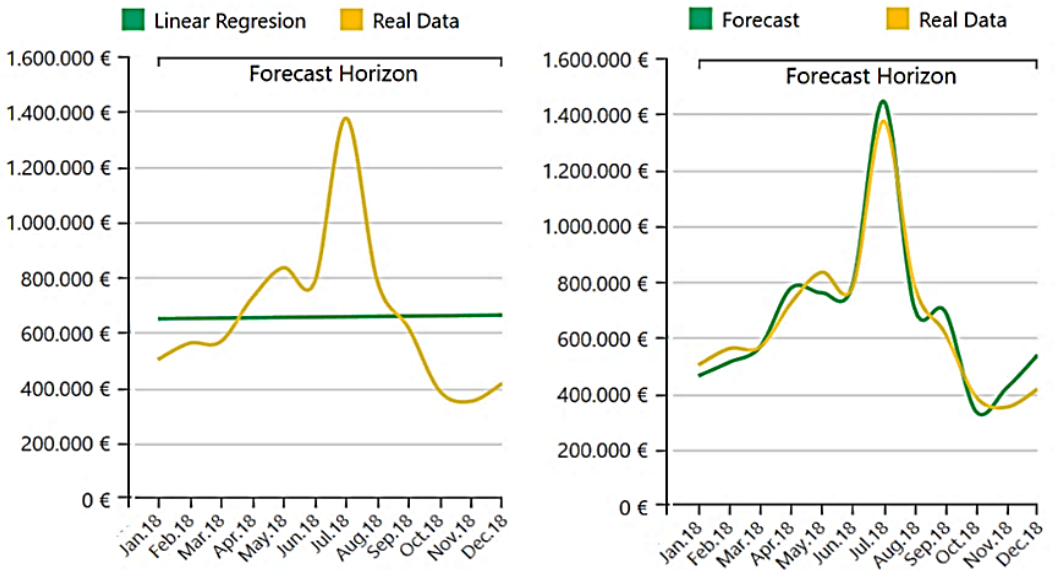
Timeline : Forecast Horizon			
Jan.18	649.460	502.442	29,26 %
Feb.18	650.672	560.924	16,00 %
Mar.18	651.884	566.442	15,08 %
Apr.18	653.096	723.414	9,72 %
May.18	654.308	833.906	21,54 %
Jun.18	655.520	780.648	16,03 %
Jul.18	656.732	1.373.718	52,19 %
Aug.18	657.944	787.313	16,43 %
Sep.18	659.156	616.234	6,97 %
Oct.18	660.368	388.327	70,05 %
Nov.18	661.580	351.583	88,17 %
Dec.18	662.792	416.223	59,24 %
Total Forecast Horizon	7.873.514	7.901.174	0,35 %

Figure 49. Linear Regression Accuracy

Timeline : Forecast Horizon			
Jan.18	463.455	502.442	7,76 %
Feb.18	512.133	560.924	8,70 %
Mar.18	567.146	566.442	0,12 %
Apr.18	776.692	723.414	7,36 %
May.18	760.051	833.906	8,86 %
Jun.18	787.279	780.648	0,85 %
Jul.18	1.444.908	1.373.718	5,18 %
Aug.18	708.517	787.313	10,01 %
Sep.18	693.624	616.234	12,56 %
Oct.18	336.130	388.327	13,44 %
Nov.18	420.884	351.583	19,71 %
Dec.18	537.980	416.223	29,25 %
Total Forecast Horizon	8.008.798	7.901.174	1,36 %

Figure 50. Board Forecast Analysis Accuracy

On the other hand, Figure 50 shows the result of the Board Forecast Function. Even when the macro-level has a lower accuracy comparing with the Linear Regression (1,36%), the single moths have a higher accuracy level. Only one value (December) exceeds 20%.



[a] Linear Regression Accuracy [b] Board Forecast Accuracy

Figure 51. Accuracy Level of Linear Regression and Board Forecast

Figure 51 shows a close view of Figure 52 with a particular focus on the year 2018, where it is possible to appreciate and comparing the level of accuracy of Forecast Board function and the linear regression. Board fits perfectly with the seasonality that occurs in the months with higher (July) and lower (October) sales, allowing Carioca staff to take the necessary measures to act in different situations. In addition to the option of drill-down, it is possible to know an estimate of the sales by Channel or Business Area (depend on the cube dimension) and based on them to have an approximation of every single product sales.

Finally, Figure 52 shows the graphical difference between the two models. Also, it can be seen that there is strong stationarity throughout the years. The peaks in sales are between the month of July and August, while the valleys are between the month of October and November. The second highest peaks are often in the months of March - April or August - September.

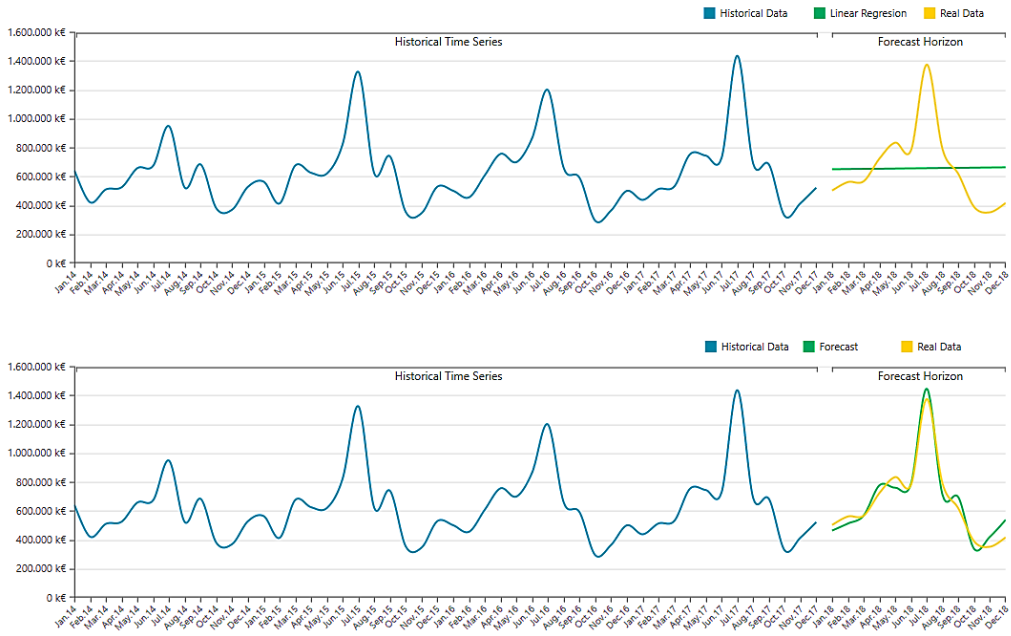


Figure 52. Graphic Analysis of the Linear Regression and Board Forecast

The behavior mentioned above is caused by the scholar calendar. Most schools in Europe start the academic period in mid-September, this cause that the most significant sales occur in the previous months (July and August) because retailers prepare for the start of classes period. The worse sales period are just after the beginning of the scholar calendar because there are few customers acquiring product in this period. And Finally, the two seconds' highest peaks are caused by the recession in the academic calendar due to the Christmas and Easter periods.

Figure 53, Figure 54, and Figure 57 expose the predictive models created through the Board tool BEAM. In these cases, unlike the previous dashboard, the models evaluate the quantities produced, and thanks to the Drill-Down option, it is possible to analyze not only the total amount produced but also every single article' quantity. Equal to sales, the quantities produced have a stationarity behavior. The company has its higher level of production in the month of July and the worse level just after starting the academic period. The main character in the three Figures is the Forecast Interval that was created with 95% confidence and predictive values.

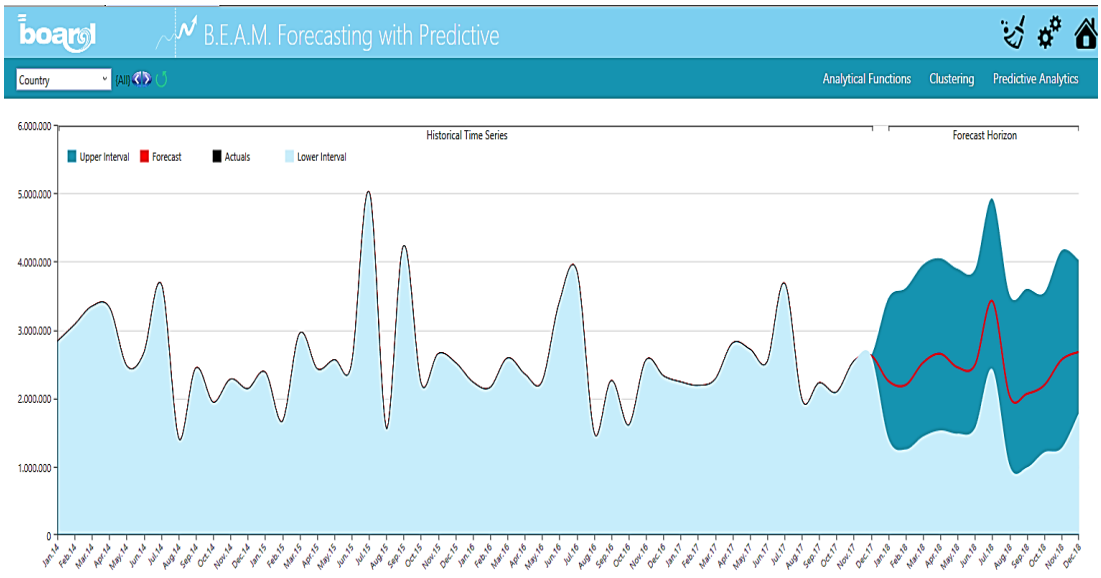


Figure 53. Production Forecast Interval

The difference between Figure 53 and Figure 54 is that in Figure 54, the user has the possibility to modify the forecast generated by Board or create his own estimation through a percentage increase of the data of the previous period. In this way, it is possible to adapt in a better way to market needs. For purposes of this work, an increase of 5% and 10% was used. In the lower right part, it is possible to see a comparison table between the real values and the values increased in the percentage chosen by the user.

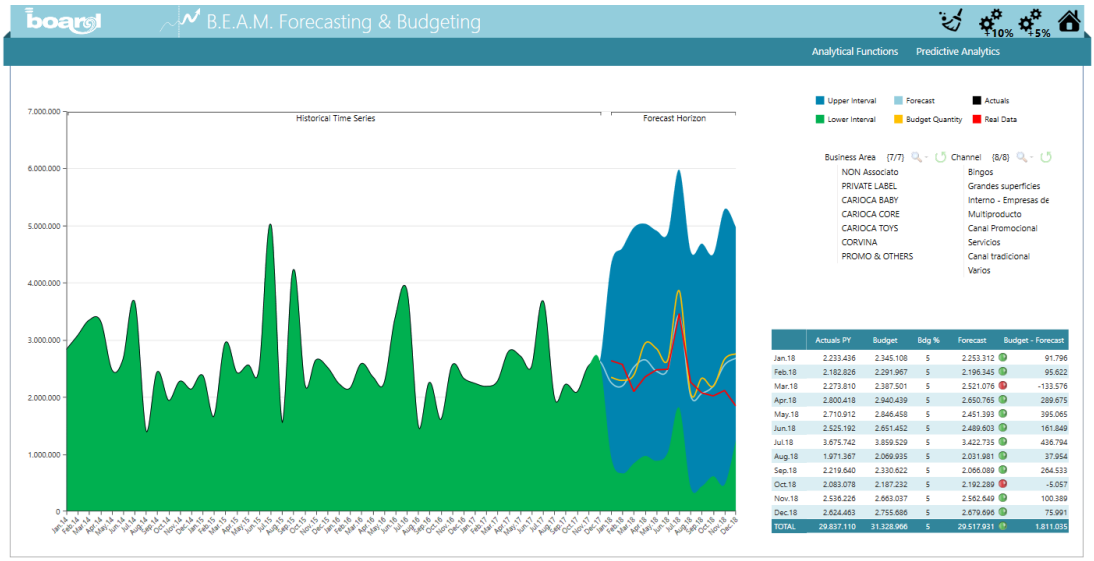


Figure 54. Production Forecast Interval with a customizable Budget

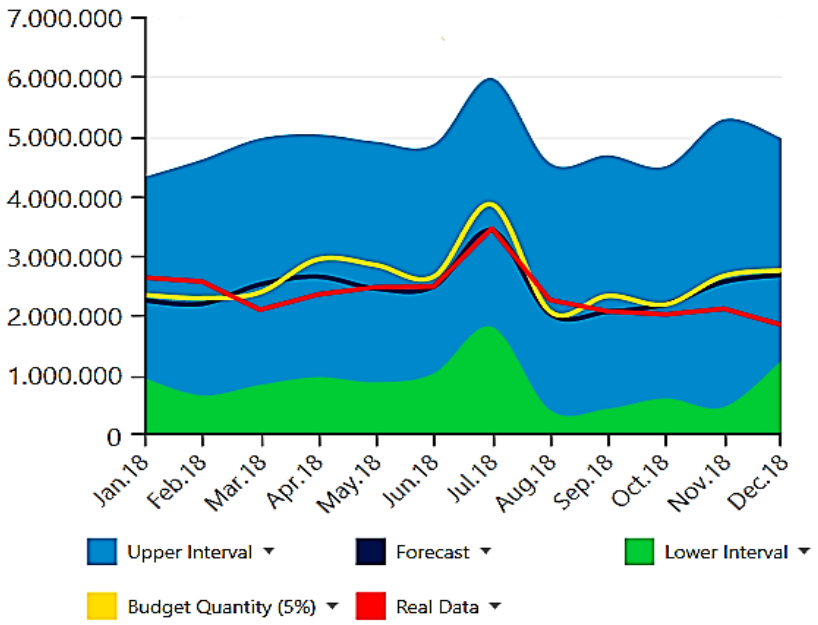
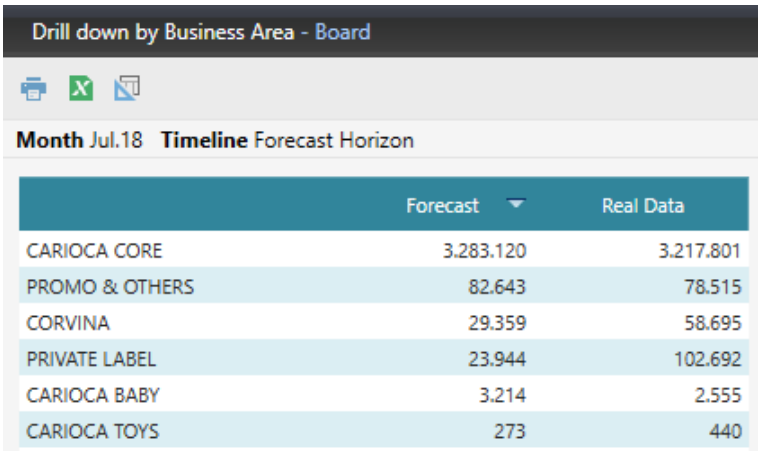


Figure 55. Production Forecast Accuracy

Figure 55. Production Forecast Accuracy has a close view of the previous image with a focus on the year 2018, which was used to analyze the level of accuracy of the predictive function of Board. In this case it can be seen how the Real Data (red) is within the confidence intervals, and in addition the Forecast line (dark blue) fits quite well with that of the current data.

On the other hand, the forecast cube, which has as dimension the Channel (which is part of the hierarchy of the client) and the Business Area (which is part of the hierarchy of the product), can be drill down in any of these entities. For example, in Figure 56, the drill-down made in July (month with the highest production) is appreciated where we see that the data is quite accurate and also with that estimate, it is possible to make the prediction of the production volumes of each Business Area.

The principal reason why the forecast cube has not directly as dimension the product and customer is due to it is complex to make a reasonable estimation with two entities that have a lot of members.



	Forecast	Real Data
CARIOCA CORE	3.283.120	3.217.801
PROMO & OTHERS	82.643	78.515
CORVINA	29.359	58.695
PRIVATE LABEL	23.944	102.692
CARIOCA BABY	3.214	2.555
CARIOCA TOYS	273	440

Figure 56. Drill-down by Business Area

Finally, in Figure 57 contains the models generated by Board, thus allowing the user to understand which model best suits their needs and the behavior of their historical data. Making a comparison between the real data and the models, we can appreciate that the model that best fits and has a higher level of accuracy is the Smooth model. It should be remembered that Board has excluded the outliers represented by the purple spots, and the variations in the school calendar have been inserted as a covariate for a better fit of the model.

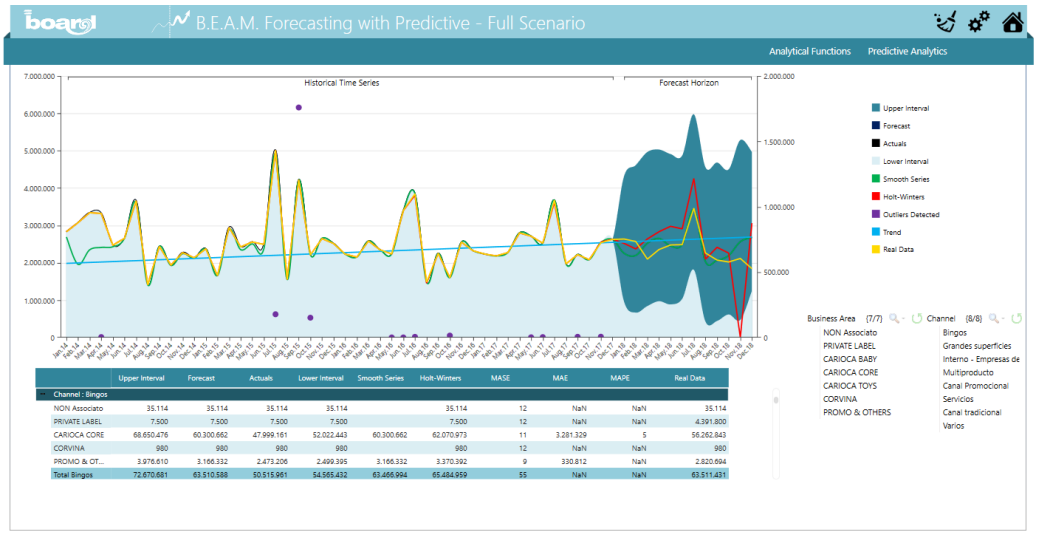


Figure 57. Full Forecast Analysis

CHAPTER 6:

CONCLUSIONS

Today most companies must analyze in a meticulously way their historical data to understand the past and the current behavior of their customers. The principal problem of this lies in the immeasurable amount of information that is usually generated every day. In the Big Data age, it is compulsory the use of tools like Business Intelligence in order to manage the data.

The goal of the thesis was to develop a Business Intelligence environment that allows the users to create analysis capable of predicting with the significant accuracy the needed production level and the sales behavior. This Business Intelligence solution will be used as a decision support system for the management of the company Carioca Spa. All dashboard contains valuable data represented in different tables or charts. The interface is user-friendly, and it is structured in such a way that any user could navigate through it easily.

This thesis fulfilled its aim and created a solution that allows making data-driven strategic decisions. Furthermore, the Business Intelligence solution presented in this work is sustainable, that means that every report exposed in its interface will be available along the time and even in the future when new data is generated only will be necessary to update the Data Reader. It is possible to schedule an automatic updating through the task scheduler.

With the implementation, it is modeled a system in which the historical data is used in order to predict future events. Thanks to that, the company can prepare

the required actions to face different possibilities. Also the solution could help the finance department at the moment of creating budget and forecast for future periods. Carioca can use Board and the application designed in order to satisfy all the current needs and possible futures requirements. The solution evidences the advantages and easily of using new technologies in the predictive analytics field through business intelligence tools.

Based on the results, it possible to say that Board Forecasts function has a highly accurate level and could improve the decision-making process of Carioca. The production rate and sales have, in general, a degree of accuracy of 4% and 1,36%, respectively. The most customers of the company are retail sellers that usually start the acquisition of products a couple of months before the beginning of the academic year of the Italian schools (generally early September). For this reason, both production and sales have stationary behavior, and July is the month with the higher sales and production rates.

Regarding the above, one possible solution to eliminate the seasonality in the sales and production could be to open new markets in countries with a different scholar academic calendar (ex. Latin-American countries). However, it is mandatory the evaluation of the opportunity cost of making a big decision like that because probably will require a high investment

The results obtained within this work shown a good accuracy level to the company sales and production behavior; nevertheless, it does not reflect it in a hundred percent. The future forecast values should be taken as just a suggestion for the company, based on data analysis, to perform its activities. However, it is its decision to follow what the software proposes or to use different criteria for the decision-making process.

BIBLIOGRAPHY

Azma, F., & Mostafapour, M. (2012). Business intelligence as a key strategy for development organizations. *Procedia Technology. 1*, pp. 102 – 106.

Bios Management. (n.d.). *L'azienda*. Retrieved from Bios Management: https://www.biosmanagement.com/it/l-azienda

Board International. (n.d.). *About us: International Recognition*. Retrieved from Board International: https://www.board.com/en/international-recognitions

Board International. (n.d.). *Capsule and Process Modelling*. Retrieved from Board: https://www.boardmanual.com/

Board International. (n.d.). *Database Design and Management*. Retrieved from Board: https://www.boardmanual.com/

Board International. (n.d.). *Predictive Analytics (B.E.A.M.)*. Retrieved from Board: https://www.boardmanual.com/

Bose, R. (2015). Advanced Analytics: opportunities and challenges. *Industrial Management and Data Systems*, pp. 155 - 172.

Brinkmann, D. (2015). *Strategic capability through business intelligence applications.*

Businska, L., & Supulniece, I. (2011). Toward Systematic Reflection of Data, Information and Knowledge. *Scientific Journal of Riga Technical University*, pp. 12-18.

Carioca. (n.d.). *About us*. Retrieved from Carioca: https://carioca.com/gb/about-us

Chee, T., Chan, L.-K., Chuah, M.-H., Tan, C.-S., Wong, S. F., & Yeoh, W. (2009). Business Intelligence Systems: State-of-the-art Review and Contemporary Applications. *Symposium on Progress in Information and Communication Tech*.

Chen, J. (2019). *Autoregressive Integrated Moving Average (ARIMA)*. Retrieved from Investopedia: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp

Columbus, L. (2018). *The State Of Business Intelligence, 2018*. Retrieved from Forbes: https://www.forbes.com/sites/louiscolumbus/2018/06/08/the-state-of-business-intelligence-2018/#16f3fd657828

Frankenfield, J. (2019). *Data Warehousing*. Retrieved from Investopedia: https://www.investopedia.com/terms/d/data-warehousing.asp

Gartner. (2017). *Gartner Says Worldwide Business Intelligence and Analytics Market to Reach $18.3 Billion in 2017*. Retrieved from Gartner: https://www.gartner.com/en/newsroom/press-releases/2017-02-17-gartner-says-worldwide-business-intelligence-and-analytics-market-to-reach-18-billion-in-2017

Gartner. (2019). *Magic Quadrant for Analytics and Business Intelligence Platforms*.

Gonçalves Teixeira, A. I. (2016). Evaluation of Business Intelligence Project Success in Tintas Robbialac, SA: An Exploratory Case Study. *Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa*.

Groves, D. G., Fischbach, J. R., Bloom, E., Knopman, D., & Keefe, R. (2013). *Adapting to a Changing Colorado River*.

Ika, L. (2009). Project Success as a Topic in Project Management Journals. *Project Management Journal, 40(4)*, pp. 6 - 19.

Isik, O., Jones, M. C., & Sidorova, A. (2013). Business Intelligence (BI) Success and the Role of BI Capabilities and Decision Environments. *Intelligent Systems in Accounting Finance & Management*, pp. 161 - 176.

Khang, D., & Moe, T. (2008). Success Criteria and Factors for International Development Projects: A Life-Cycle-Based Framework. *Project Management Journal, 39(1)*, pp. 72 - 84.

Kříž, J. (2015). Business Intelligence řešení pro společnost 1188. *Brno University of Technology*.

Llanos Ferraris, D. R. (2010). *Fundamentos de informática y programación en C*. Editorial Paraninfo.

Maheshwari, A. K. (2015). *Business Intelligence and Data Mining*. New York: Business Expert Press.

Marudappa, P., & Selvaraj, P. (2018). A Survey of Predictive Analytics Using Big Data with Data Mining. *International Journal of Bioinformatics Research and Applications. 14*, pp. 269.

Moss, L. T., & Atre, S. (2003). *Business intelligence roadmap: the complete project lifecycle for decision-support applications*. Boston, Massachusetts.

Olszak, C., & Ziemba, E. (2007). Approach to Building and Implementing Business Intelligence Systems. *Interdisciplinary Journal of Informations, Knowledge and Management, vol. 2*, pp. 135 - 148.

Pelegrini, F. G. (2015). Fatores Críticos Na Gestão De Projetos: Um Estudo De Caso Numa Empresa De Grande Porte. *Gestão & Produção, 22(2)*, pp. 280 - 294.

Ponniah, P. (2001). *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. Paulraj Ponniah.*

Rechenthin, D. (2013). *Project Intelligence.* Newtown Square, PA: Project Management Institute.

Schmarzo, B. (2014). *Business Analytics: Moving From Descriptive To Predictive Analytics*. Retrieved from https://infocus.dellemc.com/william_schmarzo/business-analytics-moving-from-descriptive-to-predictive-analytics/

Serra, C. E., & Kunc, M. (2014). Benefits Realization Management and its influence on project success and on the execution of business strategies. *International Journal of Project Management, 33(1)*, pp. 53 - 66.

Siegel, E. (2010). *Seven Reasons You Need Predictive Analytics Today.* IBM sponsored, Prediction Impact, Inc.

Software Advice. (2019). *FrontRunners, Business Intelligence Software.* Gartner.

Syntetos, A., Babai, M. Z., & Gardner, E. (2014). *Forecasting Intermittent Inventory Demands: Simple Parametric Methods vs. Bootstrapping.*

Tableau. (2019). *Data visualization beginner's guide: a definition, examples, and learning resources*. Retrieved from Tableau: https://www.tableau.com/learn/articles/data-visualization

Vizgaitytė, G., & Rimvydas, S. (2012). Business Intelligence in the Process of Decision Making: Changes and Trends. *Ekonomika, 91(3)*, pp. 147-157.

White, D. (2010). *Predictive Analytics: The Right Tool for Tough Times.* An Aberdeen Group white paper.

Wilensky, H. (1967). *Organizational Intelligence: Knowledge and Policy in Government and Industry.* New York: Basic Books.