TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Economics

Business Intelligence Tools for Supporting Modern Corporations

Master Thesis

Study programme: N0413A050030 International Management
Author: **Bc. Petr Havel**
Thesis Supervisors: Ing. Athanasios Podaras, Ph.D.
Departmens of Informatics

Liberec 2022

TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Economics

Master Thesis Assignment Form

Business Intelligence Tools for Supporting Modern Corporations

Name and surname:	**Bc. Petr Havel**
Identification number:	E20000319
Study programme:	N0413A050030 International Management
Assigning department:	Departmens of Informatics
Academic year:	**2021/2022**

Rules for Elaboration:

1. Goal, objectives, role of business intelligence in achieving strategic business targets.
2. Business intelligence for improving business operations and definition of a business case.
3. Design, analysis and implementation of the proposed solution using selected data.
4. Conclusions and implication for further improvements.

Scope of Graphic Work:	
Scope of Report:	65 normostran
Thesis Form:	printed/electronic
Thesis Language:	English

List of Specialised Literature:

- KOHTAMÄKI, Marko, 2017. *Real-Time Strategy and Business Intelligence: Digitizing Practices and Systems*. California: Palgrave Macmillian. ISBN:978-3-319-54845-6
- MARCEL, Patrick and Esteban ZIMÁNYI, 2017. *Business intelligence*. Cham: Springer International Publishing. ISBN: 978-3-319-61163-1.
- VAISMAN, Alejandro, 2014. *Data Warehouse Systems: Design and Implementation*. Berlin Heidelberg: Springer-Verlag. ISBN 978-3-642-54654-9.
- PROQUEST, 2021. *Databáze článků ProQuest [online]*. Ann Arbor, MI, USA: ProQuest. [cit. 2021-09-26]. Dostupné z: http://knihovna.tul.cz

Consultant: Ing. Pavel Doležal, Technical Leader, G- housing, s.r.o.

Thesis Supervisors:	Ing. Athanasios Podaras, Ph.D. Departmens of Informatics
Date of Thesis Assignment:	November 1, 2021
Date of Thesis Submission:	August 31, 2023

L.S.

doc. Ing. Aleš Kocourek, Ph.D.
Dean

Ing. Petr Weinlich, Ph.D.
Head of Department

Liberec November 1, 2021

Declaration

I hereby certify, I, myself, have written my master thesis as an original and primary work using the literature listed below and consulting it with my thesis supervisor and my thesis counsellor.

I acknowledge that my master thesis is fully governed by Act No. 121/2000 Coll., the Copyright Act, in particular Article 60 – School Work.

I acknowledge that the Technical University of Liberec does not infringe my copyrights by using my master thesis for internal purposes of the Technical University of Liberec.

I am aware of my obligation to inform the Technical University of Liberec on having used or granted license to use the results of my master thesis; in such a case the Technical University of Liberec may require reimbursement of the costs incurred for creating the result up to their actual amount.

At the same time, I honestly declare that the text of the printed version of my master thesis is identical with the text of the electronic version uploaded into the IS/STAG.

I acknowledge that the Technical University of Liberec will make my master thesis public in accordance with paragraph 47b of Act No. 111/1998 Coll., on Higher Education Institutions and on Amendment to Other Acts (the Higher Education Act), as amended.

I am aware of the consequences which may under the Higher Education Act result from a breach of this declaration.

June 15, 2022 Bc. Petr Havel

Název práce česky

Anotace

Tato práce se zabývá významem businessové inteligence v moderních podnicích. Nejprve budou popsány konkrétní technologie a nástroje patřící do businessové inteligence. Prezentován bude i jejich význam pro celkovou image a jeho pomoc pro lepší komunikaci mezi jednotlivými odděleními společnosti. Pro vyhodnocení bude vytvořena anketa, která bude zaměřena na lidi, kteří mají něco společného s moderními podniky a bude jim položena otázka, jak dalece rozumí nástrojům businessové inteligence a jejich významu pro moderní podniky samotné. Na základě tohoto průzkumu lze zjistit mezeru ve znalostech v moderních podnicích a do jaké míry ovlivňuje firmy a jejich funkčnost. Na konec budou na základě analýzy průzkumu předloženy možná řešení, silné a slabé stránky.

Klíčová slova

businessová inteligence, databáze, data mining, datový sklad, moderní podniky

Title of the Thesis in English

Annotation

The present thesis deals with the importance of business intelligence in modern enterprises. At first specific technologies and tools in Business Intelligence will be described. Presented will also be their importance for the overall image and its help for better communication among each department of the company. For the evaluation, a survey will be created which will be aimed at people who have something in common with modern enterprises and they will be asked how far they understand Business Intelligence tools and their importance for the modern enterprises themselves. Based on this survey might be found a gap of knowledge in modern enterprises and how far it affects the enterprises and their functionality. Possible solutions, strengths and limitations based on the analysis of the survey will come up in the end.

Key Words

business intelligence, databases, data mining, data warehouse, modern enterprises

Content

List of Figures

List of Abbreviations

BI	Business Intelligence
CRM	Customer Relationship Management
CX	Customer Experience
DMS	Data Mining Suites
DOLAP	Desktop On-Line Analytical Processing
DSS	Decision Support System
EIS	Executive Information Systems
GIS	Geographic Information Systems
HOLAP	Hybrid On-Line Analytical Processing
ID	Identity Document
IT	Information Technology
KM	Knowledge Management
MLE	Maximum Likelihood Estimation
MOLAP	Mobile On-Line Analytical Processing
MOLAP	Multidimensional On-Line Analytical Processing
OLAP	On-Line Analytical Processing
ROI	Return on Investment
ROLAP	Relational On-Line Analytical Processing
SOLAP	Spatial On-Line Analytical Processing
WOLAP	Web On-Line Analytical Processing
XML	Extensible Markup Language

Introduction

The main focus of this diploma thesis is to explain the true importance of business intelligence in modern enterprises, the tools of business intelligence and why it should not be left out. Technologies of business intelligence are becoming more and more popular, and they are the future of enterprise systems. Even though it is part and parcel of almost every company, many people are still underestimating its true potential and importance.

The majority still mostly deals with how to present the company in the right way, how to catch more customers, meet their desires and whole marketing in general, but business intelligence is still in some sort of way out of the debate.

In the second part, this assumption will be either confirmed or rejected thanks to the questionnaire spread among people who have some overall knowledge about modern enterprises. They either worked in them or at least studied the modern enterprise's problematics. Results are going to be analysed and evaluated with some final summarization.

In the last part of the diploma thesis, there is going to be a conclusion of collected results and some recommendations for the following research and business intelligence professionals.

Goal and Structure

The thesis is divided into two parts. The first part is the theoretical part. The primary objective of the theoretical part is to delineate key business intelligence concepts, indicate the value of business intelligence tools and technologies, describe business intelligence architecture and refer to the most important elements of a business intelligence solution based on background literature. Moreover, the reference to the utilised tools and methods is included.

The objective of the practical part is the design and implementation of a research procedure in which primary data has been collected via a questionnaire which was prepared by the author, the collection and dataset formulation for facilitating data analysis and, the use of statistical and data mining approaches to interpret, analyse and present the inferred results related to the value of bi from to the participants' viewpoint. This can be seen in the research flow design in Figure number 1. At the end of the thesis, there will be an evaluation and recommendations for future research. The whole thesis is going to end with a conclusion.

Objectives

The most important assumption is that people who do not have that much experience and overall knowledge in BI do not find it that crucial and cannot fully understand its importance for modern enterprises and only people who have greater experience in this sector find it part and parcel of every today´s enterprise. Thanks to the survey it will be possible to evaluate this assumption.

Research Flow Design

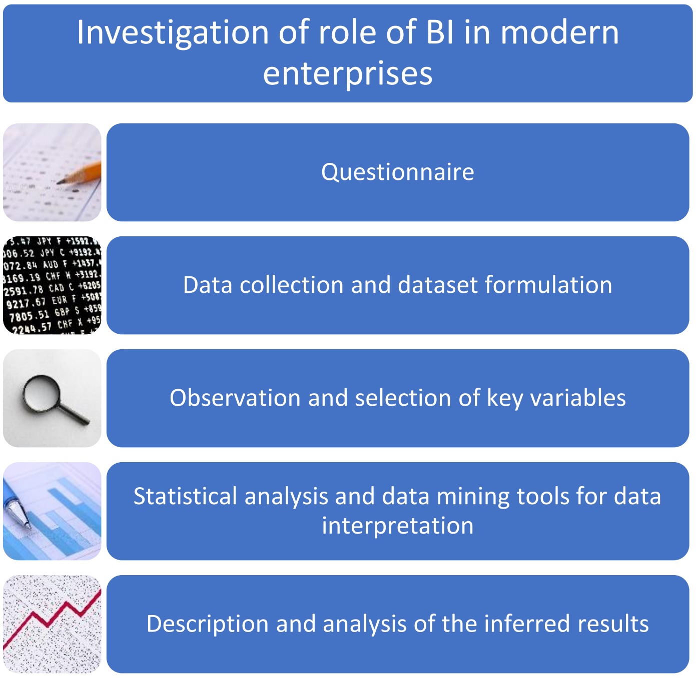

Figure 1: Research Flow Diagram
Source: Own contribution

As can be seen in Figure 1, at first the questionnaire is going to be spread among the target audience with specific preferrations. As soon as we get enough answers, data are going to be put in the preferred form. With the right form of data, they are going to be observed and key variables are going to be picked up. With the key variables, we will be able to analyse key aspects with statistical evaluation. In the end, the outcome of the analysis will be presented with the graphs.

1. Business Intelligence

Business intelligence (BI) is a data-driven decision-support system that integrates data collection, storage, and knowledge management with analysis to deliver information to decision-makers. The phrase was coined in 1989, but many of its qualities were already present in executive information systems at the time. Business intelligence focuses on analysing vast amounts of data regarding a company's operations. As a subset, it includes competitive intelligence. In computer-based systems, business intelligence uses a big database as its source of data and as the foundation for complex analysis, which is often housed in a data warehouse or data mart. Simple reporting to slice-and-dice, drill down, responding to ad hoc inquiries, real-time analysis, and forecasting are all examples of analyses. A vast number of companies offer analytical software. The dashboard is maybe the most useful of them. Business performance measurement (BPM), business activity monitoring (BAM), and the growth of BI from a staff tool to being used by people across the organization are all recent advancements in BI (BI for the masses). In the long run, business intelligence methodologies and insights will be ingrained in corporate processes (NEGASH, Solomon a GRAY, Paul, 2008).

Business intelligence is gaining traction among businesses as a way to cut costs, increase service quality, and improve decision-making processes. However, despite the fact that BI has been there for a long time, it is still struggling to reach its full potential, according to experts in the area. Disparities in how business intelligence constructs are defined and understood should be investigated because they may obstruct corporate executives' and researchers' comprehension of what BI represents (FOLEY, Éric a GUILLEMETTE, G., Manon, 2010).

In today's competitive world, businesses face significant challenges. Information has unquestionably become a crucial source of competitive advantage in today's corporate world. The primary goal of business intelligence is to aid managers in their decision-making. Simply said, greater data and information are required by managers in order to make better judgments. Managers can use business intelligence to make educated and sensible decisions about their company's operations. Informed decisions result in better, more efficient procedures in the workplace, resulting in a significant competitive advantage. Company intelligence is a crucial factor that both business and IT managers should be aware of and exploit to get a competitive advantage (FOLEY, Éric a GUILLEMETTE, G., Manon, 2010).

One of CIOs' top ten goals for the next five years has been identified as business intelligence. Even though the term "business intelligence" has been around for about two decades, corporations have just recently grown increasingly interested in learning more about it. Executive information systems (EIS) were first introduced in the early 1980s to aid upper-level managers and executives in their decision-making. Reporting and analysing capabilities have progressed from static to dynamic multidimensional reporting systems, trend analysis, drill-down capabilities, and artificial intelligence analysis since then. Many BI solutions now provide these elements to help employees make better decisions (FOLEY, Éric a GUILLEMETTE, G., Manon, 2010).

This diversity of BI technology, as well as the number of developments and concepts associated with the business intelligence concept, pose significant obstacles in defining the new concept. This position presents special challenges for managers, such as formulating a clear definition of BI, reaching a consensus on BI-related business rules, establishing quality requirements that define success, and more broadly managing people and resources. Indeed, it has been found that academics and managers alike struggle to define and understand the extent of business intelligence. When discussing BI, terms like competitive market intelligence, strategic intelligence, data warehouse, business performance management, and data mining are frequently employed. These terms are sometimes used interchangeably with BI (FOLEY, Éric a GUILLEMETTE, G., Manon, 2010).

Furthermore, it has been noticed that BI has been defined in a variety of ways by academics in the field, each with a different focus that best suited their research. Various stakeholders, including consultancies, software suppliers, practitioners, and the scientific community, have used the phrase "business intelligence" to define methods and systems committed to the methodical and purposeful examination of an organization's competitive environment.

Business intelligence collects data from many sources which are used for making a whole, final picture for the business intelligence professionals. It is a part and parcel of every modern enterprise. Every employee should pay big attention to its importance and try to use its full potential for the company. Business intelligence is created from many tools which are going to be described in the following paragraphs. Almost every modern enterprise uses databases. Databases are a key element for every company, and they are part and parcel of business intelligence as well. Business intelligence is basically every intelligent software behind the companies, therefore it should stay in focus of everyone who runs a modern enterprise.

1.1 Business Intelligence for Improving Business Operations

In the twenty-first century, business intelligence is one of the hottest subjects. BI solutions have become the go-to tool for businesses in the digital age because they take into account real-time data, analyse it, and then provide value-backed judgments. However, other businesses are hesitant, questioning if business intelligence is worth the price tag (MURIYIL, Anilkumar, Thekkan 2021).

1.1.1 True Importance of Business Intelligence

Entrepreneurs have relied on gut instinct or intuition to make business decisions for a long time. While this helped companies weather the downturn, the new-age firm demands a more sophisticated approach for evaluating data and becoming data-savvy. This is where business intelligence (BI) solutions come into play. Business intelligence is the process of combining data from one or more sources, analysing, and processing it, and then converting it into useful information. Simply said, the major goal of BI tools and comparable apps is to assist businesses in making better real-time decisions. Furthermore, it aids in the reduction of operational costs while also increasing efficiency (MURIYIL, Anilkumar, Thekkan 2021).

1.1.2 How Business Intelligence Helps Business Operations

Improves the decision-making

The business world has shifted from focusing on products to focusing on customers. Meaning, that businesses that are successful in responding to client inquiries gain trust and are more likely to develop in the present and future. To do so, we must first determine what customers expect from our company and then provide them with just that (MURIYIL, Anilkumar, Thekkan 2021).

With the introduction of BI technologies, businesses can now collect data from a variety of sources, analyse it, and use it to make business choices. The beauty of this is that businesses can now use their inept, raw, and unstructured data to their advantage (MURIYIL, Anilkumar, Thekkan 2021).

Furthermore, these technologies are computer-assisted and can function without the need for human participation. As a result, enterprises are able to reduce the overall cost of acquiring data and analysing it in order to gain meaning from it (MURIYIL, Anilkumar, Thekkan 2021).

Learn everything there is to know about our customers

Customer experience, or so-called CX, has become a critical marketing metric. In reality, marketers increasingly compete primarily on their company's customer experience score. The greater the number, the better (MURIYIL, Anilkumar, Thekkan 2021).

This is most likely due to a dynamic shift in how customers perceive company entities. When it comes to choosing a company, they no longer emphasise items. Customers nowadays are looking for convenience and comfort. Marketers can simply measure and manage customer interactions with BI technologies (MURIYIL, Anilkumar, Thekkan 2021).

This gives them a complete picture of the customer's journey and allows them to fine-tune marketing efforts to convert them into buyers. In other words, business intelligence technologies enable marketers to gain a better understanding of their customers and take steps to improve customer satisfaction (MURIYIL, Anilkumar, Thekkan 2021).

The efficiency of customer services

Following decision-making is the principle of providing excellent service to our consumers on a daily basis. Given the fierce rivalry that exists in today's corporate world, success is defined by providing clients with the best-in-industry service rather than perfect products. The key is in the level of service offered and the level of satisfaction acquired as a result of it. It's fascinating to see how business intelligence systems can document all of a customer's data, find insights, and track their purchasing habits in order to promote corporate growth. Using the information above, business professionals may simply develop and implement plans to reach out to potential clients and close deals (MURIYIL, Anilkumar, Thekkan 2021).

Productivity improvement

Business intelligence tools and/or software are created to function without the need for human intervention. That is, the procedure is mostly automated. This guarantees that all business processes run smoothly, resulting in increased end-to-end productivity (MURIYIL, Anilkumar, Thekkan 2021).

Furthermore, the company can track data at any time, removing bottlenecks and streamlining corporate processes. The dashboard function allows the concerned expert to quickly and easily see and analyse data. Artificial Intelligence tools have the additional benefit of being cloud-hosted. This makes it easier for the entire team to access the information without difficulty. All of this adds up to a more refined approach that boosts end-to-end corporate productivity (MURIYIL, Anilkumar, Thekkan 2021).

Optimisation of return on investment

Finally, BI systems allow for accurate measurement of business ROI and further optimisation for the better. The return on investment is the amount of money a company makes through marketing campaigns and/or events. Business intelligence systems provide detailed information about the sales process, making strategic decisions easier for marketing professionals. The tool also generates reports based on the information. Businesses can ta use of the aforementioned to increase sales efficiency. Furthermore, it boosts conversion rates, resulting in a higher overall return on investment (MURIYIL, Anilkumar, Thekkan 2021).

1.2 Business Case

There is a why at the start of any project. Why should the company undertake a new project? What are the advantages of doing so? And are the expenses justified? These are the questions that a business case aims to address. It explains why the project is beneficial, as well as the costs, benefits, and risks associated with the new investment. The advantage could be monetary, but it could also be based on other factors like reduced workload or higher satisfaction. A business case relies on data to back up its claims, such as the predicted cost savings expressed as a percentage or a cash flow account. The project proposal can then be approved or rejected by the decision-makers (microTOOL, 2016).

The business case can be used not only to get a project approved but also as a source of continuous commercial rationale. This is how the PRINCE2 project management approach specifies it. The business case is updated at regular intervals, and it is determined whether the project is still worthwhile in its current stage based on this information. This ensures that stakeholders and sponsors are always up to date on the project's status (microTOOL, 2016).

1.2.1 How to Create a Business Case

A complete business case analysis should be performed before creating a business case. The research is "simply" the basis for the business case. This study determines why a new project or investment is required, as well as the predicted benefits and expenses. Internal financial information, case studies of similar projects, and industrial evaluations are examples of data sources that can be used to support a business case. Alternatives to the project, as well as the risks connected with it, are considered in a business case study. The analysis must then be turned into a business case. Simply put, this is merely a recommendation at the start (microTOOL, 2016).

1.2.2 Structure of Business Case

For business cases, there is no standard template. However, there are numerous recommendations; actual guidelines are only provided by PRINCE2. The business case, according to this project management style, is divided into the following sections:

- Summary of the report: The summary, like previous reports, represents the most significant elements of the business case: decision-makers will look at it right away. The summary is usually written at the end.
- Reasons: The reasons for completing a project differ depending on the type of project. If a new legislative provision, for example, has an impact on the firm, this is a good cause to start a new project. Alternatively, if we want to develop new software, it has to be explained why in this area of the business case.

- Optional business ventures: A project suggestion is always one potential solution to a problem. Alternatives are always available. Three possibilities must always be considered in the business case, according to PRINCE2:
    - Don't do anything
    - Make the bare minimum effort
    - Make a move.

This scenario is assessed based on quantitative and qualitative criteria. That is, what are the financial implications of these options? What other consequences will there be that can't be expressed in numbers? The possibilities are numerous. Doing the bare minimum and making a move are frequently subjected to a business case study.

- Benefits to be expected: The project's intended benefit must be quantifiable so that the outcomes may be compared when it is completed. Aside from that, it should be specified when we expect the benefit.
- Expected drawbacks: Fill in the negative impacts of the project as perceived by the stakeholders. It is important to note that negative impacts are not risks. They represent the unavoidable repercussions of a particular action. A risk, on the other hand, isn't always linked to a probability – it could happen, but it doesn't have to.
- Timeline: In this section, specify how long the project will endure and when the advantages will be realised. Aside from that, we should also state when the project will be completed. As a result, this section contains information that extends beyond the project's real length.
- Biggest dangers: Danger, in addition to costs and benefits, are a significant decision-making tool. As a result, the most significant risks and their likelihood of occurrence, as well as the countermeasures, must be detailed in the business case. A summary is sufficient here; additional details about threats are provided in other publications.
- Costs: All of the project's costs are listed in this section of the business case. It also takes into account the costs that the project will incur after it is completed, such as maintenance costs.
- Appraisal of an investment: We assess the project's benefits and predicted the negative consequences to the project's risks and expenses under investment. Hence, we apply techniques such as cash flow analysis and capital value estimation. The purpose is to be able to confirm the project's investment worth. The facts from the

business case's cost and benefits sections serve as the foundation for the computation (microTOOL, 2016).

1.3 OLAP

OLAP stands for On-Line Analytical Processing which provides multidimensional analysis of data for business problematics, complex calculations and data modelling. It is a technology used by decision-makers in business intelligence interfaces. Thanks to that data discovery is much easier, it also helps in “what if” scenarios by finding the best options (OLAP.com, N. D.).

1.3.1 How Does OLAP Work?

It extracts data from data sources which are cleaned and transformed. These data are uploaded to the OLAP server where it is pre-calculated for further analysis (TAYLOR, David, 2022).

1.3.2 Analytical Operations in OLAP

1. Roll-up

It either reduces dimensions or climbs up the concept hierarchy.

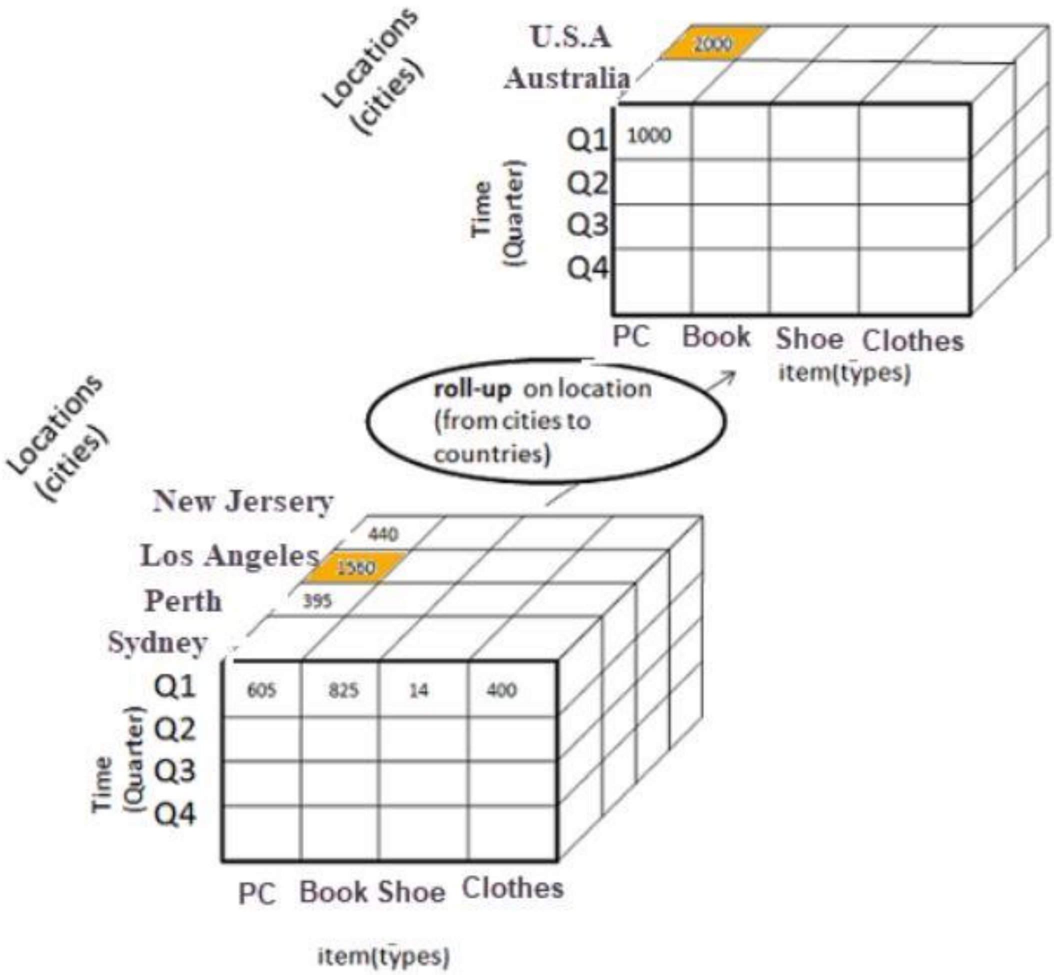

Figure 2: Roll-Up
Source: TAYLOR David, 2022

In Figure 2 we can see a roll-up from sub-dimension New Jersey, Los Angeles to the U.S.A and sub-dimension Perth and Sydney to Australia inside the cube (TAYLOR, David, 2022).

2. Drill-down

Drill-down function data is fragmented into smaller chunks, and it works in the opposite direction of the roll-up function, moving down the concept hierarchy and increasing the dimensions (TAYLOR, David, 2022).

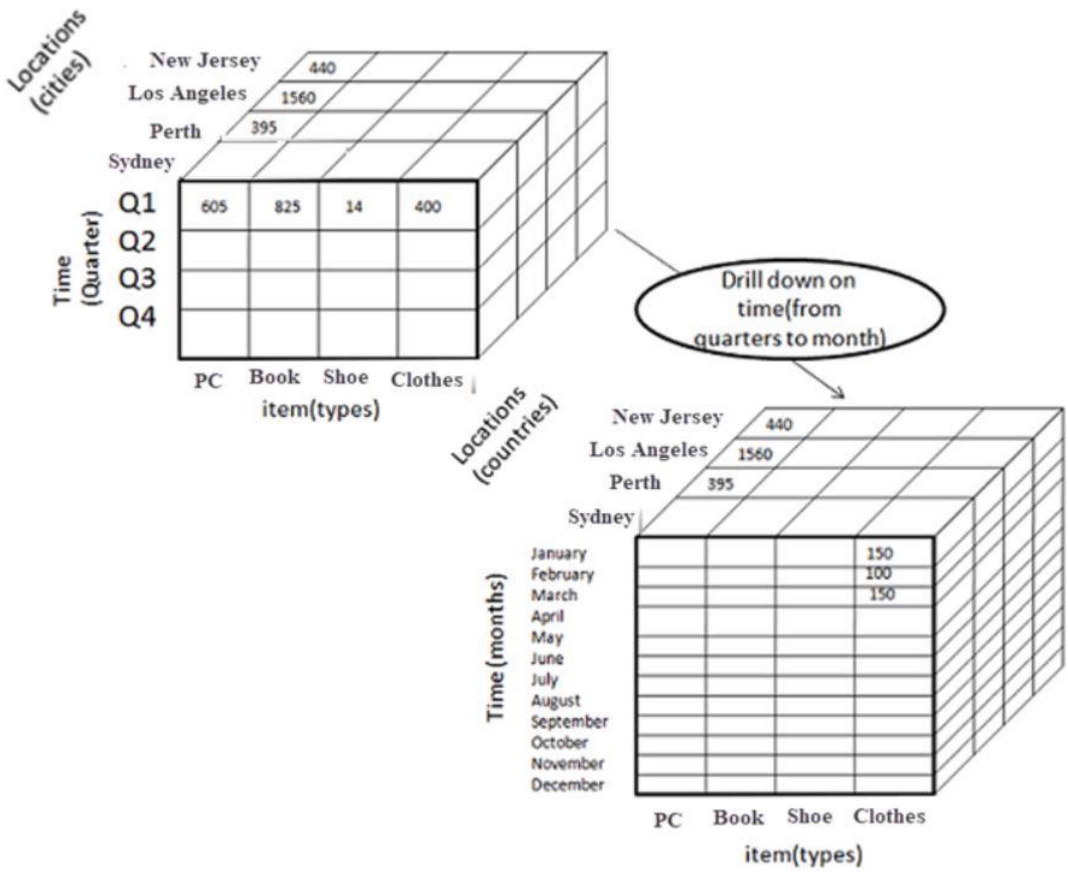

Figure 3: Drill-Down
Source: TAYLOR David, 2022

In Figure 3 we can see an example of the drill-down function by extending the quarters to specific months. Drill down function is basically just extending some type of dimension to more specific dimensions for a better overview.

3. Slice and dice

The slice function creates a sub-cube from one dimension of the previous cube. Function dice is very similar. The only difference is that dice creates a sub-cube from two dimensions.

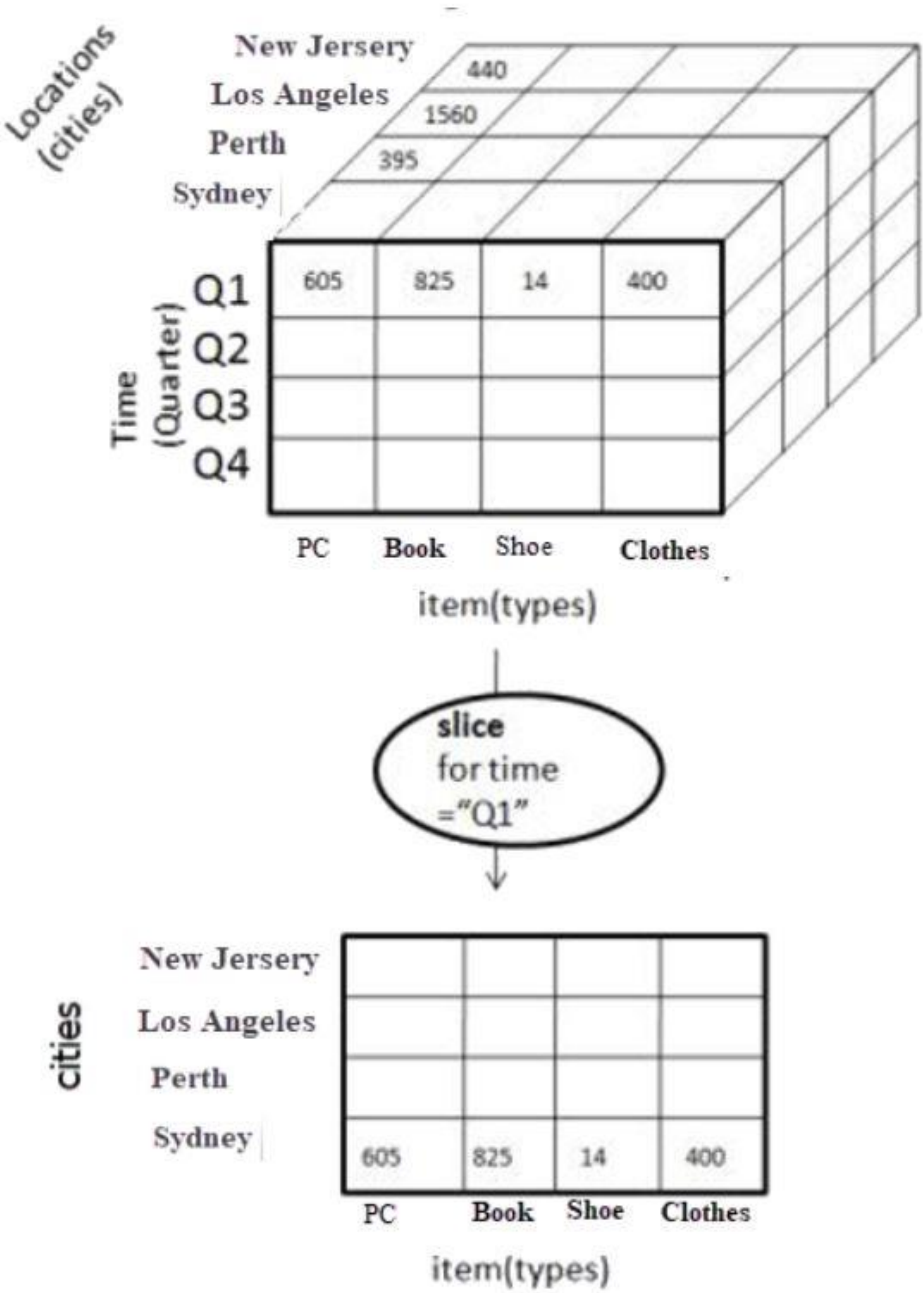

Figure 4: Slice
Source: TAYLOR David, 2022

Figure 4 is an example of a slice function where we take one specific dimension, in this example "cities" and we create a new dice out of it. This could be later on extended to another dice for example by dividing cities into city parts.

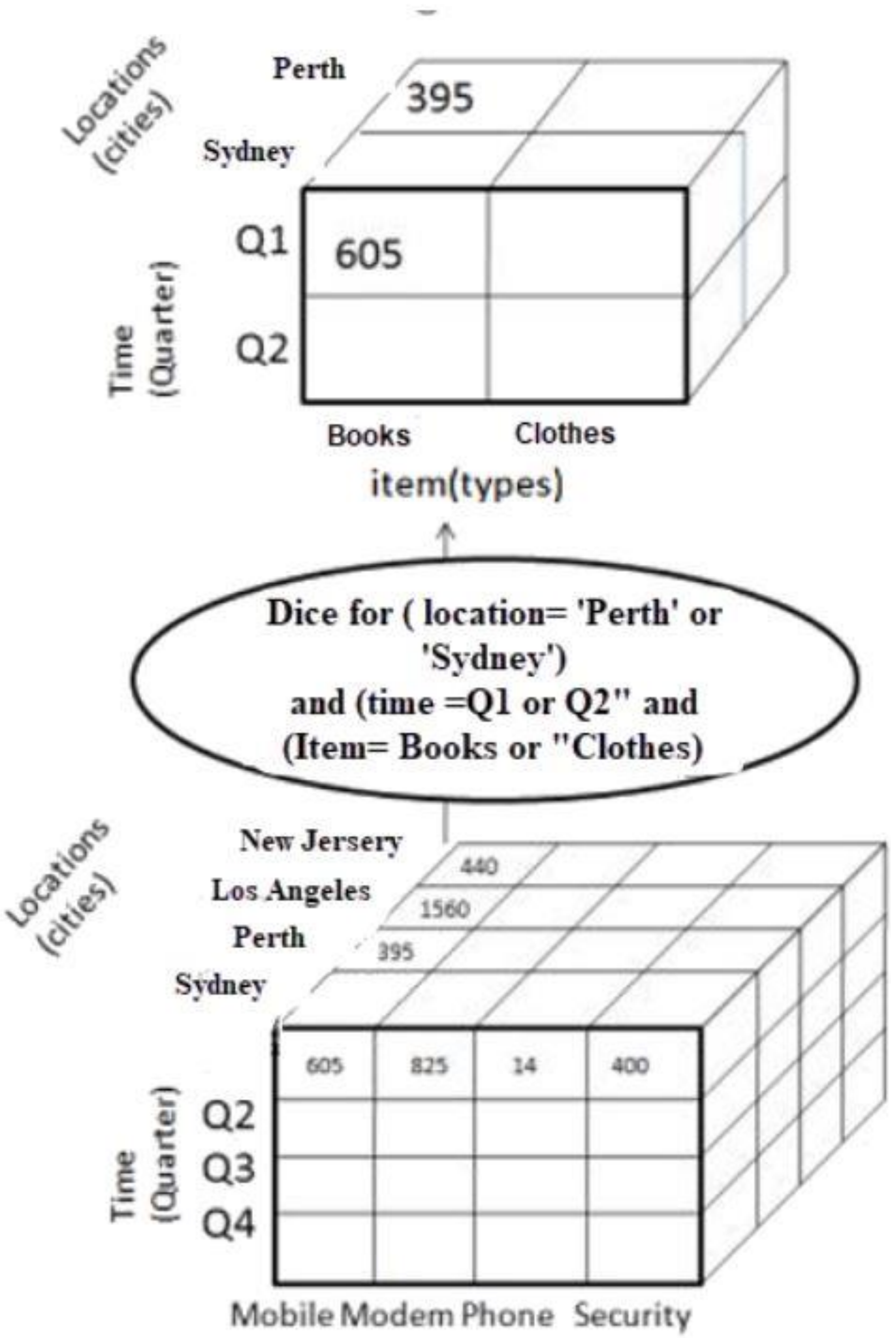

Figure 5: Dice
Source: TAYLOR David, 2022

The second example is for function dice (Figure 5) where we created new dice from dimensions, time and locations.

4. Pivot

The last function just rotates the data axes for a better overview of the data. For example, if we want to see data in a different form which would be easier to orient in for us. This example can be seen in Figure 6.

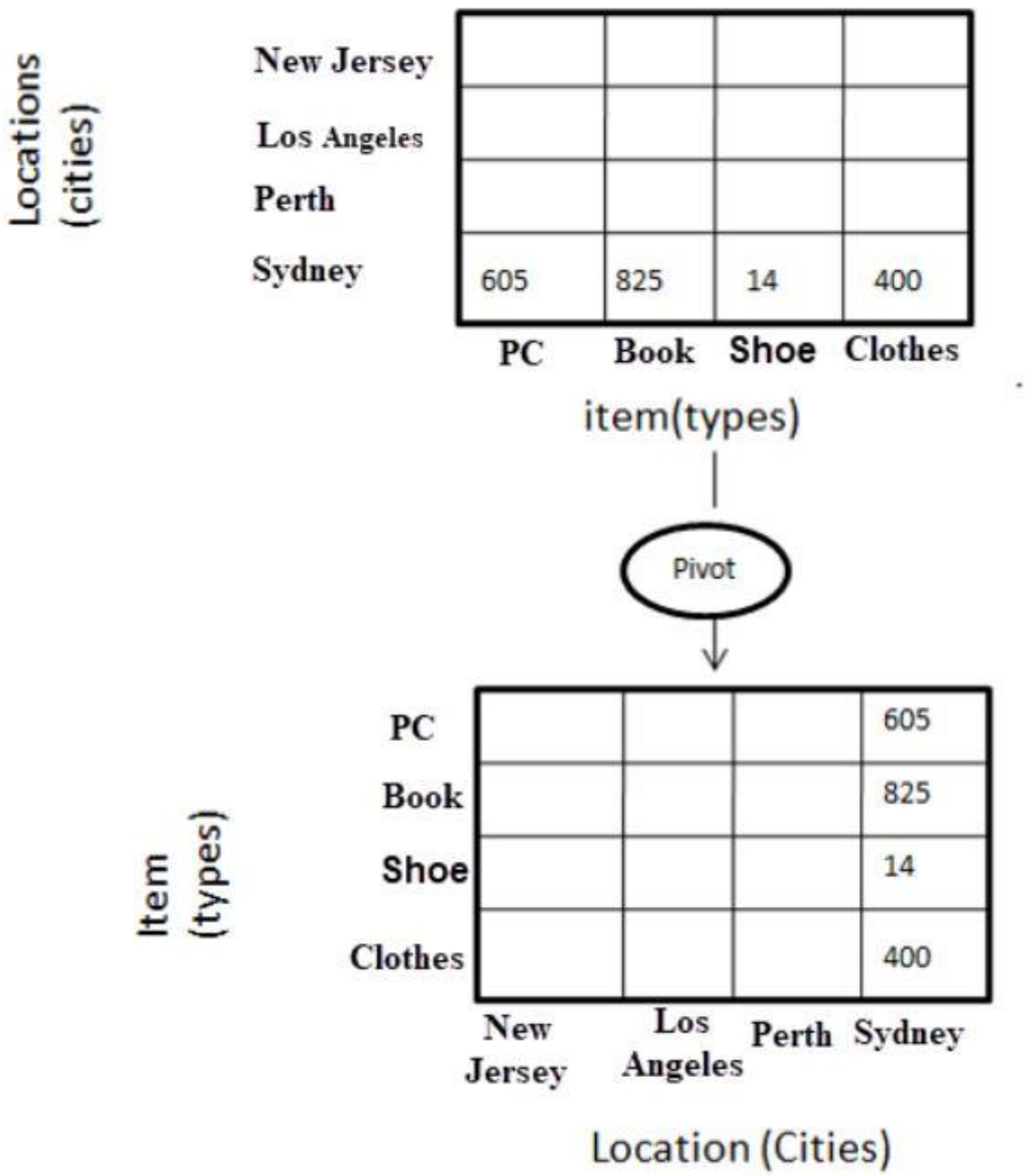

Figure 6: Pivot
Source: TAYLOR David, 2022

1.3.3 OLAP Systems

OLAP has many types of systems. Each of them provides different roles such as multidimensional data mapping, operation in multidimensional data implementation or accessing OLAP data by using a mobile phone (TAYLOR, David, 2022). The structure of OLAP systems is illustrated in Figure 7.

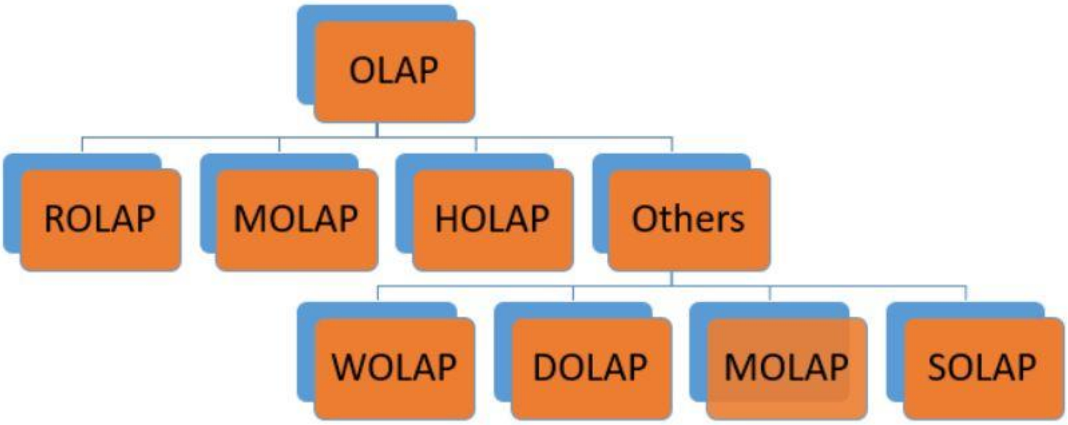

Figure 7: OLAP
Source: TAYLOR David, 2022

1.4 Data Warehouse

A Data warehouse is a process for collecting and managing data from sources to provide a better look into business insights. It is mostly used for the connection and analysis of business data from heterogeneous sources. It has a big impact on the strategic use of data. It is not designed for transaction processing but mostly for analysis and query. Its purpose is to transform data into information (TAYLOR, David, 2022).

1.5 How Does it Work?

Data comes from the transactional system or other relational databases to the data warehouse where it is transformed into information. The data might be structured, semi-structured or unstructured. The data is processed, transformed and ingested for better access for users. The

Data warehouse merges information from different sources into one comprehensive database. It also helps in the case of data mining (TAYLOR, David, 2022).

1.5.1 Types of Data Warehouse

1. Enterprise Data Warehouse

Mostly used in enterprises where it helps in cases of decision-making. It represents and organizes data.

2. Operational Data Store

Basically, it is only storing pure data when organization or transformation is not needed.

3. Data Mart

Mostly used in the case of finances, and sales. Data can be collected straight from the source. It is some sort of subset of a data warehouse (TAYLOR, David, 2022).

1.5.2 Stages of Data Warehouse

1. Offline Operational Database

This stage is copying the data system to another server

2. Offline Data Warehouse

Data are regularly updated from Operational Database then they are mapped and transformed

3. Real-time Data Warehouse

Whenever some transaction appears in the operational database, Data Warehouse is updated. A good example would be booking systems.

4. Integrated Data Warehouse

When a transaction is created in an operational system the data are updated continuously and then Data Warehouse creates a transaction which is sent back to the operational systém (TAYLOR, David, 2022).

1.5.3 Warehouse Data Components

1. Load manager

The front component is also known as the load manager. It handles all of the procedures related to data extraction and loading into the warehouse. These procedures include data transformations.

2. Warehouse manager

The warehouse manager is responsible for overseeing the warehouse's data administration. It conducts tasks such as data analysis to assure consistency, index and view development, denormalization and aggregate generation, source data transformation and merging, and data archiving and backing-up.

3. Query manager

The backend component is another name for the query manager. It handles all of the actions associated with the administration of user inquiries. Direct queries to the appropriate tables for query execution scheduling are the actions of this Data warehouse component.

4. End-user access tools

This is divided into five categories, such as

1. Data Reporting
2. Tools for queries
3. Tools for application development
4. EIS tools
5. Tools for data mining and OLAP (TAYLOR, David, 2022)

1.5.4 Who Can Use a Data Warehouse?

DWH can be good for every type of user:

- Those who make decisions based on a large amount of data
- Users who collect information from many data sources using customised, sophisticated techniques.
- It's also utilised by folks who want to access info with simple technologies.
- It's also necessary for individuals who want to take a methodical approach for making decisions.
- Data warehouse is useful if the user needs fast performance on a large amount of data, which is required for reports, grids, or charts.
- The first step is to create a data warehouse. If you're looking for 'hidden patterns in data flows and groups, you have come to the right place (TAYLOR, David, 2022).

1.5.5 What Are the Most Common Places for Usage of Data Warehouses?

1. Banks

It is commonly utilised in the banking industry to properly manage the resources available on the desk. Some banks might find it useful also for market research, product performance analysis, and operations.

2. Airlines

It is utilised in the airline system for operational purposes such as crew assignment, route profitability studies, frequent flyer program promotions, and so on.

3. Public sector

In the public sector, data warehouses are used to collect intelligence. It assists government entities in keeping and analysing tax records and healthcare records for each individual.

4. Telecommunication:

In this industry, a data warehouse is used for product marketing, sales decisions, and distribution decisions (TAYLOR, David, 2022).

1.5.6 Advantages and Disadvantages of Data Warehouse

Advantages

- Business users can access crucial data from a variety of sources all in one place thanks to a data warehouse.
- Data warehouses give consistent data on a variety of cross-functional tasks. Ad-hoc reporting and querying are also supported.
- To alleviate the load on the production system, a data warehouse might help to combine several sources of data.
- The utilisation of a data warehouse can speed up the analysis and reporting process.
- The user will find it easier to utilise for reporting and analysis after restructuring and integration.
- Users can access crucial data from multiple sources in a single location using a data warehouse. As a result, retrieving data from many sources takes less time for the user.
- A vast volume of historical data is stored in a data warehouse. This allows users to compare and contrast different historical periods and trends in order to forecast the future (TAYLOR, David, 2022).

Disadvantages

- For unstructured data, this isn't the best option.
- The creation and implementation of a data warehouse is unquestionably a time-consuming process.
- Data Warehouses can easily become obsolete.
- Changes to data types and ranges, data source schema, indexes, and searches are difficult.

- The data warehouse may appear simple, but it is really too sophisticated for most consumers.
- Users of warehouses may adopt their own business rules from time to time.
- Organizations must devote a significant amount of resources to training and implementation (TAYLOR, David, 2022).

1.6 Data Visualization

The practice of transforming raw data into a graphic format is known as data visualization. The data is represented with a specific goal in mind: to demonstrate logical unit correlations. Visualization can be done in a variety of formats depending on the sort of logical link and the data itself. So, to put it plainly, any analytical report will include instances of data interpretations, such as charts or comparison bars (AltexSoft, 2019).

Visuals are typically generated by hand using related software, such as Photoshop. However, its primary application is in the area of artificial intelligence. As a result, data visualization, often known as dataviz, has become a standard method of presenting information to consumers via the BI interface (AltexSoft, 2019).

1.6.1 Visualization of Data in Case of BI

The data goes through a lengthy procedure before it can be used to create visualizations. It can be summarized into 4 steps.

1. In the first step the data sources and data types that will be used should be defined. Then, techniques of conversion and database quality are established.
2. In the second step the data comes from its original storages, such as Google Analytics, ERP, CRM, or a supply chain management system.
3. The data is transported to a staging area where it is changed via API channels. Cleaning, mapping, and standardizing data to a consistent format are all part of the transformation process.
4. Clean data can also be moved into a storage system, such as a database. The original basic language of datasets can also be modified for tools being able to read the data.

While the visualization process isn't entirely automated, there's no need to produce the images by hand. All of the BI interfaces, in general, include templates that we may use. By putting up the necessary data characteristics, these can be updated and edited. In some circumstances, graphics can react to changes in data and display them by automatically changing graphs and tables. This is accomplished mostly through the use of data visualization libraries (AltexSoft, 2019).

1.6.2 Types of Data Visualization

Visuals are making everything more understandable. Thanks to them, data are not just a pure random numbers and letters, but it can show us everything important in the pictures. For a better understanding of the visuals, let's take a look at some examples which would create a better picture for the understanding of this problematic.

1. **Bar Charts**

One of the easiest ways of comparing data in BI are bar charts. One type of the bar chart can be seen in Figure 8, where we have scores of 4 teams compared in 2 periods.

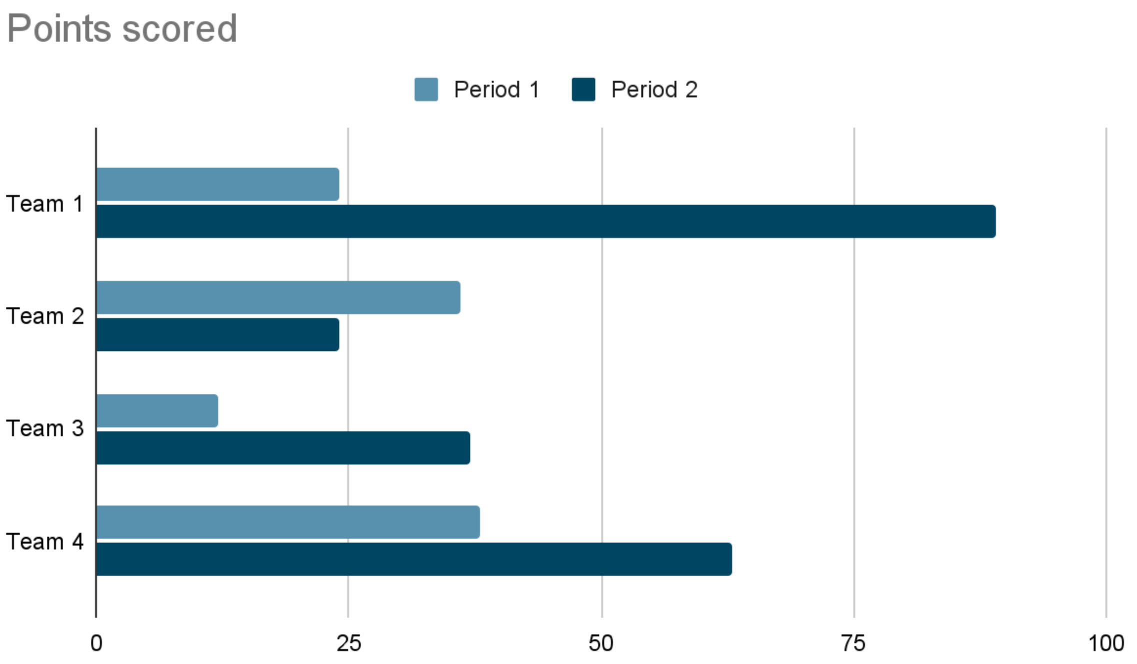

Figure 8: Bar Chart
Source: Own contribution

This type of chart can easily compare two variables by showing differences in bars. It is one of the most understandable visual tools. Bar charts are adaptable enough to display increasingly complicated data models. To illustrate distribution across market segments or subcategories of commodities, the bars might be stacked or clustered. The same is true for horizontal bar charts, which are better suited to long data labels on the bars (AltexSoft, 2019).

2. Pie Chart

Another very popular and probably most used one is a pie chart. This chart can be seen almost everywhere. It can easily compare units and objects thanks to that it is widely used across market and sales segments. A typical pie chart can be seen in Figure 9 where we can see an overview of scores for each of the 4 teams (AltexSoft, 2019).

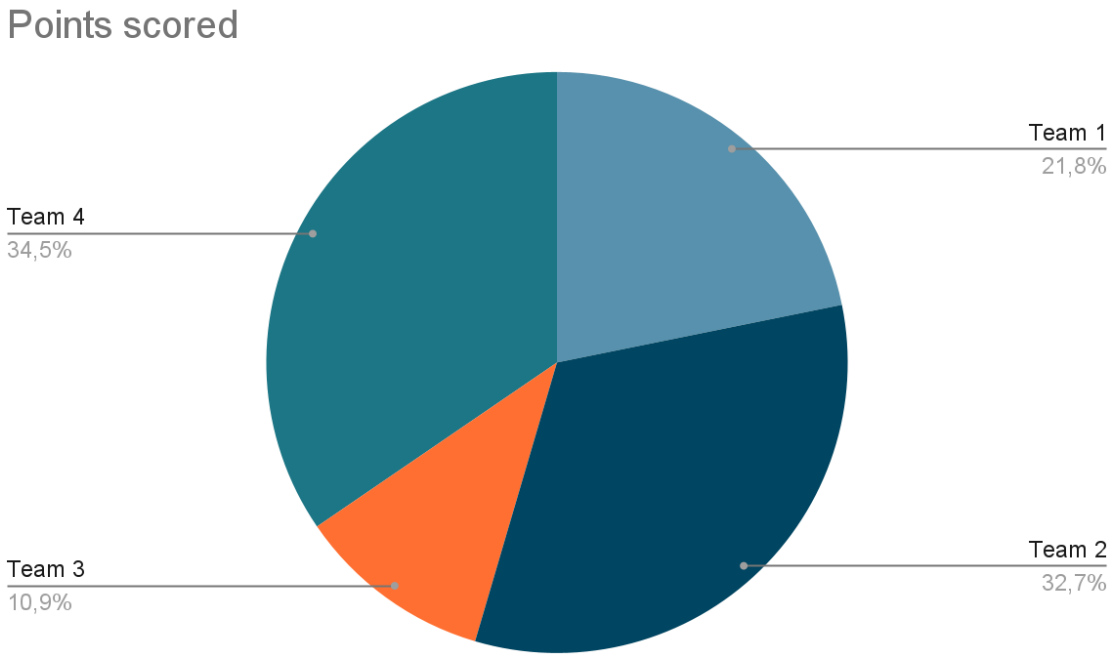

Figure 9: Pie Chart
Source: Own contribution

3. Line Graph

A line graph is commonly used in the case of finances. It can for example tell us how the value of some object has been changing over some period of time. A typical line graph can be seen in Figure 10 (AltexSoft, 2019).

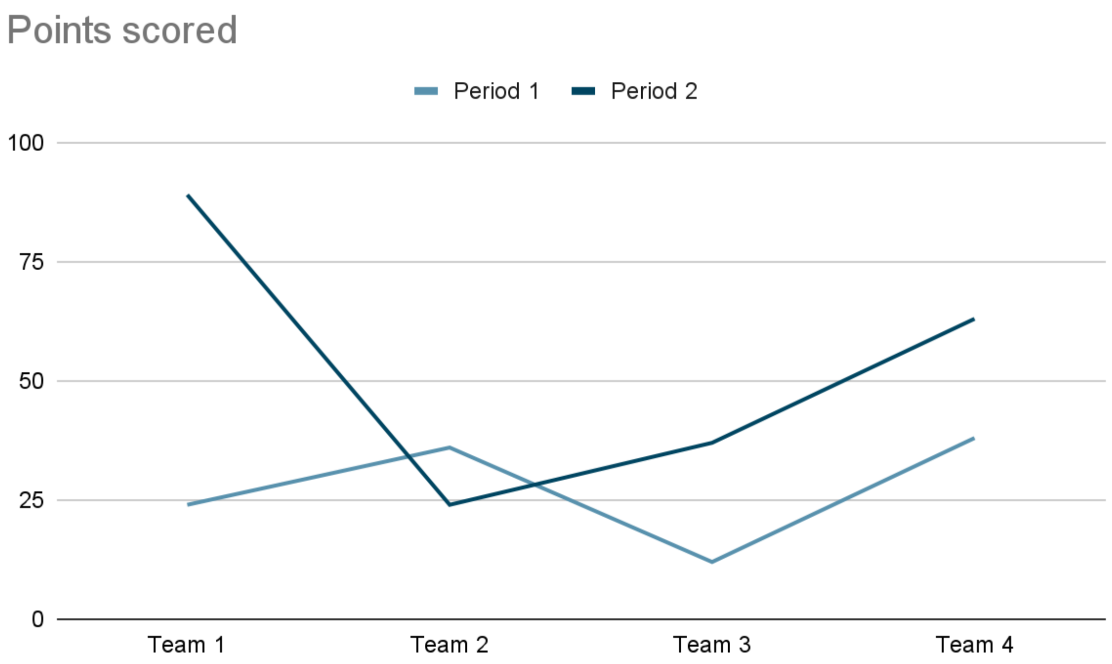

Figure 10: Line Graph
Source: Own contribution

4. Box Plot

This type of graph is a little bit more complicated and less understandable but unlike the others, it can show more sophisticated functions like the median between some quartiles. An example of a Box plot will present us Figure 11 (AltexSoft, 2019).

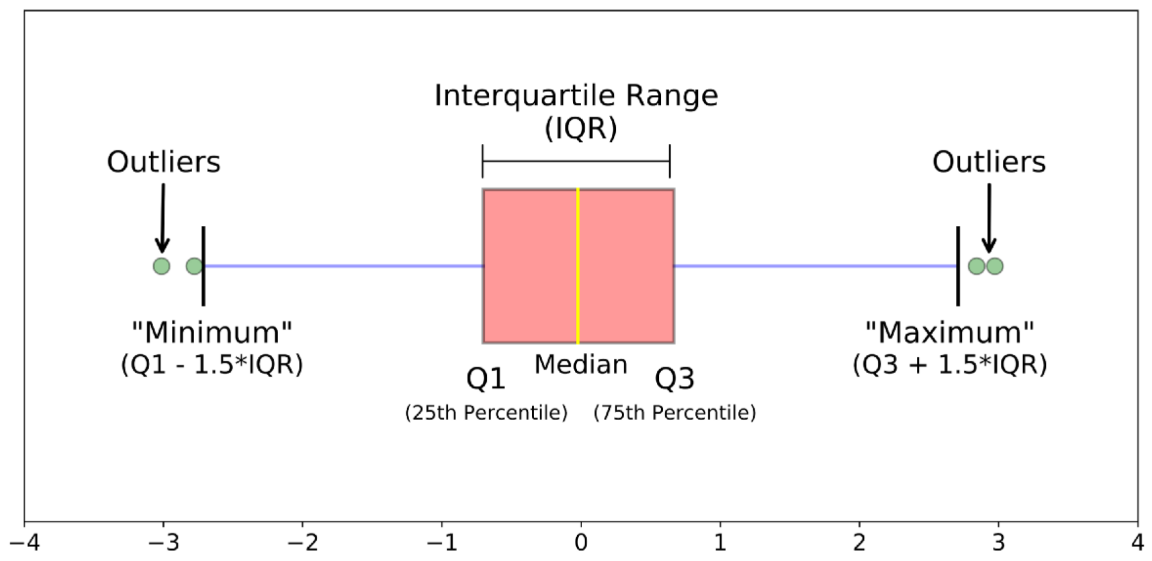

Figure 11: Box Plot
Source: GALARNYK, Michael, 2018

5. Density Map

Density/dot maps are created by dots situated over the map. It may show us the position of some specific unit. A dot might represent a single unit (for example, a marketplace) or a group of things in a certain area. Density is easy to observe in this format, but it may provide zero benefits if precise values are necessary. In Figure 12 we can see towns being replaced by dots. This can show us where most towns are situated in North America (AltexSoft, 2019).

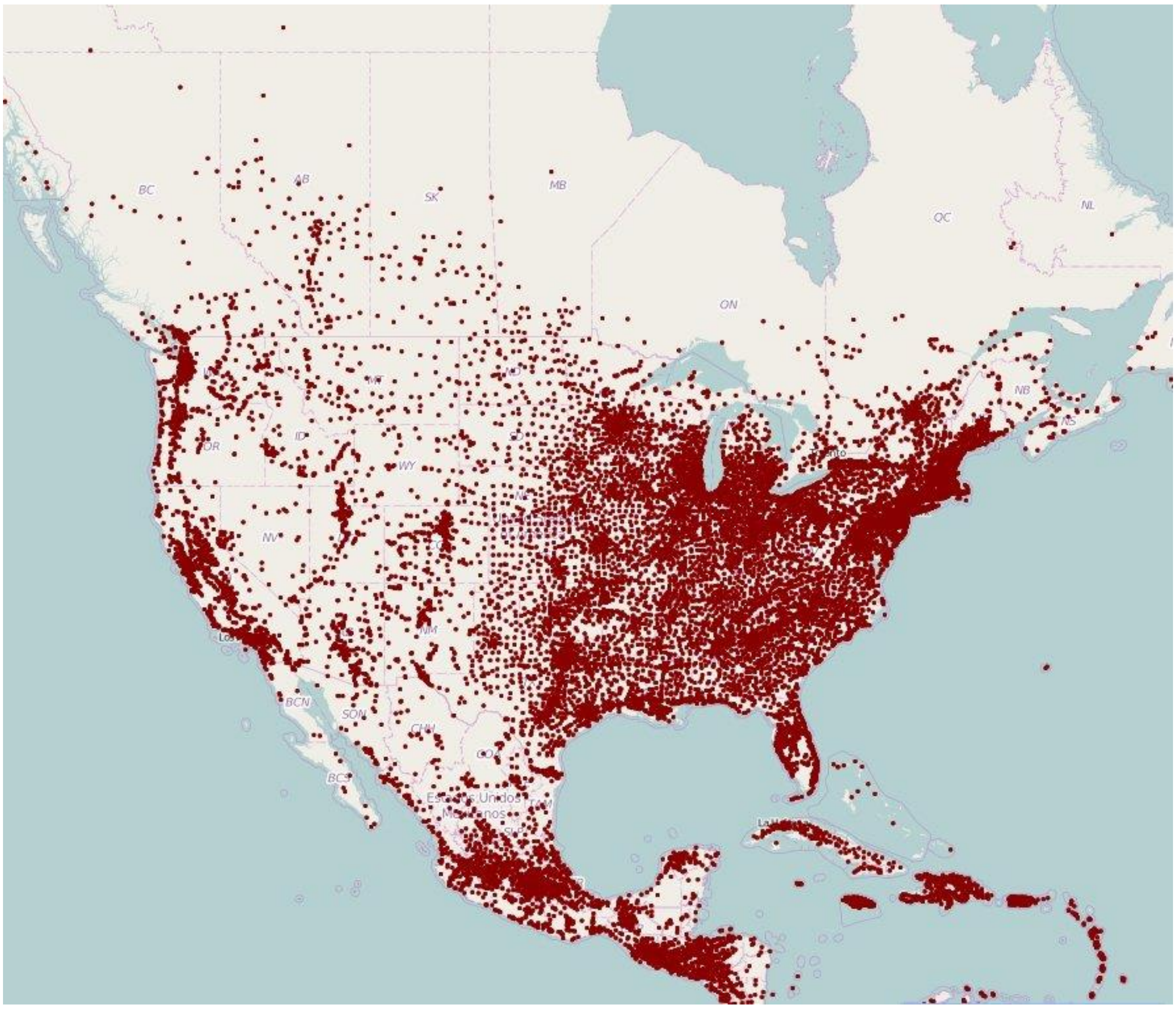

Figure 12: Density Map
Source: DESJARDINS, Jeff, 2018

1.7 Data Mining

The growth of information technology has resulted in a significant number of databases and vast amounts of data in a variety of fields. Database and information technology research has resulted in a method for storing and manipulating this valuable data for future decision-making. Facts mining is the process of extracting meaningful data and patterns from large amounts of data. Knowledge discovery, data mining, knowledge extraction, and data/pattern analysis are all terms used to describe this process (RAMAGERI, Bharati, M., 2010).

Data mining is a logical process for searching through enormous amounts of data to locate meaningful information. The purpose of this technique is to discover previously unknown patterns. Once these patterns have been discovered, they may be used to make specific decisions about how to grow their enterprises (RAMAGERI, Bharati, M., 2010).

1.7.1 History of Data Mining

With origins in statistics, artificial intelligence, machine learning, and database research, data mining have a lengthy history. 1 & 2 The term 'data mining' was first used in an article by Lovell in the 1980s. Advances in this discipline were paralleled by the development of related software tools, beginning with mainframe programs for statistical analysis in the early 1950s and leading to today's service solution, which includes a wide range of stand-alone, client/server, and web-based software (MIKUT, Ralf a REISCHL Markus, 2011).

There are many conventional data mining approaches accessible today. From a historical standpoint, these strategies have distinct origins. The focus was shifted from the proof of known hypotheses to the development of new hypotheses, which was one of the first methodologies taken from classical statistics. Methods from Bayesian decision theory, regression theory, and principal component analysis are only a few examples (MIKUT, Ralf a REISCHL Markus, 2011).

Artificial intelligence provided another set of methods, such as decision trees, rule-based systems, and others. Support vector machines and artificial neural networks are examples of machine learning approaches. Fuzzy logic, artificial neural networks, and evolutionary algorithms are examples of diverse and sometimes overlapping categorizations that are grouped together as computational intelligence (MIKUT, Ralf a REISCHL Markus, 2011).

The typical life cycle of new data mining methods starts with theoretical articles based on in-house software prototypes, then successful algorithms are distributed as research prototypes via public or on-demand software. The algorithms are then either developed into unique commercial or open-source packages comprising a family of comparable algorithms, or they are integrated into existing open-source or commercial packages. Many companies have attempted to advertise their own stand-alone packages, but only a few have succeeded in gaining significant market share. Some data mining tools have a very limited life cycle (MIKUT, Ralf a REISCHL Markus, 2011).

Internal marketing decisions and acquisitions of specialist enterprises by larger companies are common reasons for rebranding and integrating product lines (MIKUT, Ralf a REISCHL Markus, 2011).

The most significant commercial successes came from the gradual integration of data mining methods into existing commercial statistical tools. Since the 1970s, companies such as SPSS, which was created in 1975 with roots dating back to 1968, and SAS, which was founded in 1976, have offered statistical tools for mainframe computers. For larger customers, these technologies were later extended to personal PCs and client/server solutions. As data mining became more popular, algorithms like artificial neural networks and decision trees were integrated into primary products and specialised data mining companies like Integrated Solutions Ltd. (bought in 1998 by SPSS) were acquired to gain access to data mining tools like Clementine. Renaming of tools and firm mergers played a significant role in history throughout these years; for example, the tool Clementine (SPSS) was renamed PASW Modeler and is currently available as IBM SPSS Modeler following IBM's acquisition of SPSS in 2009. Tools from this statistical branch are becoming increasingly popular among users in business applications and applied research (MIKUT, Ralf a REISCHL Markus, 2011).

Data mining has evolved into a technique in its own right over the last 10–15 years, is well established in business intelligence (BI), and continues to grow in relevance in the technology and life sciences sectors. In genetics, for example, data mining was a major component in enabling methodological discoveries. It's a promising technology for fields like text mining and semantic search engines, autonomous systems like humanoid robots and cars, chemoinformatics, and others (MIKUT, Ralf a REISCHL Markus, 2011).

1.7.2 Data Structures

The dimensionality of the underlying raw data in the processed dataset is an important requirement.

The initial data mining applications were designed to work with two-dimensional feature tables of data. A dataset in this traditional format consists of N samples (for example insurance business clients) with s features (for example income, age, number of contracts, and so on) that contain real values or usually integer-coded classes or symbols (for example income, age, number of contracts, and so on). Almost all available utilities support this format. The dataset can be sparse in some circumstances, with only a few nonzero attributes like a list of s purchasing products for N different consumers. If a tool takes use of this sparse structure, it can save time and memory (MIKUT, Ralf a REISCHL Markus, 2011).

The same dimensionality characterizes some structured databases. Most text mining issues, for example, express sample texts as the frequency of words or so-called n-grams (a collection of n consecutive letters in a document) (MIKUT, Ralf a REISCHL Markus, 2011).

Time series are used as elements in the most common format with increased dimensionality, resulting in dataset dimensions ranging from one to three. Forecasting future values, identifying typical patterns in a time series, and clustering to find related time series are all common tasks. Time series analysis is important in a variety of applications, including stock market forecasts, forecasting of energy consumption and other markets, and quality control in manufacturing, and is supported by most data mining tools (MIKUT, Ralf a REISCHL Markus, 2011).

Different types of structured data exist with similar dimensions, such as gene sequences (spatial structure), spectrograms or mass spectrograms (organized by frequencies or masses), and others. Only a few programs expressly accept these forms of structured data, however, some time series analysis tools can be modified to deal with these issues. There is also a relatively new trend in data mining for pictures and videos. Besides ImageJ and ITK, graphs that can be represented as adjacency matrices, which describe the link between different nodes of a graph, are another format that leads to image-like dimensions. Graph mining has a wide range of applications, including analysing social networks and chemical structures, yet, only a few tools, such as Pegasus and Proximity, exist (MIKUT, Ralf a REISCHL Markus, 2011).

Data mining software is classified into two types: standalone and client/server. Client/server solutions are the most common, especially in business-oriented products. They're available for a variety of operating systems, including Windows, Mac OS X, Linux, and mainframe supercomputers. There are an increasing number of platform-independent JAVA-based solutions available for users in research and applied research. Nevertheless, both of these trends may potentially be a danger to privacy policies because data protection is a hard task and most of the companies are very cautious with sensitive data (MIKUT, Ralf a REISCHL Markus, 2011).

1.7.3 9 Types of Data Mining Software Tools

There are plenty of data mining software tools, which are helping users in learning about customers´ needs, reduction of costs, increasing profit and gaining valuable information about business insights. Those tools have a variety of functions, each one serving its´ own purpose. One of the examples can be the R Project which will be later used in the analysis of received data. It is mainly used for data analysis, statistical calculations, and it can present how data mining tools are used in practice. But now let’s take a look at nine general types of data mining software tools.

1. **Data mining suites**

Data mining suites are primarily focused on data mining and include a variety of methodologies. They support feature tables and time series, and text mining tools are occasionally available. Although coupling to business solutions, model import and export, reporting, and a range of different platforms are supported, the application focus is broad and not limited to a specific application field, such as business applications. In addition, the makers offer services for adapting the tools to the customer's workflows and data structures. Although DMS is primarily a commercial product with a high price tag, there are some open-source alternatives, such as RapidMiner. IBM SPSS Modeler, SAS Enterprise Miner, Alice d'Isoft, DataEngine, DataDetective, GhostMiner, Knowledge Studio, KXEN, NAG Data Mining Components, Partek Discovery Suite, and STATISTICA are some of the most common examples (MIKUT, Ralf a REISCHL Markus, 2011).

2. Business intelligence packages

Data mining is not a major focus of business intelligence packages, but they do incorporate basic data mining functionality, particularly for statistical methods in business applications. Large feature tables are supported in BIs, but they are frequently limited to feature tables and time series. They have a well-developed reporting feature as well as solid support for education, handling, and adapting to the customer's operations. They have a heavy focus on database coupling and use a client/server architecture to implement them. The majority of business intelligence software tools is commercial (IBM Cognos 8 BI, Oracle Data Mining, SAP Netweaver Business Warehouse, Teradata Database, IBM DB2 Data Warehouse, and PolyVista), however, there are a few open-source options (Pentaho) (MIKUT, Ralf a REISCHL Markus, 2011).

3. Mathematical packages

Data mining is not a priority of mathematical packages, but they do include a vast and flexible range of algorithms and visualization routines. They support feature tables, time series, and picture import formats, at the very least. Programming skills in a scripting language are frequently required for user engagement. Because data mining methods may be quickly implemented in the form of extensions (EXT) and research prototypes, MATs are appealing to users in algorithm development and applied research (RES). Commercial (MATLAB and R-PLUS) and open-source (MATLAB and R-PLUS) MAT packages are available (R, Kepler). In theory, table computation software tools like Excel might be included here as well, although it is not covered in this work. The majority of tools are cross-platform, yet they have database coupling flaws (MIKUT, Ralf a REISCHL Markus, 2011).

4. Integration packages

Integration packages are expandable bundles of many different open-source algorithms that can be used as standalone software tool (mostly based on Java; for example, KNIME, the GUI-version of WEKA, KEEL, and TANAGRA) or as something of a bigger enhanced version package for MAT type tools (for example, Gait-CAD or PRTools for MATLAB). Standard formats are supported for import and export, although database support is limited. The majority of programs have a graphical user interface and are available for a variety of platforms. When open-source integration packages are built on commercial solutions of the MAT type, licence types are mixed. The tools are appealing to algorithm creators and users in applied research because of their expandability and rapid comparison with other tools, as well as their ease of integration of application-specific methods and import options (MIKUT, Ralf a REISCHL Markus, 2011).

5. EXT

EXT are little add-ons for other products like Excel, Matlab, R, and others, with limited yet helpful capabilities. Only a few data mining algorithms, such as artificial neural networks for Excel (Forecaster XL and XLMiner) or MATLAB, are used here (Matlab Neural Networks Toolbox). There are commercial and open-source versions, but fundamental tool licences must be provided. The user interface is similar to that of the basic tool, such as utilizing a programming language (MATLAB) or embedding the extension in the menu (Excel) (MIKUT, Ralf a REISCHL Markus, 2011).

6. Data mining libraries

Data mining libraries are a collection of functions that implement data mining algorithms.
An Application Programming Interface (API) for the interaction between the software tool and the data mining operations might be implemented in other software tools. Although there is no graphical user interface, some functionalities can help with the integration of specific visualization tools. They're usually written in JAVA or C++, and the solutions are cross-platform. For support vector machines, open-source examples include WEKA (Java-based), MLC++ (C++-based), JAVA Data Mining Package, and LibSVM (C++ and JAVA-based). Neurofusion for C++ is a commercial example, whereas XELOPES (Java, C++, and C) uses alternative licence types. Data mining library tools appeal to users in algorithm development and applied research because they allow data mining software tools to be embedded into bigger data mining software tools or particular solutions for restricted applications (MIKUT, Ralf a REISCHL Markus, 2011).

7. Specialities

Specialities are comparable to DMS tools in that they solely implement one type of method, such as artificial neural networks.

For such methods, they include a number of advanced visualization tools. When compared to other tools, SPECs are easier to use, which makes them more useful in the classroom. For decision trees, use CART; for Bayesian networks, use BayesiaLab; for rule-based systems, use C5.0, WizRule, Rule Discovery System; for association analysis, use MagnumOpus; and for artificial neural networks, use JavaNNS, Neuroshell, NeuralWorks Predict, RapAnalyst (MIKUT, Ralf a REISCHL Markus, 2011).

8. RES

The first implementations of new and revolutionary algorithms (unfortunately not very stable) are frequently RES. They only have one or a few algorithms, and they don't have any graphical or automated assistance. Import and export functionality is limited, and database coupling is either absent or ineffective (MIKUT, Ralf a REISCHL Markus, 2011).

The majority of RES tools are open source. Users in algorithm creation and applied research, particularly in very inventive sectors, find them very appealing. GIFT for content-based image retrieval, Himalaya for maximal frequent itemsets mining, sequential pattern mining, and scalable linear regression trees, Rseslibs for rough sets, and Pegasus for graph mining are other examples. Today's prominent tools, such as WEKA and RapidMiner, began in this category and then moved to other categories, such as DMS (MIKUT, Ralf a REISCHL Markus, 2011).

9. Solutions

Solutions are a collection of tools tailored to specific application disciplines, such as text mining (GATE), image processing (ITK, ImageJ), drug discovery (Molegro Data Modeler), microscope image analysis (CellProfilerAnalyst), or mining gene expression profiles (Partek Genomics Suite, MEGA). The good support of domain-specific feature extraction techniques, evaluation measures, visualizations, and import formats is an advantage of these systems. The amount of support for data mining methods varies from very basic (especially in image processing) to highly evolved algorithms. In some circumstances, more generic

DMS or INT tools (KNIME, Gait-CAD for peptide chemoinformatics) can additionally assist specialised domains. There are numerous commercial and open-source options available. For better understanding were created tables with specific tools with their type and the link. They can be found in the appendix (MIKUT, Ralf a REISCHL Markus, 2011).

1.8 CRM

With the rapid growth of e-commerce, the intensification of worldwide market rivalry, and the increasing diversity of consumer demand, many businesses have shifted their management focus from 4P to customer relationship management. CRM (Customer Relationship Management) has been the subject of enterprise management study. In general, a standard CRM system requires the construction of an information system on its own, which comprises of marketing management, sales management, and customer service management subsystems, with the system's function centered on processes (HE, Li a XIN, Guan a YUFENG, Gong 2008).

The analytical CRM system in the organization has flaws in its design, which makes it difficult to collect and evaluate historical data and insufficient for extensive customer research. In order to design CRM systems in an integrated environment, businesses must leverage Business Intelligence technologies to build an integrated information platform. The CRM system's design will improve the structure and functionality of traditional CRM systems while also strengthening the application of Business Intelligence on customer behaviour analysis and decision support capabilities. Constructing a new CRM system architecture based on Business Intelligence that allows for optimal client resource management and maximum utilization of client resources to mine customer value (HE, Li a XIN, Guan a YUFENG, Gong 2008).

1.8.1 CRM System Framework in BI

The CRM system should be able to meet the marketing, sales, and service needs of departments, as well as improve market decision-making capabilities, sales unified management, and customer service quality. Second, the CRM system should create lines of communication between sales, marketing, and service. Finally, the CRM system should work well with other business systems to capture all elements of business data, optimise production processes based on customer requests, and meet management requirements. As a result of the demand research, this article develops a business intelligence-based integrated CRM system (HE, Li a XIN, Guan a YUFENG, Gong 2008).

1.8.2 Subsystems in CRM

Customer Contact Subsystem, Foreground Operation Subsystem, and Background Decision-Making Subsystem should make up the CRM system framework based on business intelligence. Customers can get customer data through Call Centers, EMAIL, Enterprise Information Portals, Fax and Mail, Face-to-Face communication with customers, Ecommerce, Wireless Access, and other customer contact methods. The data format for CRM foreground operation Subsystem is then standardized and integrated using XML (HE, Li a XIN, Guan a YUFENG, Gong 2008).

CRM business portion, CRM analytical part, CRM Basic Information Database, EAI Application Integration Interface, and Data warehouse make up the Foreground Operation Subsystem. The CRM operation functions include marketing automation, sales automation, and customer service automation, and it collaborates with other information systems and exchanges data using EAI middleware technology (HE, Li a XIN, Guan a YUFENG, Gong 2008).

The CRM database contains the integration data from the CRM business part and the customer contact centre. Some of these dates may be directly applicable to CRM analysis. However, mass data should be extracted, cleaned, transformed, and loaded, then deposited in the Data Warehouse of the customer theme, which is part of both the foreground and background decision-making subsystems. Three Data Marts have been settled based on the Customer Data Warehouse theme: Customer Data Mart, Product Data Mart, and Customer Interaction Data Mart. The CRM analytical portion, based on these data marts, performs operations such as online analytical processing, product sales analysis, marketing analysis, and market analysis. When integrating data and information from the Contact subsystem, the foreground operational subsystem should transmit it back to the Contact subsystem. Customer theme data warehouse, Data Mining Tool, and Business Model / Rule base make up the Background Decision-Making subsystem. To mine data and accomplish Customer

Knowledge Discovery in the database (CKDD), the Data Mining Tool will combine with model and technique bases. The system may communicate the created consumer knowledge to many departments while also meeting corporate needs (HE, Li a XIN, Guan a YUFENG, Gong 2008).

1.8.3 CRM-Oriented Data Warehouse and Its Design

Customer is the subject of data warehouse, in addition to the client but also to customers, it's the analysis of sales, which also involves the product, and the domain of data warehouse is the client-centred theme, according to the operational requirements of the CRM system. As a result, there is a client-centred theme in this data warehouse domain, as well as a product and customer interaction behaviour theme domain. Second, a data warehouse application structure has to be created; used will be the application structure "data warehouse + data mart" Data Mart, which includes three things: Data Mart, also known as a customer data mart, a product data mart, or a customer interaction behaviour data mart, is the formation of intelligence departments within businesses to meet the requirement for unique management decisions (HE, Li a XIN, Guan a YUFENG, Gong 2008).

Customers, products, and consumer behaviour are the three themes that have been defined while planning a data warehouse. There is a design concept model in the customer, product, and customer behaviour themes, and used can be Dot Modeling for creating a conceptual model, with the customer theme as an example (HE, Li a XIN, Guan a YUFENG, Gong 2008).

Theme domain-oriented can be based on the data mart's business requirements. Partially store on multiple servers so that relevant departments can readily access them. The data, on the other hand, is still shared throughout the many departments. Data marts in various business sectors are still planned and designed as a client-oriented theme subset of a full data warehouse in CRM systems (HE, Li a XIN, Guan a YUFENG, Gong 2008).

1.8.4 Data Flow Among the CRM System Parts

The data warehouse of customer motif-oriented CRM system architecture is built on business intelligence. The data warehouse is surrounded by a coordinated operation system and an analytical system. Customer data from CRM operations, customer contact centres, and other information systems are all integrated into the data warehouse in an enterprise application integration environment. Then, for the customer-oriented data warehouse, use business intelligence analysing technology like OLAP online analytical processing and DW data mining to analyse and conclude data, dig out customer knowledge and feed it back to every department, to improve customer happiness, improve production and operation flow, and support the marketing decision of the firm, such as marketing, sales, production, human resource, decision making, and so on (HE, Li a XIN, Guan a YUFENG, Gong 2008).

After the information integration, the customer data and information are stored in the data warehouse, after which the customer knowledge is extracted by crossing the customer intelligence, and finally, these valuable information and customer knowledge are fed back to the enterprise's departments, resulting in the creation of an information-circulating system based on customer knowledge, and achieving the ring-closed form of the enterprise's customer (HE, Li a XIN, Guan a YUFENG, Gong 2008).

1.9 DSS

The term "decision support system" refers to a broad range of systems, techniques, and technology. Some believe the word DSS is obsolete and that it has been superseded by a "new type" of the system known as on-line analytical processing, or OLAP. Others appear to consider knowledge-based decision support systems as the "state-of-the-art" in decision support systems. As the "actual" DSS, operations researchers typically focus on optimisation and simulation models. The phrase decision support system, abbreviated DSS, is, in my opinion, a useful and inclusive name for many sorts of information systems that support decision making (POWER, Daniel, J. 1997).

It's important to remember that if a computerized system isn't an OLTP (online transaction processing system), someone will call it a DSS. Someone will most likely refer to a software program that runs on a PC and can assist management in making a choice as a DSS. DSS includes EIS, ESS, geographic information systems (GIS), OLAP, software agents, knowledge discovery systems, and group DSS (POWER, Daniel, J. 1997).

So, how can IT and business leaders talk about creating the DSS that the executives actually want? How do we make sense of all the jargon around the term "decision support system"? My first piece of advice would be to learn everything we can about DSS. The second suggestion is to participate in a DSS design and development process that includes both future executive users and IT administrators. Knowledgeable business executives should talk openly with IS professionals about deliverables, capabilities, outcomes, needs, and which decisions a proposed system should enable. To achieve these objectives, we'll look at two types of DSS: enterprise-wide DSS and desk-top DSS (POWER, Daniel, J. 1997).

Enterprise-wide DSSs are linked to massive data warehouses and service a company's numerous managers. Single-user desk-top DSSs are modest systems that run on a single manager's PC. These two categories include a wide variety of capabilities. Let's employ buzzwords like a data warehouse, OLAP, executive information systems (EIS), knowledge-based DSS, intranets, and client-server architectures (POWER, Daniel, J. 1997).

The advent of enterprise-wide DSS with very huge data warehouses, which are meant to help decision-makers find out nearly anything about their firm in a matter of seconds, has generated a lot of buzz. Drill down, slice and dice, graph and chart corporate and external data for decision-makers. The Decision Maker's Workbench (DMW), developed by Mervyn's Department Stores and MicroStrategy in 1994, is a well-known example. Its claims are, luckily, more modest than those made for certain data warehouse/OLAP systems (POWER, Daniel, J. 1997).

DSS can range from relatively simple systems to massive data-intensive and analytically sophisticated executive information systems. We can identify enterprise-wide DSS, which are primarily file drawer systems that provide for instant access to data objects, based on Steven Alter's (1980) definition. At a more advanced level, data analysis systems make it simple to manipulate data using computerized analytical tools such as statistics packages, data mining, and so on. The most advanced enterprise-wide analysis systems give users access to a number of decision-oriented databases or data marts, predefined models and visualizations, and triggers and alerts tied to events or variables in the corporate data warehouse (POWER, Daniel, J. 1997).

The most advanced enterprise-wide DSS built on and expanded on the executive information systems (EIS) that were offered in the late 1980s. To give executives easy online access to current information on a company's status, EIS uses "state-of-the-art" visuals, communications, and data storage systems. We now take for granted much of the "state-of-the-art" from 1987, and our systems are more powerful than what most suppliers intended to deliver in 1987 (POWER, Daniel, J. 1997).

We observe bridges between enterprise-wide DSS, data warehouses, and desktop DSS in most enterprises. Some DSS consultants, for example, envision an enterprise-wide DSS that functions essentially as a file drawer system for supplying data that is then evaluated on a PC. One type of DSS architecture is the one-way bridge. What data is stored where and how it will be processed and displayed are both important considerations. DSS architecture is a difficult concept to grasp quickly. Let us conclude for the time being that DSS can be used both enterprise-wide and on a single user's PC. Data and analyses can be moved back and forth from the client desktop and related DSS tools to server storage and server-based DSS tools using a client-server architecture. In an enterprise, DSS and data can be found everywhere (POWER, Daniel, J. 1997).

Desktop, single-user DSS don't get the same press and attention as enterprise-wide DSS, but they're just as useful. For desktop studies or to construct specific DSS applications for individual managers, we occasionally employ spreadsheet programs like Excel or Lotus 1 2 3. We occasionally purchase specialised DSS packages for a single PC or a server. Expert Choice is a specialised software that functions as a desktop DSS and can be used here as an example (POWER, Daniel, J. 1997).

The analytical hierarchy method is implemented by Expert Choice. This Windows software program can help us structure difficult problems, define priorities and rank alternatives, measure the consistency of judgments, allocate resources, and conduct a cost-benefit analysis, among other things. The tool aids in the organization of problem-related data into a hierarchical model that includes a goal, possible situations, criteria, and options. Expert Choice allows the decision-maker to make logical decisions about the relative importance of criteria and the preference for alternatives over criteria (POWER, Daniel, J. 1997).

There is also a selection of desktop DSS accessible. On a single executive's PC, file drawer DSS is implemented in Microsoft Access. Accounting and financial models can be used as a desktop DSS in Microsoft Excel or as programmed components in an enterprise-wide DSS.

Analysts in some companies use desktop tools to prepare financial analyses and then publish the results on the company intranet or EIS (POWER, Daniel, J. 1997).

Simulation, another DSS tool, is frequently included in desktop packages. Single-user desktop packages are typical for optimisation software packages and DSS created using them. In some cases, however, a DSS optimisation model may calculate using live or "real-time" data received across a local or wide area network. Model of suggestion Knowledge-based systems, or DSS, is frequently implemented as a single-user desktop program (POWER, Daniel, J. 1997).

Expert systems are another term for knowledge-based systems. These computer systems employ symbolic logic to evaluate data, have an explicit knowledge base, and can explain conclusions in a user-friendly manner. Knowledge-based systems can help novice management make a complex decision by reminding an experienced decision-maker of possibilities or issues to consider (POWER, Daniel, J. 1997).

So, what exactly is a DSS? A decision support system (DSS) is a computer-based interactive system that assists managers in making decisions. A decision support system (DSS) assists management in retrieving, summarizing, and analysing decision-relevant data. It could be predominantly a data-driven or model-driven DSS. It could be an enterprise-wide DSS that supports a large number of managers in a networked, client-server system with a customized data warehouse or a single-user DSS on a manager's PC. We have to remember that identifying what type of computer-based system we're trying to construct, not what we label it, is the first step in creating an effective decision support system (POWER, Daniel, J. 1997).

1.10 Knowledge Management

Scholars have proposed various definitions of knowledge management, and there is still no consensus. The concept of knowledge management was first proposed by Harvard University professor Peter F. Drucker. Knowledge management, he argues, is about providing knowledge and existing knowledge to identify effective ways to use knowledge to generate the best results. And he emphasised the need for knowledge management being based on a learning organization. Narrow knowledge management focuses on knowledge production, acquisition, processing, storage, dissemination, and application. Generalised knowledge management, on the other hand, encompasses not just knowledge management but also the administration of other knowledge-related resources. A knowledge management system (KMS) is a platform that organizes and implements KM, as well as management tools for explicit and tacit information. It can gather, collate, transfer, and communicate knowledge, as well as convert tacit knowledge to explicit knowledge via socialization, internalization, externalization, and combination, resulting in new knowledge (CHENG, Lei a CHENG, Peng, 2011).

1.10.1 Similarities of Business Intelligence and Knowledge Management

They are both information technology-based. The Internet, computer gear, software, database storage, and network communication technology are all used in BI and KM.

Both collect, collate, share, and use information and knowledge in business operations, and they rely on information and knowledge to perform their functions. KM and BI work in tandem and compliment one another.

Knowledge is the object of KM, and it is primarily concerned with humans who master knowledge, their culture, and their behaviour. It emphasises the significance of knowledge innovation and whether it is properly used. People generally employ quantitative analysis of technical expertise to solve business problems with the help of business intelligence systems, so the applied effect of BI is tightly tied to users' skills (CHENG, Lei a CHENG, Peng, 2011).

1.10.2 How BI and KM Differ

Connotation

Transaction processing systems, administrative information systems, management information systems, decision support systems, and so on were all used to develop BI. In the evolution of the knowledge economy era, KM refers to management theories and approaches. It was emphasised that knowledge is the most valuable resource and strategic capital and that knowledge generation, diffusion, and utilization are critical to a company's competitive advantage. The goal of knowledge management is to capture, share, and distribute unstructured text and graphical data (CHENG, Lei a CHENG, Peng, 2011).

Focus

BI is mostly concerned with information resources. The entire BI process is rather closed and independent, and it pays greater on the combination and integration of the exterior morphology of information, as its goal is to make information resources orderly and structured. While KM systems deal with knowledge resources, their key goals are knowledge sharing and innovation. It emphasises the interaction of explicit and tacit information, promotes knowledge innovation, and works to establish a corporate culture of knowledge sharing and operational processes. BI systems deal with objective information in the real world for organizations, whereas KM systems deal with subjective and personal knowledge, which includes both explicit and easily copied knowledge in the real world and hidden knowledge stored in the individual mind that is hardly directly accessible, such as people's thoughts, skills, experiences, and perspectives (CHENG, Lei a CHENG, Peng, 2011).

Technologies

Data warehousing, online analytical processing, data mining, and enterprise information portals are among the basic technologies used by business intelligence. Document management, groupware technology, text mining and retrieval technology, enterprise knowledge portals, and other technologies are among the key technologies of KM (CHENG, Lei a CHENG, Peng, 2011).

Relationship between KM and BI

Richard T. Herschel and Nory E. Jones investigate BI as a subset of knowledge management. They argue that BI is more concerned with explicit knowledge, whereas KM encompasses both tacit and explicit knowledge. Both are effective in facilitating learning, decision-making, and comprehension (CHENG, Lei a CHENG, Peng, 2011).

Knowledge management and business intelligence, according to Cook and Cook, are both drawn from enterprise management theory and backed by technology. Although BI aids organizations in maximizing the value of information, it is unable to incorporate non-quantitative data. Unstructured or semi-structured information makes up 80% of corporate business data. It is more valuable to the business to extract meaningful information and expertise from unstructured material (CHENG, Lei a CHENG, Peng, 2011).

According to J. Okkonen, knowledge management and business intelligence overlap to some level. BI transforms data into information and intelligence, while KM systems give that information and intelligence meaning (CHENG, Lei a CHENG, Peng, 2011).

KM and BI, according to Kadayam, should be merged. According to him, significant technological advancements are bridging the gap between KM and BI, and their integration will deepen and widen the search for knowledge and information while also increasing the value, movement, and investment returns of intelligence (CHENG, Lei a CHENG, Peng, 2011).

Currently, some businesses are working on developing an integrated KM and BI platform. Although KM is not as mature as BI, IBM believes that as time passes, the systems will learn from one another and encourage new techniques of use, resulting in a collaborative effort known as KMBI (CHENG, Lei a CHENG, Peng, 2011).

Only by integrating KM and BI, according to Nemati, can corporate decision-making and behaviour be improved. They argued that the next generation of knowledge systems should capture, screen, store, organize, and transmit business knowledge as well as data and information. They propose that the knowledge warehouse concept may be extended from the data warehouse model. It will help with knowledge capture and coding, as well as information reuse and sharing within a company (CHENG, Lei a CHENG, Peng, 2011).

Conclusion

KM and BI are complementary because they grow as systems in their own right, with separate functions and various system frameworks.

As a result, their integration should be based on their shared qualities. The integrated KMBI system may increase user productivity, give the best service to enterprise customers, enable enterprise decision-makers to analyse and apply a range of data, information, and knowledge, improve the quality and speed of decision making, and boost competitive advantage (CHENG, Lei a CHENG, Peng, 2011).

1.11 GIS

"Over the past decade, geographic information systems (GIS) have become increasingly important as a supporting and enabling technology within the business environment and especially within the sphere of business intelligence (BI)." (BANK, Standard a POSTHUMUS, Rudi, 2008)

Globalisation, greater customer knowledge and complexity, and more rivalry are all causing the business environment to become more challenging. We are witnessing a fast-changing economic landscape, with international competitors joining the markets and major changes in our population's income and lifestyle profiles (BANK, Standard a POSTHUMUS, Rudi, 2008).

Companies engage in information technology to stay competitive, corporate activities grow more automated, and measurement tools get more complex. As a result, businesses frequently collect huge amounts of internal data about customers, corporate operations, and financial performance. The BI professional must save, locate, and analyse data in order to offer useful information to decision-makers, and this data overload presents a challenge. Clients have physical addresses, and branches are positioned to service certain trade areas, thus all company operations take place somewhere in geographic space (BANK, Standard a POSTHUMUS, Rudi, 2008).

Many business intelligence officers see the capacity of GIS technology to connect data from many business units and source systems by using the location element as a common denominator. Perhaps even more importantly, GIS offers considerable advantages in spatial analysis and information visualisation to reveal patterns and trends (BANK, Standard a POSTHUMUS, Rudi, 2008).

1.11.1 Why Are GIS Important for BI

While GIS and BI have traditionally served independent purposes, they have now merged, allowing business analysts to expand their picture of the firm by combining geographic and business data. At the same time, increased consumer knowledge of mapping information for commercial purposes is raising company awareness of location exploitation. As a result, combining GIS with BI technologies yields results that go beyond merely visualising data on a map (BANK, Standard a POSTHUMUS, Rudi, 2008).

It helps business users to examine and analyse links between geographic data and business data by combining the analytical power of databases with the geographic capabilities of maps. While BI tools are great for analysing who, what, and when, they fall short when it comes to queries about where, such as the relationship between where customers reside and where they shop (BANK, Standard a POSTHUMUS, Rudi, 2008).

Businesses function in two environments: internal and external.

Making the right strategic and tactical decisions requires an understanding of the wider picture of what the internal and external environments look like and how they interact. Data warehouses and management information systems are meant to collect, store, and transmit internal data to company decision-makers. When these systems are functioning properly, they can provide vital information that allows the organisation to maximise revenue from existing sources while also improving operating efficiency and lowering costs. In reality, many large organisations' business units function in silos, with internal data that varies greatly in format, nature, and amount (BANK, Standard a POSTHUMUS, Rudi, 2008).

Even with MIS and BI systems in place, data transmission between silos may be inefficient, leaving managers only a partial picture of the company's internal environment. By using the common denominator of the location to connect together data of various types from many source systems, GIS may help bridge the gap between silos, speeding up the process of discovering, filtering, and comparing data to find solutions (BANK, Standard a POSTHUMUS, Rudi, 2008).

Understanding the external environment is a more difficult task. The external environment is where new clients and possibilities can be found, as well as where competitors can be a threat. Many businesses struggle to find and integrate data from the outside world into their existing information systems, which are often built to report structured internal data. This makes gaining a true competitive edge and developing a comprehensive "whole perspective of business" challenging (BANK, Standard a POSTHUMUS, Rudi, 2008).

The role of GIS in bridging the gap between internal and external data assets is critical. When analysed with internal data in a similar framework, vector and raster topographic maps, aerial images, census and other demographic data, and even satellite imagery can become incredibly important business information. Managers are under constant pressure to accomplish more with fewer resources. They must also comprehend increasing amounts of data related to increasingly complicated business concerns. Reporting is a cornerstone of BI, with "dashboards" gaining traction in place of management reports. The dashboard is a well-known graphic tool that allows business users to view important performance indicators and can be used to better understand what happened, when, and why (BANK, Standard a POSTHUMUS, Rudi, 2008).

By including mapping visualisation into these dashboards, firms can make spatial analysis more accessible to a broader audience, transforming it from a tool for GIS and BI specialists exclusively to a real tool for all business users. A map's ability to simplify information and compare many variables of a problem is unmatched by any other reporting style. GIS is especially well adapted to improving management reports and "dashboards" with intuitive maps and images that take less time to assess and are easier to comprehend. In today's fast-paced business world, the company that finds the solution to an issue first can sometimes gain a considerable competitive edge. In the corporate world, information is only valuable if it reaches the correct decision-maker at the right moment and in an easily understandable format. By connecting various internal and external data "pipes" through location, GIS can help BI professionals find more data and information faster (BANK, Standard a POSTHUMUS, Rudi, 2008).

Because corporations keep huge amounts of data on practically every conceivable area of their business operations, mistakes are almost certain to occur, lowering the value and credibility of information. Many data types, such as huge client databases, might make it tough or even impossible to spot mistakes in tabular database forms. Some of these mistakes can only be noticed when the data is imported into a GIS through a procedure like geocoding and mapped or spatially analysed (BANK, Standard a POSTHUMUS, Rudi, 2008).

2. Methodology

As a survey method was used a questionnaire which will be later on evaluated through figures, graphs and analysis.

2.1 Questionnaire

As the main tool for the questionnaire was used google form. This questionnaire contains 14 questions which were spread online via Facebook, Messenger WhatsApp and email communication. Thanks to the snowball, the effect author of the thesis hopes for something around 50 filled questionnaires. 50 answers should be enough for good evaluation and meaningful outcome. Part of the questions are aiming for finding out if the respondents are familiar with terms like business intelligence, business intelligence tools, databases, data mining etc. The second part of the questions is about the working experience in business intelligence, whether respondents do find business intelligence as a part and parcel of modern enterprises and if they understand the true importance of it.

2.1.1 Target Audience

The questionnaire was spread among either young or older people who have some experience in modern enterprises and who are in some way working with BI tools. The Group of respondents that was aimed to were not specified by gender or specific country. Aimed groups according to the job were modern enterprises and people working in those were preferred, in specific we also targeted people who are working in the education, manufacturing, banking and health sector. IT companies were not excluded. Unemployed respondents were also an option, but they had to have some knowledge about modern enterprises.

2.1.2 Expectations

We are expecting that the majority will know about business intelligence problematics, some of them will have some experience and many of them will be willing to learn more about it. But we also expect that even though all our respondents already know something about modern enterprises, many of them will not know what they are actually working with – business intelligence tools.

2.2 Association Rules

One of the most fundamental machine learning topics utilized in market basket research is association rules. In a store, all veggies are grouped together, all dairy items are grouped together, and cosmetics are grouped together once more. Investing time and resources in purposeful product placements like this minimizes a customer's shopping time while also reminding the customer of relevant products she/he would be interested in purchasing, allowing retailers to cross-sell. Association rules aid in the discovery of all such connections between objects in large databases (GARG, Anisha, 2018).

To detect relationships, rules do not tie back a user's various transactions over time. A group of products with unique transaction IDs is investigated. This is useful for product placement on aisles. Collaborative filtering, on the other hand, links all transactions to a user ID in order to find similarities between users' preferences. Various metrics have been established to assist us in determining the strength of the link between the variables (GARG, Anisha, 2018).

2.2.1 Support and Confidence

Two key characteristics are used to assess the strength of an association rule: support and confidence. The frequency with which a specific rule appears in the database being mined is referred to as support. The amount of times a given rule turns out to be true in practice is referred to as confidence. A rule may appear to have a strong correlation in a data collection because it appears frequently, but it may emerge much less frequently when applied. This would be a situation where there is a lot of support but not a lot of confidence (LUTKEVICH, Ben, 2020).

In contrast, a rule may not stand out in data collection, but further research reveals that it occurs frequently. This is a situation where there is a lot of confidence but not a lot of support. These metrics aid analysts in distinguishing causation from correlation and determining the usefulness of a rule (LUTKEVICH, Ben, 2020).

The lift value, or confidence to support ratio, is a third value component. There is a negative correlation between data points if the lift value is negative. There is a positive correlation if the value is positive, and there is no correlation if the ratio is equal to 1 (LUTKEVICH, Ben, 2020).

2.2.2 Algorithms

AIS, SETM, Apriori, and variations of the latter are examples of popular algorithms that use association rules. As the AIS algorithm reads the data, itemsets are created and counted. The AIS method analyses which large itemsets comprised a transaction in transaction data and creates new candidate itemsets by extending the large itemsets with other items in the transaction data (LUTKEVICH, Ben, 2020).

Although the SETM algorithm generates candidate itemsets as it searches a database, it accounts for them at the end of the scan. New candidate itemsets are created in the same way as with the AIS method, but the generating transaction's transaction ID is recorded in a sequential data structure with the candidate itemset. The support count of candidate itemsets is calculated by aggregating the sequential structure at the end of the pass. Both the AIS and SETM algorithms have the drawback of being able to generate and count a large number of small candidate itemsets (LUTKEVICH, Ben, 2020).

Candidate itemsets are generated utilizing only the large itemsets from the previous pass with the Apriori algorithm. The preceding pass's huge itemset is linked with itself to generate all itemsets with a size greater by one. After that, each created itemset with a small portion is removed. The candidates are the remaining itemsets. Any subset of a frequent itemset is considered as a frequent itemset by the Apriori algorithm. Using this method, the algorithm minimizes the number of candidates investigated by only looking at itemsets with a higher support count than the minimal support count (LUTKEVICH, Ben, 2020).

2.3 Chi-Square Test

In R, a Chi-Square test is a statistical tool for determining if two category variables have a significant connection. The two variables are drawn from the same sample. These variables are then classified as Male/Female, Red/Green, Yes/No, and so on. For example, we could create a dataset containing observations about people's cake-buying habits. Also, try to find a link between a person's gender and the cake flavor they favor. However, if a link is discovered, we can prepare for an adequate stock of flavors by knowing the number of people who will be visiting by gender (data-flair, 2018).

We must check the p-values in this test in particular. Furthermore, we treat this test as if this was a null hypothesis and an alternate hypothesis, as we do with all statistical tests. The important point is that we reject the null hypothesis if the p-value in the result is less than a preset significance level, which is commonly 0.05.

H0: The two variables are mutually exclusive.

H1: The two variables have a relationship.

A chi-square test is used to assess the two independent variables in the situation of a null hypothesis (data-flair, 2018).

2.4 Logistic Regression

This statistical model (also known as the logit model) is frequently used in classification and predictive analytics. Based on a collection of independent variables, logistic regression calculates the likelihood of an event occurring, such as voting or not voting. The dependent variable is confined between 0 and 1 because the outcome is a probability. A logit transformation is performed to the odds in logistic regression, which is the probability of success divided by the probability of failure. This logistic function is also known as the log odds, or the natural logarithm of odds. This logistic function is given by the following formulas:

Logit(pi) = 1/(1+ exp(-pi))

(Equation 1)

ln(pi/(1-pi)) = Beta_0 + Beta_1*X_1 + … + B_k*K_k

(Equation 2)

Logit(pi) is the dependent or response variable in this logistic regression equation, while x is the independent variable. Maximum likelihood estimation is typically used to estimate the beta parameter, or coefficient, in this model (MLE). This method uses numerous iterations to test different beta values in order to find the best match for log odds. The log likelihood function is created by all of these rounds, and logistic regression aims to maximize this function to get the optimal parameter estimate. The conditional probabilities for each observation can be calculated, logged, and averaged together to provide a forecast probability once the optimal coefficient (or coefficients, if there are more than one independent variable) has been identified. A probability less than 5 predicts 0 in binary classification, while a probability larger than 0 predicts 1. It's best practice to evaluate how well the model predicts the dependent variable after it's been computed, which is known as the goodness of fit. The Hosmer–Lemeshow test is a widely used method for determining model fit (IBM, 2020).

2.4.1 Binary Logistic Regression

The response or dependent variable in binary logistic regression is dichotomous in nature, with only two possible outcomes (e.g., 0 or 1). Predicting whether an e-mail is spam or not spam, or if a tumor is malignant or not malignant, are two common applications. This is the most often used technique in logistic regression, and it is also one of the most widely used classifiers for binary classification in general (IBM, 2020).

2.4.2 Multinomial Logistic Regression

In this form of the logistic regression model, the dependent variable contains three or more possible outcomes, but there is no predetermined order for these values. To sell films more efficiently, movie studios, for example, aim to anticipate what kind of film a viewer is likely to see. A multinomial logistic regression model can assist the studio figure out how much an individual's age, gender, and dating status influence the type of film they prefer. The company can then target a specific movie's advertising campaign at a group of people who are likely to see it (IBM, 2020).

2.4.3 Ordinal Logistic Regression

When the response variable includes three or more alternative outcomes, but these values have a predetermined sequence, this sort of logistic regression model is used. Ordinal responses include grading systems ranging from A to F and rating scales ranging from 1 to 5 (IBM, 2020).

2.4.4 Qualitative and Quantitative Evaluation

Qualitative and quantitative evaluations are the two sorts of evaluations. Asking a user or group of users whether the outcome of an information retrieval system provides a satisfactory answer is referred to as qualitative evaluation. The focus of the qualitative evaluation is on one or more users' experiences with a system. Quantitative evaluation entails having a method for numerically quantifying the findings of an information retrieval system (DALIANIS, Hercules, 2018).

2.4.5 Metrics

Evaluation can serve a variety of goals. Different constraints may apply to the amount of data used for training and evaluation. In some circumstances, strong recall takes precedence over high precision, whereas in others, the converse is true. A development set is used when creating a system. A development set can be data for creating artefact rules or training material for a machine learning system. The development set, also known as the training set in machine learning, is used to train the machine learning system. Part of the training set can be set aside for algorithm error analysis, and the machine learning algorithm can be tweaked based on the faults, a process known as parameter tuning. The development test set is an element of the training set (DALIANIS, Hercules, 2018).

A test set is set aside to test the artefact; this test set is frequently referred to as held-out data because it is not used for development or training. When data is scarce, a method known as k-fold cross-validation is used, which involves dividing the entire dataset into k folds, with the k-1 folds being used for training and the remaining one, the 1 fold, being used for evaluation: the folds are switched until all folds have been trained and tested on the remaining k-1 folds, after which an average is calculated. The most common method is 10-fold cross-validation (DALIANIS, Hercules, 2018).

Precision and recall are two criteria used to assess a retrieval system's performance. Precision is defined as the number of correct instances recovered divided by the total number of instances retrieved. The number of accurate examples recovered divided by the total number of correct instances is called recall. Instances can be individual characters in a text or a whole document in a corpus of documents that have been retrieved. A confusion matrix is frequently used to describe the various entities (DALIANIS, Hercules, 2018).

Depending on the weight function, the F-score is defined as the weighted average of precision and recall. The harmonic mean of precision and recall is the F1-score. The F-measure is another name for the F-score. Different indices can be used to weight precision and recall in the F1-score (DALIANIS, Hercules, 2018).

Positive predictive value (PPV), which corresponds to precision, is another widely used statistic in medical and clinical systems. Another metric is accuracy, which is defined as the percentage of true cases (both positive and negative) among all examples retrieved. Precision and inverse precision are weighted arithmetic means of accuracy. When accuracy is good but precision is poor, the system functions well but the outcomes are slightly skewed; compare this to striking the bullseye, which requires both high accuracy and precision (DALIANIS, Hercules, 2018).

3. Evaluation and Presentation

This chapter will talk about evaluation of questions and presentation of the answers, their analysis, and the results.

3.1 Evaluation of the Questions

Thankfully 90 answers were collected through the questionnaire. This amount of answers should be enough for the analysis.

Question number 1: What is your gender?

Possible answers:

- Male
- Female

This question was aimed for equality between male and females' part of the respondents. This aim was fulfilled, and it gave us the ability to find valuable information from both respondents' genders.

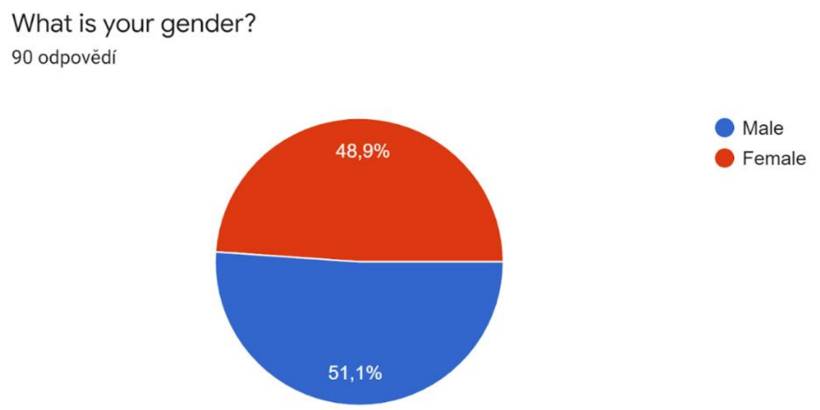

Figure 13: Gender
Source: Own contribution

According to the answers, 48,9 % of the questionnaire was filled by females and 51,1 % by males. We wanted to gather information from both genders because we wanted to see opinions from every possible angle and every possible type of modern enterprise worker.

Question number 2: *What is your age?*

Possible answers:

- Under 18
- 18 - 23
- 24 - 28
- 28 - 35
- Over 35

We wanted to divide our respondents into people who are

1. Under 18 years old – they do not have that much working experience and are most likely still studying.

2. 18 - 23 years old- these people are either starting to work or they are studying at universities with some possible part-time jobs.

3. 24 - 28 years old- for those who are already working for some period of time, or they are finishing their university studies so it is the beginning for their working career and they are looking for job offers.

4. 28 - 35 years old – the majority of these people are not studying anymore because they are getting pretty experienced in their jobs so they will probably have more overview about modern enterprises systems and tools.

5. Over 35 years old – those people are very experienced in their jobs and they are most likely very experienced with modern enterprise systems and tools.

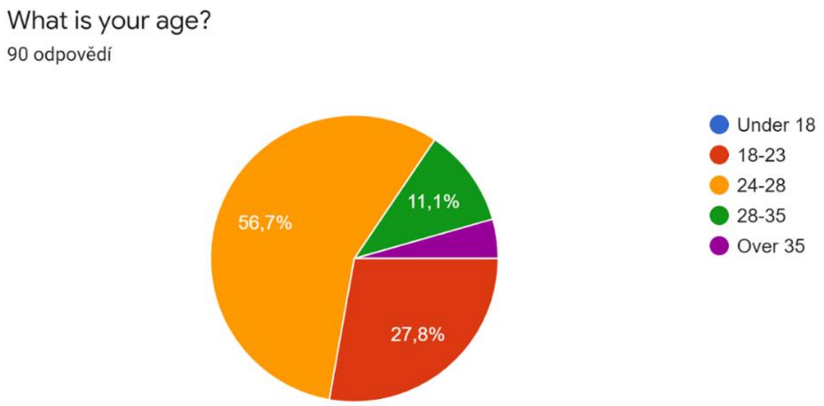

Figure 14: Age
Source: Own contribution

Question number 3: Are you currently....

Note: respondents could pick more than one answers

Possible answers:

- Studying
- Employed
- Part-time job
- Running own business

By this question, we wanted to find out which of our respondents are having most of their modern enterprise experiences from their studies, who is already employed, who is having a part-time job and who is running his own business.

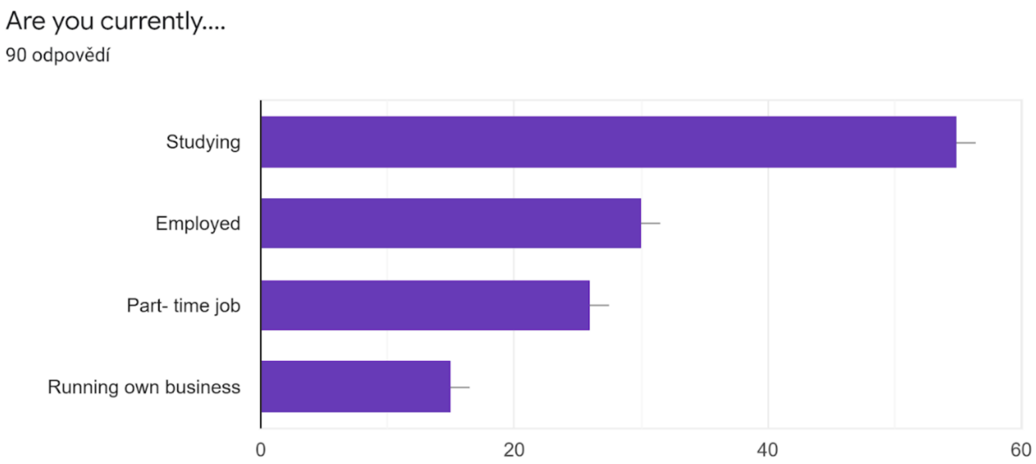

Figure 15: Employment Status
Source: Own contribution

Question number 4: Do you know what "business intelligence" stands for?

Possible answers:

- Yes
- No

The main purpose of this question was to find out how many people are actually familiar with connections between modern enterprises and business intelligence and whether they actually know about this term.

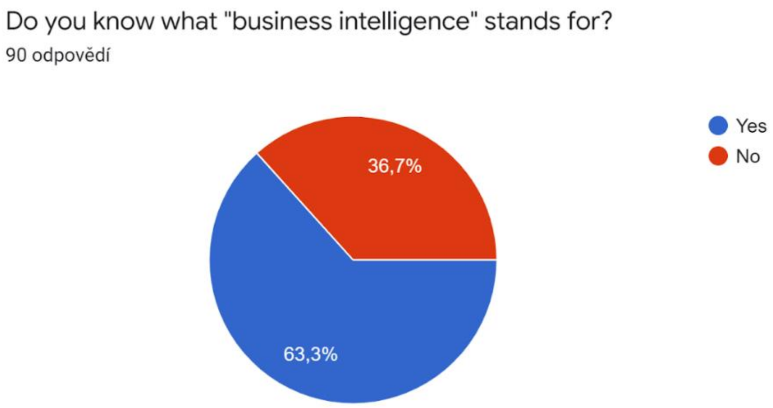

Figure 16: Business Intelligence Knowledge
Source: Own contribution

Unfortunately, thanks to those answers we found out that even though all the respondents do have some experience in modern enterprises problematics 36,7 % of them do not actually know about the connections between modern enterprises and business intelligence itself. This was very shocking and very unexpected. Nevertheless, it will still be analysed later- on in section "Regression analysis of questions number 4 and number 6" and in section "Analysis without statistics".

Question number 5: Do you have any experience with business intelligence?

Possible answers:

- Yes
- No

By this question, we found out what percentage of our respondents have already worked with business intelligence tools or with business intelligence in general. In some way, it was an extension of question number 4. If the respondents do know what business intelligence stands for, did they actually work with business intelligence tools?

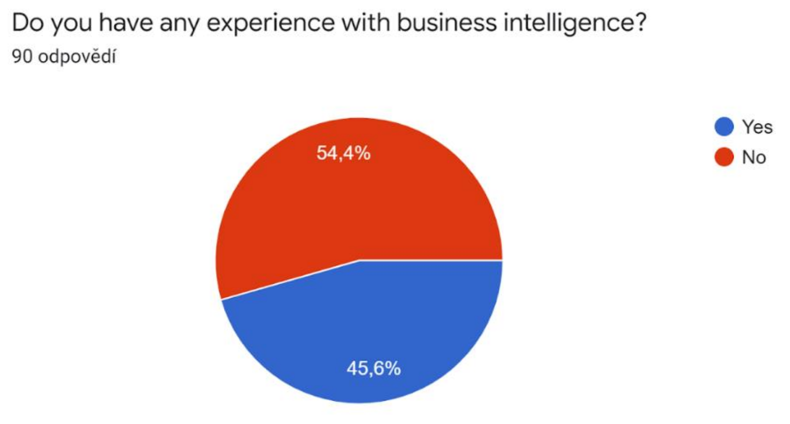

Figure 17: Experience with Business Intelligence
Source: Own contribution

These results were pretty much expected. 63,3 % of our respondents did know therm business intelligence but just 45,6 % of them actually worked with it. The majority of our respondents never worked with business intelligence tools, or they just did not know about it.

Question number 6: *Do you find business intelligence as a crucial part of today´s companies?*

Possible answers:

- Yes
- No

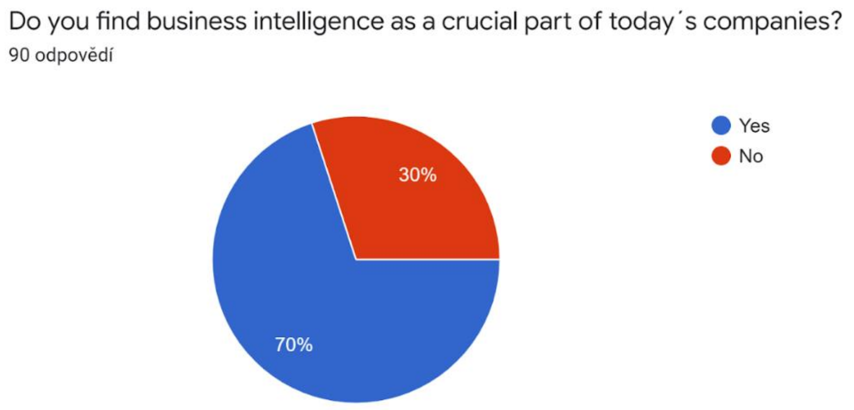

Figure 18: Business Intelligence as a Crucial Part
Source: Own contribution

This question is later-on analysed in the section "Regression analysis of questions number 4 and number 6" and in the section "Analysis without statistics" for a deeper look.

Question number 7: *Are you working in the public or private sector?*

Possible answers:

- Public
- Private
- I am not working

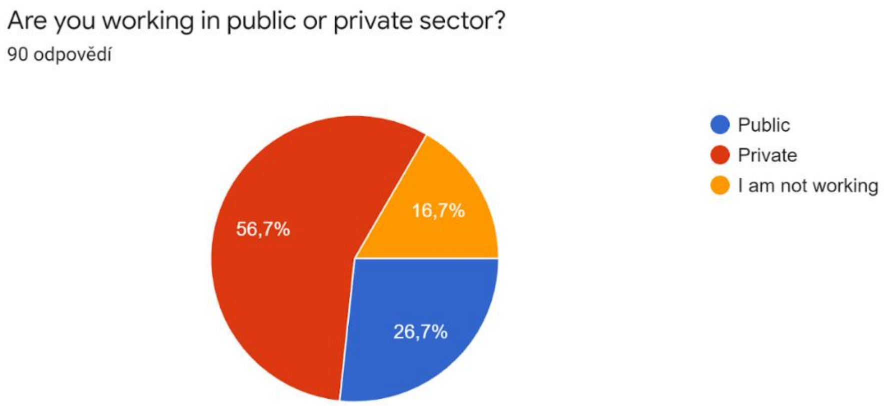

Figure 19: Sector of Employment
Source: Own contribution

These results were expected. We can see that 56,7 % of our respondents work in the private sector, 26,7 % in the public sector and 16,7 % are not working. Those who are not working are most likely just students.

Question number 8: *Which is the domain of the company or organization you are employed?*

Note: respondents could pick more than one answers

Possible answers:

- Banking
- Insurance
- Education
- Health
- Manufacturing
- Other

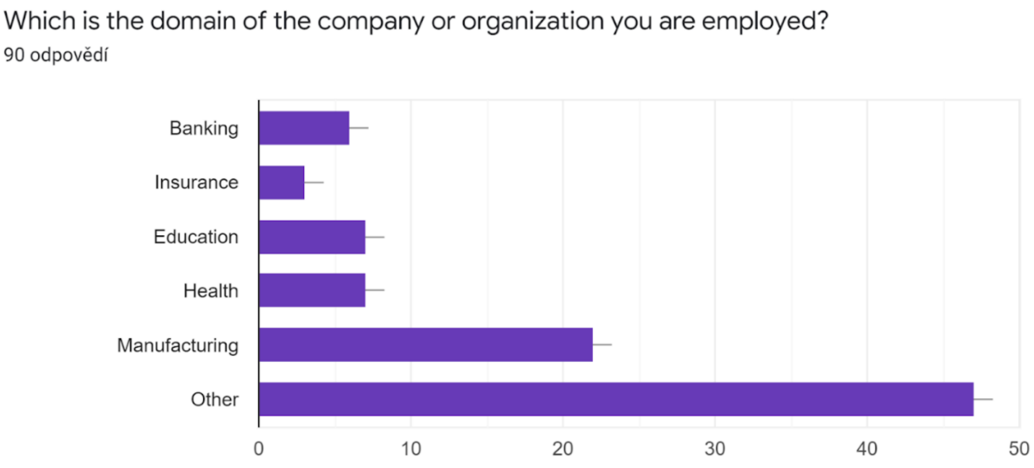

Figure 20: Domain
Source: Own contribution

The majority of answers in section “other” were expected and section “manufacturing” taking second place was expected as well but we also wanted to see how many of respondents are working in the education sector or health because these sectors are using business intelligence tools as well such as databases of patients, students, or customers in general.

Question number 9: *Are you working with business intelligence systems?*

Possible answers:

- Yes
- No

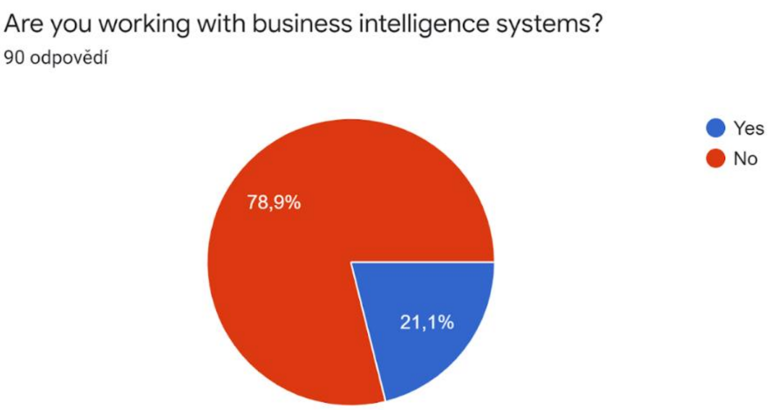

Figure 21: Business Intelligence Involvement
Source: Own contribution

This result was surprising. A much bigger number of people answering “Yes” was expected especially because the majority of sectors of any kind are using at least databases which are part of business intelligence. Our expectations were either very out of range or many people do not know what is hiding under the term business intelligence systems”.

Question number 10: *How many years of experience do you have as BI employee (if any).*

Possible answers:

- None
- < 2 years
- 2 - 5 years
- 5 - 10 years
- > 10 years

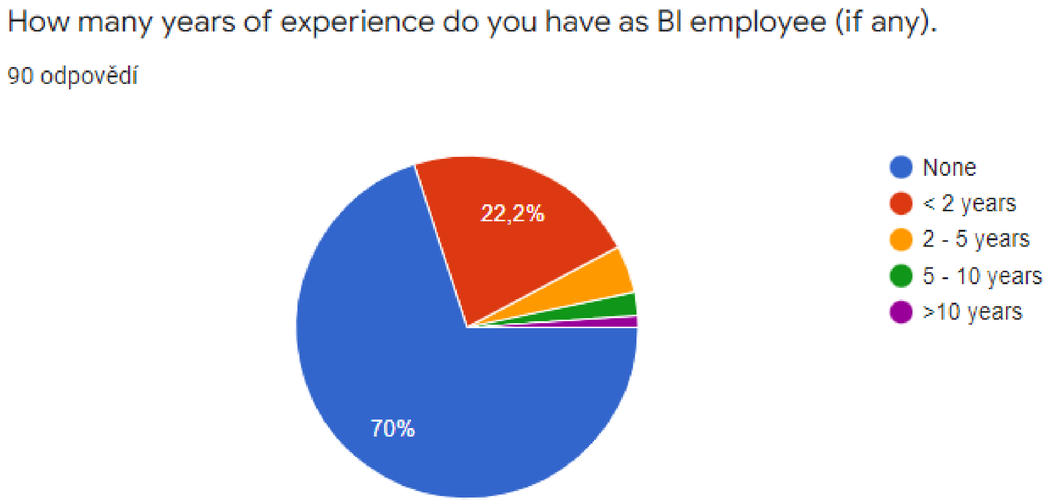

Figure 22: Years of Experience with Business Intelligence
Source: Own contribution

From this graph we can see that 70 % of our respondents do not have any experience as BI employees, 22,2 % of respondents have some small experience, 4,4 % of respondents do actually have more experience in BI, 2,2 % of respondents are very experienced, and 1 respondent has over 10 years of experience in business intelligence.

Question number 11: Would you consider becoming a BI specialist?

Possible answers:

- Yes
- No

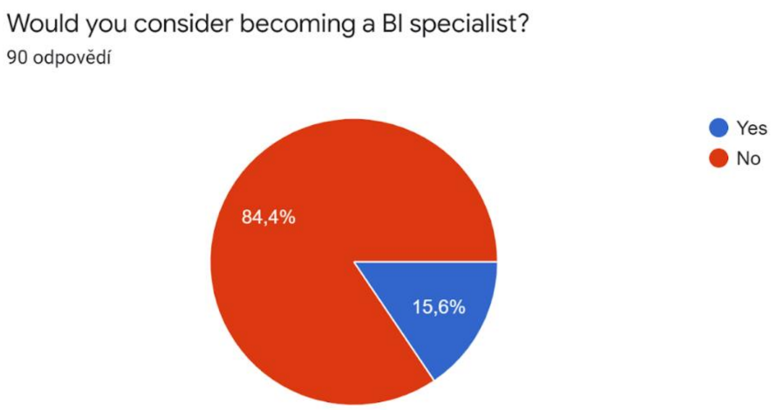

Figure 23: Business Intelligence Specialist - Interest
Source: Own contribution

Thanks to this question we found out that 14 respondents would actually consider becoming business intelligence specialists. This number is very pleasing.

Question number 12: *Do you have any knowledge with databases and data mining?*

Possible answers:

- Yes
- No

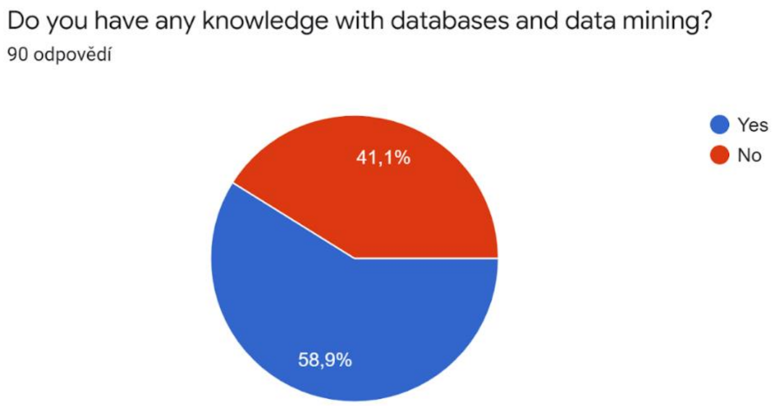

Figure 24: Knowledge of Databases and Data Mining
Source: Own contribution

58,9 % of our respondents do have some knowledge in data mining or with databases. That is a pretty big number even though the databases are part of almost every modern enterprise system. Thanks to that we know that the majority of our respondents are not just using databases, but they do actually know more about it.

Question number 13: Would you attend a seminar to acquire some knowledge in data mining or database tools?

Possible answers:

- Yes
- No

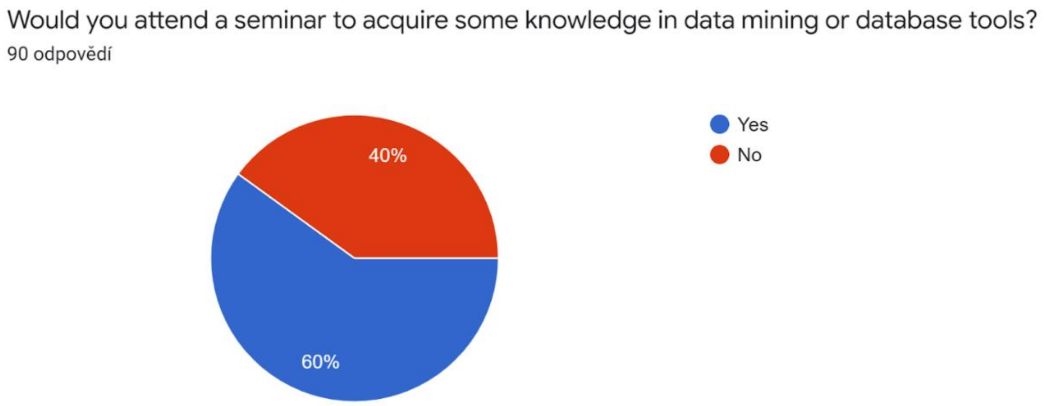

Figure 25: Interest in Business Intelligence Seminars
Source: Own contribution

The majority of our respondents are also willing to gain more knowledge in data mining and database tools.

Question number 14: *Name some reasons why would you attend this type of seminar…*

Note: respondents could pick more than one answers

Possible answers:

- Better salary
- Better working challenge
- More working opportunities
- Data analysis is the future
- Improve my skills
- Obtain better decisions

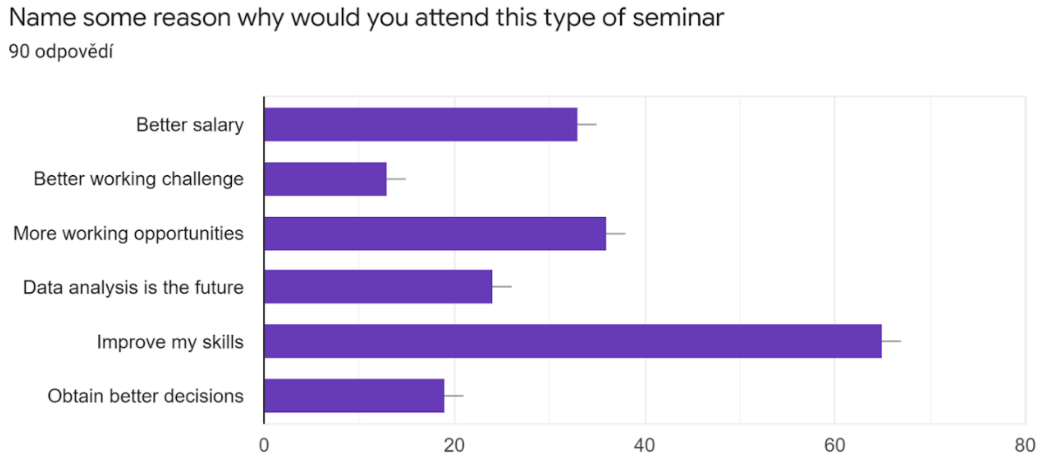

Figure 26: Reasons for Attendance in Seminars
Source: Own contribution

The last question gave us an overview of the reasons why respondents would attend this type of seminar. To be honest, the most expected reason was a better salary, but this reason actually took the third place. Most of our respondents would attend the seminar for improving their skills. The second place took more working opportunities, on the fourth place is data analysis is the future, obtaining better decisions is on fifth place and last place took the better working challenge.

3.1.1 Logistic Regression

For finding the most important attribute, logistic regression was used.

accuracy 0.8064516

Precision

[1] 0.7738095
Recall
[1] 0.7323232

f1-score

[1] 0.7472826

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.7815     1.0629   1.676 0.093719 .
AGE          -1.1866     0.5647  -2.101 0.035610 *
BI            2.6944     0.7703   3.498 0.000469 ***
```

Results

By stars, we can see the importance of the attribute for the question whether our respondents find BI as a crucial part of modern enterprises. As we can see age did not play such a big role but most important was knowledge in BI. Logistic regression was made with the supervisor of the thesis.

Limitations

It is not ideal to use logistic regression when all of our input variables are categorical and none of them are continuous. Also, duplicates were not removed because the dataset would be way too small. So, for this type of data, the regression model is not that great option. Therefore, the following analysis are going to be made.

3.1.2 Chi-Square Test

For the questions „Do you know what business intelligence stands for“ and „Do you find business intelligence as a crucial part of today´s companies“ was analysed by the Chi-Square test.

Variable 1= BI_Experience (yes, no), Variable 2 = BI_Usefull (yes, no)

The frequency of paired values table:
> table(dataset$BI_EXP, dataset$BI_VAL)

HO: Independence between two variables
H1: Dependence between two variables

p-value = 1.962e-06 (see below) so we reject H0

```
     0   1
 0  25  24
 1   2  39
> chisq.test(dataset$BI_EXP, dataset$BI_VAL, correct=FALSE)

    Pearson's Chi-squared test

data: dataset$BI_EXP and dataset$BI_VAL
X-squared = 22.632, df = 1, p-value = 1.962e-06
```

Results

This test was made for dependency between those two variables. We rejected H0 and accepted H1, so these two variables are indeed dependent. A Chi-square test was made with the supervisor of the thesis.

3.1.3 Association Rules

For association rules analysis was used Rstudio with aRules and aRulesViz packages. Code for analysis can be seen below and graphs of output will be presented by Figure 27 and Figure 28.

Code to install (always # are comments for the reader):

```
#association rules

install.packages("arules")
library("arules")
install.packages("arulesViz")
library("arulesViz")

#RULES 1: from all data we want to report the top 10 (most popular) rules with minimum
SUPPORT=0.01 and min Confidence=0.5 or 50% but when output is BI_Usefulness=YES
```

#THE SCRIPT IS:

BI_USEFULNESS_rules1 <- apriori(data=dataset, parameter=list (supp=0.01,conf = 0.5), appearance = list (rhs="BI_USEFUL=Yes"))
inspect(head(sort(BI_USEFULNESS_rules1, by = "confidence"), 10))
##RESULT

#RULES 2: from all data we want to report the top 10 (most popular) rules with minimum SUPPORT=0.01 and min Confidence=0.5 or 50% but when output is BI_Usefulness=NO
#THE SCRIPT IS:

BI_USEFULNESS_rules1 <- apriori(data=dataset, parameter=list (supp=0.01,conf = 0.5), appearance = list (rhs="BI_USEFUL=No"))
inspect(head(sort(BI_USEFULNESS_rules1, by = "confidence"), 10))
##RESULT (SEE ATTACHEMENT 10 MOST POPULAR RULES TXT)

By application of this analysis, we got these results for the 10 most popular rules with input being BI_USEFUL=Yes:

[1]

Lhs:	{EMPLOYMENT=Studying, BI_KNOWLEDGE=Yes}
Rhs:	{BI_USEFUL=Yes}
Support:	0.1014493
Confidence:	1.0000000
Coverage:	0.1014493
Lift:	1.500000
Count:	7

[2]

Lhs:	{AGE=18-23, BI_KNOWLEDGE=Yes}
Rhs:	{BI_USEFUL=Yes}
Support:	0.1304348
Confidence:	1.0000000
Coverage:	0.1304348
Lift:	1.500000
Count:	9

[3]	
Lhs:	{DOMAIN=Other, BI_EXP=Yes}
Rhs:	{BI_USEFUL=Yes}
Support:	0.1159420
Confidence:	1.0000000
Coverage:	0.1159420
Lift:	1.500000
Count:	8

[4]	
Lhs:	{BI_KNOWLEDGE=Yes, DOMAIN=Other, BI_EXP=Yes}
Rhs:	{BI_USEFUL=Yes}
Support:	0.1159420
Confidence:	1.0000000
Coverage:	0.1159420
Lift:	1.500000
Count:	8

[5]	
Lhs:	{BI_EXP=Yes}
Rhs:	{BI_USEFUL=Yes}
Support:	0.3623188
Confidence:	0.9259259
Coverage:	0.3913043
Lift:	1.388889
Count:	25

[6]	
Lhs:	{BI_KNOWLEDGE=Yes, BI_EXP=Yes}
Rhs:	{BI_USEFUL=Yes}
Support:	0.3623188
Confidence:	0.9259259
Coverage:	0.3913043
Lift:	1.388889
Count:	25

[7]	
Lhs:	{AGE=24-28, BI_EXP=Yes}
Rhs:	{BI_USEFUL=Yes}
Support:	0.1739130
Confidence:	0.9230769
Coverage:	0.1884058
Lift:	1.384615
Count:	12

[8]

Lhs:	{AGE=24-28, BI_KNOWLEDGE=Yes, BI_EXP=Yes}
Rhs:	{BI_USEFUL=Yes}
Support:	0.1739130
Confidence:	0.9230769
Coverage:	0.1884058
Lift:	1.384615
Count:	12

[9]

Lhs:	{AGE=24-28, BI_KNOWLEDGE=Yes, DOMAIN=Other}
Rhs:	{BI_USEFUL=Yes}
Support:	0.1304348
Confidence:	0.9000000
Coverage:	0.1449275
Lift:	1.350000
Count:	9

[10]

Lhs:	{DOMAIN=Manufacturing, BI_EXP=Yes}
Rhs:	{BI_USEFUL=Yes}
Support:	0.1159420
Confidence:	0.8888889
Coverage:	0.1304348
Lift:	1.333333
Count:	8

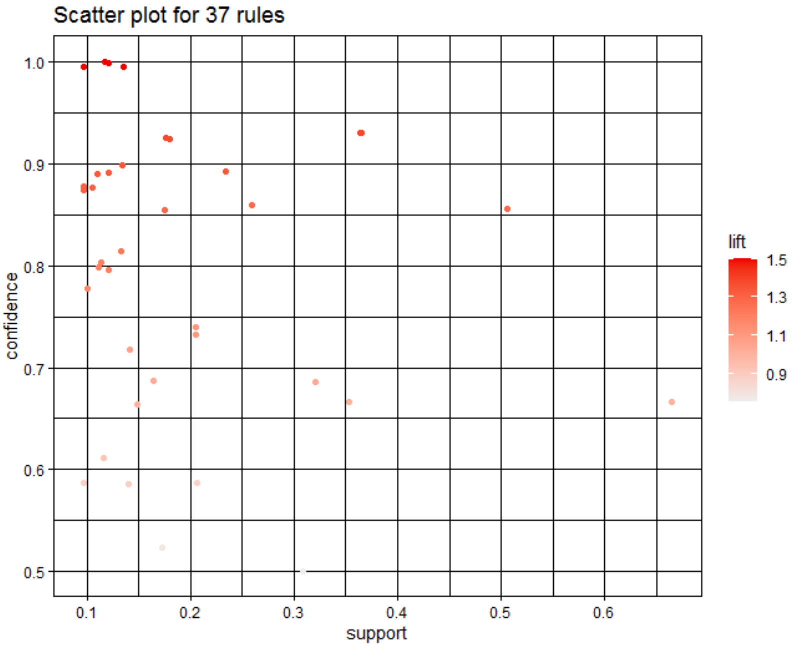

Figure 27: Association Rules 1
Source: Own contribution

Another 7 rules we received when input was BI_USEFUL=No:

#but here the output is BI_Usefulness=NO, minimum confidence=0.5
#code/script
BI_USEFULNESS_rules1 <- apriori(data=dataset, parameter=list (supp=0.1,conf = 0.5), appearance = list (rhs="BI_USEFUL=Yes"))
inspect(head(sort(BI_USEFULNESS_rules1, by = "confidence"), 10))

[1]

Lhs:	{BI_KNOWLEDGE=No}
Rhs:	{BI_USEFUL=No}
Support:	0.2463768
Confidence:	0.6071429
Coverage:	0.4057971
Lift:	1.821429
Count:	17

[2]

Lhs:	{BI_KNOWLEDGE=No, BI_EXP=No}
Rhs:	{BI_USEFUL=No}
Support:	0.2463768
Confidence:	0.6071429
Coverage:	0.4057971
Lift:	1.821429
Count:	17

[3]

Lhs:	{AGE=24-28, BI_KNOWLEDGE=No}
Rhs:	{BI_USEFUL=No}
Support:	0.1304348
Confidence:	0.6000000
Coverage:	0.2173913
Lift:	1.800000
Count:	9

[4]

Lhs:	{AGE=24-28, BI_KNOWLEDGE=No, BI_EXP=No}
Rhs:	{BI_USEFUL=No}
Support:	0.1304348
Confidence:	0.6000000
Coverage:	0.2173913
Lift:	1.800000
Count:	9

[5]

Lhs:	{BI_KNOWLEDGE=No, DOMAIN=Other}
Rhs:	{BI_USEFUL=No}
Support:	0.1159420
Confidence:	0.5714286
Coverage:	0.2028986
Lift:	1.714286
Count:	8

[6]

Lhs:	{BI_KNOWLEDGE=No, DOMAIN=Other, BI_EXP=No}
Rhs:	{BI_USEFUL=No}
Support:	0.1159420
Confidence:	0.5714286
Coverage:	0.2028986
Lift:	1.714286
Count:	8

[7]

Lhs:	{BI_EXP=No}
Rhs:	{BI_USEFUL=No}
Support:	0.3043478
Confidence:	0.5000000
Coverage:	0.6086957
Lift:	1.500000
Count:	21

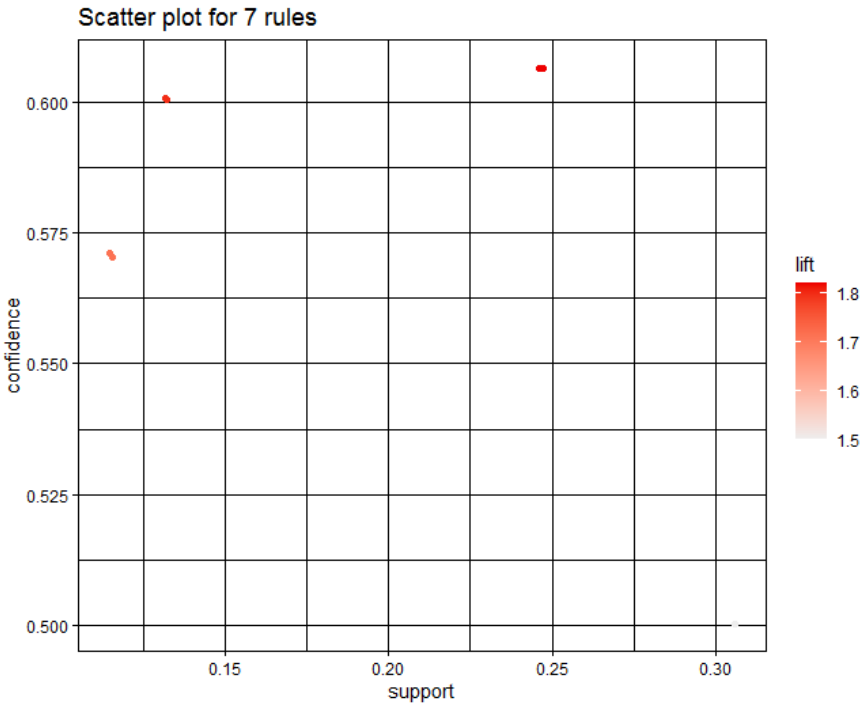

Figure 28: Association Rules 2
Source: Own contribution

Results

Biggest confidence we received from rule number 1 for output “BI_USEFUL=Yes”. This rule ensures us that those respondents who are currently studying and have some BI knowledge are very-likely going to know the true importance of business intelligence for modern enterprises. Also, other rules justified our logistic regression. Every rule is showing us that people who do have knowledge in BI or do have experience with BI are most likely going to know the true importance of BI and those who do not have any experience and knowledge do not understand Bi´s importance. It might sound obvious that people who do not have knowledge about it, will not find it important but after finding out that some people with connection to the modern enterprises told us in the questionnaire that they do not know what business intelligence stands for, which was very surprising, we had to make sure that our predictions were right by using statistical analysis. Association rules analysis was made with the supervisor of the thesis.

3.1.4 Analysis without Statistics

If we take a look at the table without a statistical point of view, we can easily calculate that 89 % of people who answered “Yes” to “Do you know what “business intelligence” stands for?” also answered “Yes” to “Do you find business intelligence as a crucial part of today´s companies?”. On the other hand, just 36 % of people who answered “No” to “Do you know what “business intelligence” stands for?” answered “Yes” to “Do you find business intelligence as a crucial part of today´s companies?”. From this, we can see a true relationship between these two questions, and we can say without a doubt that answers to the question “Do you find business intelligence as a crucial part of today´s companies?” were affected by answers to the question “Do you know what “business intelligence” stands for?”.

For a better overview two graphs for the question “Do you find business intelligence as a crucial part of today´s companies?” were created. The first one (Figure 29) for those who answered “Yes” to the question “Do you know what “business intelligence” stands for?”

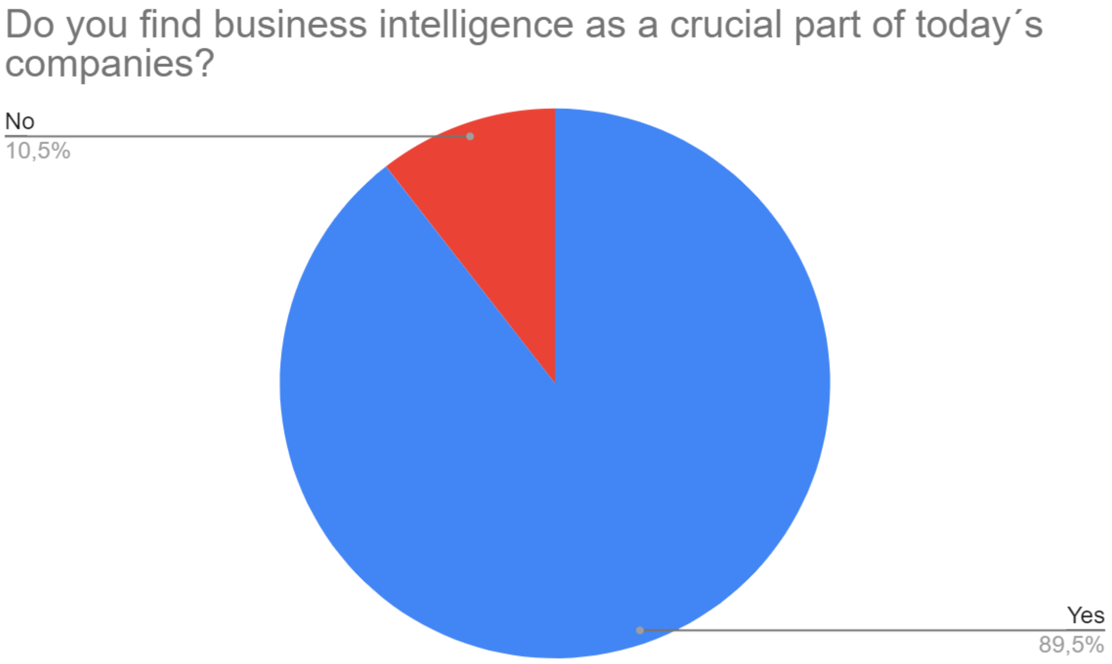

Figure 29: Analysis for Answers Yes
Source: Own Contribution

The second graph (Figure 30) for those who answered “No” to the question “Do you know what “business intelligence” stands for?”

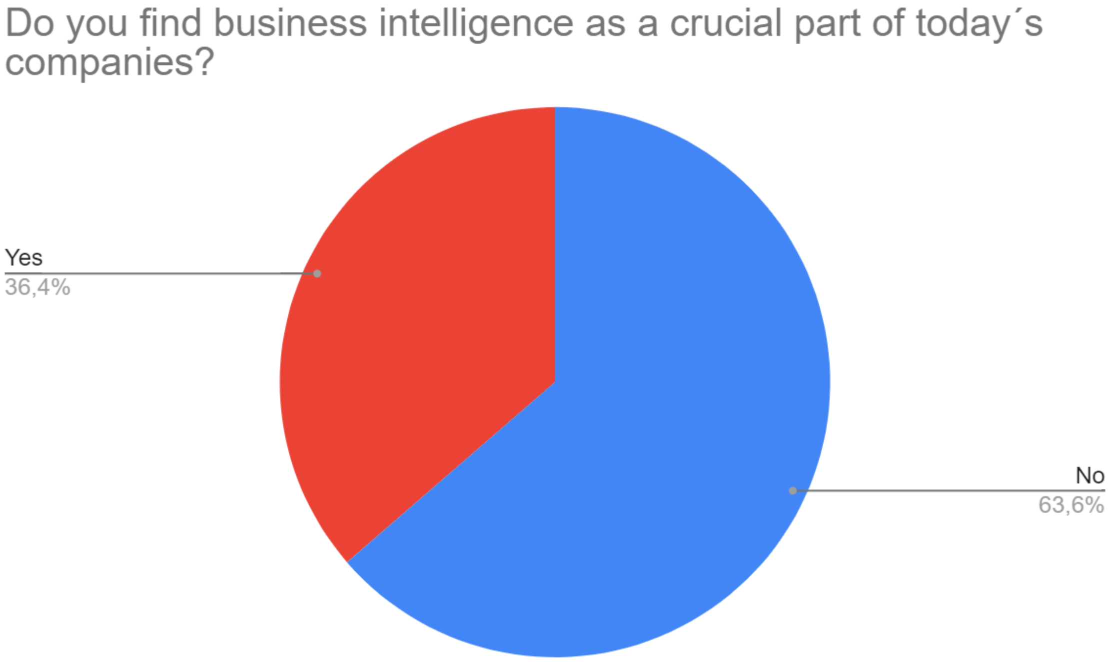

Figure 30: Analysis for Answers No
Source: Own Contribution

Calculations:

$51/57*100 \doteq 89$

Equation (3)

$12/33*100 \doteq 36$

Equation (4)

4. Summary and Further Recommendation

From the answers to the questionnaire, it is obvious that even employees or people who have something in common with modern enterprises do not know everything about business intelligence tools or business intelligence itself. Some of them are interested in acquiring more knowledge in this problematic and they are willing to attend seminars of this kind.

The main aim of this questionnaire was to find what is the level of knowledge of our respondents in the case of business intelligence and some value-added information about every specific respondent and that was fulfilled.

For further evaluation, analyses were made. The main aim of the analysis was to find whether there is a connection between question number 4 and number 6. This was important for us at the point where we wanted to see if people who do not have deeper knowledge in business intelligence will still know how important it is for modern enterprises or if only the people who have bigger knowledge in business intelligence will truly understand its importance and potential. The predictions were, unfortunately, true and people with less knowledge did not understand the importance of business intelligence in modern enterprises.

The future of business intelligence lies on the knowledge of modern enterprise employees and business intelligence professionals therefore they should be trained, and seminars should be provided to every single person who is interested in it. In the opinion of the diploma thesis's author, the knowledge about business intelligence should be spread among modern enterprise employees and the requirements for this kind of knowledge should be higher.

For future research, we would recommend including more not- categorical questions for the bigger potential of evaluation and statistical analysis. Also, the specific age of respondents could be better than just the scales.

Conclusion

The aim of the thesis was to tell the audience about business intelligence technologies, talk about business intelligence in general, describe specific aspects of business intelligence and most importantly give arguments why business intelligence is part and parcel of every modern enterprise. Thanks to the description of each business intelligence tool, we were able to show how it affects enterprises in every possible aspect and why it is important to constantly learn about either old or new technologies.

With the technologies and aspects being described the analysis comes into place. The survey was made by usage of the questionnaire. This questionnaire was spread among the respondents. After getting 90 answers, the questionnaire was closed. We had to close the questionnaire because we needed enough time for analysation of the answers. Then we were looking for key elements of answers. After finding those elements the analysis could start.

We evaluated the answers using logistic regression, Chi-square test and association rules analysis. The outcome was helpful and showed us where we should turn our attention to. We found that many respondents did not know about business intelligence at all even though they were familiar with the terms of modern enterprises.

In the end evaluation of questions gave us an overview about the knowledge of our respondents about business intelligence, their age, the domain of work, their experience and their interest in gaining new knowledge.

References

AltexSoft, 2019. A Complete Guide to Data Visualization in Business Intelligence: Problems, Libraries, and Tools to Integrate, Free Data Visualization Tools [online]. AltexSoft [cit. 2022-04-7]. Available on: https://www.altexsoft.com/blog/data-visualization-tools-types-techniques/

BANK, Standard a POSTHUMUS, Rudi, 2008 GIS as a tool in business intelligence [online]. PositionIT [cit. 2022-04-12]. Available on: https://www.ee.co.za/wp-content/uploads/legacy/AppT-GIS%20as%20a%20tool.pdf

CHENG, Lei a Peng CHENG. Integration: Knowledge Management and Business Intelligence. In: *2011 Fourth International Conference on Business Intelligence and Financial Engineering* [online]. IEEE, 2011, 2011, s. 307-310 [cit. 2022-04-14]. ISBN 978-1-4577-1541-9. Available on: doi:10.1109/BIFE.2011.172

DALIANIS, Hercules. Evaluation Metrics and Evaluation. DALIANIS, Hercules. *Clinical Text Mining* [online]. Cham: Springer International Publishing, 2018, 2018-05-15, s. 45-53 [cit. 2022-05-15]. ISBN 978-3-319-78502-8. Available on: doi:10.1007/978-3-319-78503-5_6

Data-flair, 2018. Chi-Square Test in R [online]. DataFlair [cit. 2022-04-14] Available on: https://data-flair.training/blogs/chi-square-test-in-r/

DESJARDINS, Jeff, 2018. Mapped: Population Density With a Dot For Each Town [online]. Visual Capitalist [cit. 2022-05-20]. Available on: https://www.visualcapitalist.com/mapping-population-density-dot-town/

FOLEY, Éric a Manon G. GUILLEMETTE, 2010 What is Business Intelligence? International Journal of Business Intelligence Research [online]. 1(4), 1-28 [cit. 2022-04-24]. ISSN 1947-3591. Available on: doi:10.4018/jbir.2010100101

GALARNYK, Michael, 2018. Understanding Boxplots [online]. Towards Data Science [cit. 2022-04-22]. Available on: https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51

GARG, Anisha, 2018. Complete guide to association rules [online]. Towards Data Science [cit. 2022-04-22]. Available on: https://towardsdatascience.com/association-rules-2-aa9a77241654

HE, Li, Guan XIN a Gong YUFENG, 2008. Study on the Design of CRM System Based on Business Intelligence. In: First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008) [online]. IEEE, 2008, 2008, s. 185-189 [cit. 2022-05-02]. ISBN 978-0-7695-3090-1. Available on: doi:10.1109/WKDD.2008.107

IBM, 2020. Logistic regression [online]. IMB [cit. 2022-06-05] Available on: https://www.ibm.com/topics/logistic-regression

KOHTAMÄKI, Marko, 2017. Real-Time Strategy and Business Intelligence: Digitizing Practices and Systems. California: Palgrave Macmillian. ISBN:978-3-319-54845-6

LUTKEVICH, Ben, 2020. Association rules [online]. TechTarget [cit. 2022-05-02]. Available on: https://www.techtarget.com/searchbusinessanalytics/definition/association-rules-in-data-mining

MARCEL, Patrick, and Esteban ZIMÁNYI, 2017. Business intelligence. Cham: Springer International Publishing. ISBN: 978-3-319-61163-1.

microTOOL, 2016. What is a business case [online]. microTOOL [cit. 2022-04-02] Available on: https://www.microtool.de/en/knowledge-base/what-is-a-business-case/

BARTSCHAT, Andreas, Markus REISCHL a Ralf MIKUT, 2019. Data mining tools. WIREs Data Mining and Knowledge Discovery [online]. 9(4) [cit. 2022-03-05]. ISSN 1942-4787. Available on: doi:10.1002/widm.1309

MURIYIL, Anilkumar, Thekkan, 2021. How can business intelligence improve business growth? [online]. euclidz.ai [cit. 2022-04-10]. Available on: https://euclidz.ai/blog/how-business-intelligence-improve-business-growth/

NEGASH, Solomon a Paul GRAY. Business Intelligence. BURSTEIN, Frada a Clyde W. HOLSAPPLE. Handbook on Decision Support Systems 2 [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, 2008, s. 175-193 [cit. 2022-06-1]. ISBN 978-3-540-48715-9. Available on: doi:10.1007/978-3-540-48716-6_9

NEGASH, Solomon, 2004. Business Intelligence. Communications of the Association for Information Systems [online]. 13 [cit. 2022-05-05]. ISSN 15293181. Available on: doi:10.17705/1CAIS.01315

OLAP.com, 2022 What is the definition of OLAP? [online]. OLAP.com [cit. 2022-04-10]. Available on: https://olap.com/olap-definition/

POWER, Daniel, J. 1997. What is a DSS? [online]. DSSResources.COM, 1.3: 223-232 [cit. 2022-04-21]. Available on: https://dssresources.com/subscriber/password/papers/whatisadss.pdf

PROQUEST, 2021. Databáze článků ProQuest [online]. Ann Arbor, MI, USA: ProQuest. [cit. 2021-09-26]. Available on: http://knihovna.tul.cz

RAMAGERI, Bharati, M., 2010. Data mining techniques and applications [online]. Indian Journal of Computer Science and Engineering 1(4) [cit. 2022-04-21]. ISSN: 0976-5166 Available on: https://www.researchgate.net/publication/49616224_Data_mining_techniques_and_applications

SYKES, Alan, O. 1993. An Introduction to Regression Analysis [online]. Coase-Sandor Institute for Law & Economics Working Paper No. 20 [cit. 2022-04-25]. Available on: https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1050&context=law_and_economics

TAYLOR, David, 2022. What is Data Warehouse? Types, Definition & Example [online]. Guru99 [cit. 2022-04-07]. Available on: https://www.guru99.com/data-warehousing.html

TAYLOR David, 2022. What is OLAP? Cube, Analytical Operations in Data Warehouse [online]. Guru99 [cit. 2022-04-07]. Available on: https://www.guru99.com/online-analytical-processing.html

VAISMAN, Alejandro a Esteban ZIMÁNYI. Data warehouse systems: design and implementation. Heidelberg: Springer, [2014]. Data-centric systems and applications. ISBN 978-3-642-54654-9.

List of Appendices

Appendix A - Commercial Tools for Data Mining

Program	Category	Link
ADAPA (Zementis)	Data mining suites	www.zementis.com
Alice (d'Isoft)	Data mining suites	www.alice-soft.com
Bayesia Lab	Specialities	www.bayesia.com
C5.0	Specialities	www.rulequest.com
CART	Specialities	www.salford-systems.com
Data Applied	Data mining suites	data-applied.com
DataDetective	Data mining suites	www.sentient.nl/?dden
DataEngine	Data mining suites	www.dataengine.de
Datascope	Data mining suites	www.cygron.hu
DB2 Data Warehouse	Business intelligence packages	www.ibm.com/software/data/infosphere/warehouse
DeltaMaster	Business intelligence packages	www.bissantz.com/deltamaster
Forecaster XL	EXT	www.alyuda.com
GhostMiner	Data mining suites	www.fqs.pl/business intelligence/products/ghostminer
IBM Cognos 8 BI	Business intelligence packages	www.ibm.com/software/data/cognos/data-mining-tools.html
IBM SPSS Modeler	Data mining suites	www.spss.com/software/modeling/modeler
IBM SPSS Statistics	Mathematical packages	www.spss.com/software/statistics
iModel	Data mining suites	www.biocompsystems.com/products/imodel

InfoSphere Warehouse	Business intelligence packages	www.ibm.com/software/data/infosphere/warehouse
JMP	Data mining suites	www.jmpdiscovery.com
KnowledgeMiner	Specialities	www.knowledgeminer.net
KnowledgeStudio	Data mining suites	www.angoss.com
KXEN	Data mining suites	www.kxen.com
Magnum Opus	Specialities	www.giwebb.com
MATLAB	Mathematical packages	www.mathworks.com
MATLAB Neural Network Toolbox	EXT	www.mathworks.com
Model Builder	Data mining suites	www.fico.com
ModelMAX	Solutions	www.asacorp.com/products/mmxover.jsp
Molegro Data Modeler	Solutions	www.molegro.com
NAG Data Mining Components	Data mining libraries	www.nag.co.uk/numeric/DR/DRdescription.asp
NeuralWorks Predict	Specialities	www.neuralware.com/products.jsp
Neurofusion	Data mining libraries	www.alyuda.com
Neuroshell	Specialities	www.neuroshell.com
Oracle Data Mining (ODM)	Data mining suites	www.oracle.com/technology/products/bi/odm/index.html
Partek Discovery Suite	Data mining suites	www.partek.com/software
Partek Genomics Suite	Solutions	www.partek.com/software

PolyAnalyst	Data mining suites	www.megaputer.com/polyanalyst.php
PolyVista	Business intelligence packages	www.polyvista.com
Random Forests	Specialities	www.salford-systems.com
RapAnalyst	Specialities	www.raptorinternational.com/rapanalyst.html
R-PLUS	Mathematical packages	www.experience-rplus.com
SAP Netweaver Business Warehouse (BW)	Business intelligence packages	www.sap.com/platform/netweaver/components/businesswarehouse
SAS Enterprise Miner	Data mining suites	www.sas.com/products/miner
See5	Specialities	www.rulequest.com
SPAD Data Mining	Data mining suites	eng.spadsoft.com
SQL Server Analysis Services	Data mining suites	www.microsoft.com/sql
STATISTICA	Data mining suites	www.statsoft.com/products/data-mining-solutions/G259
SuperQuery	Data mining suites	www.azmy.com
Teradata Database	Business intelligence packages	www.teradata.com
Think Enterprise Data Miner (EDM)	Data mining suites	www.thinkanalytics.com
TIBCO Spotfire	Data mining suites	spotfire.tibco.com
Unica Predictive Insight	Data mining suites	www.unica.com
WizRule and WizWhy	Specialities	www.wizsoft.com

XAffinity	Specialities	www.exclusiveore.com

Source: MIKUT, Ralf and REISCHL, Markus, 2011

Appendix B -Table of Free and Open-Source Tools

Program	Category	Link
ADaM	Data mining libraries	datamining.itsc.uah.edu/adam
CellProfilerAnalyst	Solutions	www.cellprofiler.org/index.htm
D2K	Data mining suites	alg.ncsa.uiuc.edu
Gait-CAD	Integration packages	sourceforge.net/projects/gait-cad
GATE	Solutions	gate.ac.uk/download
GIFT	RES	www.gnu.org/software/gift
Gnome Data Mine Tools	Data mining suites	www.togaware.com/datamining/gdatamine
Himalaya	RES	himalaya-tools.sourceforge.net
ImageJ	Solutions	rsbweb.nih.gov/ij
ITK	Solutions	www.itk.org
JAVA Data Mining Package	Data mining libraries	sourceforge.net/projects/jdmp
JavaNNS	Specialities	www.ra.cs.uni-tuebingen.de/software/JavaNNS/welcome e.html
KEEL	Integration packages	www.keel.es
Kepler	Mathematical packages	kepler-project.org
KNIME	Integration packages	www.knime.org
LibSVM	Data mining libraries	www.csie.ntu.edu.tw/ cjlin/libsvm
MEGA	Solutions	www.megasoftware.net/m distance.html
MLC++	Data mining libraries	www.sgi.com/tech/mlc
Orange	Data mining libraries	www.ailab.si/orange
Pegasus	RES	www.cs.cmu.edu/ Pegasus

Pentaho	Business intelligence packages	sourceforge.net/projects/pentaho
Proximity	Specialities	kdl.cs.umass.edu/proximity/index.html
PRTools	EXT	www.prtools.org
R	Mathematical packages	www.r-project.org
RapidMiner	Data mining suites	www.rapidminer.com
Rattle	Integration packages	rattle.togaware.com
ROOT	Data mining libraries	root.cern.ch/root
ROSETTA	Specialities	www.lcb.uu.se/tools/rosetta/index.php
Rseslibs	RES	logic.mimuw.edu.pl/ rses
Rule Discovery System	Specialities	www.compumine.com
RWEKA	Integration packages	cran.r-project.org/web/packages/RWeka/index.html
TANAGRA	Integration packages	eric.univ-lyon2.fr/ ricco/tanagra/en/tanagra.html
Waffles	Data mining libraries	waffles.sourceforge.net
WEKA	Data mining suites, data mining libraries	sourceforge.net/projects/weka
XELOPES Library	Data mining libraries	www.prudsys.de/en/technology/xelopes
XLMiner	EXT	www.resample.com/xlminer

Source: MIKUT, Ralf and REISCHL, Markus, 2011

Appendix C – Data for Analysis

Do you know what "business intelligence" stands for?	Do you find business intelligence as a crucial part of today´s companies?
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes

Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	No
Yes	No
Yes	No
Yes	No
Yes	No
Yes	No
No	No
No	No
No	No
No	No
No	No
No	No

No	No
No	No
No	No
No	No
No	No
No	No
No	No
No	No
No	No
No	No
No	No
No	No
No	No
No	No
No	No
No	Yes
No	Yes
No	Yes
No	Yes
No	Yes
No	Yes
No	Yes
No	Yes
No	Yes
No	Yes
No	Yes
No	Yes

Source: Own contribution

Appendix D - Questionnaire

1. What is your gender?
- Male
- Female
2. What is your age?
- Under 18
- 18-23
- 24-28
- 28-35
- Over 35
3. Are you currently....
- Studying
- Employed
- Part-time job
- Running own business
4. Do you know what "business intelligence" stands for?
- Yes
- No
5. Do you have any experience with business intelligence?
- Yes
- No
6. Do you find business intelligence as a crucial part of today´s companies?
- Yes
- No
7. Are you working in public or private sector?
- Public
- Private
- I am not working
8. Which is the domain of the company or organization you are employed?
- Banking
- Insurance
- Education
- Health
- Manufacturing
- Other
9. Are you working with business intelligence systems?
- Yes
- No
10. How many years of experience do you have as BI employee (if any).
- None
- < 2 years
- 2 - 5 years
- 5 - 10 years
- >10 years
11. Would you consider becoming a BI specialist?
- Yes
- No
12. Do you have any knowledge with databases and data mining?

- Yes
- No

13. Would you attend a seminar to acquire some knowledge in data mining or database tools?

- Yes
- No

14. Name some reason why would you attend this type of seminar

- Better salary
- Better working challenge
- More working opportunities
- Data analysis is the future
- Improve my skills
- Obtain better decision