BUSINESS INTELLIGENCE REQUIREMENT ANALYSIS IN SMALL AND MEDIUM ENTERPRISES

A MASTER'S THESIS

in

Industrial Engineering

Atilim University

FATIH ISIK

MAY 2016

BUSINESS INTELLIGENCE REQUIREMENT ANALYSIS IN SMALL AND MEDIUM ENTERPRISES

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

BY

FATIH ISIK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF INDUSTRIAL ENGINEERING

MAY 2016

Approval of the Graduate School of Natural and Applied Sciences, Atılım University.

_____

Prof. Dr. İbrahim Akman

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Assoc. Prof. Dr. Turan Erman Erkan

Head of Department

This is to certify that we have read the thesis "Business Intelligence Small and Medium Enterprises Requirement Analysis" submitted by "Fatih IŞIK" and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Supervisor

Assoc. Prof. Dr. Turan Erman Erkan _____

Examining Committee Members

Asst. Prof. Dr. Uğur Baç _____

Asst. Prof. Dr. Salih Tekin _____

Date: 04.05.2016

I declare and guarantee that all data, knowledge and information in this document has been obtained, processed and presented in accordance with academic rules and ethical conduct. Based on these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Fatih Işık

Signature:

ABSTRACT

BUSINESS INTELLIGENCE REQUIREMENT ANALYSIS IN SMALL AND MEDIUM ENTERPRISES

Isik, Fatih

Industrial Engineering Department

Supervisor: Assoc. Prof. Dr. Turan Erman Erkan

May 2016, 71 pages

This thesis reviews and discusses benefits of the Business Intelligence usage by the requirement analysis on various small and medium enterprises in Turkey.

ÖZ

KÜÇÜK VE ORTA ÖLÇEKLİ İŞLETMELERDE İŞ ZEKASI İHTİYAÇ ANALİZİ

Işık, Fatih

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Yöneticisi: Doç. Dr. Turan Erman Erkan

Mayıs 2016, 71 sayfa

Bu çalışmada, Türkiye'deki çeşitli küçük ve orta ölçekli işletmeler üzerinde yapılan ihtiyaç analizi kapsamında İş Zekası kullanımının faydaları üzerinde irdeleme yapacak şekilde bir uygulama gerçekleştirilmiştir.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Business Intelligence History

Even though the date of appearance is before the Christmas, Business Intelligence is still a new and wise solution for today. A famous Chinese author, Sun Tzu, was the first who stated Business Intelligence at his book in accordance with the military strategy. He was thinking that to win of a war, you must know the strong and weak sides of both you and your enemy. For the Business Intelligence today, that is a kind of cornerstone.

If we have a look at the recent years, Howard Dresner, who is known as a founder of the term BI because he stated some techniques to take the decision making forward via usage of the required data sources.

So the people can have a general conclusion states that it is not enough only to understand your current customers, that is, you should also know your competitors. Otherwise, the success is far away from you. According to these general ideas, the BI tools have been developed very quickly as they give an opportunity to understand the customers in a better way.

In relation with the Business Intelligence, some milestones can be given as below in a chronologic calendar:

1865: Richard Millar Deven's Cyclopaedia of Commercial and Business Anecdotes contains the first use of the term "business intelligence".

1958: IBM researcher Hans Peter Luhn defines Business Intelligence: **"The ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal."**

1989: Future Gartner Group analyst Howard Dresner defines Business Intelligence: **"Concepts and methods to improve business decision making by using fact-based support systems."**

1993: Commercialization of the internet begins.

1996: RENFE, the Spanish railroad company creates the first public facing BI application.

1997: The term Business Intelligence becomes widely used.

2003: Facebook launches.

2004: Cindi Howson launches BI Scorecard.

2005: Youtube and Twitter launch.

2007: Apple releases iPhone.

2008: Android launches.

2010: Adoption of BI grows and 35% of organizations employ pervasive BI when 67% of "best in class" companies have some self-service BI.

2011: Top BI trends appears as; big data, cloud computing, data visualization, mashups, mobile, predictive analytics, search, online customer experience and social networking.

2012: Technological features on the rise: ability to perform federated queries and ability to generate reports from in-memory database.

Business Intelligence is still growing up, and what started as a back-office function is now something consumers use every day (Harris C., 2012). The battle between BI companies today is to provide speed, affordability and high capacity storage. And the BI market is expected to reach $20.8 billion by 2018.

1.2 Definition and Use of Business Intelligence

Throughout a strategical planning, the "Business Intelligence" term represents systems and tools which play an important role. These systems bring an opportunity to a company to access, collect and to analyze the obtained datas in order to use for decision-making. Generally, we face such systems of Business Intelligence in the subjects like marketing research/segmentation, consumer profile and support, inventory and/or statistical analysis etc.

For the company managers, BI provides a capacity to analyze the capability for discovering the verified data in order to support a current decision and to explore some daily business activities. For instance, if we take a look at the IBM for nowadays, BI is a concept for the application of some set of technologies via usage of an amount of data to receive significant information (Turban E., Sharda R., Delen D., King D., 2010).

The systems which are stated above are the guides for important decisions in the business. It is also important to react quickly and to find out the best decisions at the right time. If BI is used in a right way, it can support the decision making as well.

And with the use of BI systems, it is also possible to have some gains such as the exhaustive communications between the departments, well-coordinated activities as well as the express reactions and decisions depending on the changes.

Mainly, for the environment of today's world, BI is one of the valuable technologies among the IT tools and at the same time, it helps to get a better performance while usage of the other systems of IT as well (Vercellis C., 2009).

For the company directors and managers in order to have quicker and wise decisions, the technologies such as Business Intelligence have a very important role, especially for the enterprises that update the information the managers/directors already have. In the last 20 years, there has been a big growth for both of the services and products in the industry. Business men are generally looking for advanced data analyze methods to get the rightest business decisions and for providing the best services to the consumers. Due to the fact that, BI is used for manufacturing, for the shipments/transportations, for any kind of support and/or feedback services etc.

BI works are generally applied over a data which is integrated to a data warehouse. Such a data warehouse is managed with some servers and by special database management systems. In the last years, some kinds of data optimization techniques and processes have been developed. These techniques are generally used in order to execute SQL queries when we have a big amount of data.

In addition to these traditional servers, new "in memory BI engines" are also used nowadays. So the big memory sizes are operated in a good way.

Enterprise search engines are also usable tools of BI in the last years due to the fact that we can note the other BI technologies such as web analytics, text analytic and/or data mining engines etc. Likewise, CRM are commonly used and it supports the functionality for the segmentation of the customers (Turban E., Sharda R., Delen D., King D., 2010).

In BI, generally it is needed to have a search over some different types of data among the enterprises. For instance, a professional sales man who will have an appointment with a valuable customer, would like to have a detailed customer information before the upcoming meeting. This kind of information is formed by different sources which are; CRM database, e-mails and some type of spreadsheets and/or documents etc. In such a situation, it is very important to rank and group the information with using a keyword while the search in BI system. The main focus while such an enterprise search should be using a keyword that is familiar in the BI systems. Nowadays, as we know, some vendors such as Google's Search provide us an opportunity to enterprise the capability of our search.

It generally takes a long time to investigate and develop a BI project. Regarding their sizes and scopes, some BI projects may be complex as well. Due to their complexity, we can face some kind of problems while the project and there is a must to have information requirements to overcome these problems. It will probably reduce the possibility of failure to get the right set from the requirements.

During the last decade, a lot of new methods were found in order to analyze the requirements. Some of the current ones are; "Supply-driven, Demand-driven, Goal-driven, Model and Ontology-driven information requirement analysis". A very

important detail for all methods is that; BI systems have to be capable and/or flexible in order to be compatible with the changing current requirements. And the best exercise which can be done by engineers and/or consultants is finding a current number of cases for developing a new method in order to verify the information requirements. So that, it is wise to use some tools and/or methods such as the mock-ups, prototype setups and wire frames as a solution in the BI (Vercellis C., 2009).

It's an important criteria to understand the clients for the marketing departments as it is also very important to understand how much a campaign is successful. When there are some current tasks of the data mining, a market sales man could plan his own strategy. But one of the biggest obstacles for a clear usage of BI is that generally the market sales man may not see the real profits of a system that he faces while meeting his objective.

A major consumption for the marketing department is "advertisement". Due to the nature of the available number of the clients, it is very important to understand which advertisements are effective for what kind of clients the most (Dyche J., 2002). Logically, that increases the take-up rates of the applied campaigns of advertisement as well. Here, of course, IT has an important role as it states the key insights of BI. Thus, it's also important to discuss what kind of BI technique should be applied during this process.

There are different various techniques of data mining and BI. These techniques are used as a tool for minimizing the marketing expenses and to maximize the profits of a campaign. Here it's also important to define the potential clients and save the clients that already exist. And of course, there is need to state what people need and what they expect (Hillier, F., S., Lieberman, G., J., 1995). While application of the data mining methods for marketing, it could be useful to mention about that the data mining tasks which can be used are; clustering, classification, pattern association, summarization, predictive modeling, social network analysis and the link analysis. In the third part, we'll have a look at these tasks and will describe the clustering analysis in details.

Mainly, BI usage could be categorized to the following categories:

A. Business operations reporting: One of the most common form of BI is the business operations reporting. That contains the actuals and says how the actuals are stacking up against our goals. This type of BI is manifesting itself often in the standard weekly and/or monthly reports which have to be developed.

B. Forecasting: Many of us have no doubt while running to the requirements for the forecasting and we all will agree that the forecasting is a science and at the same time an art. Forecasting is an art as we can never be sure about what the future brings. What happens if the competitors decide spending a big amount of their money for advertising? And what happens if the price of oil rises up to $250 for a barrel? At the same time, it's a science as we could extrapolate from the historical data and it isn't a guess totally (Dresner, H., 2007).

C. Dashboard: The primary purpose of the dashboard is conveying the knowledge at a view. For such an audience, there is a little requirement for drilling down for the

data. At the same time, the presentation as well as the ease of use is very important for the dashboard for being useful.

D. Multidimensional analysis: The Multidimensional analyze is a kind of slicing and/or dicing of the given data. It offers a good insight for the numbers at a granular level. And that requires a good data warehousing for the business analysts to get the required data.

E. Finding correlation among different factors: That is diving so deeply into the BI. And the questions are like: How are the different factors correlating to one another? Are there significant time trends which could be anticipated?

1.3 Business Intelligence Tools

There are some methods for creating dashboards by using the common tools like Excel, Access and proprietary systems which are embedded at the databases. However, these approaches could lead to that the dashboards aren't very user-friendly as well as difficult for being updated.

Business intelligence tools help us taking the dashboard ideas into a next level. More than a simply and graphically displayed static data, they also offer us trend analysis, drilldown capabilities and forecasting which could expand our insights into the program performance (Duda R.O, Hart P.E, 2001). Via a good business intelligence tool, we are able to combine the data which are coming from different sources and we can view it from multiple perspectives as well as can distribute it much more easily.

Beyond the simple reporting, business intelligence tools also allow a more comprehensive analyze of the data of our organization. Every organization doesn't need a business intelligence tool, however if we have a solid strategy for evaluating a program and to monitor and capture the raw data which we need, but struggling for analyzing the data and make the use of it, one of such tools may be a good fit. Let's take a look at what they could offer us and how they are working.

The BI tools have grown in its popularity with nonprofits including a number of reasons. The most prominent of these reasons are; flexibility, better data integration and easier distribution of data as well as the other good visuals. Some organizations have a data which can be compartmentalized in some separate information groups that we can have a donor information in one of the databases, customer information at a second available database and a financial information in the third database. Business intelligence tools let us integrating the data from
some numerous sources and we can even pull the data from an Excel spreadsheet and/or Access databases as well as from any database which has an "Application Programming Interface" that is a way for the programmers in order to access and/or export the data (Levin N, Zahavi J, 1999).

Most of the business intelligence tools are moving the data to a virtual storage space and/or to a different data warehouse via developing a different database. That also allows the data being manipulated for the analysis without affecting our data defined at the original database. Data could be also loaded manually and set for being loaded automatically at the determined times and/or when a certain action occurs. If we

want to have a more flexible reporting, then we could get it from our existing systems and the business intelligence tools may offer even a more flexibility than something such as Excel and/or Access and may also display data dynamically from an amount of perspectives near the real-time. As the data is being loaded from some original sources to the warehouses, it isn't exactly "real time," but it can be assumed as "close" (Peppers D., & Rogers M., 2004). For instance, it could be possible to get a report at the amount of the people who are admitted to our program and to see our success rate at some stages. Then we face a trend analyze of our success rates over the time. They also provide much more diverse methods to access and distribute the reports and/or the dashboards. Then we can e-mail the staff "up to the minute" program data on a scheduled day and/or time and can view the dashboards from a smartphone and/or a tablet where we have the internet connection. That also lets us setting up the custom portals for some certain audiences who has the unique Access for viewing the predefined performance measures and/or the reports. All of the data in the union wouldn't help if it couldn't be displayed in a method which can easily be interpreted. Business intelligence tools are going far beyond the dials as well as the bar graphs and could provide some sophisticated graphics such as the scatter plots which are moving with time and sparklines which show a lot of data points and/or forecasts which are assuming different scenarios (ReibStein D., Philip E. Pfeifer, Neil Bendle, Ronald T. Wilcox, Paul Farris, Rajkumar Venkatesan, Cesar Brea, 2014). These graphics do not only display more information, but they could also be customized with the colors and/or the themes which can match the brand of our organization.

Although the business intelligence tools can vary in cost, most of the organizations could find out an affordable system for the specific price range that satisfies their requirements. However, there are also some other things for being considered before we are investing any time and/or money including our organization strategy, data collection as well as the personnel (Gartner, 2008). Our organization should have a well defined strategy for the evaluation in place. Here the question is; do we know that how our organization wants analyzing its programs and what metrics for track? A business intelligence tool can give us the answer and spending the time that are outlining the requirements in our process helps to save much time and/or money later. These metrics should be very specific. We should also outline the data fields which we want to use and how we hope analyzing them. However, we are choosing to define them, make sure that we know the basics what we want to analyze and/or how we are planning to make it before beginning the shopping for a business intelligence tool even though that means consulting with somebody who is able to help defining an effective as well as an appropriate plan in order to measure our data (Koch I, 2013). After implementing a business intelligence tool, some organizations can struggle via operationalizing the data for building a powerful performance culture in other words, finding some ways for improving the performance based on our data. Our organization might want to change the processes as it begins to analyze the program data from some different angles. But if the data in the business intelligence tool isn't organized the same as in our system, such changes could be very difficult for enacting. The data warehousing develops different data models and could result in making the key data points very difficult for operationalizing. Here if one of our goals is purchasing a business intelligence tool, we might want to consider re-organizing our data model the first. If the difference in the data models is greater, that means a more difficult or complex project (Foley, J. and Bates, T., 2007). We

also should weigh the human side of this equation. The main question here is; what skills do the staff of our organization already has? Every kind of new tool would require learning new skills as well as training. A consultant could help for getting our organization up and processing, but should also be ready for investing in IT staff training for configuration and/or continuing maintenance. That is, during the implementation process it could engage and help them learning from the mentioned consultants.

There are some kinds of business intelligence tools in the market and some new ones are being developed as well. Many of these systems are adopting the web-based platforms and/or using an in memory technology that means offering an increased accessibility as well as responsiveness over the installed systems, but could also suffer while scaling to a big dataset, and require a good internet connection (Daroczi G., 2015). So that, these tools tend to be not very expensive when compared to the other types, but all are complex systems which require somebody with the data expertise to set-up and/or maintain.

GoodData's: That is a cloud based platform offering good visuals, a quick data engine and data pivoting (which is a good method to summarize the data depending on different variables) from multiple datasets that are updated in a near real-time. That tool also offers us canned reports that may not be applied to all non-profits and brings a "reports and dashboards sharing" feature for the simple collaboration between the colleagues. GoodData's tool is very fast for deploying, but the set-up and administering the data model could require a technical expertise and accessing the data by using a cloud-based user interface can require a good internet connection (Habul A., Amila Pilav-Velic, 2012). If we are looking for a cloud-based solution which allows us to have a quick adhoc analysis and some more technical people on the staff, that might be a great option.

iDashboards: That is a slick data visualization tool which is allowing our organizations to view and to analyze the programs as we are also developing what if scenarios for the strategical planning (Habul A., Amila Pilav-Velic, 2012). The user interface allows the staff personalizing an individual dashboard by a little bit training. A built-in wizard also helps us to facilitate it's connections to different data sources and/or automatic refreshes every moment. If our organization wishes the staff to develop personalized dashboards with a great visualization for analyze and/or planning, iDashboards could be a good option.

Birst: That is a strong and flexible tool including a responsive interface which stores the data in a memory or on disk, that is allowing it scaling quickly when easily performing the sophisticated calculations contain a large dataset. Birst is coming pre-packaged with standard reports and these may not meet the requirements of the most nonprofits, but such a tool allows the end users making practical Excel like calculations by themselves. Set-up is comparatively simple as Birst's new graphic interface for the logical modeling and/or the data warehouse automation technology that is automatically creating a data scheme (Greenberg P., 2009). Birst is a good option for mid-sized and large nonprofits with a dedicated IT staff that work with much data and looking scalability and/or flexibility in the deployments.

QlikView: That is a tool which is very quick for deploying and relatively simple for administering. Such a tool is developed for exploring the data via leveraging their an associative model which allows the users clicking on a piece of data and viewing all the relevant data that is associated with it (Habul A., Amila Pilav-Velic, 2012). This type of visualization is good for the analytical end users that may not have a pre-defined question. QlikView is very fast as the tool uploads and compresses our data in its own memory for allowing a quick analyze as well as a drilldown performance. Our organizations can expect to send the IT staff for training due to the complex tool uses a propriety scripting and a unique interface design tool as well as a management consol. If we are looking for a tool which is very fast for deploying, that is very appropriate. It is cost effective, very fast and requires a powerful IT staff that is prepared for attending a week training, QlikView is a wise option and provides the software free to nonprofits, containing training and/or consulting.

Tableau: That is able to process straight from our database without an additional modeling requirement assuming that all data which we wish to report is at a single database. That also allows to quick deployments in case of that our current data structure can be assumed sufficient. Our organizations could also choose re-building the data models in memory. Tableau tool is extremely visual and allow the organizations to see the data quickly while answering the questions. That tool's proprietary language that is called VizQL translates our data to the graphics which could be manipulated at the drag and/or drop interface. Tableau also provides all its training materials as online and for free and also hosts regular webinars for keeping everybody updated. Mainly, that's an excellent fit for the organizations which don't have big IT departments, but if they require a tool which offers a powerful visualization suite and/or a fast deployment (Collica Randy S., 2011). And Tableau provides a reduced costly licensing for the nonprofits.

BI tools could mainly relieve much headaches that are associated via monitoring and/or evaluating the programs as well as the outcomes. Our organizations' needs depend on what the systems are currently using and how much sophisticated we want our analytic capabilities to achieve. Every organization don't need a business intelligence tool but if we've a good strategy for program evaluation and/or monitoring and capture the raw data we need, but struggling for analyzing the data and making the use of it in our organization, these tools could be wise options for us (Brandt, S., 1999).

Every tool provides us connectivity and visuals but "the devil is in the details as always". Set-up, customization as well as the scalability could greatly affect the ease of usage and can cause more headaches (Boslaugh S., 2012). That's why, we should make sure that our organization thought through its data strategy, evaluating the tool completely and purchasing the type that totally fits the requirements of the future data analyze and as well as the current capacity of IT departments.

Mainly the business intelligence tools empower our organizations facilitating an improved business decision. Business intelligence tools also help the users throughout an extended enterprise accessing the company information as well as reporting and/or analyzing the critical data in an efficient manner. It is not just like delivering reports from the data warehouse, it is also providing a large number of people secure and/or easy access to the right knowledge so that they could make wise

decisions. The best business intelligence tools allowing the employees enhancing their productivity as maintaining a big degree of the self sufficiency.

The business intelligence tools could empower our organization with:

1- Processing the requests fastly, efficiently as well as intelligently
2- Allowing the operational employees and/or clients accessing to strong interactive reports and/or dashboards
3- Helping the employees for responding faster and efficiently to change the conditions affecting our organization
4- Allowing the developers, power users and analysts customizing their dashboards and applications depending on their specific requirements
5- Making the greatest information part for our organization's "natural culture"
6- Transforming a raw data into intuitive and illustrative enterprise reports and/or charts easily and quickly
7- Enhancing the reporting systems via incorporating a transactional form and data maintenance and/or some update capabilities

1.4 Business Intelligence Technology and Architecture

BI is a concept of working with IT tools to get the competitiveness of business, the perception of risk that may happen in the environment, and the probability of action. That is a concept which is including the combination of the traditional business functions with the IT communications and the capacities that are provided. Intent of the BI is a combinated and coordinated application of business information in order to improve the services, products, profits and the healthy life of a company for a long-term. BI as well as the Competitive Intelligence are the processes for treatment, collection, analysis and usage of the strategical information for a company. So it's all about the information that is really important for a company. The definition is a very important segment as well of the treatment results and information, in such a case that the information is not usable for decision makers. The meaning of this; there is not any difference between the data and information. Also, it's useful to mention the difference between the Competitive Intelligence and the BI. The first statement is much more detailed and involves BI, that is much more focused on software and IT tools. And effective BI system allows to data saving from all the departments within a company including their analysis, preparation of the required reports and addressing the users who are the most important. In this sense, each user has the information that is the subject of interest and of course, which is required in order to apply its business activities. BI applications can accelerate the timing of making business decisions depending on the quality of data that should be located in a central place, such as a data warehouse. It results in useful and qualified information. On the contrary, a unique report system makes contribution to highly qualified and faster decision making as well, but it also improves all the business processes. Specifically, the system of BI integrates the information about the suppliers, consumers, competitors, and overall operations. BI improves the strategical and operational planning, analyze, control and optimization of the business operations as it ensures the successful targeting clients and via tracking much more competitive activities. That helps us to predict the future trends as well (Habul A., Amila Pilav-Velic, 2012).

When we look at the architecture, the BI architecture includes three main constituents: Data extraction&integration, data mining and the enterprise report systems. So that, in the most important sources of the data such as OLAP, SCM, CRM, ERP and various databases, the documents arise from the business processes and activities in the company (Brandt, S., 1999). The above mentioned systems are used in the processing and saving of initial data. Transaction and ERP systems are very important in order to automate transactions and to capture the data that is generated. But most importantly, they don't supply information that is convenient to analyze and for the comparisons. However, the above mentioned datas form multiple systems that can improve the performances of the companies and could make profitable decisions and new steps for them.

Because of the different sources, the datas could be in various formats and they may not be standardized as well. Due to the fact that, the next step involves data standardization – ETL (which equals to Extracting, Transformation, Loading) processes of the data collection and processing for the systems of BI. During the ETL process steps, the quality of data and so that, the quality of knowledge is critical which is the one contains data transformation that takes much time. The data could be used for reporting, analyzing and etc. Collected and saved datas should be available and have to be integrated within the organization. Regarding this, the employees could make wise decisions. For instance, BI applications enable employees in the customer relations department to receive information about all the transactions with specific customer in a short term. They are able to provide offers according to the customers' requirements and aspirations. By evaluating this goal, it is possible to be achieved via generating a functional Data Warehouse that produces new knowledge.

It is also important to provide an available data to external parties in real time. Reporting systems help to prepare quantitative datas in a report format that can include numbers, graphs and charts. They also provide standard, ad hoc and/or summary reports. Ad hoc reports are generally used during the circumstances are required as a quick insight for operations of the company without any previous preparations. And the Summary reports present the business at a certain time.

Standard reports which are generated via the applications allow a defined flexibility in the construction that they have. Their content is the most common application with accompanying charts. In addition to the reporting tools, there are also dashboards and scorecards. Dashboards supply an easy access to data and they support decision making. On the contrary, the scorecard technology is used for determining and to realize the business strategy that are linking the objectives with physical measures (Burnett K., 2001).

The significant BI components are: Data Mining, ad hoc, planned queries and the analysis tools. These tools help users to create queries and enable them to analyze the data. Users are able to investigate the data and learn from it. Furthermore, Data Mining is a process that can use statistical, mathematical, artificial intelligence and at the same time the machine-learning techniques for extracting and identifying useful information from the data warehouses (Habul A., Amila Pilav-Velic, 2012).

1.5 CRM and Business Intelligence Approach for Customer Satisfaction

1.5.1 Business Intelligence and Customer Relationship Management

In today's economy, the priority of business activities works like a communication chart between the suppliers and the consumers. This communication chart is based on the interests of both sides: the companies that seek to profit and trying to grow, and the clients who want to achieve a good service. The most successful companies nowadays are those which build their business processes according to the customer expectations. New business trends, of course require company's orientation to the clients. These primarily help to grow the manufacturing capacities that cause the differentiation of the services and products regarding the different consumer's aspirations which bring larger volume of sales. Also, there are new technologies that simplify selection of the companies that the clients can show their confidence. This shows a simple comparison by competitive offers, costly low transition to the other customers and suppliers, important time savings throughout continuous provision of the required information and offers personalized products and/or services. In such circumstances, to achieve a competitive advantage is a difficult job. The value of the clients becomes invaluable for the company. Development of IT, importantly improve the mutual relations of the company and the clients. Also modern IT is a strong medium to establish good contacts between the company personnel and the clients (Dresner, H., 2007). Possibility of creating direct relations between the companies and clients can lead to their personalization. Multi-channel communication is a key for the establishment of more sophisticated and good relationships easier. But it brings more difficulties for integrating data and to create unique images of clients. Companies that do not have an integrated information for the clients, don't know the potential consumers and can't offer specialized services and/or products. Because of the lack of data integration, clients require each contact with the company to give back their data, which could cause their dissatisfaction as well. The Customers feel ignored to that company. And due to fact that, it is significant that the knowledge about the clients is stored in one place and could be available for everybody in the organization. That could be achieved via creating Customer Data Warehouse (Holland PC, Naude P, 2004).

When there is no functional integration of Data Warehouses, there are not any successful relations with the customers or no acquiring new information in the form of BI, which means providing qualified information. So that, we are able to discuss the importance of integrating client information for achieving customer satisfaction, especially in the terms of "multi-channel" communications and contact centers as well as role of the databases for creating distinctive images of them.

One of the most important usage of data warehouses is while analyzes and forecasting the basis of converting information to knowledge and to BI, that is of particular importance according to the concept of CRM. Regarding this, Data Warehouse makes the core of BI systems, because data and it's availability in the warehouse could state the quality and determine how much the business decisions are effective. Furthermore, we could be able to discuss the DSS (Decision Support Systems) while the process of achieving the satisfaction of the client, via focusing on their significance with the analysis and foresight of their behaviors (Habul A., Amila Pilav-Velic, 2012). The quality knowledge management helps us for improving the

quality of the products and/or services, promotional sales and amount of the profits as well as the customer satisfactions. That's why, the main goal of IT is increasing the customer satisfaction. BI is the main purpose of knowledge management and it is a way for delivering the right information in a right format and at the right time. It doesn't involve the formation of larger quantities, but it contains the provision of qualified information. Regarding this, we can say that the data mining techniques have been adopted. These techniques contain the algorithms for obtaining useful information. The discovered knowledge denotes a solid base for the clients' profiling and represents the segmentation depending on their attitudes, habits and behaviors.

Data mining and machine learning techniques can help us to have large amounts of data, so that they may simplify decision making processes and stimulate automatic formation of the clients' models. That is a significant BI characteristic considering the information about the clients' grows and that transforms with each consumer's interactivity. So that, progressive CRM is activated via BI applications and solutions as the basis of a successful client relationship strategy. Thus, the most distinctive advantage of BI for CRM is it's rooting in personalization. Personalization of the relations between the company and the clients is helping to company to understand and respond the requirements of each client. This ensures that each customer has exactly what they want and at the right time. BI applications support the application of different channels of communication with the clients; such as internet, e-mailing, telephone etc. Clients in accordance with their preferences, try to find the channel which mostly suits them and by whom they will get a personalized message.

Analytical CRM means a kind of processing, extraction, collection, processing, reporting, storage and interpreting the customer data. The advantage of all these applications is that the data can be used from multiple sources and their interpretation with available procedures is dependent on the requirements and aims that we try to achieve (Duda R.O, Hart P.E, 2001). The generated contacts constitute an operational data that represents the foundation of their analysis through the analytical CRM.

As we will notice, connection between the BI applications and CRM is the most pronounced within analytical CRM as it is given that consumers related data analyze tools need management support. A logical and complete image of the client depends on the integrated data warehouse and that is called the Customer Data Warehouse. Customer Data Warehouse contains a big amount of the detailed data. It also requires a sufficient and well designed ETL process as well as a quantifiable data warehouse (Dyche J., 2002).

The integration between CRM and BI can be seen at Figure 1. below. However, the integration of modern BI technologies and CRM systems, provides a path to the client's loyalty. Via possession of the information of whole transactions and customer's experience, companies could increase the total delivered value. On the contrary, BI helps to develop and complete an image that ensures the client for a quicker, easier and better decision making. Supplementary usage of CRM systems and BI provides an integrated approach to the customers that include improvements while customer profiling, easier detection value for the clients and let us to measure the success of company by satisfaction of its customers. Usage of CRM and BI solutions develops a comprehensive CRM and not just an inactive solution for their

requests, as well as a prediction and shaping of their behaviors. BI could detect different incentives for increasing the sales and revenue, such as a quicker conversion of the current potential into actual customers. That also reduces the total number of outgoing clients and increases the sales to already existing customers. Thus, in the modern business, CRM should not be evaluated separately from the BI. They form a unique model that helps companies to forecast the customer behaviors and for making decisions based on the forecasts, as well as creates profitable customer relations.

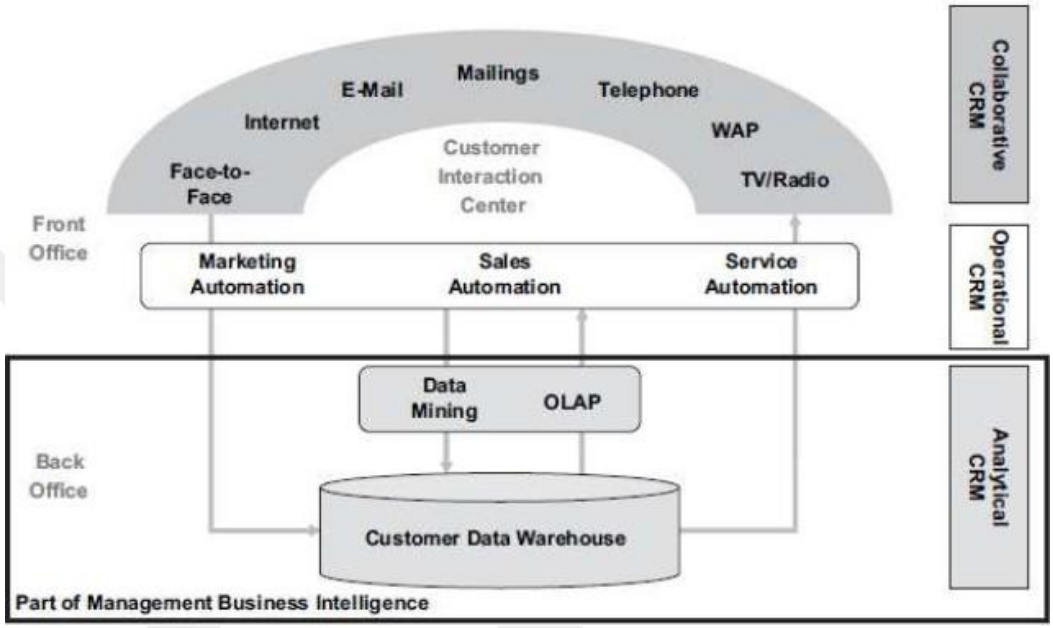

Figure 1: Customer relationship management and Business intelligence (Vercellis C., 2009)

Web based and modern business requires finding out new ways to create and maintain relationships with the customers. CRM contains a recurring process by which the knowledge collected for the clients is converted into positive relations with them. Determining direct and two-way communication is also important for the global competitive environment. In terms of the hard competition and worldwide availability of the knowledge, products and/or services, the customers expect a consistent communication with the company and a good compliance with their needs. Thus, a company that is the part of an e-story, has to try to offer high quality and customized services and/or products, and a range of extra benefits that may attract new and the existing consumers as well. Due to the fact that, companies need to manage relationships with the consumers have to increase their competitiveness in the market. The development of IT has importantly improved the relations between the companies and the clients. In fact, modern IT is a strong instrument that increases the communication between the company employees and the clients. It also increases the effectiveness of the management within this company (Foley, J. and Bates, T., 2007). The most distinctive IT feature in terms of the CRM is enabling the interactive communication. That also improves the value creation for the consumers. Multi-channel communication simplifies the foundation of more sophisticated and two-way relations although it makes more difficult to integrate datas and develop the unique images of the consumers. That’s why, the collected and saved data about the

customers has to be available and combined within all the organization. That purpose could be realized via creating the functional customer data warehouse. Regarding this, the data warehouse represents the information cores of the CRM systems. The data quality and availability in the warehouses helps us to offer qualified products and/or services to the clients.

When there is no functional data warehouse, the integration of data as the basis of successful relationships with the clients or acquiring new information in the form of BI is not possible. The most important advantage of BI applications for CRM is their foundation in personalization. Furthermore, via possession of the complete knowledge regarding all performed transactions and/or clients' experiences, the companies are able to deliver added value and they are also able to get their satisfaction. On the contrary, BI applications improve development of a real and complete image of the consumer, that provide quicker, wise and simple decision making. Supplementary use of the CRM systems and BI, provides a holistic approach to the clients and that also involves improvements in client profiling, simple detection value for the consumers, measuring ability for the success of a company in satisfying the clients and developing a comprehensive CRM (Habul A., Amila Pilav-Velic, 2012).

1.5.2 Analytical CRM

Some markets actually suffers as much from the undefined supplier marketing messages as the market for Analytical CRM solutions (like software vendors) each have a definition of what the analytical CRM comprises. In a new market, all could claim to be the market. The complexity of an ACRM project is like that the buyers are especially careful while they are purchasing any products from a high cost.

ACRM (Analytical CRM) can be defined as the process through which enterprises are transforming the customer data that are gathered through the operational CRM into an actionable client insight (Carlberg C., 2013).

Analytical CRM takes place at the heart of CRM. The call Centers, automation software and operational software for sales, service and/or marketing represent the working parts of CRM as the Analytical CRM can be defined as the thinking part as well. Applied Analytical CRM also identifies some trends and/or patterns of behavior and extending these to the predictive analysis. ACRM, rather than a single product and/or service, is the composition of many components. This process can be supported via an amount of tools that enable the analysis itself such as the data warehousing, data quality as well as the tools which enable accessing the information. These are preparing the data for analyze (data warehousing and data quality), perform the analyze itself (OLAP tools) and providing an access to information (reporting and query) (Bruce Peter C., Nitin R. Patel and Galit Shmueli, 2010).

An example of the Analytical CRM in action can be in the telecoms provider. Through billing, the enterprise collects information regarding the historical use for a particular client. Through the call center software, the call center interactions would be recorded. Through engineering, enterprise would have an idea about the network utilization. All these information would be stored in data repositories. Analytical CRM would also allow the operator identifying patterns of a behavior and finding

out similar clients as well as being able to create some predictions about their behaviors (Blili and Raymond, 1993). Once this is modeled, opportunities for up selling new products or adjusting service plans could also be acted upon.

The basis for analytics is data. So the firms should invest operational CRM systems firstly for the customer data analysis. Operational CRM is like the fuel for ACRM engine, in a sense. ACRM provides a means of derivation more value from the investments.

The level of decision making is changing. Several years ago, IT departments were the owners for not only the infrastructure but also the data. So the users would request some particular reports. Today the marketing department is in charge of its own data and needs data not only for the strategic decisions but also at micro levels for a tactical daily decision. Individual executives can create their own reports and they can run their own campaigns as well. This need of data for the micro-decision making is one of the major drivers for accessing the analytics within an enterprise.

The analytics are used for being the preserve of scientists via some complex algorithms. For the last years, the IT departments have started to control it. And recently, it is moved to be a part of the marketing departments' gun warehouse. Such an evolution of the ownership also mirrored in an evolution of the usability and that interfaces became easier, better and smarter. Such smarter tools would encourage bigger user groups in the enterprise and an enlargement in the market (Gartner, 2008).

The competitive industries such as financial services would be the most inclined for purchasing the technology to provide them a competitive edge. However, some groups of users in an organization also compete with the each other for a bigger budget. IT budgets are generally being cut and the marketing budgets are slashed in times of recession. Analytics can be unusual in that buying points may lie down within the IT, analyst or marketing functions and provides the supplier different opportunities for every sale.

The CRM application started to be globally flat and shrank in the beginning of the 21st century more or less. All the suppliers have been aware of that through the partnering or acquisition, they would have an advantage in a growing area. That also raises the profile of analytics within all the potential clients. This is due to the operational CRM suppliers have far more visibility with clients when compared to the analytics providers. Via pushing the Analytical CRM message, they are now able to create the awareness for what the analytics could do.

It can be seen that a lot of customers didn’t achieve any return on the investment in operational CRM implementations as they are trying to understand how to maximize the past investments and/or the existing infrastructures. That’s why some customers are those who are in the process of purchasing and began realizing that the ongoing and maintenance cost might be higher than expectations once the consultants became immersed for the nuts and/or the bolts of integration (Foley, J. and Bates, T., 2007). Some of the analytics providers are reporting that the clients are often those who are rapidly requiring to get out of the CRM expenditure hole with bringing in a small piece of software which could deliver an immediate ROI and can be reduced against an ongoing “nosebleed” while running the CRM strategy.

Analytics is a complex beast fundamentally. Via dumbing it down for making it easier to promote, it also ceases being analytics. Its selling points are also difficult that it primarily provides the enterprise information upon which it might or might not choose acting. Unfortunately, such smacks of vision marketing ailed the technology sales for some years, so that, so many suppliers (especially the smaller ones) have difficulty while selling their version of analytics.

In the European Union, there are already an amount of directives that are affecting the transportation and/or manipulation of the client data (Bernstein, P. and Haas, L., 2008). A very important of these is "Privacy Directive" that is effectively prohibiting the manipulation of consumer data without a clear consent. Other issues that have an effect can be said as the "Cross border data flow" that is prohibiting to transfer the data to any country which is not deemed safe. Currently, the United States isn't considered as safe by European Union standards, for instance (Bernstein, P. and Haas, L., 2008). The "Distance Selling Directive" that is placing a strict regulation for the selling via telephones of any service and/or product, and placing several restrictions for selling the financial services and/or products.

It can be possible grouping the propensity to purchase Analytical CRM technology within a particular industry via assessing the data volume, data per customer, a function of customer volume as well as volume of the transactions. The degree of commoditization is a function of the competitiveness in the nature of the goods and/or the service which is sold.

The industries that are selling undifferentiated goods and/or services and that possess a high data volume would be in bigger requirement of Analytical CRM for creating an analytical advantage.

There are many consumer data generated worldwide and the industry is being more competitive every day. That's why, a big deal of the analytical advantage could be gained through the Analytical CRM. However, the client loyalty is also too difficult to maintain, so there is a limited usage for the analytics.

Analytical CRM is using the data in the warehouse for precisely segmenting and defining the consumer groups as well as targeting the outbound marketing. Without a solid data, it will not work effectively. DQ (Data Quality) is the process for consolidating and/or cleaning every byte of the data within the enterprise. Quality of the data is directly affecting the effectiveness of the outbound marketing, customer loyalty and analytics. And more fundamentally, it is affecting the ability for doing business. A general misconception is that; data quality is an address verification and more accurately, a kind of consumer identification.

The suppliers who are offering analytics without emphasizing DQ and its general importance might be misleading for the market. Clients have to be careful that they are not going to purchase a solution which will work with 100% effectiveness as the suppliers should also be careful that they are not becoming embroiled while the long projects where the data quality issues can be insurmountable.

1.5.3 Approach to Successful Business Intelligence

BI is referred to as decision support and/or CRM analytics while it's aligned with a CRM software system and client strategies. It also empowers the decision makers

understanding, forecasting, analyzing and impacting the business performance. BI software tools help us transforming the raw data from different sources to the useful information and distributing this insight into the people who could use it, when they do need it, for improving decision making timeliness and/or the accuracy (Boslaugh S., 2012).

As with the other enterprise software applications, implementing BI, decision support and/or CRM analytic software tools with no accompanying strategy, business process support as well as the IT alignment would risk the implementation, the challenge adoptions and most possibly not achieve objectives and/or ROI (return on investment). Such kind of an advisory proposes us a 10 step strategic approach for successfully planning, justifying and deploying a business intelligence and/or CRM analytics solution (Blili and Raymond, 1993).

The management ability for making accurate business decisions is extremely influential for determining if the company surpasses or become surpassed by the competitors. For many of the business executives, decision making is like an art which becomes strong with the experience. And experience is what we have while we do not achieve what we expected. The decision makers can find themselves while trying to make decisions depending on an incomplete, irrelevant and inaccurate information (Carlberg C., 2013).

Excel appears like a stop gap measure on the way of the BI. However, natural flexibility of the spreadsheets is as a sword but also leads to the lack of data integrity via an easy transposition error and/or while a sum formula misses the row, or eventually different types of the truth as the team members are working on a project argue over where the data can be called as much more accurate. Questionable reliability is leading the users to distrust the data and look for the alternate solutions. Spreadsheets are strong tools that might compliment the decision making solutions, but they put decision making at a risk while it is pretending to be a system record.

Decision makers are recognizing their needs for graduating from the spreadsheets and leveraging a more strategic technology tool for accessing much more data more timely as well as with a greater integrity. Managers and other decision makers are also looking for the automation solutions that can allow them spending less time while retrieving and/or compiling the historical knowledge but spending more time while analyzing the knowledge which is supporting their business initiatives, allowing them planning the future in a better way, quickly identifies some areas where they need an attention and delivers the insight into contribute for the improved decisions systemically (Brandt, S., 1999).

Organizations normally can possess a lot of valuable data, although in some different stores which resemble the silos. Prospect data in the SFA (sales force automation) systems, can lead the data for a lead management system, the service history at a call center, the client data in a CRM system, the product and/or the sales data in the ERP (enterprise resource planning) system, the sales performance data in a compensation system, the business plan and/or budget data in the spreadsheets etc.

For executing the business strategies and/or outperforming the competitors, business leaders are generally pursuing the combination of systemic processes and/or the decision support software tools that help to have better source, contextualize, aggregate and deliver business insight into the information workers, decision makers

as well as the operational managers throughout the enterprise (Bernstein, P. and Haas, L., 2008).

Successful business intelligence deployments could be driven via a pressure change. Only as the opportunity that's associated with wise decisions and/or pain of the bad decisions is recognized and that exceeds the shared effort of cash expense, dedicating the resources and business breakdown would exist to go the distance via a business intelligence solution. An attempt for implementing a business intelligence solution as well as legitimating the intention, as the decision makers cannot recognize a decision making contradiction is a hard fight which would meet with the change management resistance like decreasing the sponsorships, bring IT reluctances as well as some user adoption challenges. If we find out ourselves at such a position, we should probably start with an education agenda and gather if we are gaining a momentum or not.

While determining shareholder objectives, it can be necessary to define our shareholders the first. As all of the enterprise software deployments as well as a visible and executive sponsorship is a need, very soliciting team expectations would be great. Furthermore, if we are starting with a departmental and/or line of business project, our shareholders would likely contain these business unit directors and/or a managerial staff as well as the line directors and the support staff that are held for performancing the standards that can materially affect the departments' measured objectives. Each these roles that are holding the relevant firsthand knowledge, are the keys for a successful deployment and should be afforded the opportunity for identifying their objectives. All objectives might not make the project scope, but all must be surfaced, considered and heard.

When the executives and/or the decision makers are looking for BI solutions that could be configured, exercised and tailored with less IT involvement, there can be a mistake while selecting or attempting for the implementation of a business intelligence product without IT participation. Even with the SaaS (Software as a Service) business intelligence solutions, that can easily be provisioned on-demand and much more quickly deployed, business intelligence requires data staging, data cleansing and transfer as well as integrating and/or combining data from an amount of usually different information systems and/or technologies (Harris C., 2012). Those tasks would definitely benefit from the technical talent on our IT teams.

Similarly, business intelligence solutions should be compatible with the business objectives, focusing on the business requirements and delivering business and/or an operational insight. That's why, business intelligence deployments cannot be information technology driven. BI is well accomplished via leveraging high and the best skills from both of the information technology and the business staff as well as ultimately results in a collaborative relationship as each of the sides is dependent upon other for being successful.

Business intelligence tools aren't only for the top executives. The purpose is connecting as a lot of operational decision makers via the operational data as possible. Larger participation leads to the wise decisions for more levels in the organizations, an increased operational alignment including the company's top strategical goals and/or a culture of learning (Bruce Peter C., Nitin R. Patel and Galit Shmueli, 2010).

Identification of the right performance metrics cannot be a “onetime" event. As the business plans, management directives and the budgets are changing, the performance metrics are also compatible with these objectives. With the changing business conditions, new opportunities as well as the competitive threats, the business leaders have to implement a process of review for re-affirming, adjusting and/or replacing the metrics and looking into that the measured metrics are optimal to achieve the missions for which they’re compatible. It is throughout such a review process that a lot of directors could identify the newest business drivers for the first time.

The performance metrics can be unique and mostly dependent on the individual business goals. However, some performance indicators are also common for the industries. Marketing metrics might contain campaign performance indicators like response and conversion rates as well as the Return on Investment (ROI) as the sales might be compatible with the pipeline quality, win rates and the forecast accuracy. Client service is sure for measuring the strategical metrics like the consumer satisfaction and/or retention as well as some operational metrics like FCR (first call resolution) rates and/or the “up-sell” conversions (Burnett K., 2001). Its key is beginning with fewer and more useful metrics as opposed to measuring everything initially which could be measured.

After identification of the needed data, we can create an inventory of the data sources. For Customer Relationship Management (CRM) analytics, bulk of the data is likely residing in the CRM system. However, the additional data stores might include a marketing automation system, the groupware systems and e-mails, a CMS (Content Management System), an ERP (Enterprise Resource Planning) application, an accounting software system and shadow systems and/or the Excel spreadsheets.

As the internal transaction based applications are holding a wealth of the structured data, it is an unstructured data reside in an external social media as well as the other repositories which present both of the opportunity and challenge. An unstructured social content reside on the social networks, rating, blogs and review sites and the online communities are possessing a valuable contributory data which could be appended for the existing information in order to be added as a further perspective and/or an insight for the decision makers (Daroczi G., 2015).

Software technology enablers should transform the raw data to the business insight. A Customer Relationship Management analytics and/or BI platform generally contains a data warehouse and/or data marts that is storing, aggregating and correlating the data; some integration and/or middleware tools that is extracting, transforming and loading (ETL) data through the disparate sources; and the data visualization technologies like dashboards, online analytical processing (OLAP) and scorecards that are permitting reporting, query as well as the interactive analysis.

Consolidation of the data to a central system is also an analytics of the best practice to achieve the system of records and avoiding inherent problems that are associated with the disparate and redundant data, containing a conflicting data and/or some versions of the truth. A data warehouse or a small amount of the domain specific data marts would ease data extraction, inquiry, deliverability and reporting as well as the system administration.

We don't have to deliver the information in real time as much as it works at the right time. Real time might be necessary for capitalizing on the short term opportunities, but at the same time, a less frequent periodical data refresh might be fine in order to monitor more strategical efforts.

BI solutions can be available in a variation of the software technologies and/or the deployment options. If we consider a SaaS (Software as a Service) business intelligence while the IT resources are at minimum value, the budgets could be tight and the capital expenditures can be unrealistic and/or we look for the proof of a concept for validating the solution. Software as a Service analytical solutions can be defined via excel while the business requirements are straight-forward as the data sources are also relatively contained and the management wants going deep in the particular domain areas (Duda R.O, Hart P.E, 2001). Software as a Service analytics is very popular for the functions like text mining, website analytics, pipeline analysis and speech analytics.

Evaluating the open source business intelligence software could be wise while our business intelligence needs require customization as we are seeking control to modify the tools. This option could also be attractive for the independent software vendors (ISVs) looking for embedding a business intelligence solution in our software products. The open source business intelligence experiences its highest adoption at the certain industries like a government as well as some certain regions like the Southeast Asia, Eastern Europe and Latin America etc. (Dyche J., 2002).

Business intelligence software solutions are varying in terms of the target market, the scope, delivery models and the value proposition. A proper and simple software selection project should be our best assurance for making sure that we are acquiring the analytics system which is matching with our business requirements and/or objectives the best.

Dirty data is one of the top reasons of the business intelligence project delays. Some organizations are anticipating data quality issues and subsequent the data cleaning requirements before initiating their data population phases. For avoiding such a mistake, they are sampling each of the data sources early for determining the data quality and allowing for the right time for scrubbing the data before it is imported to their data warehouse.

Once we start to seed the data warehouse with a clean data, we shouldn't stop there. We should be looking for the tools that could automate the clean data acquisition and/or the ongoing maintenance. For instance, CRM systems could be using the tools like an address verification, account merging and spell check. Getting the data rightly at the source is speeding the data consolidation as well as reducing the data maintenance and lowering the costs (Duda R.O, Hart P.E, 2001).

Once for the production, we would require a formal process that is supplemented via enabling the software tools for both cleaning and enriching the data on the go-forward basis. We should never lose the idea that the quality data is a "pre-requisite" for our qualified decisions.

BI and CRM analytics programs are similar to the other enterprise software applications as some well defined and multiple-phase projects are lowering the risk. Such an approach is allowing the early lessons learned for shaping the future roll

outs and permitting the project directors for publicizing the victories and facilitating the user adoption.

We should start with a small business unit and after achieving the success, we can methodically expand our roll-outs for including much more business units and/or integration with much more information systems. We could also recognize that the performance variables which appear seldom and the factors even for the departmental objectives are residing entirely within the department. Instead of that, the process flows can be traversing the departments. Even with a goal for beginning small and advance in phases, the project should be commencing with an end in the mind. The decision makers also have to understand the "enterprise-wide" impact of the current decisions.

The measurement should start via tracking staff utilization of the business intelligence tools. User adoption would be growing overtime so tracking the person who will access the tools as well as the volume of users overtime would be providing an early indicator for the Return on Investment (Greenberg P., 2009).

The information analysis is revealing the learning and/or the insight but it is also raising many more questions than the answers and thereby requires extending data models for the interrogation of the data as well as inserts new measurements and/or dimensions for discovering some new relationships, for displaying the data for some various roles and for contextualizing the data for making it much more actionable.

Business unit operational goals, business strategies and the functional targets can be fluid so the metrics as well as the business intelligence solutions that are supporting them should continually advance for remaining relevant.

The top business achievers are maintaining a direct link among the business objectives and the operational performance that are the needs for achieving these goals. The businesses are advancing their growth objectives while the metrics which are driving these objectives forward are being measured and/or being improved as needed. Many companies are recognizing this connection and the process of alignment as well as religiously measurement and implementing the corrective actions which are remaining elusive. For the sustained success companies, there should be implemented a formal process whereby the performance metrics are correctly measured and modified and/or adapted according to that learning as a business shift (Collica Randy S., 2011).

Measurement also contains calculating the business intelligence project ROI (return on investment). The enterprise software projects are not complete without the periodic evaluation as well as to determine whether the slated goals are being accomplished. For this end, the information management costs and/or benefits should have been calculated ahead of the business intelligence project for having a comparison point. Business intelligence return on investment is being determined periodically as the additional phases could be completed, the information is growing and the adoption rates are increasing as well.

Making wise business decisions across the enterprise is a kind of perpetual journey. Some of the additional follow on initiatives might contain the grant information dissemination to much more people or increasing the cross functional information exchange between more lines of the business. Here the purpose is democratizing the

business intelligence analysis across the business. It also contains improving the speed and/or timeliness of the information delivery. But at the same time, the information needs not to be real time, so we could be aligning the information types by the need for speed.

Implementing the alert notifications is important as well. Such notifications are generally delivered in a real time based on a performance metric's threshold value that is being exceeded as well as could be providing management of the opportunity for remediation of a performance deviation before it might get out of the hand (Brandt, S., 1999).

Tailoring data visualization via roles and an experiment with new information delivery tools of the new forms of dashboards and/or tools that are permitting some new methods of slicing and/or dicing the data is very important as well and possible via dragging and/or dropping the data results with some new measurements.

We should be stepping up into some predictive modeling and/or analysis. These types of analyze are using a technology for discovering the hidden patterns and supporting the what-if scenarios and/or the proforma modeling. So we will be able to forecast the affects of some new or proposed efforts with a powerful tool by allocating the budget and scarce resources between many of the competing alternatives (Bernstein, P. and Haas, L., 2008).

Appending much more unstructured and/or social media data with an existing repository is also important. Social CRM tools acquire a more unstructured client data. However, for most of the companies, such a data is remaining as isolated from the CRM records.

We should be moving in another direction and into some new social CRM tools and/or some mashups that permit the users developing analytical dashboards, scorecards, charts and some graphs and distributing them to some social sites such as the wikis, blogs or Facebook etc.

Hereby, we should be compliment our CRM data which is historical as well as a lagging indicator for the consumer behavior with a more current and/or dynamic data which is generated from the client surveys, consumer loyalty programs and voice of some client programs. The data from these measures could also be used as a leading indicator and generally telling us the true consumer relationship.

It could also be wise to implement training and/or coaching programs that are aiding uses in interrogating, reviewing and acting based on the data.

We should be considering the implementation of an information management solution for centralizing and betterly sharing policies, training as well as the education materials, advanced capabilities, instructional procedures and the best practices.

It is a good idea to check out self service business intelligence solutions that are powerful but also easy for using, browser-based query and/or reporting tools, wizard driven creation and interactive graphics to the hands of the staff. That could be an ideal method for keeping up with the users' growing information demands and/or relieving the under staffed IT resources.

Mainly, information is a kind of fuel for the intelligent organizations. However, the managements' traditional usage of reviewing generic, static and historic reports that are generally only pulling data from a single source is not sufficient any more in the competitive markets (Boslaugh S., 2012). Such reports are being designed for the passive viewing for advancing the business initiatives as at the shortest cycles as possible. What is required here are the strategies and the software tools for helping the decision makers for advancing from looking in a rear view mirror to see what is ahead.

Business intelligence solutions are increasing their popularity, and starting to be embedded with some traditional CRM software solutions. These tools are offering us an increasing value as extended across the business. Strategically, data driven projects such as the annual business budgets, plans, forecasts, incentive compensation plans, period-end reporting packages, product price elasticity exercises, regulatory compliance, workforce analysis etc. could currently benefit from the automation and/or an analyze via the business intelligence solutions (Bruce Peter C., Nitin R. Patel and Galit Shmueli, 2010).

BI tools have a historically displayed data as a means for viewing and analyzing that what is happening with interrogating the data with some different dimensions and measurements. However, with strong business intelligence tools that deliver an increased data mining and/or predictive analytics, much more types of data for increasing the confidence levels and enhancing increased learning as well as with the new technologies for simplifying and accelerating the business intelligence deployments, the analytics are not just for the enterprise firms with the deep IT resources anymore.

The directors, executives and the staff in the organization are responsible for making the best decisions they could according to the knowledge they have. If the relevancy, timing and the insight from that knowledge improve, their decisions would improve as well.

While pursued strategically, Customer Relationship Management analytics and/or the BI solutions applying much more relevant information to the decisions and involving more contributors while the decision making processes for reaching wise decisions in a less time. If we improve the quality of our decision making processes, we would also improve the execution of the business objectives.

CHAPTER 2

SURVEY EVALUATION

2.1 Typical Market Research Surveys

Some of the main reasons and purposes of the market research surveys is obtaining specific information from a population which is being surveyed and an endpoint of the project is to write a report and/or a presentation etc. in order to disseminate the information which is obtained from the research, survey as well as the associated analytics of the survey responses. When a report or presentation conveys the objectives of our study, the study parameters and the results as well as conclusions for the most market research projects end at this point. Here we can need demonstrating that when designed properly, the market research survey could provide the report and/or presentation, along with an analytic model which allows the results of the survey to be extended to the consumer or prospect database.

The market research is usually a broad topic. And the area which we will focus on is the market research of a company that may use to derive insights on the clients or prospects which they can't obtain in the dataset and they typically have in purchase and/or the sales data, sales operational systems or the contact information. The market and survey researchers generally gather information about what the people can think on a topic. The market analysts who are analyzing the surveys help the companies understanding that what types of services and/or products the people want and determining who will buy them as well as the price which they are willing to pay for such products and/or services. The research analysts formulate some methods as well as processes in order to obtain the data which they need via designing the surveys in order to assess the client preferences and choices as well. Most of these surveys are conducted via using the internet, e-mails or phone. However, sometimes the methods such as customer focus can group the discussions and personal interviews could also be used. The survey researchers also gather information about the people's opinions and/or attitudes on various issues of interest such as organizations and/or companies or the governments. The survey researchers generally focus on their design efforts on the set of questions as well as how they should frame these questions in order to ascertain the desired information about opinions of the people. Generally, a requirement in their survey design is to reach the target population in which to collect their responses. Responses need to represent a larger population if inferences about the population should be made with this market research analytics.

Many of the market research surveys focus on a set of questions to ask the survey recipients as well as the analysis of their survey responses and the final research report that summarizes the findings of the researches. However, after the research final report and/or presentation, most of the market research ends. That helps us to understand how to extend the researches so that the results of the survey analytics could be applied to a larger set of the clients or prospects from which the survey recipients have been drawn initially.

Some key items that need to be done during the research design phase of a project for ensuring that the consumer or prospect data could be matched back to the database for model development as well as deployment of the scoring results (Collica Randy S., 2011).

2.2 Matchback of Survey Responses

In order to match the survey responses to our clients or prospect data, we need to design the survey specially and select the population which includes a unique identification field. This unique identifier can be used after the responses are gathered to matchback to the consumer or prospect database. Such an identifier key can be used not only to matchback responses but also for any analytics which are done on the responses.

Figure 1-A shows a flowchart showing the design of the research survey's beginning and the final result of the analytical model for deploying the segmentation that could be derived from our survey responses.

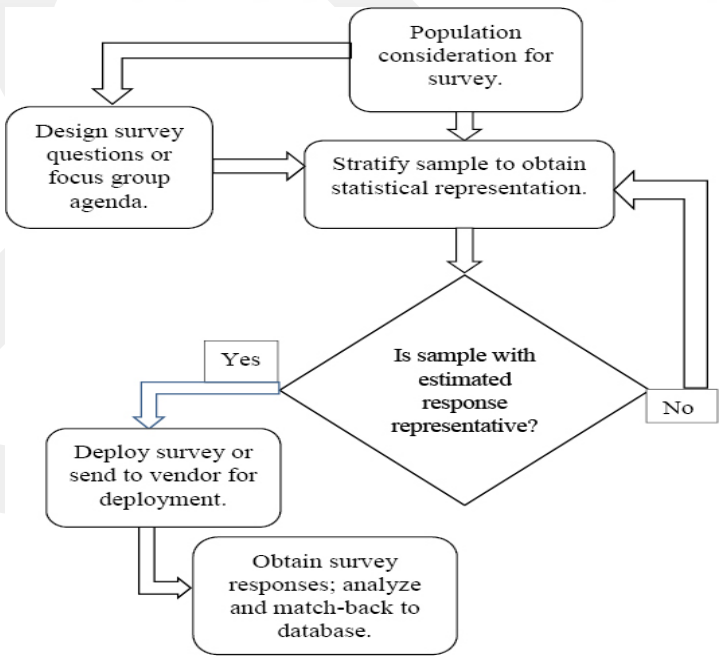

Figure 1-A: Flowchart of Typical Market Research Survey Design

The flowchart in Figure 1-A describes a general process flow; a specific flow for our business or organization may be somewhat different, but that should note us a wise general idea (Collica Randy S., 2011).

2.3 Analysis of Survey Responses: An Overview

The analysis of responses of the surveys is the key for most of the market research activities in order to ascertain a desired research objective. However, the area for a special focus is the resulting segments which could be derived from the statistical techniques like the discrete choice modeling, factor and/or discriminant analysis, covariance analysis and maximum difference preference scaling. Statistical analysis, regarding the market research methods, can also be categorized into two basic groups: statistical inference and descriptive statistics. For descriptive statistics, the basic measures such as the variance, mean, frequency counts and the
Distributions which are used for characterizing as well as to profile the question responses and maybe the other information that are collected from the survey such as the company size, or in the client area items such as the household income and/or other demographic attributes. In statistical inference, a main hypothesis or set of hypotheses regarding the population can be tested via using the sample data. The claims regarding these hypotheses are what we'd like to test and to see if they are true or false.

As in most types of researches, the market research can be tested for ensuring it's reliability and it's ability for being applied to a larger set of population as well as its validity. Here generalizability is the capability for making the inferences from a smaller sample to a population in general. And reliability is the extent for which a measurement can produce a consistent set of the results. In a set of statistical analyses, we can have confidence that if we should repeat the survey, we would achieve to a similar set of conclusions (Collica Randy S., 2011).

CHAPTER 3

CLUSTER ANALYSIS

3.1 Introduction to Cluster Analysis

Cluster Analysis is an exploratory technique that partitions observations into different clusters and/or groups. In medicine, psychology, biology, marketing as well as in finance, multivariate measurements of objects and/or individuals are the current data of interest (Greenberg P., 2009). In biology, the human blood cells might be the objects that we want to analyze. Cells that have alike multivariate responses can be grouped together when the cells whose responses differ from each other are partitioned into some kind of different clusters. The analysis of cells from an amount of individuals can result to different cluster patterns as well. These differences are of course, informative for the biologists and allow them to state conclusions about the onset and/or progression of a disease or a patient's current response to the applied treatment.

Clustering techniques can be applied whenever an amount of data needs to be grouped into meaningful piles. In some applications, we can estimate that the data fall into two groups. But in some cases, the number of clusters may not be known. The goal of Cluster Analysis is determining the cluster allocation for each observation and establishing the number of clusters.

For some clustering methods such as k means clustering, we have to specify the number of clusters before the application of the method. This is not always simple and if additional information doesn't exist about the number of clusters, that typically converges to different values and looks at very potential interpretations for our clustering results.

The central for any kind of clustering approach is the idea of similarity for two randomly chosen vectors. We can measure the degree of similarity for two different and multivariate observations by using a distance measure. As estimated, an Euclidean distance between two vectors can be defined, and this is typically the first as well as one of the most common distance that appears in Cluster Analysis. For a high dimensional data, the cosine distance could be more meaningful and that can yield better results than the Euclidean distance.

If we think about the multivariate measurements as if they are continuous random variables, but the attributes of the objects like a color, shape and/or species are relevant, they should be integrated into an analysis as well. For a data, an additional variable that defines color or a numerical value can be appropriate.

In medical applications, the pathologists know there might be four different cancer types. If such an extra information is available, it also helps their analysis and may guide them to choose the number of clusters.

The strength of Cluster Analysis can be explained by its exploratory nature. Depending on the number of clusters and the measured distance, different cluster patterns can appear. These patterns can help us to understand the structure of the data. For many data sets, the statistician should work closely with a collector and/or owner of the data as they know detailedly about the data. Different cluster patterns may indicate that unexpected substructures can exist, which leads to further investigations of the data. Because of that, the interpretation of a cluster analyze has to contain a subject expert (Koch I, 2013).

Cluster analysis is also used for forming groups and/or clusters of the similar records based on some type of measurements which are made with the mentioned records. The main idea is characterizing the clusters in a way that will be useful for the aims of our analysis. This main idea is applied in a lot of areas such as astronomy, medicine, archaeology, education, chemistry, psychology, sociology and linguistics etc. For instance, biologists made the extensive usage of classes and/or subclasses in order to organize species. A very important success of the clustering at chemistry was the periodic table of the elements belongs to Mendeleevo (Hillier, F., S., Lieberman, G., J., 1995).

Another popular use of cluster analysis is for the market segmentation. Here the consumers can be segmented depending on the demographic and/or transaction history information. And a marketing strategy may be applied for each of the segments. It can also be used for the marketing structure analysis for identifying groups of similar service and/or products according to their competitive measures of similarity. In marketing as well as for the political forecasting, clustering of neighborhoods using the postal zip codes in the USA was used successfully for grouping the neighborhoods by lifestyles.

Information of lifestyles could be used for estimating the potential demand for products and/or services. In finance, the cluster analysis is used in order to create the balanced portfolios. A given current data on a variety of investment opportunities, one can find out the clusters depending on the financial performance variables (such

as a daily, weekly or monthly return), volatility and other characteristics such as the industry and/or market capitalization.

Selecting securities from between the different clusters could also help to get a balanced portfolio. Another application area of the cluster analysis is finance for the industry analysis. For a known industry, we would like to find groups of similar firms based on the measures like the growth rate, market size, profitability, product range and the presence in different international markets. These groups can also be analyzed for understanding the industry structure and in order to determine that who is our competitor.

A different and interesting application of the cluster analysis is described in Berry and Linoff (1997) regarding the design of a new set of sizes for the army uniforms of women in the USA Army. The study has come up with a new clothing size system including 20 sizes only as the different sizes can fit a lot of different body types. These twenty sizes have been the combinations of five measurements which are their chest, neck and the shoulder circumference as well as the sleeve out seam and neck-to-buttock length. This is very important as it shows how a new insightful view could be gained via examination of the clusters of records (Bruce Peter C., Nitin R. Patel and Galit Shmueli, 2010).

Cluster analysis can also be used for a huge amount of data. For example, internet search engines are using the clustering techniques for the cluster queries that the users submit. These could then also be used in order to improve the search algorithms.

Typically, a basic data which is used for forming the clusters can be supposed as a table of measurements on some different variables. Here each column represents a variable and each row means a record. Our goal should be forming the groups of records, so that the similar records can be in the same group. The number of clusters can be determined from the data as well (Bruce Peter C., Nitin R. Patel and Galit Shmueli, 2010).

3.2 Definition of Clustering

As a general definition in many books, clustering is defined as a process for separating the datas to some different classes. With this way, it is possible to understand the similarities and differences between our clusters. So mainly, we can say that clustering is a used technique for analyzing a statistical data.

Clustering is generally preferred when the customers are not well-segmented. Via using the clustering analysis, it is possible to find out the characteristics in order to understand that which advertisement campaigns could be applied for what kind of clients.

Cluster analysis could also be applied for an automatic identification of the different natural groupings. That’s why, clustering can also be named as a kind of segmentation technique. In such a technique, the data instances that are similar and/or near to each other can be grouped into one cluster. Similarly, the data instances which are very different and/or far away from each other can be separated into different clusters. Clustering is an “unsupervised” learning technique due to

there aren't outputs or dependent variables for which we may compute a right and/or wrong answer (Levin N, Zahavi J, 1999). A correct number of the clusters or any definition of such clusters cannot be undefined ahead of time. Clustering techniques may just suggest to the user that how many clusters could make sense from the characteristics of a provided data. The user is able to state a different, larger and/or smaller number of clusters founded on if they can make a business sense. Thus, a cluster analysis technique can define many different clusters according to the analysis of the data. However, there can also be good and wise cluster definitions according to that how much close the cluster parameters would fit our data.

Clusters are well defined patterns and include a lot of kinds of patterns. Some clusters can be traditional such as the data points which appear like hanging together. However, there are also some clusters where all the cluster points represent the circumference of a defined circle. There may also be concentric circles containing the points of different circles as they represent different clusters. Clusters are differing according to their sizes, shapes and their density as well. If there is a "noise" in the data, then it is much more difficult to detect the clusters. Mainly, an ideal cluster should be defined as a set of points and it should be isolated and/or compact. In reality, a cluster is a subjective entity for the interpretation and for the significance of this cluster, it is required to know a domain information. Let's have a look how many different clusters we can visualize according to the following data:

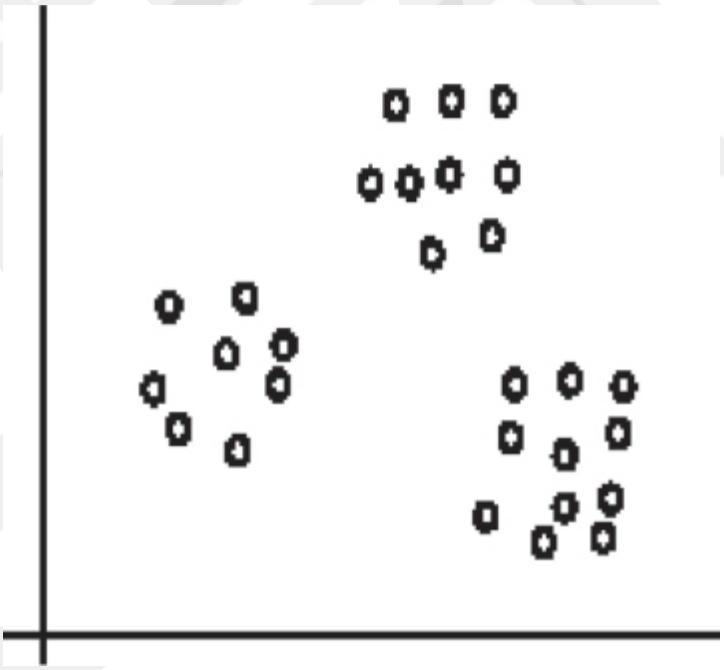

Figure 2: a Clusters sample

Regarding the above figure, there are three different clusters and they have approximately equal sizes. However, they can also be shown as two clusters as well depending on how we draw the dividing line. There is not an optimal way in order to calculate it. Heuristics can be generally used for stating the number of the clusters (Collica Randy S., 2011).

3.3 Applications and Tecniques of Cluster Analysis

Cluster analysis is used for almost all fields where we face a big variety of transactions. It can help us to provide a definition, tags and the characterization for a defined population. It also helps to identify natural groupings of the consumers, products and/or services and patients etc. Furthermore, it helps us to notice outliers at a specific domain and that's why, decreases the size as well as the confusions of our problems. Another very important application of the cluster analysis is the market research. Clients can be grouped into some clusters according to their characteristics, price sensitivity, requirements and geography etc.

Here are some examples of clustering:

Market segmentation: Grouping the consumers according to their similarities and differences. For example, their common wants and tendency to pay, their requirements etc. can help to define the targeted marketing.

Product portfolio: People who have similar sizes could be categorized in order to make small, medium and large sizes for some of their clothing wearing.

Text mining: Clustering could help the organization of a well known collection of text documents according to their content differences and similarities for clusters of some related topics (Collica Randy S., 2011).

Cluster analysis could also be defined like a machine learning technique. The quality of a clustering result changes according to the distance function, algorithm as well as the application. Firstly, we should evaluate a distance function. Most of the cluster analyze methods are using a distance measurement in order to calculate the closeness between the pairs of items. Here we can talk about two major measures of distances. One of them is the Euclidian distance which is one of the most intuitive measure and the other one is Manhattan (or rectilinear) distance where we can only notice the orthogonal directions. The Euclidian distance is hypotenuse of a triangle when the Manhattan distance is a sum of two sides of our triangle.

The main objective of a clustering algorithm is that interclusters distances should be maximized and the intraclusters distances should be minimized.

There are a lot of types of algorithms in order to get the clusters. There are some top down and/or hierarchical methods which are starting by developing a known number of the clusters. We can also mention about the bottom up methods that is starting via identifying the clusters naturally occurring. One of the most popular clustering algorithm is "K means algorithm". It is a top down as well as a statistical technique that can be applied via minimizing our least squared distances from the provided center points of the clusters. Machine learning techniques (such as the neural Networks) could also be used for clustering. Comparing the cluster algorithms is so difficult due to the fact that there is no single and/or right number of the clusters.

A general pseudo code is given for clustering hereunder:

1. Choosing arbitrary groups or segments.
2. Starting with the initially and randomly chosen center values of these arbitrary groups.
3. Classifying the instances to our closest groups.
4. Calculating new values of our group centers.
5. Repeating the steps #3 and #4 until our groups converge.
6. If the clusters are not satisfactory, returning to step #1 and choosing some different number of groups or segments.

Mainly, the clustering exercise should be proceeding forward with a defined different number of clusters or with different locations of the points (Collica Randy S., 2011).

3.4 Clustering Exercise and Types of Clusters

Here is a simple exercise identifying the clusters from the data. X and Y are the dimensions. And we want to determine the number of clusters as well as the center points of these clusters.

X	Y
2	4
2	6
5	6
4	7
8	3
6	6
5	2
5	7
6	3
4	4

Figure 3: Clusters from data

The plot of ten data points that are shown in 2 dimensions show they are distributed randomly (Figure 4). With the help of bottom up technique, number of the clusters as well as their centroids can be found. These points are distributed fairly randomly and enough for being considered as only one cluster. Here the circle represents the

central point or the centroid of these points. However, there we see a big distance between the points (2,6) and (8,3). That's why, this data should be separated into two different clusters. The three points (at the bottom right) form one cluster and the other seen seven points form another cluster. These two clusters will look like Figure 5. And the circles will be like the new centroids. The bigger cluster will seem like very far away. So that, the four points (on the top) could also form a separated cluster. The three clusters will look like at Figure 6 as well. Mainly, this solution has three clusters. The cluster which is on the right side is far away from the other clusters. Furthermore, its centroid isn't so close to all the data points. The cluster at the top looks like very tight fitting although it is a good centroid. And the third cluster is spreading out at the left, so it is not very useful.

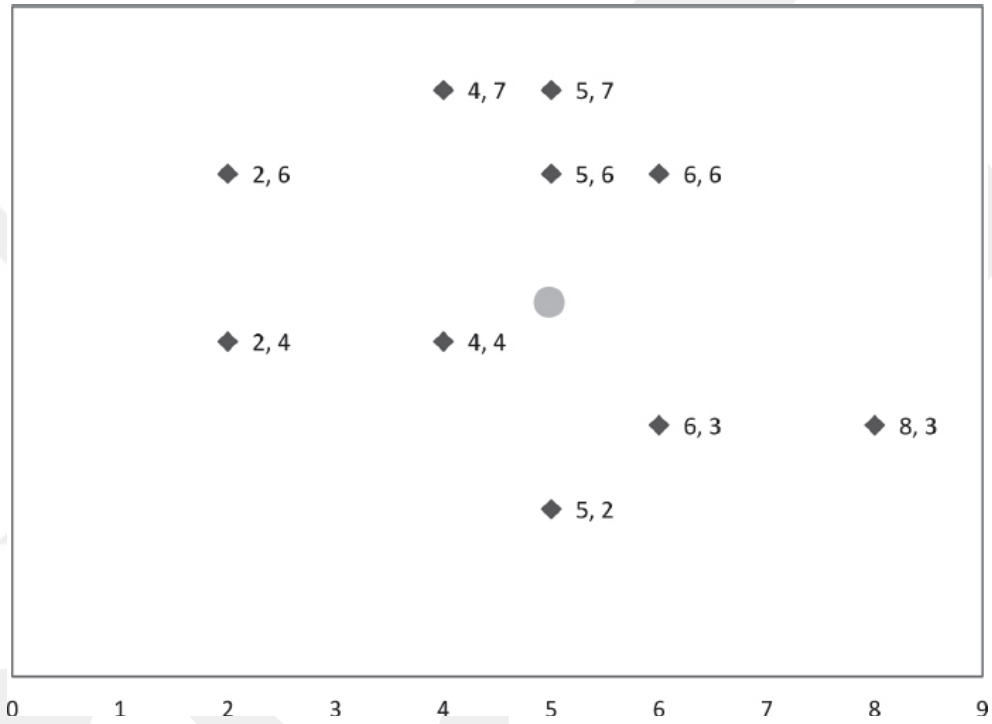

Figure 4: Initial data points and centroid

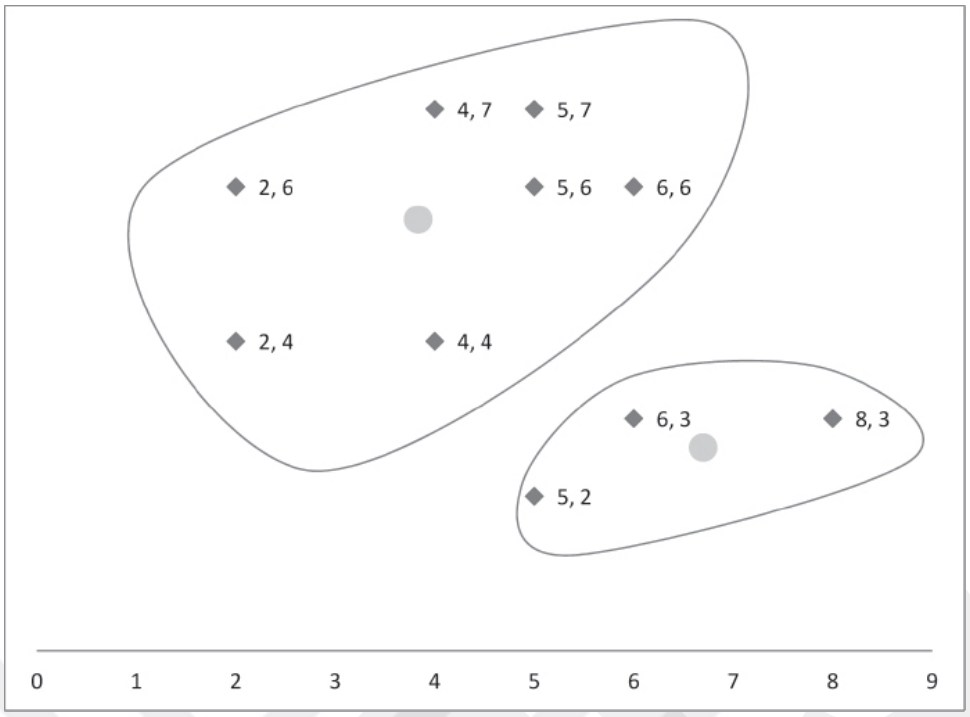

Figure 5: Dividing into two clusters

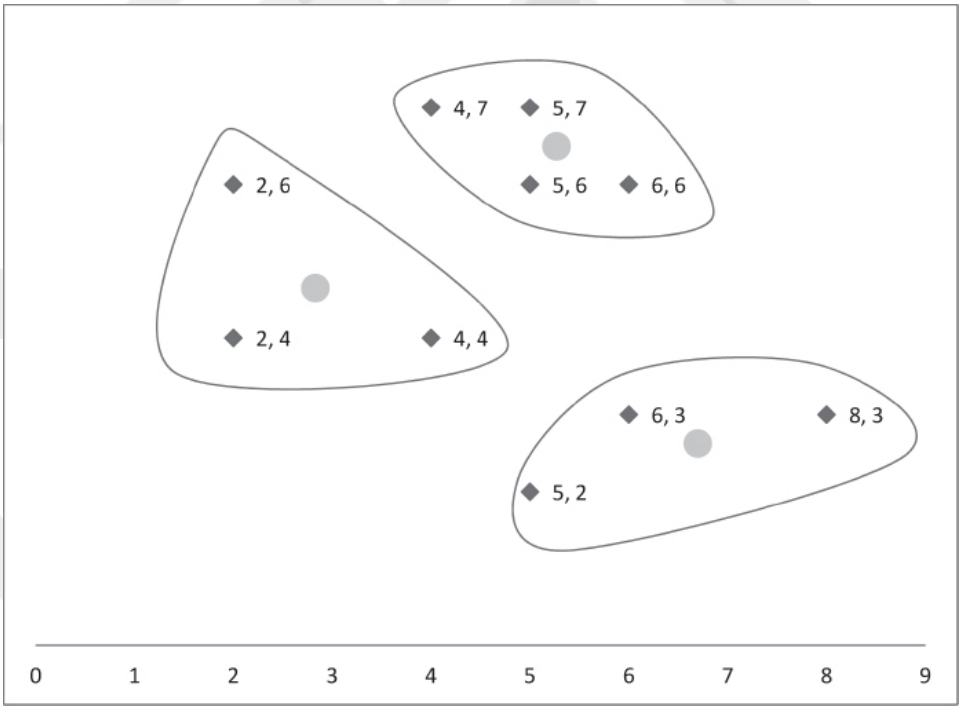

Figure 6: Dividing into three clusters

This is a simple exercise in order to create 3 best fitting cluster definitions that is obtained from a defined data. The number of clusters would of course be dependent on the data and the application for which we are going to use the data (Peppers D., & Rogers M., 2004).

Originally, the clusters and high-performance computing have always been synonymous. However, today, the meaning of a cluster has expanded beyond high-performance for containing the high availability clusters and the load balancing clusters (Blili and Raymond, 1993). Practically, there is a considerable overlap among these. But while focusing primarily on these high-performance clusters, it would be worth to take a brief look at the high availability and/or load balancing clusters as well.

High availability (HA) clusters are generally used in some mission critical applications. If you cannot afford the lost business which is resulting from having your web server as it goes down, you can want to implement it using a highavailability cluster. The main idea to get a high availability is the redundancy. Here the highavailability cluster is the composition of multiple machines and is a subset that can provide the appropriate service. In the purest form, only a single machine and/or a server is directly available when all the other machines is at standby mode. They monitor the primary server for insuring that it will remain operational. If the primary server fails, a secondary server should take its place.

The main idea of a load balancing cluster is providing a better performance via dividing the work among different computers. For instance, while a web server is being implemented by using load balancing clustering, the different queries can be distributed among the different computers in the clusters. So this may be accomplished via using a simple roundrobin algorithm. For instance, a RoundRobin DNS can be used in order to map the responses to DNS queries to some different IP addresses. So that, while a DNS query is made, the local DNS server is returning the addresses of the next defined machine in the cluster and visiting the machines in a defined roundrobin fashion as well (Sloan Joseph D., 2004). However, this kind of approach can also lead to dynamic load imbalances. More sophisticated algorithms use the feedbacks from some individual machines in order to determine which machine is able to handle the next task the best.

We should keep in mind that the term load balancing can mean different things for the different people. A high-performance cluster can be used for scientific calculations and a cluster which is used as a web server might approach load balancing in very different ways. So each application has some different and critical requirements as well.

Any cluster can provide redundancy, an improved performance and scalability regardless of its own classification. As the load balancing can provide a greater availability, it's usual to see that both of the load balancing and high availability are in the same cluster. The Linux Virtual Server Project is a simple example of the combination of these approaches. Such a Linux Virtual Server Project is a high availability server which is implemented via distributing some tasks among an amount of the real servers (Sloan Joseph D., 2004).

3.4.1 K-Means Clustering Algorithm and Number of Clusters

The K-means clustering belongs to "nonhierarchical class" of the clustering algorithms. It's one of the most popular algorithms which is used for clustering practically as it is quite simple and quick. It's also considered as more robust for

different types of variables and more appropriate for large data sets which are commonly seen in marketing. Furthermore, it is less sensitive for some clients who are outliers who are extremely different from the others.

For the K-means clustering, the user should specify the number of required clusters before starting the clustering algorithm. The basic algorithm for K-means clustering is like below:

1. Choosing the number of clusters = k.

2. Generating the cluster centroids with k random points.

3. Assigning each point to the closest cluster centroid.

4. Re-computing the new cluster centroid.

5. Repeating the 3rd and 4th steps until a convergence. Usually the convergence criterion is like that; if the assignment of the clients to clusters is not changing over some multiple iterations.

A cluster centroid is the average of all points in the cluster. It's coordinates can be found from the arithmetic mean of each dimension separately over all of the points in our cluster. Let's consider that as Sam, Joe and Sara in an example. If we represent them depending on their importance ratings on the premium savings and the neighborhood agent like: Sam = {3,4}, Joe = {4,7}, Sara = {5,3}. Then If we assume they are belonging to the same cluster, the center for their cluster can be obtained as:

Cluster Centroid $Z = (z_1, z_2) = \{(4+3+5) \div 3, (7+4+3) \div 3\}$

Here Z_1 could be measured as the average of the ratings of Sam, Joe and Sara on the premium savings. Similarly, z_2 can be measured as an average of their ratings on the neighborhood agent. Below, Figure 7 provides representation of the K-means clustering (ReibStein D., Philip E. Pfeifer, Neil Bendle, Ronald T. Wilcox, Paul Farris, Rajkumar Venkatesan, Cesar Brea, 2014).

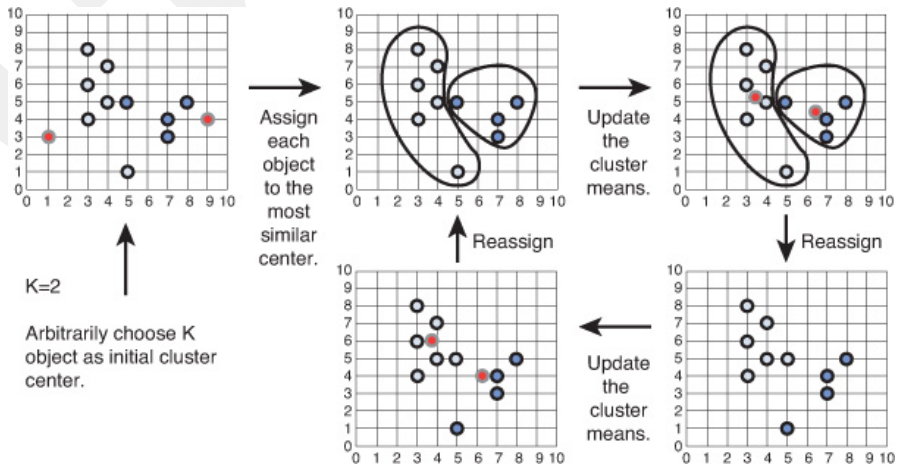

Figure 7: Visual representation of Kmeans clustering

One of the main issues of the K-means clustering is that it doesn't provide an estimation of the number of clusters that is existing in the data. K-means clustering must be repeated several times with different number of clusters in order to determine the number of clusters which are appropriate for the data. A commonly used method for determining the number of clusters is called as elbow criterion.

The elbow criterion notes that you have to choose a number of clusters so that, adding another cluster will not add a sufficient information. The elbow can be identified via plotting the ratio of the "within cluster variance to between cluster variance" versus "the number of clusters". The within cluster variance is the estimation of an average of our variance in the variables that is used as a basis for segmentation between the clients who are belonging to a particular cluster. The between cluster variance is also the estimation of our variance of the segmentation basic variables between the clients who are belonging to the different segments. The objective of cluster analysis is minimizing the "within cluster variance" and maximizing the "between cluster variance". That's why, as the number of clusters increases, the ratio of the "within cluster variance" to the "between cluster variance" would decrease. But at some point, our total marginal gain from adding an additional cluster will be dropping, and giving an angle graphically called the elbow. In Figure 8, such an elbow is indicated via the circle. Therefore, the chosen number of clusters should be 4 (ReibStein D., Philip E. Pfeifer, Neil Bendle, Ronald T. Wilcox, Paul Farris, Rajkumar Venkatesan, Cesar Brea, 2014).

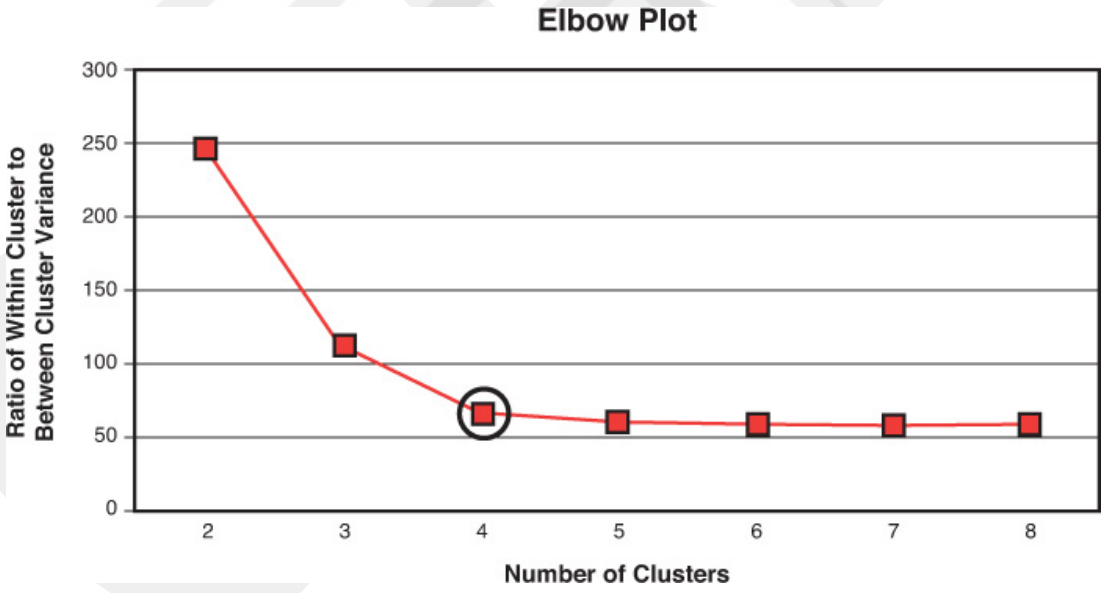

Figure 8: Elbow plot for determining number of clusters

It has to be noted that the initial assignment of a cluster seeds can have a bearing on the final model. Here some common methods in order to ensure the stability of the results can be obtained from the K- means clustering:

• Run the algorithm several times with some different starting values. While using the random starting points, run the algorithm several times in order to ensure a different starting point at each time as well.

• Split the data as randomly into two halves, and run the cluster analysis separately at the each half. The results can be robust and stable if the number of the clusters and the sizes of our different clusters are similar at the both of the halves.

3.4.2 Motivation for Clustering of Customer Attributes and Implementing the Top Algorithms for Clusters

If we have a look at the motivation for using the clustering techniques for CRM, we could note that the Segmentation Using a CellBased Approach is a method of segmentation that can be done via using a Self Organizing Map Neural Network in order to accomplish a type of clustering (Prahalad C.K., Ramaswamy P.B., Katzenbach J.R., Lederer Ch., & Hill S., 2002). Hereby, we should also have a look at the techniques of clustering. One of the best basic understanding ways to see how the clustering is measured is; Distance; the Basic Measures of the Similarity & Association. Here the concept of distance and similarity comes into play when we are attempting to find out the groups and/or segments within our database of clients, patients, fraud cases etc. and the ones that share some sort of similarities with each other.

Clustering is a technique which is represented as an undirected data mining. It is undirected as there is no variable on the data set which we try to predict and there is no response variable. Sometimes, the data sets are rather complex and they are not apparent pattern that seems to appear using the other techniques such as: Decision Trees, Regression and/or a Neural Network. One method of a clustering technique is allowing the clustering algorithm to use only some specific set of variables which the data miner would like to use and that will force the algorithm using these variables of interest while the clustering session. The natural tendency of people when they are faced a complex task is attempting to break it down into the smaller sized pieces. Then each is hopefully simpler than the entire data set as a whole. In the context of CRM, the work of finding groups of the clients inside our database which are similar in some way is a specific marketing and some of the sales programs could be designed just for them (Bernstein, P. and Haas, L., 2008).

Many cluster segments can exist for each data set of the clients based on what should be selected for the measurement of similarity. For instance, if a desired objective in a CRM project is profiling the clients in a database, then the measure of similarity may be the variables and/or combinations thereof that help to describe how the clients are different and/or similar to each other. For the consumers, type of the industry, size of the company in amount of the employees and/or the amount of revenue etc. may be measured.

Other metrics of the interest can be how many times the consumers have responded to a marketing campaign during the last six months, or what kind of web pages the customers visited on your web site and how long they remain online, whether the clients just searched information or they visited product and/or service areas at your web site. All of these types of metrics can help classifying a client according to some types of behavior of interest. A single data set can possibly contain a lot of varying cluster segments based on what has been used as inputs in order to measure the similarity. This can also be useful in a business context as the different type of segmentations would be performed on the same set of the consumer records. The

combinations of these segments can also be used in conjunction with the each other for accomplishing a specific business purpose (Coliica, 2011).

The clustering can also be stated as an unsupervised classification of the datasets. The objective of the clustering algorithm is dividing a given dataset contains the set of points or objects into some groups of data instances and/or objects with distance and probabilistic measures. The members in the same group are closer by distance and/or similarity or via some other measures. In other words, that is a way of maximizing the similarity of the intraclusters as well as minimizing the similarity of the interclusters (Zadeh, L., A., 1974).

A clustering algorithm can be used for preparation of the further analysis as the other objective is for understanding the nature of our dataset. One of the most popular clustering processes is illustrated at the diagram (Figure 9).

The main steps in such a process include the following:

- **Feature selection:** At this step, we choose the visible features from original dataset
- **Clustering algorithm design:** At this step, we design an appropriate algorithm according to available clustering algorithms, as that just builds one from the scratch
- **Cluster validation:** At this step, we evaluate the clusters and supply a degree of confidence regarding the result
- **Result interpretation:** This step gives us an intrinsic main idea about the input data

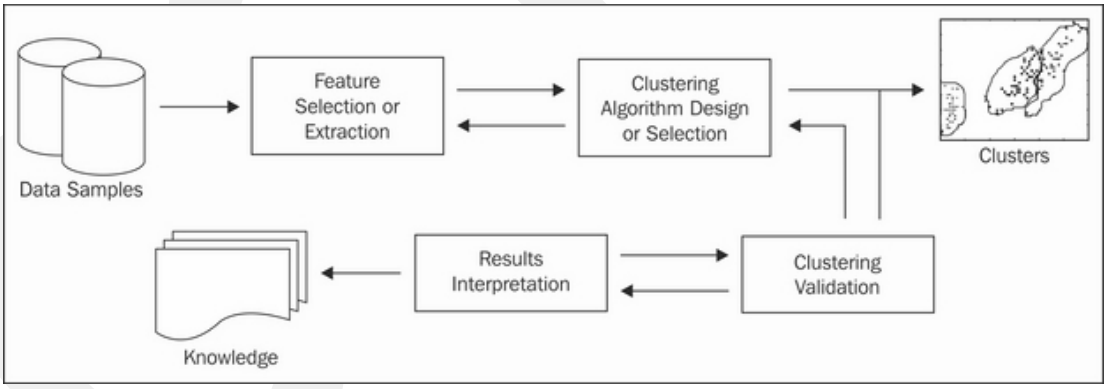

Figure 9: The most common clustering process

There are a lot of categorization methods for categorizing the clustering algorithms; density based methods, partition based methods, hierarchical methods, grid based methods, spectral methods etc.

Every clustering algorithm has its own limitations and the best practices for some current and certain conditions and/or datasets. Once an algorithm is chosen, parameters and the distance measurements regarding that algorithm require a careful consideration (Makhabel B., 2015).

3.4.3 Classification, Discriminant Analysis and Logistic Regression

Cluster analysis is one of the best as an exploratory technique. We can deploy the cluster analysis when we have a number of records which are ranging from around one hundred to even hundreds of thousands. But we don't know how to classify them or what the classifications should be. We have an access to the other variables which describe the records' behavior and the other descriptive measurements like the height, cholesterol levels, weight etc.

If we know what our categories should be, then we can assign the each record to a different category. In this case, we can run a discriminant analysis or a logistic regression analysis. Using the logistic regression, our purpose could be determining that how the predictor variables should combine in order to predict a category using the maximum likelihood techniques. Or via using the discriminant analysis, our purpose can be discovering that how our predictor variables should work together in order to bring the categories which we observe for each of the records.

In either case we will have information about the category membership before we are undertaking the analysis. Here we may establish our equations via using the data that contains the category information and apply them by using the data whose category information can be missing. But at some point, the analysis can take note membership of the records and that ties the predictor variables to the knowledge regarding the category membership.

The things could be different in the cluster analysis. We may not have the information about membership in some different categories and so that, we cannot deploy either discriminant analysis or logistic regression. Or we may have the category information but it could be not useful for the purpose that we have in our mind. It's also possible that we could be interested to determine whether our predictor variables can group the records on their own accord as the principal component analysis combined with a rotation of some sort and that enables the underlying factors to appear from a mass of the measurable variables.

That is what the cluster analysis does and how we can get away without noticing what categories may be hidden at an undifferentiated mass of the measurable variables. The main idea is that if we examine how close, distant and different records are from one another on the measurable variables, the clusters emerge inevitably.

The question which remains is if the clusters mean anything. Did the measured variables result in meaningful groups of our records? Or are these groups and/or clusters are just the artifacts of some randomly chosen distances between the records which cannot be grouped? It is possible to ask such questions because the cluster analysis is generally considered to be an exploratory technique and distinct from the techniques which are used in order to confirm an existing hypothesis. Definitely, the techniques such as the discriminant analysis and logistic regression are used for exploring patterns in our dataset. But due to we have the actual grouping data in our hand while we are undertaking a discriminant analysis or logistic regression, the exploring we do is usually limited with learning more about how the predictor variables can work in order to define the groups.

To the contrary, the cluster analysis is appropriate while we don't know what the categories are or even though there is a meaningful category. And that's why, it is wise for exploration. Of course, if we have enough records in order to arrange a preliminary analysis as well as a subsequent cross validation, it should also be possible to view the cluster analysis as a means of the confirmation of a finding. That's wise, of course. And we are always on a safer ground as we have enough data for establishing a finding and to test if the other data supports this finding or not (Carlberg C., 2013).

The cluster analysis can be called as a set of techniques that lets the cases being grouped according to their values for one variable or more variables. Some of the cluster analysis techniques distribute the cases to groups by partition and the other techniques provide hierarchical trees that have a taxonomic relationship among the groups. A related technique, discriminant function analysis can also be used in order to develop some rules for assigning the cases into groups depending on an understanding of the parametric structure of our groups; discriminant function analysis (DFA) is better for predicting a group membership than the cluster analysis. Generally, these two techniques can be used in conjunction with the each other. The cluster analysis may also be used while the number of groups is not known initially. Here once this number is established, DFA could be used for prediction of the individual group membership.

Cluster analysis is also useful for two scenarios. Firstly, we may already know how many groups we expect finding out in the data, so we should pass this number of groups into the algorithm and allow it taking care of the allocation. Alternatively, if we may not know how many groups exist, in this case we can ask the algorithm estimating how many groups there are in actual fact (Boslaugh S., 2012).

The statistical methods are limited while we deal with a categorized data, for instance while analyzing the surveys.

Even if some methods try to convert discrete variables to the numeric ones, such as via using a number of dummy or some indicator variables, in most of the cases, it can be better to think about our research design goals instead of trying to use the old methods in the analysis.

We could also replace a categorical variable with a number of the dummy variables via defining a new variable for the each label of our original discrete variable, and then we assign "1" to the related column and assign "0" to all the others as well. Such values could be used as the numeric variables in some statistical analysis such as with the regression models.

While we are analyzing a sample and a target population with categorical variables, usually we wouldn't be interested in the individual cases. But instead, we could be interested in similar elements and/or groups. Similar elements could be defined as some rows in a dataset with the similar values in our columns (Daroczi G., 2015).

CHAPTER 4

ANALYSIS STUDY AND CONCLUSION

4.1 Analysis Results: Cluster Dendrograms and T-Test Application

A questionnaire given in the Appendix has been applied to 20 companies including some small and medium enterprises in Turkey. According to the answers that we got regarding this survey, we aimed to evaluate these companies depending on their BI awareness levels as well as to understand how effective they use the business intelligence when these companies' sizes are compared.

For questions 5, 6, 7, 8 and 9 in the questionnaire; we intended to have answers such as "Very Low, Low, Medium, Good and Very Good" which give us an idea about the awareness levels of the companies. Regarding the answers of the respondents whose company have a defined BI strategy, we had an opportunity to compare the BI usage and awareness levels of these companies. And with this occasion, it could be possible to compare the BI awareness levels of the companies by:

1- **Question 3:** Comparing the companies who are using the business intelligence for each of these different purposes: **"Reporting, Analyzing, Alerting or Predictive Modeling".**
2- **Question 4:** Comparing the companies who are using **"ERP, CRM, ERP&CRM or all of these together"** in order to get data for the business intelligence management that they have.
3- **Question 10:** Comparing the companies depending on the type of analysis such as "**Customer analysis, Supplier analysis or Annual Turnover analysis"** which they use the most for the defined BI management in their companies.
4- **Question 12:** Comparing the companies depending on that who utilizes the outcomes of BI in these companies who are using business intelligence: **"The top and the middle management or the sales & development staff"**.
5- **Question 13:** Comparing the companies depending on that which BI tool they use the most for their defined BI management: **"Query tool, Reporting tool or OLAP".**
6- **Total number of the company staff:** Comparing the BI levels of the companies depending on the total company staff that they have: **"Total number of Staff<100, 100<Total number of Staff<1000 and Total number of Staff>1000".**

4.1.1 Analysis Results: Different Purposes of the Business Intelligence Usage for the Awareness Test

At the third question in our survey (Appendix), we asked to the respondents that for which purpose they use BI in their companies and we asked for a ranking from them. Then we analyzed the responses by comparing the companies according to their 1st rank choice; Reporting, Analyzing, Alerting or Predictive Modeling. So that, we grouped the companies based on their 1st rank choice and we got 4 different pie charts each are representing one of the Reporting, Analyzing, Alerting or Predictive Modeling that are possibly chosen by the respondents at the 1st rank.

Mainly, each pie chart represents the answers of the respondents for questions 5, 6, 7, 8 and 9. For example, the Reporting pie chart below denotes the responses among "Very Low, Low, Medium, Good and Very Good" options of the questions 5, 6, 7, 8 and 9 only between the respondents for whom the Reporting is at the 1st rank for question 3.

In the pie charts, the first number denotes the answer to questions 5, 6, 7, 8 and 9 from "1" for "very low" to "5" for "very good". The second number denotes the number of that answer in our data and the last number denotes the percentage of this answer between all the data.

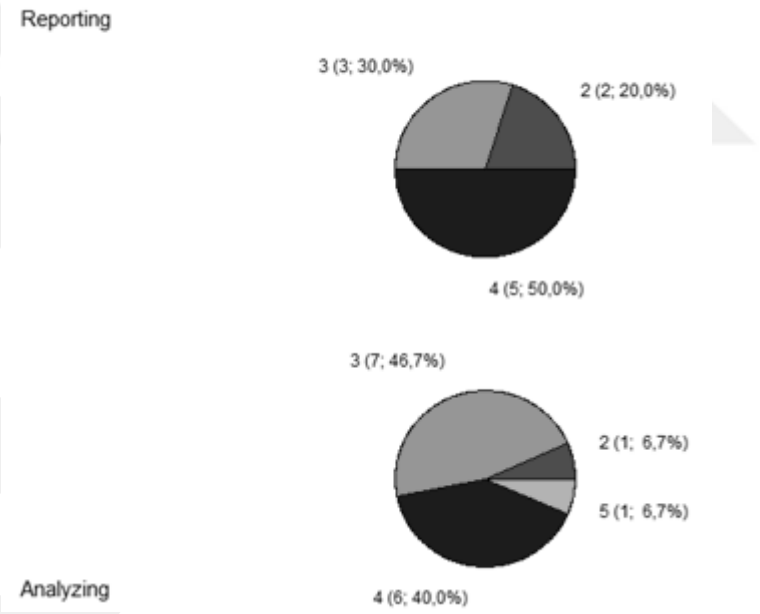

Figure 10: Pie chart-1 of the analysis

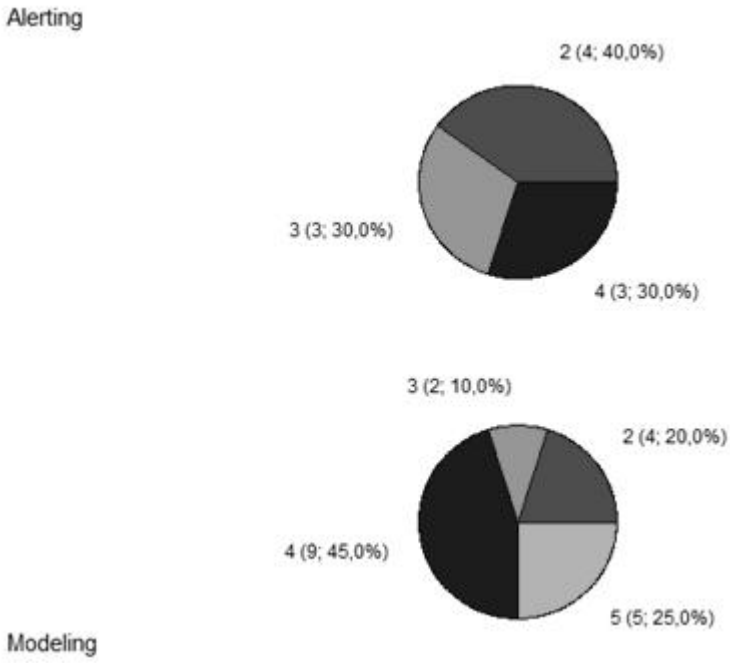

Figure 11: Pie chart-2 of the analysis

Based on the data sets we have shown with the above pie charts, we firstly applied a T-test in order to see the Mean (mean values of all N data sets), StDev (standard deviations), SE Mean (standard error mean) as well as the histograms of all N data sets. The results and all histograms of our T-test can be seen hereunder:

Table 1: T-test-1 of the analysis

Variable	**N**	**Mean**	**StDev**	**SE Mean**
Reporting	10	3,300	0,823	0,260
Analyzing	15	3,467	0,743	0,192
Alerting	10	2,900	0,876	0,277
Modeling	20	3,750	1,070	0,239

At the following histograms, the vertical diagram shows the frequency that is the total number of responses which contain "Very Low, Low, Medium, Good and Very Good" options of the questions 5, 6, 7, 8 and 9 from "1" for "very low" to "5" for "very good" when the horizontal diagram represents these responses graphically as well. ***For instance, at the first histogram below, we see the responses of the respondents who marked "Reporting" at the "1st" rank in the questionnaire. And we can say that between the respondents for whom the "Reporting" is at the "1st" rank, there are only "Low, Medium and Good" options chosen as we face only "2, 3 and 4" at our histogram shown below.***

All histograms are the results of our T-test which is done by Minitab 13,0 and have been drawn with 95% t-confidence interval for the mean.

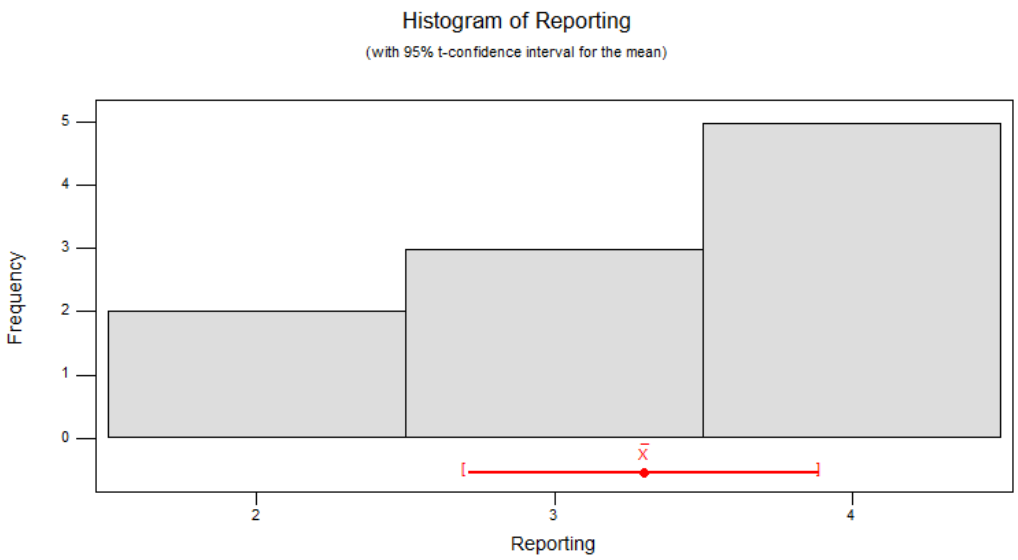

Figure 12: Histogram-1 of the analysis

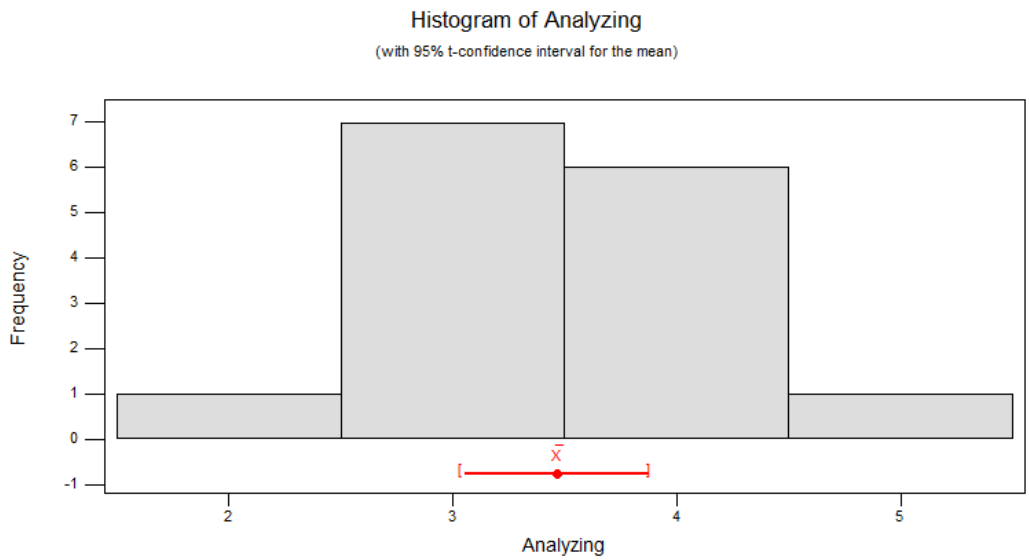

Figure 13: Histogram-2 of the analysis

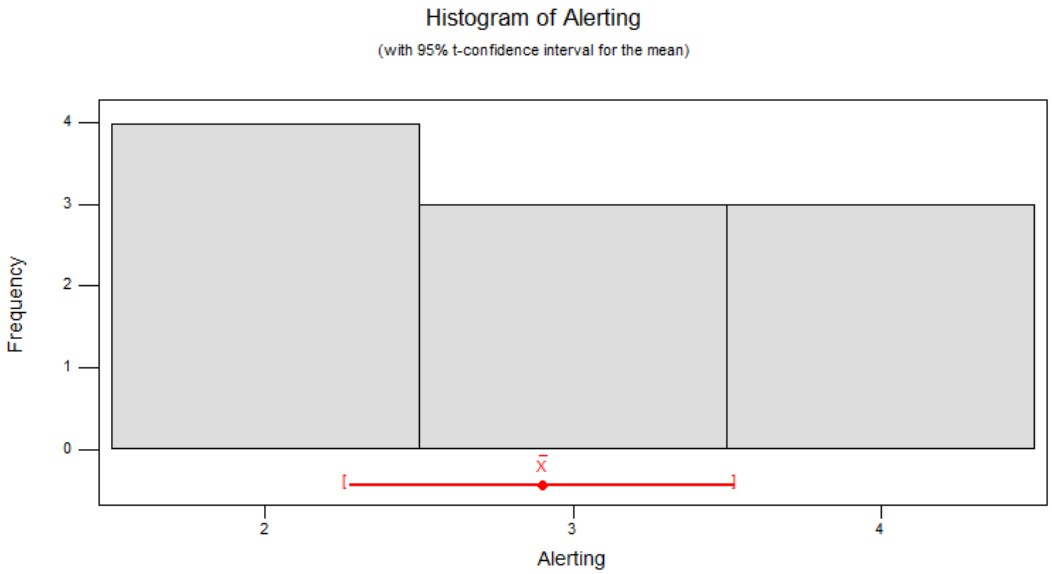

Figure 14: Histogram-3 of the analysis

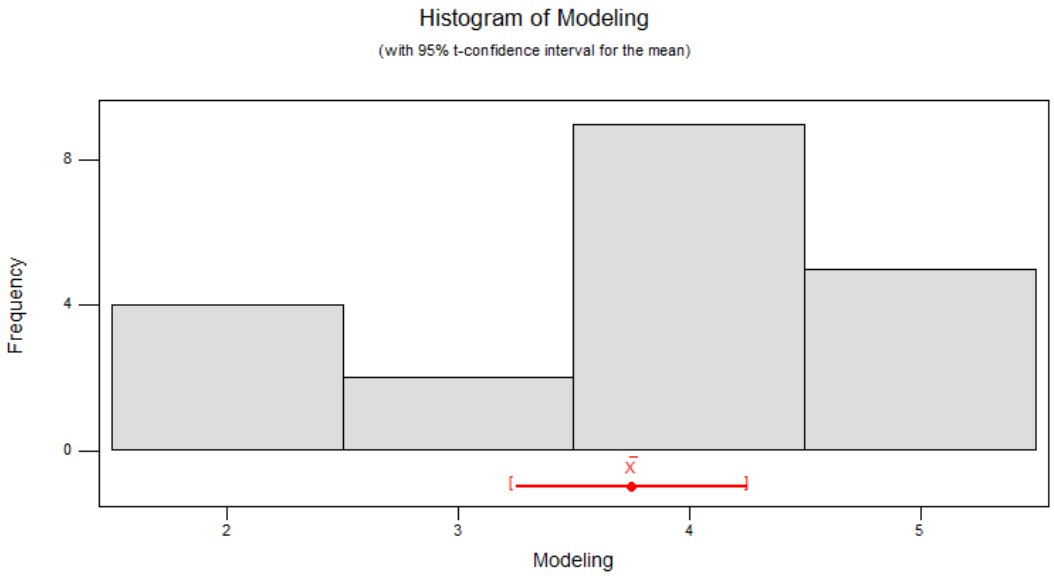

Figure 15: Histogram-4 of the analysis

After the T-test, we have also applied a cluster analysis for all data which are shown above by the pie charts. According to the analysis results; the amalgamation steps, number of the determined clusters at each step, the similarity and distance levels between the variables, the joined and new clusters at each step as well as the number of observations in the new clusters are given below.

Table 2: Cluster-1 of the analysis

Step	No. of clusters	Similarity level	Distance level	Clusters joined	New cluster	No. of obs. in cluster
1	3	79,49	0,410	3 4	3	2
2	2	60,02	0,800	1 3	1	3
3	1	46,81	1,064	1 2	1	4

Based on the cluster analysis we applied by Minitab, we get a dendrogram seen below showing the similarity diagrams of all our variables in the analysis.

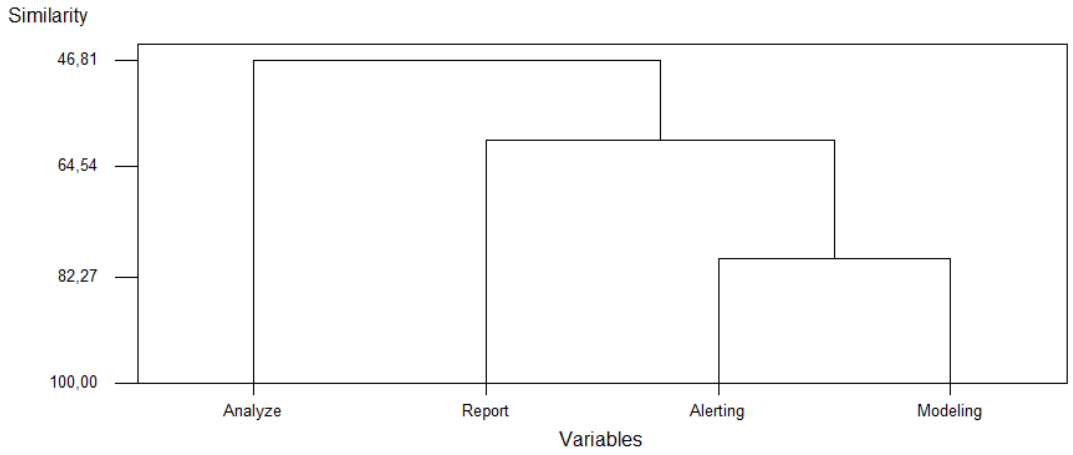

Figure 16: Dendrogram-1 of the analysis

As a notification seen on the above dendrogram, for example, we can note that ***the respondents for whom "Alerting" is at the "1st" rank and the respondents for whom "Modeling" is at the "1st" rank have a high similarity when we compare their answers to the questions 5, 6, 7, 8 and 9 from "1" for "very low" to "5" for "very good".*** Such comparisons can also be made between the other variables depending on their answers to the questions 5, 6, 7, 8 and 9.

4.1.2 Analysis Results: Awareness Test Based on ERP, CRM and SCM Systems Usage

At the next fourth question in our survey (Appendix), we asked to the respondents that from which systems their company gets the datas for the business intelligence management in their companies and we declared them that they can make multiple choices between "ERP, CRM and SCM". Then we analyzed the responses by comparing the companies according to their multiple choices among; "ERP, CRM, ERP&CRM or ERP&CRM&SCM" options. So that, we grouped the companies based on their choices and we got 4 different pie charts each are representing one of the "ERP, CRM, ERP&CRM or ERP&CRM&SCM" options that are possibly marked on our survey by the respondents.

As in the third question's analysis, each pie chart represents the answers of the respondents for questions 5, 6, 7, 8 and 9. For example, the ERP pie chart below denotes the responses among "Very Low, Low, Medium, Good and Very Good" options of the questions 5, 6, 7, 8 and 9 only between the respondents for whom "only ERP" is chosen/marked for question 4.

As in the third question's analysis again, in the pie charts, the first number denotes the answer to questions 5, 6, 7, 8 and 9 from "1" for "very low" to "5" for "very good". The second number denotes the number of that answer in our data and the last number denotes the percentage of this answer between all the data.

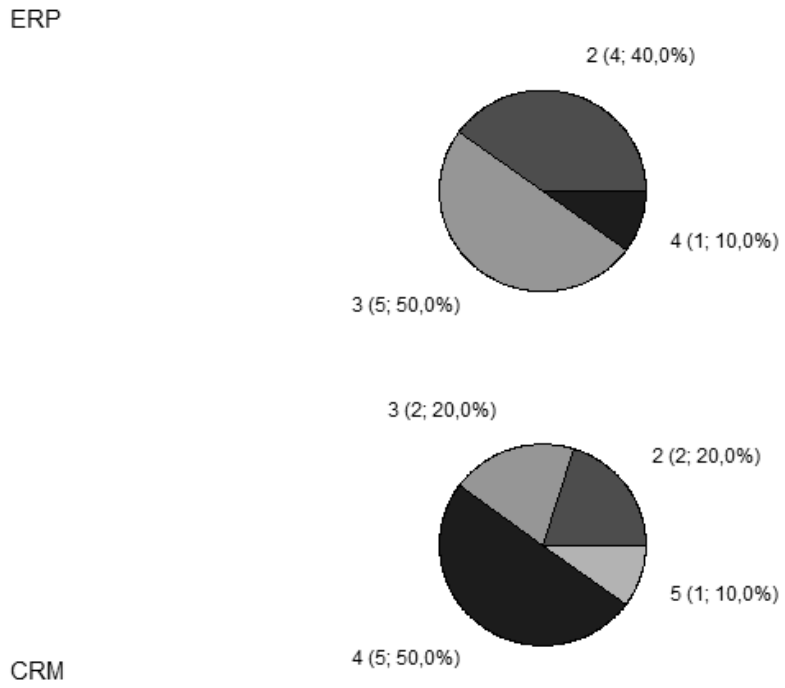

Figure 17: Pie chart-3 of the analysis

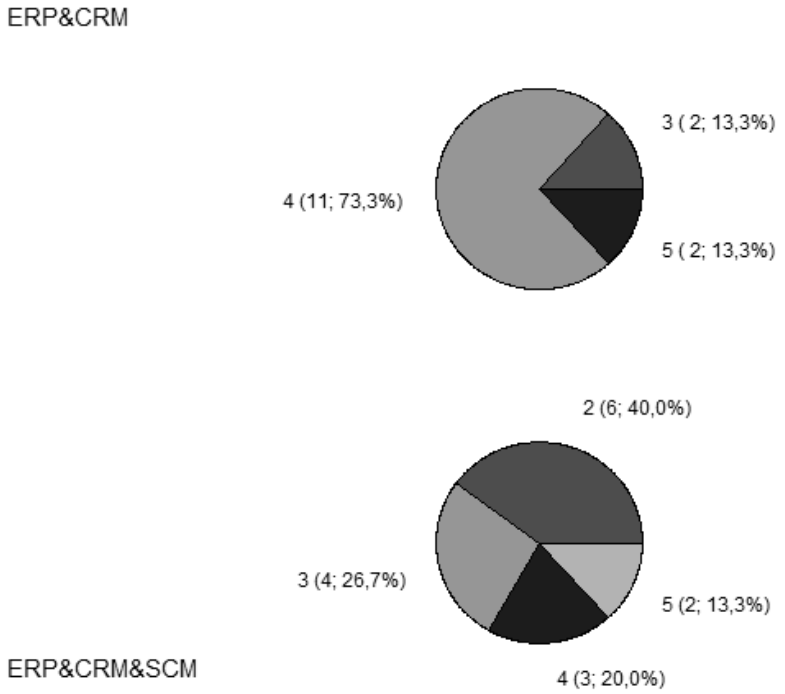

Figure 18: Pie chart-4 of the analysis

Based on the data sets we have shown with the above pie charts, again the T-test have been applied in order to see the Mean (mean values of all N data sets), StDev (standard deviations), SE Mean (standard error mean) as well as the histograms of all N data sets. The results and all histograms of our T-test can be seen hereunder:

Table 3: T-test-2 of the analysis

Variable	N	Mean	StDev	SE Mean
ERP	10	2,700	0,675	0,213
CRM	10	3,500	0,972	0,307
ERP&CRM	15	4,000	0,535	0,138
ERP&CRM&SCM	15	3,067	1,100	0,284

At the following histograms, again the vertical diagram shows the frequency that is the total number of responses which contain "Very Low, Low, Medium, Good and Very Good" options of the questions 5, 6, 7, 8 and 9 from "1" for "very low" to "5" for "very good" as the horizontal diagram represents these responses graphically.

All histograms are the results of our T-test which is again done by Minitab 13,0 and have been drawn with 95% t-confidence interval for the mean.

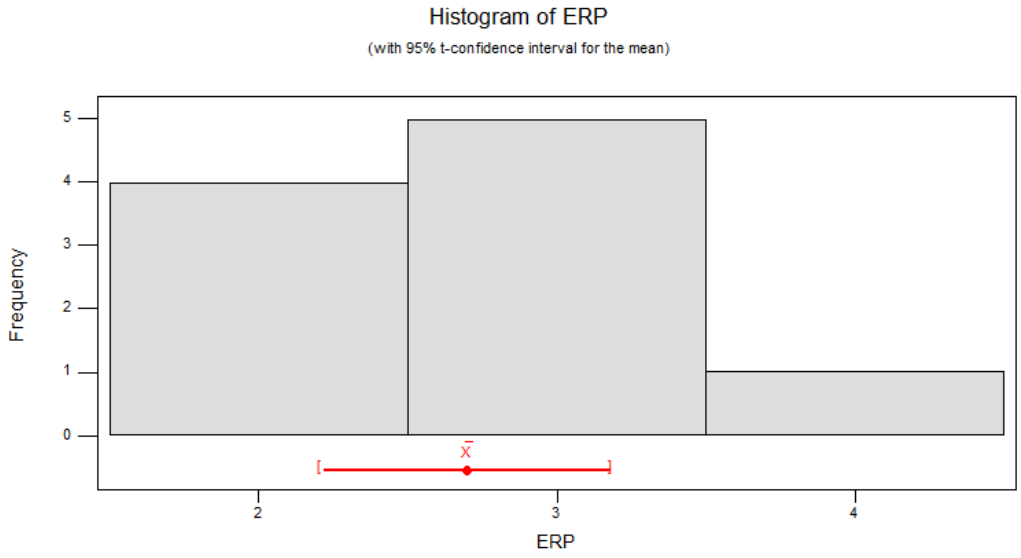

Figure 19: Histogram-5 of the analysis

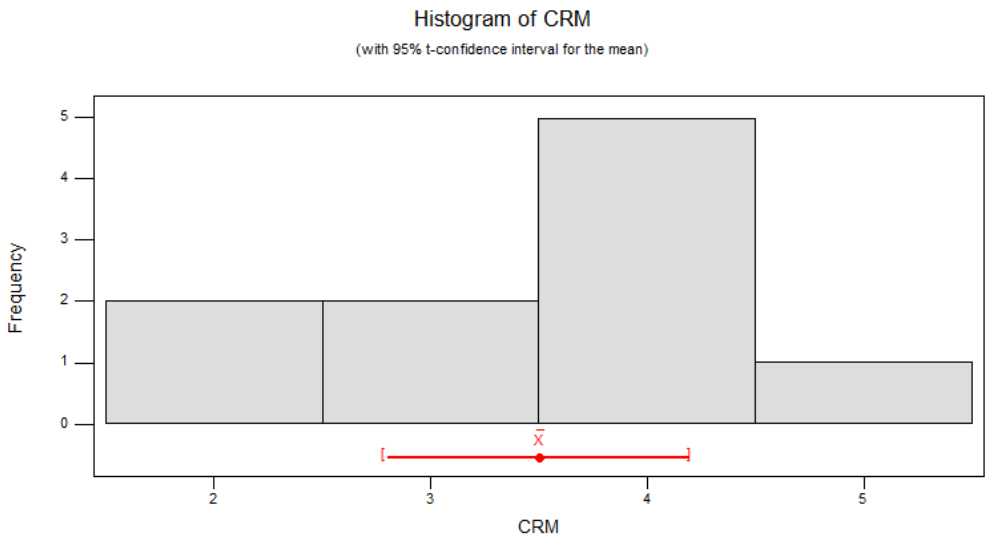

Figure 20: Histogram-6 of the analysis

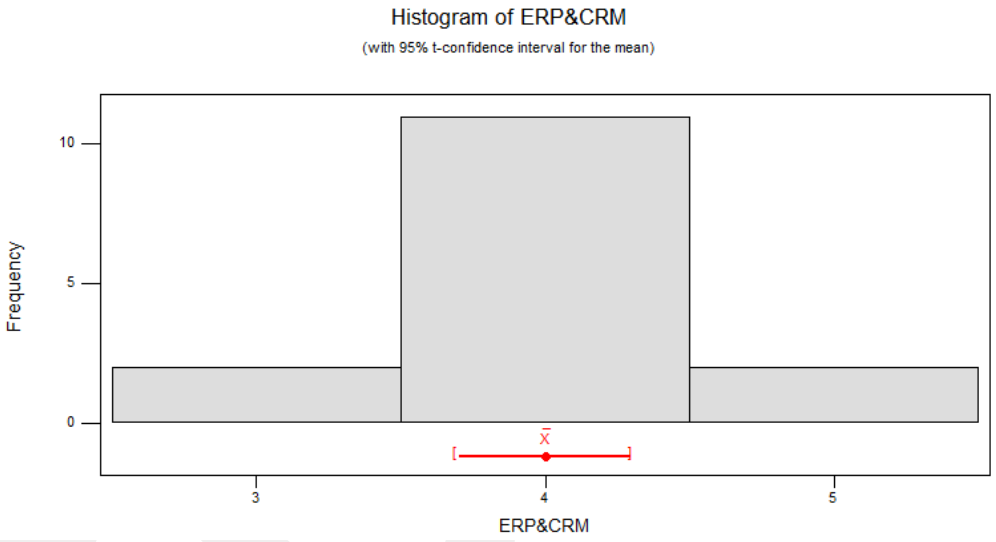

Figure 21: Histogram-7 of the analysis

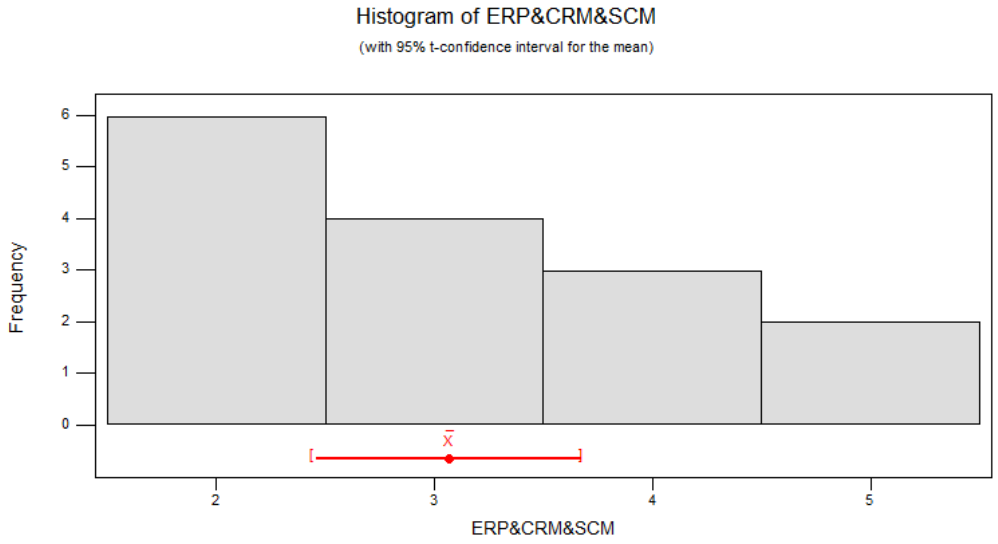

Figure 22: Histogram-8 of the analysis

After T-test, we have again applied a cluster analysis for all data which are shown above by the pie charts. According to the analysis results, the amalgamation steps are given below.

Table 4: Cluster-2 of the analysis

Step	No. of clusters	Similarity level	Distance level	Clusters joined	New cluster	No. of obs. in cluster
1	3	74,40	0,512	3 4	3	2
2	2	69,05	0,619	1 3	1	3
3	1	68,52	0,630	1 2	1	4

Finally, we get a dendrogram seen below showing the similarity diagrams of all our variables in the analysis.

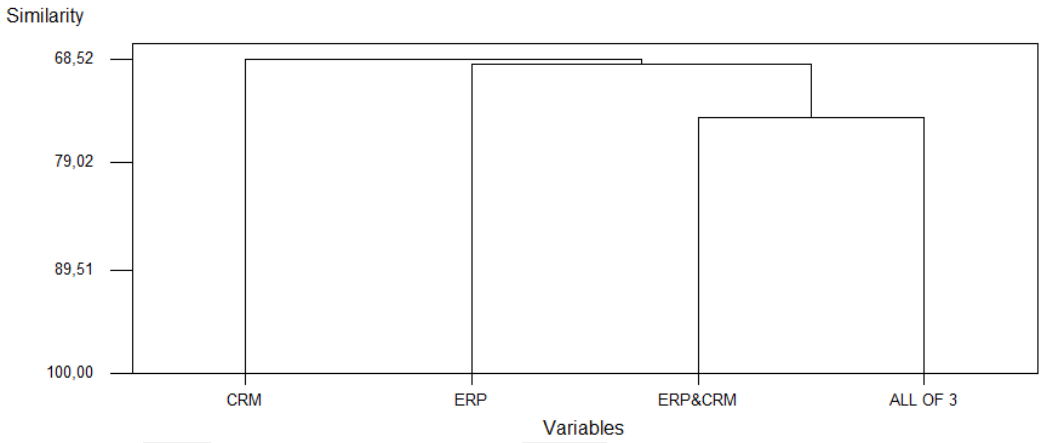

Figure 23: Dendrogram-2 of the analysis

4.1.3 Analysis Results: Awareness Test Depending on the Usage of Different Types of Analysis

At the tenth question in our survey (Appendix), we asked to the respondents that for which analysis their company use the business intelligence and we asked for a ranking from them. Then we analyzed the responses by comparing the companies according to their 1st rank choice; to analyze customers, suppliers or the annual turnovers of the company. So that, we grouped the companies based on their 1st rank choice and we got 3 different pie charts each are representing one of "to analyze customers, suppliers or the annual turnovers of the company" that are possibly chosen by the respondents at the 1st rank.

Like in the other questions' analysis, each pie chart represents the answers of the respondents for questions 5, 6, 7, 8 and 9. For example, the Customer pie chart below denotes the responses among "Very Low, Low, Medium, Good and Very Good" options of the questions 5, 6, 7, 8 and 9 only between the respondents for whom "to analyze the Customers" is at the 1st rank for question 10.

In the pie charts hereunder, the first number denotes the answer to questions 5, 6, 7, 8 and 9 from "1" for "very low" to "5" for "very good". The second number denotes

the number of that answer in our data and the last number denotes the percentage of this answer.

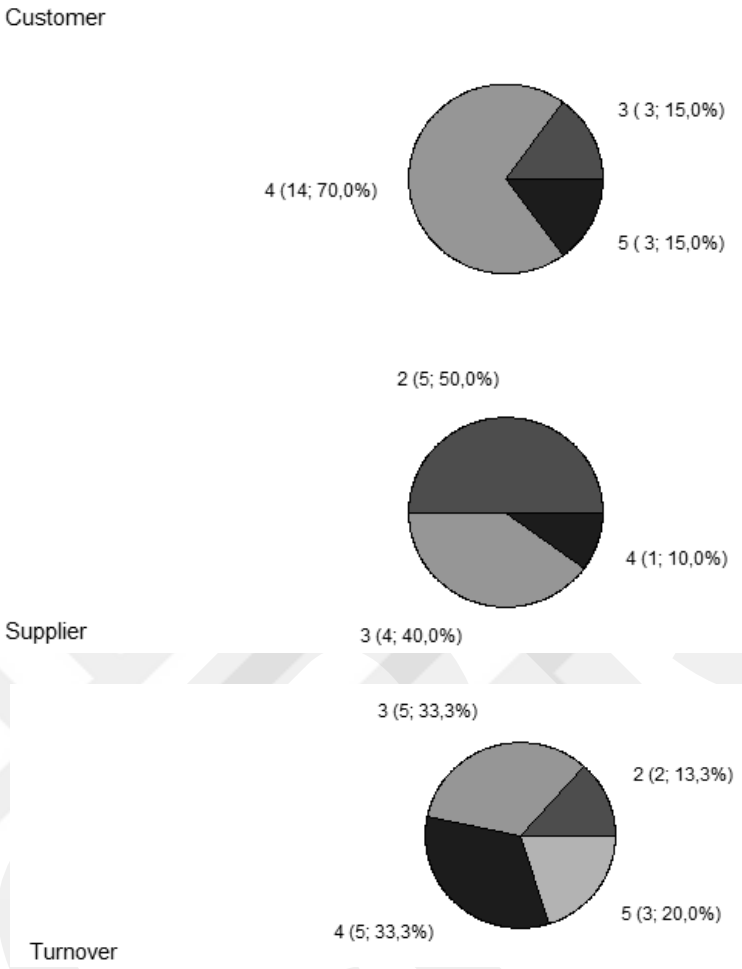

Figure 24: Pie chart-5 of the analysis

Based on the data sets we have shown with the above pie charts, the applied T-test in order to see the Mean (mean values of all N data sets), StDev (standard deviations), SE Mean (standard error mean) as well as the histograms of all N data sets gives the below results and histograms:

Table 5: T-test-3 of the analysis

Variable	**N**	**Mean**	**StDev**	**SE Mean**
Customer	20	4,000	0,562	0,126
Supplier	10	2,600	0,699	0,221
Turnover	15	3,600	0,986	0,254

At the following histograms, the vertical diagram shows the frequency that is the total number of responses from “1” for ”very low” to “5” for “very good” when the horizontal diagram represents these responses.

Figure 25: Histogram-9 of the analysis

Figure 26: Histogram-10 of the analysis

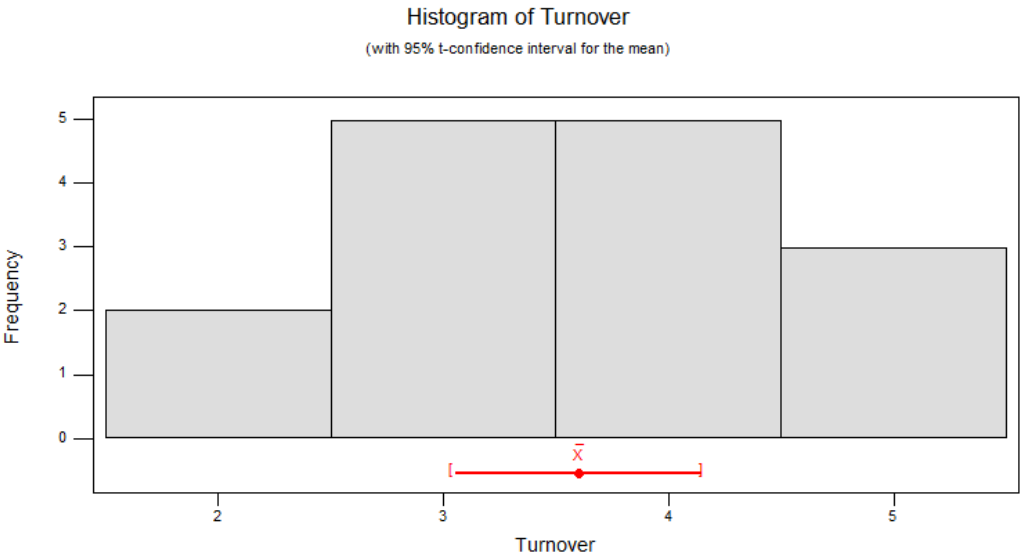

Figure 27: Histogram-11 of the analysis

After the T-test, we have again applied a cluster analysis for all data which are shown above by the pie charts. The analysis results and steps are given below:

Table 6: Cluster-3 of the analysis

Step	No. of clusters	Similarity level	Distance level	Clusters joined	New cluster	No. of obs. in cluster
1	2	86,78	0,264	2 3	2	2
2	1	74,40	0,512	1 2	1	3

Based on the cluster analysis we applied by Minitab, we get a dendrogram seen below showing the similarity diagrams.

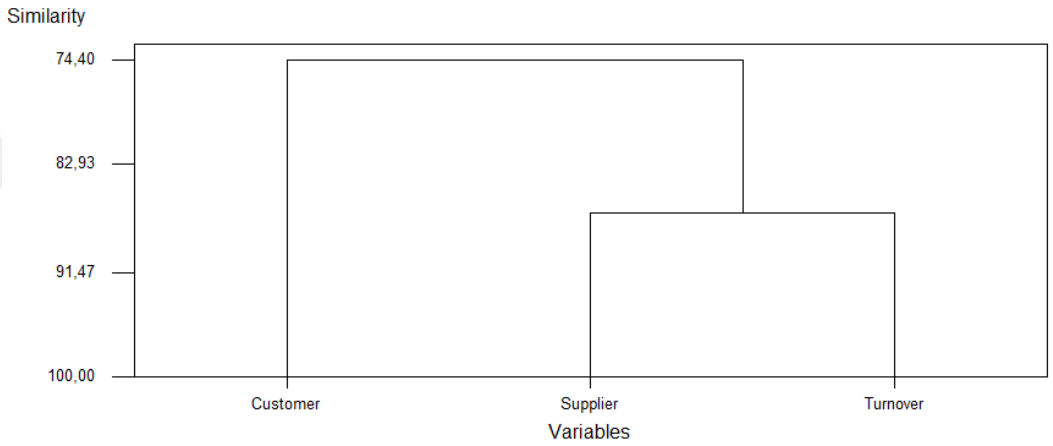

Figure 28: Dendrogram-3 of the analysis

4.1.4 Analysis Results: Awareness Test Depending on Who the Users Are

At the twelfth question in our survey (Appendix), we asked to the respondents that who utilizes the outcomes of the business intelligence systems in their companies and we noted them that they can make only one choice between "the Top Management, the Middle Management and the Specialists, the People in Sales and Development, or a Communication Unit". Then based on the available data sets that we have, we analyzed the responses by comparing the companies according to their choices among; "the Top and the Middle Managements or the Sales & Development Staff " options. So that, we grouped the companies based on their choices and we got 2 different pie charts each are representing one of the "Management or Sales& R&D" options that are possibly marked on our survey by the respondents.

As in the other questions' analysis, each pie chart represents the answers of the respondents for questions 5, 6, 7, 8 and 9. For example, the Management pie chart below denotes the responses among "Very Low, Low, Medium, Good and Very Good" options of the questions 5, 6, 7, 8 and 9 only between the respondents for whom "the Management" is chosen/marked for question 12.

As in the other questions' analysis again, in the pie charts, the first number denotes the answer to questions from "1" for "very low" to "5" for "very good". The second

number denotes the number of that answer in our data and the last number denotes the percentage of this answer.

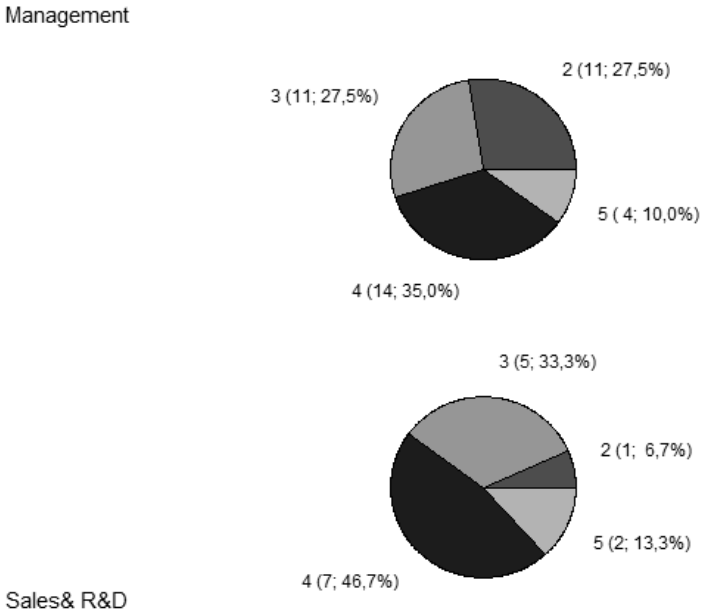

Figure 29: Pie chart-6 of the analysis

Based on the data sets we have shown with the above pie charts, after we apply the T-test, the results and all histograms can be seen hereunder:

Table 7: T-test-4 of the analysis

Variable	N	Mean	StDev	SE Mean
Management	40	3,275	0,987	0,156
Sales& R&D	15	3,667	0,816	0,211

At the following histograms, we can see the vertical diagram showing the frequency while the horizontal diagram representing these responses graphically again.

Figure 30: Histogram-12 of the analysis

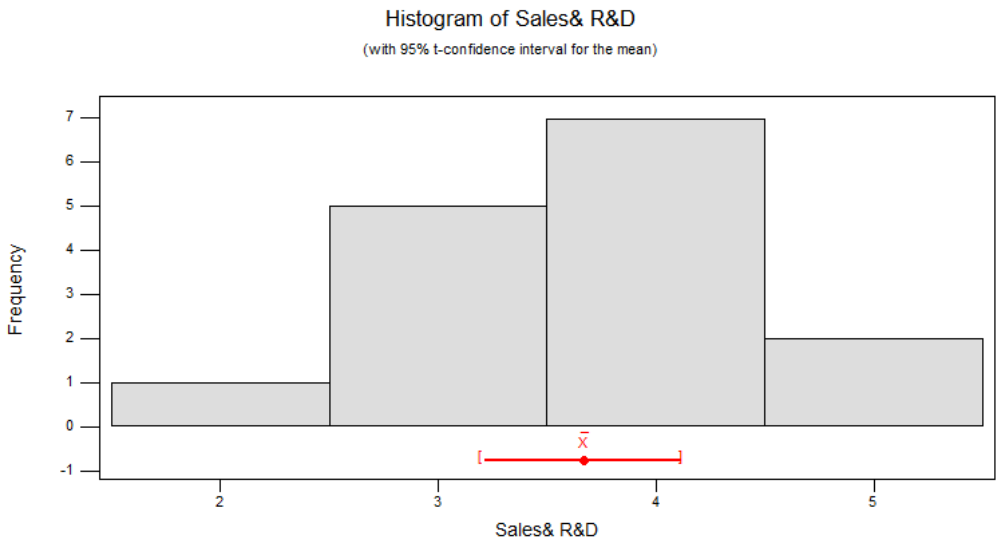

Figure 31: Histogram-13 of the analysis

After the T-test, we again applied a cluster analysis for all data which are shown above by the pie charts. According to the analysis results, all the amalgamation steps and other numerical values in the new clusters are given below.

Table 8: Cluster-4 of the analysis

Step	No. of clusters	Similarity level	Distance level	Clusters joined	New cluster	No. of obs. in cluster
1	1	70,75	0,585	1 2	1	2

As a result diagram, we again get a dendrogram seen below showing the similarity diagrams of all our variables in the analysis.

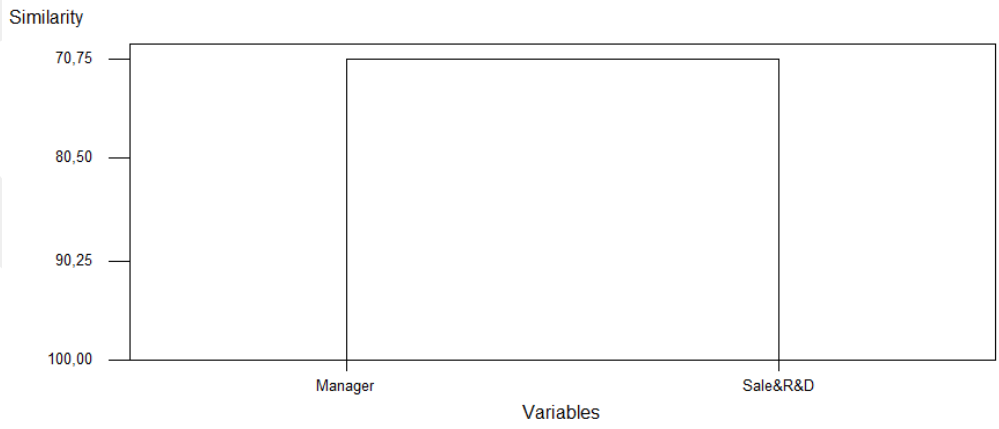

Figure 32: Dendrogram-4 of the analysis

4.1.5 Analysis Results: Awareness Test for Different Tools

At the thirteenth question in our survey (Appendix), we asked to the respondents that what kind of BI tools their company use for the business intelligence management in their companies and we asked for a ranking from them. Then we analyzed the

responses by comparing the companies according to their 1st rank choice; Query tool, Reporting tool, OLAP and Data Mining tool. So that, we grouped the companies based on their 1st rank choices and we got 3 different pie charts each are representing one of "Query tool, Reporting tool or OLAP" that are possibly chosen by the respondents at the 1st rank.

Like in the other questions' analysis, each pie chart represents the answers of the respondents for questions 5, 6, 7, 8 and 9. For example, the Query pie chart below denotes the responses among "Very Low, Low, Medium, Good and Very Good" options of the questions 5, 6, 7, 8 and 9 only between the respondents for whom "the Query tool" is at the 1st rank for question 13.

In the pie charts below, the first number is denoting the answer to questions ,the second number is denoting the number of that answer and the last number is denoting the percentage.

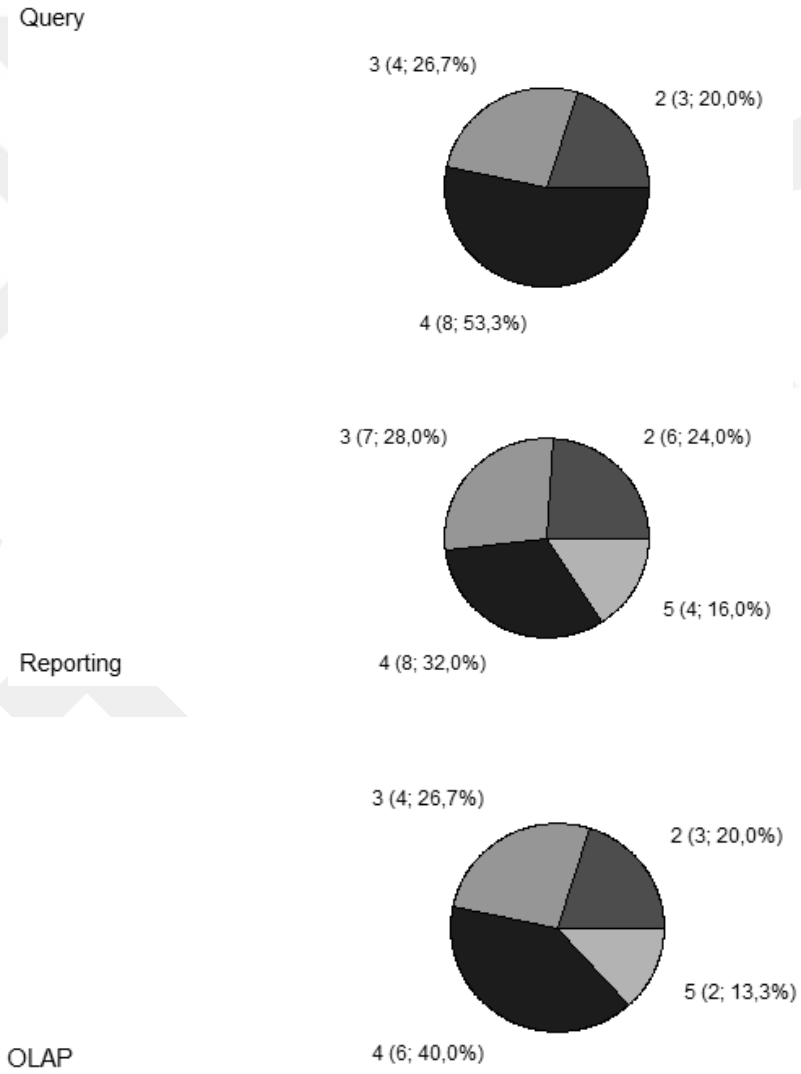

Figure 33: Pie chart-7 of the analysis

Based on the data at the above pie charts, the applied T-test gives the below results and histograms:

Table 9: T-test-5 of the analysis

Variable	N	Mean	StDev	SE Mean
Query	15	3,333	0,816	0,211
Reporting	25	3,400	1,041	0,208
OLAP	15	3,467	0,990	0,256

Figure 34: Histogram-14 of the analysis

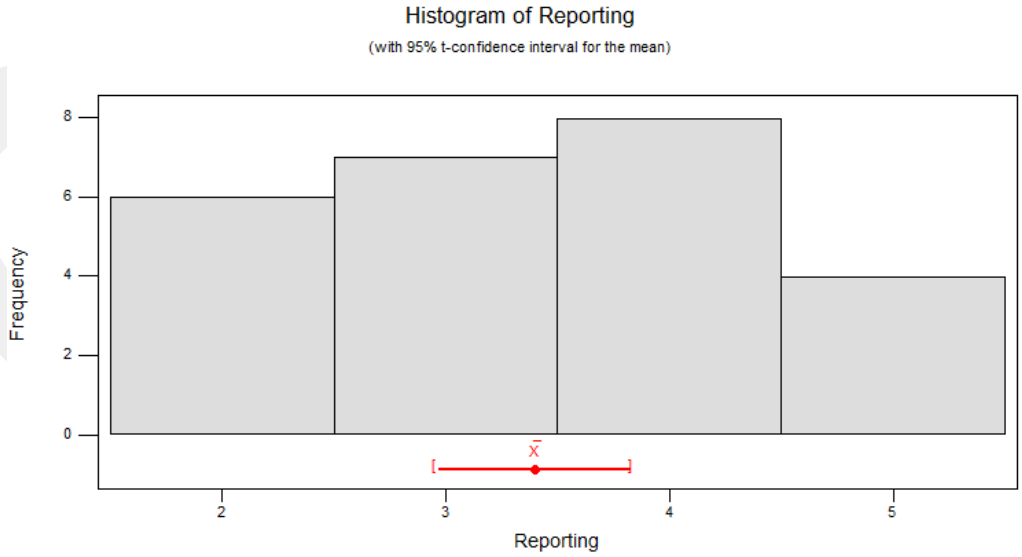

Figure 35: Histogram-15 of the analysis

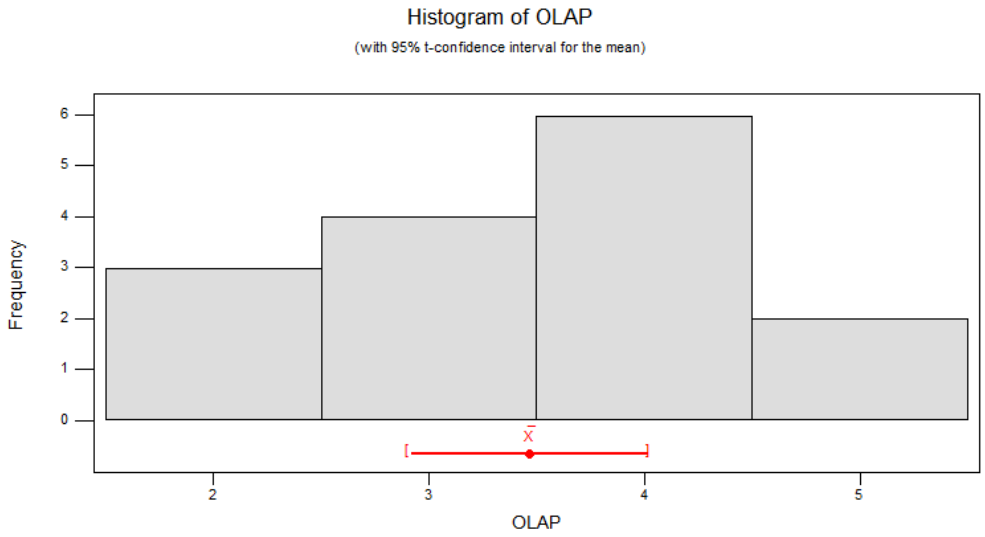

Figure 36: Histogram-16 of the analysis

Furthermore, we have applied a cluster analysis for all data which are shown above by the pie charts. The analysis results, steps and the dendrogram are given below:

Table 10: Cluster-5 of the analysis

Step	No. of clusters	Similarity level	Distance level	Clusters joined	New cluster	No. of obs. in cluster
1	2	66,19	0,676	1 3	1	2
2	1	62,98	0,740	1 2	1	3

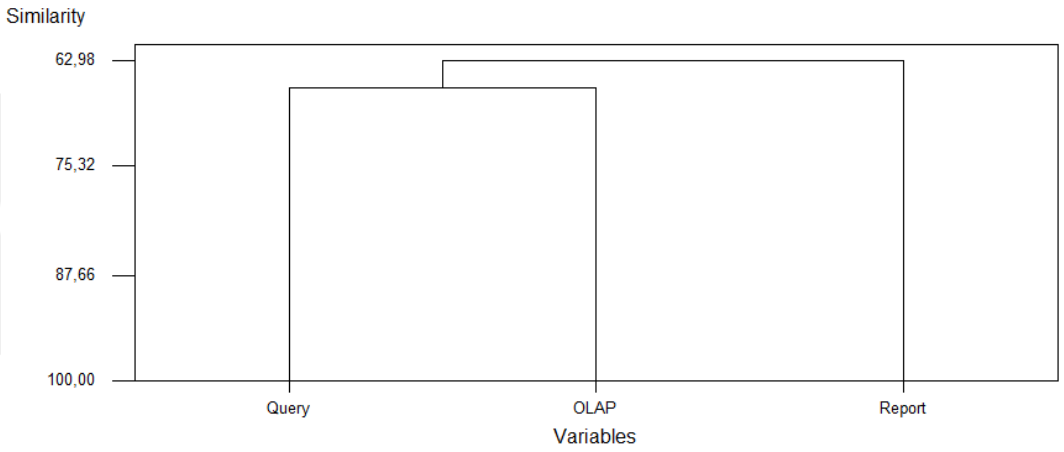

Figure 37: Dendrogram-5 of the analysis

4.1.6 Analysis Results: Awareness and Largeness Test Depending on the Total Company Staff

At the beginning of our survey (Appendix), we asked to the respondents the total staff in their companies. Then based on the available data sets, we analyzed the responses by comparing the companies according to their total staff (personnel) with; "Total Staff<100, 100<Total Staff<1000 or Total Staff>1000" options. So that, we

grouped the companies based on their total staffs and we got 3 different pie charts each are representing one of the "Total Staff<100, 100<Total Staff<1000 or Total Staff>1000" options that are noted on our survey by the respondents.

As in the other questions' analysis, again each pie chart represents the answers of the respondents for questions 5, 6, 7, 8 and 9. For example, the "TS<100" pie chart below denotes the responses among "Very Low, Low, Medium, Good and Very Good" options of the questions 5, 6, 7, 8 and 9 only between the respondents for those whose company employ totally not more than 100 staff.

Below, the first number denotes the answer to questions from "1" for "very low" to "5" for "very good". The second one denotes the number of that answer in our data set and the last number denotes the percentage of this answer in the pie charts.

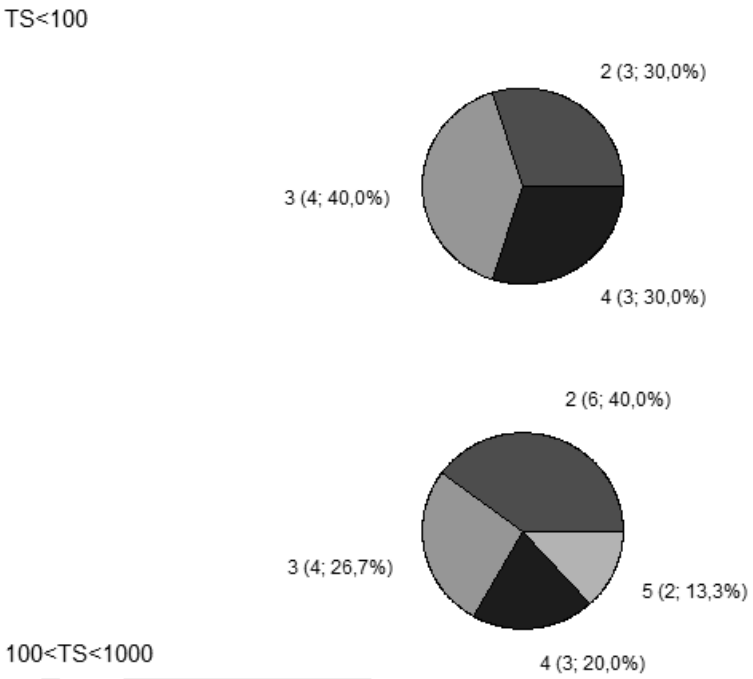

Figure 38: Pie chart-8 of the analysis

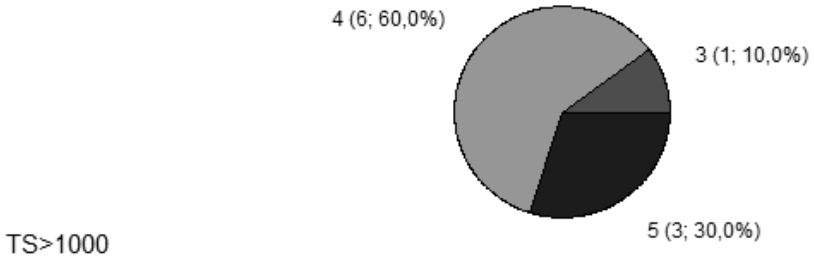

Figure 39: Pie chart-9 of the analysis

Based on the data sets at the above pie charts, the results of T-test and all histograms can be seen hereunder:

Table 11: T-test-6 of the analysis

Variable	N	Mean	StDev	SE Mean
TS<100	10	3,000	0,816	0,258
100<TS<1000	15	3,067	1,100	0,284
TS>1000	10	4,200	0,632	0,200

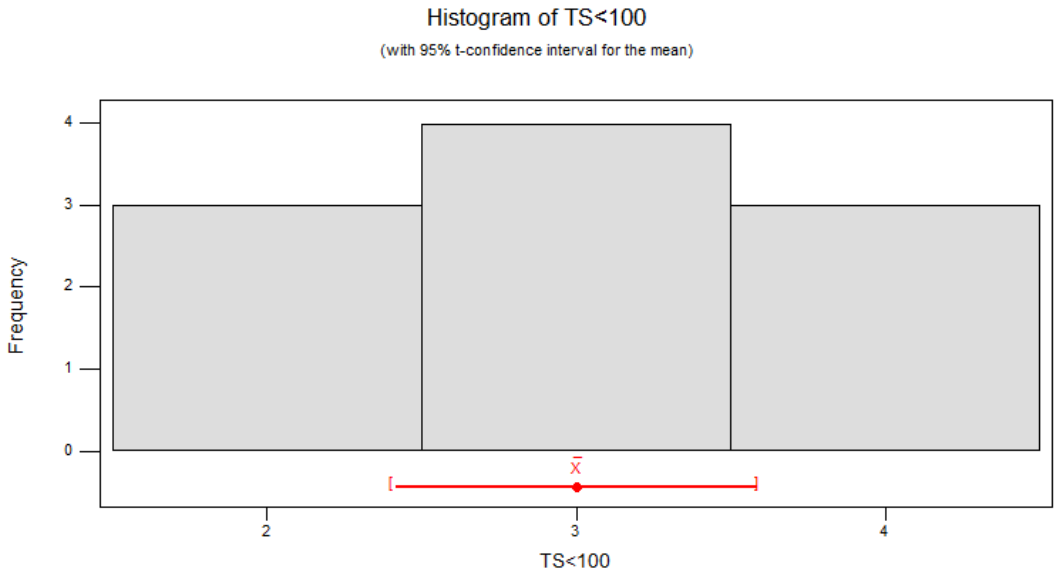

Figure 40: Histogram-17 of the analysis

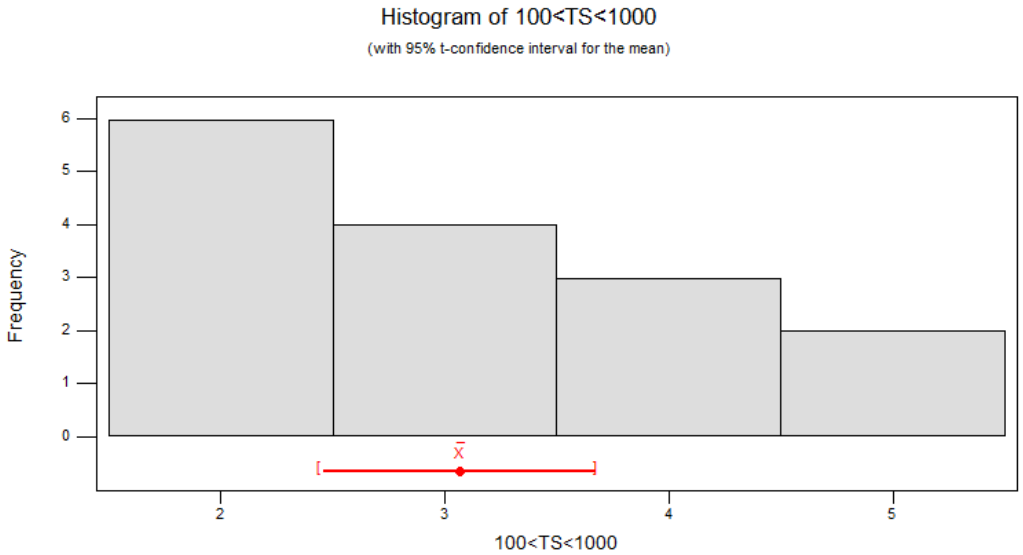

Figure 41: Histogram-18 of the analysis

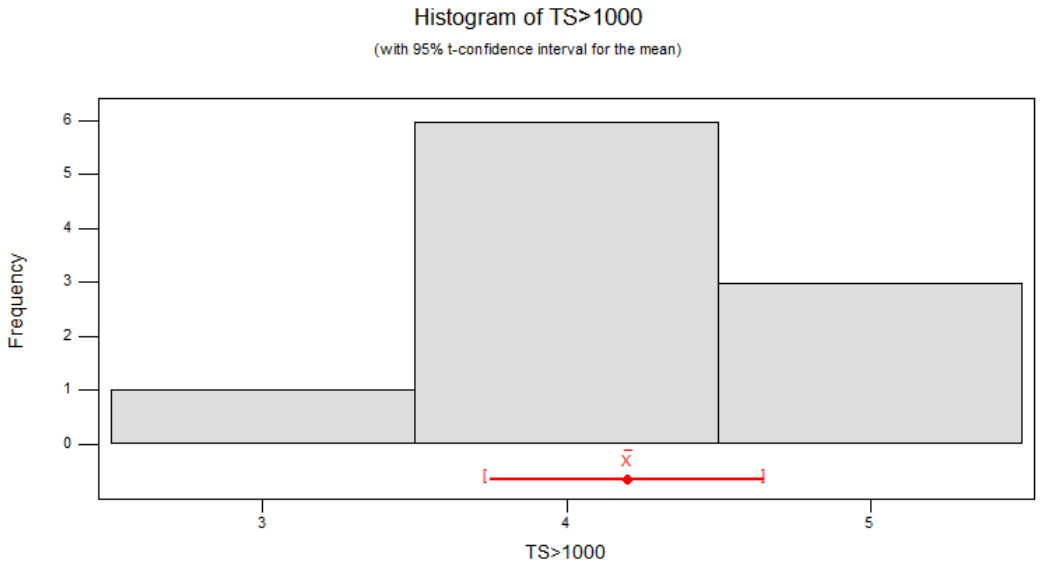

Figure 42: Histogram-19 of the analysis

After the T-test, we applied a cluster analysis for all data which are shown above by the pie charts. According to the analysis results, all the amalgamation steps and other numerical values in the new clusters are given below. As a result diagram, we again get a dendrogram seen below showing the similarity diagrams of all our variables in the analysis as well.

Table 12: Cluster-6 of the analysis

Step	No. of clusters	Similarity level	Distance level	Clusters joined	New cluster	No. of obs. in cluster
1	2	43,70	1,126	1 2	1	2
2	1	41,87	1,163	1 3	1	3

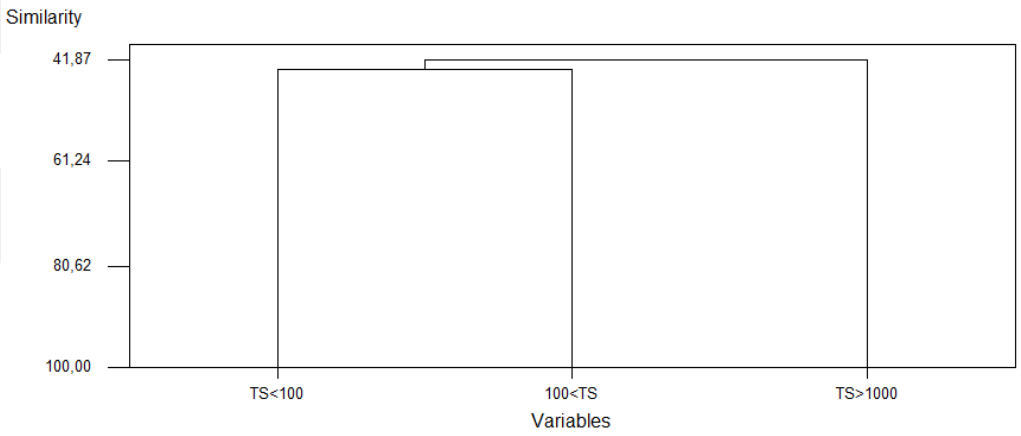

Figure 43: Dendrogram-6 of the analysis

4.2 Conclusion and Discussion

The purpose of our research was identifying a requirement analysis discussion depending on the critical success factors for BI that is implemented like an extension of the ERP systems. A conceptual framework can be developed from such a research literature that the identified factors associated with the business intelligence systems are; the purposes of the companies, different systems such as ERP, CRM and SCM, different analysis, users, BI tools as well as the largeness of the companies depending on how big staff they are working with. Depending on such a Literature Review, a series of survey questions have been developed. These questions have been investigated using a general approach for the small and medium enterprises in Turkey and that included the content analysis of BI industry with the industry practitioners. The content analysis of 20 small and medium enterprises was conducted for validating the success keys that are identified in a framework and definition of any factors which have not been identified as well. The questionnaires, conducted with the BI practitioners, are based on the findings for the requirement analysis and so as for verifying the effective factors and for determining their criticality.

The requirement analysis phase of our research identified a broad range of the BI factors. In the requirement analysis phase of the industry, we identified some factors that are relevant for the aspect of BI. After the questionnaire phase, we had a chance to discuss these factors which are most relevant. It was evident to say that while some companies had implemented the BI like an accelerator, some of these companies were even far away than placing a BI strategy as well as identifying the factors associated with this implementation. Of course, that doesn't mean that the important factors of BI which are not noticed by the companies were not relevant. However, this may be the case in some companies. If a company always has a good level of the management support, this might not be noted as a current requirement for this company. Indeed, how a company can know that the lack of a management support may negatively affect their projects? It would be wise to identify specific factors as a standard implementation practice.

This research aims to make some contributions for both of the academia and industry to understand the requirements of the companies as well as their applications for a successful BI management. Companies are increasingly using BI for improving their decision making as well as their corporate performances. Many of the companies implement BI as an extension of the ERP systems for providing a seamless integration among the systems transactions and/or the systems used for analyzing the business process performances. Companies are seeking for ways for improving the effectiveness of their BI initiatives. So the importance of requirement analysis in BI implementations should be well-documented.

According to the research results, the companies who are using the business intelligence systems for predictive modeling have a better awareness when compared to the companies who are using it for reporting, for instance. Although it is very important to have a reporting mechanism in a company, many companies prefer to make investment on predictive modeling or alerting. That can be defined like a tendency to predict the future more than reporting and analyzing today's current situation. And as a result of that, possibly we face many companies whose BI awareness is high between the predictive modeling users. Furthermore, we notice the

high similarity between alerting and predictive modeling diagrams (Figure 16) based on the cluster analysis. Both of predictive modeling and alerting possibly could be used for the future of the company. Of course, that requires a very detailed and conscious business intelligence management as well as a high awareness. And that's why, the results are not surprising for us. However, we hereby can note that the companies who are using the business intelligence mostly for reporting and/or analyzing might be preferring to managing the process with some simple tools and low awareness with a general approach of thinking that they can solve the problems when they currently face. That is acceptable but requires a discussion especially for the big companies, of course.

Another important result of our analysis (table 3) is that; the companies who are using ERP and CRM together has more awareness when compared to the companies who are using only one of ERP or CRM. That is not surprising for us as that can be called as a criteria to understand how advanced the company is managing the business intelligence. Many companies might foresee that it could be enough for them to use only one system like an ERP system for the data management. However, we hereby see that the companies becoming more aware of the business intelligence management when they use alternative systems like at least one alternative system such as CRM. Furthermore, we also see that the highly similar (Figure 23) diagrams of CRM&ERP usage and CRM&ERP&SCM usage identified by the cluster analysis dendrograms with similarity diagrams. So it is important to note that both of the cluster analysis and t-test says us the same thing and provide a valuable research feedback as well as a wise hint for the companies that are using business intelligence management systems.

When we have a look at the analysis results (table 5 & table 6) for the comparison of the companies who are using the business intelligence for different types of analysis, we notice a very valuable result. It says us that the business intelligence management and awareness is at higher levels for the companies who are using the business intelligence for analyzing the customers when compared with the companies who aim to analyze the suppliers the first. Here the first thing coming to our mind is the customer-supplier relationship one can say that the customer is always right. So if a company is looking for new customers, that requires a detailed and careful market research process, of course. And even a small mistake can cost some years for the company. Instead, if the company is looking for a supplier, sometimes it is worthy to try even if they don't have a detailed research process. Because as a customer, that company might just purchase a little amount for the beginning for testing purposes. But the supplier has only one chance; winning or losing the customer that might not even bring them a second chance.

In our research, we also studied on the awareness of two different type of companies. One of them gives the use of business intelligence to the management staff's hands as the other one gives the use to the people in sales and/or development organization. The results of the analysis (table 7 and table 8) say us that the companies where the sales& R&D staff are effectively using the business intelligence is more successful and aware of the business intelligence management. That actually supports one of the main laws in the life; the growth starts from the root. So that, if a company has a well-organized business intelligence management starting from the R&D staff, that means they are fed from the root as the most creative ideas generally come into the

mind from a sales or R&D staff. But of course, that doesn't mean the management staff don't use the business intelligence effectively in such companies as well. However, if the more effective users are the R&D staff, that is an indicator proving that the business intelligence have been integrated very well in the company that even the sales and/or R&D staff using it effectively. Some companies would also want to have a chance for collaborating on the ideas and that is something they cannot do without the technical data and staff that may not give them the ability to do it by themselves. Professionals in IT have to make sure that the created data is not only usable via the people who are not talented, but it is also shareable and easy for collaborating within the conjunction with the other departments in the company. According to many experts, the next years would be the years of the business intelligence market as it is finally moving away from being primarily IT-centric as well as it is becoming more user-friendly. Over the past years, some IT professionals have created produced and used the reports which are associated with the software, but there are also some companies looking for being involved at the ownership level as they are getting a feel for what they're using. That means business intelligence becomes much more user-friendly and usable across the year, as it is opposed to being a technical solution which has been criticized. In today's business intelligence landscape, we notice a dilution in the roles that each person is playing in an organization. That can be true in the case of the data analysis. Today's data analysts are no longer the business intelligence experts within their departments. The data is being analyzed and/or generated with the operations manager, by the supply chain executives or even by the sales staff.

While we also asked the companies that what kind of BI tools they use, we had three popular answers; query tools, reporting tools and OLAP. The results of the analysis (table 9 and table 10) are very close when we use cluster analysis as well as a t-test for analyzing. And that is acceptable when we evaluate that all these three tools are effectively used in the market. However, we should mention about an importing similarity here between the query tools and OLAP. According to the cluster analysis and the dendrogram (Figure 37) based on these analysis, we see a current similarity diagram between the query tools and OLAP. That can be explained by the definitions of each tool as OLAP is an approach to answer the multi dimensional analytical queries as well. This is very important as one of the most important hurdles which some companies face is the fact that some business intelligence tools may not be user-friendly. They could be hard for working with and are discarded and/or passed on to somebody else. This also shows the requirement for the tools which are easy for understanding and straightforward for utilizing

As we had an important question in our survey asking the total number of company staff, we also had an opportunity to discuss the awareness of the companies depending on how the company is big. Here we faced a big difference (table 11) between the companies who have more than 1000 staff and who have less than 1000 staff who are working for these companies. For the companies that have at least 1000 personnel, it was very valuable to see their high awareness level when compared to the others. Furthermore, the similarity diagrams (Figure 43) that we got with the cluster analysis also shows a similarity between the companies who have less than 100 staff and the companies who have a staff between 100 and 1000 people. That shows us if the company is quite big, their business intelligence awareness is at a high level as well. But of course, that is possible with a well defined business intelligence management. For small companies, the full scope of business

intelligence would not be required to get the aimed results. However, it could be true that they should have “all or nothing” while investing as well as that could turn many away.

Allowing the companies and buyers only purchasing the things which they want and require could be an effective strategy as well as allowing them having a business intelligence solution which is working for them at any affordable price. The decision makers in many companies might not strictly be office based and they may be out at some points during the working week. However, the big deals would always need to be made and cannot wait. Via permitting the usage of business intelligence by mobile channels, the developers would make it easier for the companies implementing a strategy in the modern world as well as uptaking can be greatly improved.

Nowadays, many companies are starting to use the certain aspects of the business intelligence, but many of them are not investing if they feel that they are being forced for taking the ownership of a whole strategy and instead of that, preferring to pick and/or choose the certain elements. Self service business intelligence has already become a popular solution among the business sector and the companies who are able to create their own business intelligence software will be the companies who do the best. And finally, age of the non IT business intelligence specialist is descending as the decision makers aims to have a hand in business intelligence now more than ever. The upcoming years are expected to be the years of consolidation and/or transformation for the BI. Regarding the rise of the internet, the integration of cloud and/or the increasing demand for the mobility is going to shape the way that the working staff consume the data as well as shaping an established industry that is worth to billions of $ per a year. The ERP systems are also being transformed from the desktop model to a mobile technology. Integration can be said as one of the top aspects for any of the big tech trends such as the mobile, digital, cloud and internet business. With the increasing requirement for doing more with the less, the enterprises would be keen for adopting agile approaches for integration. That would also make the business agility together with the flexibility for using a variety of products and/or services, the scalability for keeping pace with the business volume as well as the efficiency for keeping the costs at a minimum.

Advances in the graphical and intuitive modeling also mean a big role for the data visualization in the business intelligence. As the self-service analytics are becoming much more mainstream, the tasks such as forecasting and/or prediction would become common as well as a lot less painful. When it comes to the ERP systems, there is a mostly increasing demand for the software that are complementing the each other. Many companies are creating a big puzzle with the items of software which suit their requirements perfectly. More importantly, choice of the software becomes crucial and the major software developers are reacting quickly into the current trends as well as providing a quick interaction and a big variety of their services and/or products. That means the puzzle will get larger year by year.

REFERENCES

Bernstein, P. and Haas, L. Information integration in the enterprise, 2008, Pages 12-25

Blili and Raymond, Information Technology: Threats and Opportunities for Small and Medium-Sized Enterprises, International Journal of Information Management, 1993, Pages 9-18

Boslaugh S., Statistics in a Nutshell, Second Edition, O'Reilly Media Inc., 2012, Pages 11-16

Brandt, S., Data Analysis: Statistical and computational methods for scientists and engineers - 3rd ed., Springer- Verlag, 1999, Pages 13-28

Bruce Peter C., Nitin R. Patel and Galit Shmueli, John Wiley - Data Mining for Business Intelligence, 2010, Pages 12-32

Burnett K., The handbook of key customer relationship management: the definitive guide to winning, managing and developing key account business, Financial Times, 2001, Pages 12-26

Carlberg C., Decision Analytics: Microsoft Excel, Que, 2013, Pages 11-34

Collica Randy S., Customer Segmentation and Clustering Using SAS Enterprise Miner, Second Edition, SAS Institute, 2011, Pages 15-62

Daroczi G., Mastering Data Analysis with R, Packt Publishing, 2015, Pages 15-38

Dresner, H., A Short History of Business Intelligence and Where It's Headed, The Insights - Oracle Hyperion, 2007, Pages 13-60

Duda R.O, Hart P.E, Pattern Classification, Wiley Publications, 2001, Pages 14-18

Dyche J., The CRM Handbook: A business Guide to Customer Relationships Management, (1st edition), Addison-Wesley Professional, 2002, Pages 17-21

Foley, J. and Bates, T., Data Warehousing & Business Intelligence for the SMB, IT toolbox white papers, 2007, Pages 9-17

Gartner Research, Findings: Social Network Analysis Is coming into the Limelight, 2008, Pages 21-35

Greenberg P., CRM at the Speed of Light: Social CRM 2.0 Strategies, Tools, and Techniques for Engaging Your Customers, (4th edition), McGraw-Hill Osborne Media, 2009, Pages 14-28

Habul A., Amila Pilav-Velic, Customer Relationship Management and Business Intelligence, Intech, 2012, Pages 24-66

Harris C., Analysis and Enterprise, Information Builders, 2012, Pages 12-25

Holland PC, Naude P, The metamorphosis of marketing into an information-handling problem, The Journal of Business & Industrial Marketing, 2004, Pages 20-27

Hillier, F., S., Lieberman, G., J., Introduction to Operations Research – International Editions, McGraw-Hill, Inc., 1995, Pages 11-24

Koch I, Analysis of Multivariate and High-Dimensional Data, Cambridge University Press, 2013, Pages 14-19

Levin N, Zahavi J, Continuous predictive modeling - A comparative analysis. Journal of Direct Marketing, 1999, Pages 35-44

Makhabel B., Learning Data Mining with R, Packt Publishing, 2015, Pages 23-29

Peppers D., & Rogers M., Managing Customer Relationship: a strategic framework, (1st edition), John Wiley & Sons Inc., 2004 Pages 33-38

Prahalad C.K., Ramaswamy P.B., Katzenbach J.R., Lederer Ch., & Hill S., Harvard Business Review on Customer Relationship Management, (1st edition), Harvard Business Press, 2002, Pages 20-31

ReibStein D., Philip E. Pfeifer, Neil Bendle, Ronald T. Wilcox, Paul Farris, Rajkumar Venkatesan, Cesar Brea, The Definite Guide to Marketing Analytics and Metrics, FT Press, 2014, Pages 14-76

Sloan Joseph D., High Performance Linux Clusters, O'Reilly Media Inc., 2004, Pages 18-46

Turban E., Sharda R., Delen D., King D., Business Intelligence, Second Edition, 2010, Pages 22-54

Vercellis C., Business Intelligence: Data Mining and Optimization for Decision Making, First Edition, 2009, Pages 13-52

Zadeh, L., A., Fuzzy sets and their applications to cognitive and decision processes, Academic Press, Inc., 1974, Pages 10-29

APPENDIX

A1. QUESTIONNAIRE

Company name: **Year of foundation:**

Sector: **Company Type: LTD ☐ AŞ ☐**

Brands / Trademarks:

Number of workers: **Number of white-collar personnel:**

Number of personnel who is graduated or under-graduated:

Please make only one choice for the multiple choice questions hereunder.

For all questions without the likert chart ones, please make only one choice.

For the likert chart questions, please note a point from 1 to 4/5/7 that is ranked from the weakest / least to the strongest / highest.

1- What is the definition of BI for you?

A) Tools for managing the information B) a warehouse of data
C) Analytical applications D) a new way to do business E) Another

2- Is there any defined BI strategy in your organization?

A) There is a comprehensive BI strategy B) There is a partly defined BI strategy
C) BI is very new for our company so we don't have a defined strategy yet
D) None **(please pass directly to question No. 17)**

3- For which purpose do you use BI? (Please rank from 1 to 4)

(…) Reporting (…) Analyzing (…) Alerting (…) Predictive modeling or optimization

4- From which systems your company get the datas that are used for the BI management of your company? (You can make multiple choices)

☐ ERP **☐ CRM** **☐ SCM**

1: VERY LOW 2: LOW 3: MEDIUM 4: GOOD 5: VERY GOOD

5- Do you see your company as skilled enough to take the advantage of the BI systems?

1 o 2 o 3 o 4 o 5 o

6- How good is your data governance for BI?

1 o 2 o 3 o 4 o 5 o

7- How is the user adoption in your company for BI?

1 o 2 o 3 o 4 o 5 o

8- How proactive approaches does your company have in BI?

1 o 2 o 3 o 4 o 5 o

9- How significant benefits do you have from the BI activities?

1 o 2 o 3 o 4 o 5 o

10- Which analyses you use BI for? (Please rank from 1 to 7)

(…) to analyze customers (…) to analyze suppliers (…) to analyze competitors or other shareholders (…) to analyze advertising costs and successes (…) to analyze performance of the personnel and company(…) to analyze annual turnovers of the company (…) Other

11- Could you indicate some benefits you earned via usage of BI? (Please rank from 1 to 7)

(…) to acquire new customers / suppliers (…) to launch new products (…) to have fraud detection (…) to have more detailed market researches (…) to improve the business process (…) Costs saving would be possible I) Other

12- Who utilizes the outcomes of BI in your company?

A) the top management B) the middle management and the specialists C) People in sales and/or development organization D) a communication unit E) Other

13- What kind of BI tools do you use? (Please rank from 1 to 5)

(…) Query tool (…) Reporting tool (…) OLAP (Online Analytical Processing) tool (…) Data Mining tool (…) Other

14- According to you, could you name one factor in order to describe the most significant benefit of the BI activities?

A) It would be possible to harmonize the ways of thinking for the company personnel B) It would be possible to broad the understanding business in general C) It would be possible to strength strategic planning D) It would be possible to increase professionalism in acquisition and to have analysis to understand the meaning of information

15- What is the most important reason of the lack of a BI organization in your company before? (If you have a BI organization in your company, please skip to the next question)

A) BI is expensive B) We don't have enough skilled representatives to apply C) We couldn't succeed to apply any ERP module yet D) We think that we don't need BI E) Another

16- How do you expect the investment on BI to increase in the upcoming 5 years?

1 o 2 o 3 o 4 o 5 o

17- What is the most important change that you expect for BI in the upcoming 5 years?

A) BI operations are going to increase in importance and become more common in companies B) BI is going to be an activity for the large corporations C) BI is going to be a need-based practice D) BI is mainly going to be a technology which is only considered as a supporting and enabling tool E) Other

18- Considering the huge variety of BI solutions and techniques, which one do you think the most important for today's world?

A) Scorecarding, dashboarding and information visualization B) Business performance management C) Data mining D) Decision support systems E) Other