

Erlis Marko

Data Mining and Analysis

Metropolia University of Applied Sciences

Electronics Engineering

Bachelor in Electronics

Thesis

16 May 2018

Author Title	Erlis Marko Data Mining and Analysis
Number of Pages Date	22 pages + 2 appendices 16 May 2018
Degree	Bachelor of Engineering
Degree Programme	Degree Programme in Electronics
Professional Major	-
Instructors	Janne Mäntykoski, Senior Lecturer, Metropolia UAS Aki Heikkinen, Project Manager
<p>This work was made for Roche Oy. Roche Oy is one of the leading pharmaceutical and diagnostic companies in the world.</p> <p>The purpose of this project was to create a database for sales department of the company, which will be used by the employees in the field enabling them to report in real time their expenses. The database should be easy and friendly for the users, to minimize the errors in the calculation of the monthly and annually budget for each employee. The project supervisors of each team will have the right to follow their employees' expenses and approve these expenses every month. Every four months, the project leaders should be able to subtract from the database the necessary information needed to complete the report.</p> <p>During the project different tools were used to adapt to the company's environment. In the end the tools used for this project were Google Spreadsheet-Data Studio and Excel with Power BI.</p> <p>The project is presented in a different way from what is used in the company. Some of the codes and the names of the products, employees are changed or not present. The end result was satisfactory, and the company started to use the platform built.</p>	
Keywords	Business Intelligence, Data Mining, Google Sheet JavaScript

Contents

1	Introduction	1
2	Data Mining	2
3	Real-time Analytics	3
4	Business Intelligence (BI)	4
5	Client and the Company's Requirements	4
5.1	The Company Environment	4
5.2	Scope	5
5.3	Initial Approach	5
5.3.1	Access	7
5.3.2	Tableau Desktop	8
5.3.3	Google Spreadsheet	8
5.3.4	Data Studio	9
5.3.5	Excel	10
5.3.6	Power BI	11
6	Roche project	12
6.1	Roche Google Spreadsheet	13
7	Conclusion	21
8	Data Analysis in Industry	17
8.1	EKE –Smart Vision TM	17
	References	23

Appendices

Appendix 1: Google spreadsheet code

Appendix 2: Pictures

List of Abbreviations

BI Business intelligence. Software used to analyze data and presenting the information.

CEO Chief Executive Officer

1 Introduction

During the last decades we have seen a fast growth of computer and smart phone usage all over the world. These devices are part of our life nowadays and we use them in school, at work, during our free time or when we communicate with other people which can be in the same city or in another country. Through these devices we can get the information that we need very fast, and very cheap.

Computer are adapted by the companies since the early stages of the technology. The employees use these devices to search or collect information and reported to their managers faster and safer. In general managers and CEOs must make a lot of hard decision for they employees or products future. For making better decisions companies like to collect as much data as they can on products, clients, market, and their own expenses so in the end they can make the right decisions. The information should be spread as well to right employees, and it should be understandable by all the employees in charge of the project regardless of their background.

By collecting almost all the data that is available we tend to get a lot of information that are not always easy to be read and understood, and very difficult to be used. To make better decision for the company and the products these data should be seen and analyzed. Making possible the visualization of this information we need to use different techniques and methods.

The goal of this project was to design and create an environment that can be used by almost all the sales department employees of the company, where they can add, modify and check the data. Collect the data added from the employees and subtract the required information. The environment should be easily accessible by all the employees and supervisors and should fulfill all the requirements set by the managers.

2 Data Mining

Data mining gets the name from mining vast sets of information in databases or warehouses. The idea is the same as when we are mining metals from minerals deposit in Earth's crust. In both cases we will have to use different tools and apply different methods to get the necessary result.

In the last decades our society had been shifted toward a digital society and economy. Right now, we are living in the information age, where more and more data is generated in every aspect of our lives, and saved somewhere in some database of a company. We create data when we eat, shop, go out with friends, watch TV, go to the doctor or surf the web. This increase in data storage is a result of technology advance and adaptation of this technology by the society. There is enormous data that is already saved and stored in some database. This data in general contains important information that are not always discovered. To make possible the absorption of all the necessary information we use a method called data mining.

Data mining is the technique that is related to exploring and sorting interesting large sets of data to identify patterns and understand and predict future trends. Data mining can bring out new information and reveal new different insights that are invisible to the eyes, and it can be used as the tool that can help to interpret these data.

Data mining can help the companies, researchers and individuals to understand trends, predict customer needs, detect fraud, study illness conditions etc. To get the required and wanted data, we will have to use different mathematical and computational algorithms that will search through these large data for patterns and get the essential knowledge [1].

Most of the data that we collect are evidences that we live in the world. For example, online transactions, temperature, steps, grades, when we get paid and start spending the money, these are simple numbers or data that we have. When we start to process this data, we start to see useful information that we can use in the future. The information that we get from the data that we have collected is what we call knowledge. An example of how the data is turned into information and knowledge can be found in figure 1. The knowledge can be use in the future for better decisions in business, school, sport, traffic, technology etc.

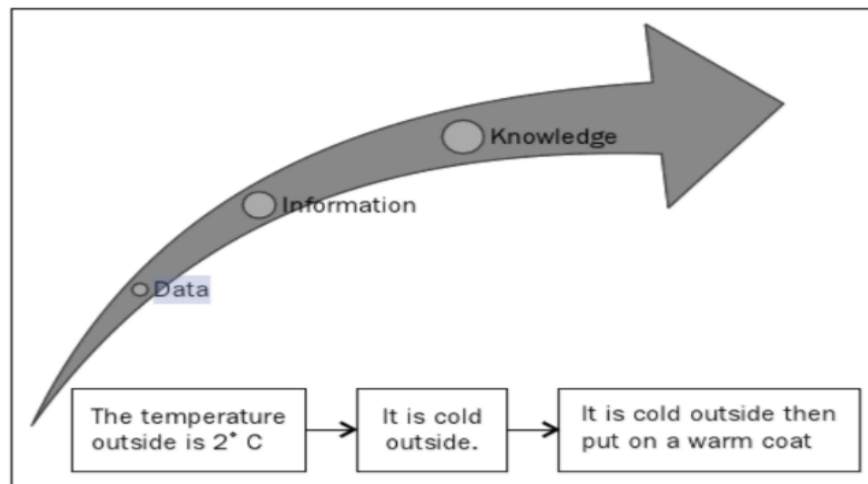


Figure 1. [3] From data to knowledge.

Data analysis is referred as the process of compiling and analyzing data in a way that it is able to clarify what has happened in the past and forecast what will happen in the future. Data analysis is not only about the numbers that we have, but is also about making the right questions and trying to find the right answers. [3]

3 Real-time Analytics

In this digital world people are generating more and more data every day. Similarly, individuals also companies generate a lot of data. Before the companies were able to create daily, weekly, monthly and in the end yearly reports, but in this highly competitive digital world this can not be possible anymore. In today's world everything has to do with time, and we should be able to visual the data as soon as possible. Some data can be stored in databases and used in the future, but there are also some data that has to be turned into information instantly to get the benefits out of it. For this process we use real-time analytics. Real-time analytics allows companies and individuals to see in real time the information, as soon as the data enters the system.

Nowadays, the implementation of real-time analytics can be combined with cloud-based environments allowing companies to access the information from all over the world. For example, a company can have the factory in another continent and the headquarters in another continent. By applying real-time analytics, headquarter can access the information of the factory in real time. With the help of cloud-based environment this can be

combined with different factories permitting companies to save time and money. The same system can be applied as well to track products, and plan their productions based on this data. [4]

4 Business Intelligence (BI)

With the development of the technology, Business Intelligence (BI) has become gradually more and more important to the academics, business and to the manufacturing companies over the last decades. Business Intelligence is able to combine both analytics, data gathering and working data, to help individuals and companies in analyzing data and presenting practical information that can be used in the future to make decisions.

The main aim of the Business Intelligence is to help businesses, individuals and company to save time and money. The Business Intelligence tools give the capability to understand, and present changes that are happening inside our data. It allows the users to detect opportunities and risks that can come while dealing with other companies or clients. Business Intelligence can be used to keep track of our goals, follow production performance, analyze data, accomplish reports and support managers and decision makers to have the necessary information based on facts. Business Intelligence is not able to keep all the data on its own, but uses data that can be internal or external and is able to pull off information from different systems like: data mining, data warehouse, knowledge management etc. Through human interaction Business Intelligence is able to convert the data that pulls off from other systems into beneficial information that can be used in the future [6; 7].

5 Client and the Company's Requirements

5.1 The Company Environment

Roche Oy is multinational company operating in across more than 100 countries. The company is divided in two divisions, pharmaceutical and diagnostics. In Finland the company operates with both divisions. The pharmaceutical division is divided into smaller teams.

The sales department of the company has five teams, each team has in average six employees. Individually every employee has around 2-3 product that is responsible and different area, the areas are located in Finland. To create the budget every employee of the department will have to know the product that he has and the area where he/she will work. Every employee estimates the budget required for the product at the beginning of the year, the supervisor of each team will have to approve this budget. These data will be used in the future to compare the estimation made by the employee and the present situation. The same data will be used to create reports, visual dashboards and to make a simple and comprehensive representation.

The employees were divided in three categories. Category one, the ones that generate the data, meaning the field workers. Category two, the ones that need to see and check this data, meaning the supervisors. Category three, and the final group is for the managers who requires only reports and dashboards.

At the beginning of the project the company used Excel to keep track of their data. Every employee had an Excel sheet where they uploaded their data. The company did not have any visual view of the data and everything was just numbers.

5.2 Scope

The purpose of this project was to create a simple and easily used database for the employees of the companies, where they can insert the information related to the product. The database should be able to easily be accessed, modified, and informative for the employees and the supervisors.

5.3 Initial Approach

At the beginning of the project there were some meetings with the project manager and project supervisors. During these meetings current situation was analyzed, how the company collect, and uses the data. What kind of data they have, and what kind of data they will like to collect. In the meeting some of the company's and employee's requirements and needs were covered. The company had a list of requirements and goal that had to

been taken into the consideration during the design phase of the project. Some of the requirements set by the company were:

- Easily accessed by the employees
- Easily modifiable
- Easily analyzed
- Cheap
- Fast

During the meetings with the managers of the project different options and methods were discussed. To meet the requirements of the company some of the tools available in the market and some that are not available were covered. Some of the tools that were handled as an option for our project are listed below and compared:

Table 1. List of tools used and comparison.

	Price	Accessibility	Analyses	Visualization
Excel	Good	Good	Good	Good
Power BI	Good	Very Good	Very Good	Very Good
SAS	Good	Very Good	Very Good	Very Good
Access	Good	Good	Good	Good
Tableau	Good	Very Good	Very Good	Very Good
Google Spreadsheet	Very Good	Very Good	Very Good	Very Good
Data Studio	Good	Very Good	Very Good	Very Good

During the first weeks of the project almost all the tools listed on table 1 were used and tried to adapt with the company environment. Even though the timetable was tight and there was not too much time, different software were used and compared one to another to get a better understanding and adapted it to the environment of the company.

5.3.1 Access

Access is a database management system part of Microsoft office family. Access is a very powerful tool that allows individuals to create collect, maintain and arrange data according to their needs. The structure of Microsoft Access makes it easy to select, sort and represent the data. Access helps you to analyze larger amounts of data, save them securely, and increase the productivity in work placement.

Access can work with all kind of data, from sales to address, phone numbers, employee's paycheck. Compared to another spreadsheet access it is friendlier, and it allows individual to get easily access to the information and give visual presentation of these data.

Databases that are formed in Access are made up of objects This means that in Access you are able to create the database one part at the time and create relationships between these parts, which makes possible the connections of different data into one database. In every table we can have cells that allow users to enter certain data and avoid other data. The relationship between different tables varies, it can be simple or very complicated. We can have three types of relationships for tables. They can be one to many, many to many and one to one. The relationships that we create depend on the tables, and data that we have. When the tables, cells, and the relationships are created we can use these data to withdraw the necessary information and represent it in a visual form. [2]

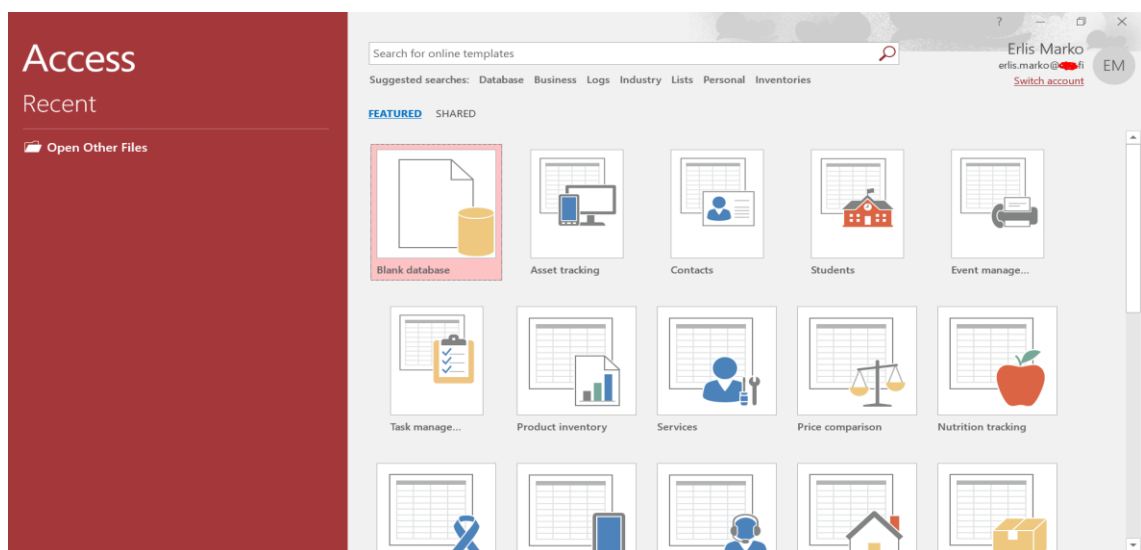


Figure 2. Access front page

5.3.2 Tableau Desktop

Tableau Desktop is a data visualization software that helps individuals and companies to see, comprehend and follow data that is available but in the same time invisible to the eye. The software is able to create fast dashboards and calculation, spot trends and identify opportunities. There are two editions available in the market, personal and professional edition.

The main idea of Tableau is to connect with other data sources and create a visualization representation of this data. Personal edition allows data to be connected with other software like Access, Excel, Odata, Tableau Data Extract. The professional edition is more for larger databases, SQL database, columnar analytics databases, data appliances.



Figure 3. Tableau Desktop [8]

5.3.3 Google Spreadsheet

Google spreadsheet is part of Google Drive family and one of three web-based software offered by Google. The program is similar to Windows Excel and uses row and column format as in Excel. It allows users to make numerical manipulation and create own code, functions and formulas. Google spreadsheet is compatible with Excel and CSV. The software allows its users to create, modify and share data with other users online. The software permits multiple users in the same time, and all the users are able to modify and

communicate with each other at the same time. All the files that are created are saved in Google's servers.

During the design phase was decided to work with google spreadsheet since it matches with most of the requirements set by the company like cheap, fast, easily modified and easily accessed. The company was familiar with Google Spreadsheet since it is used in other projects. Another bonus for using spreadsheet is the beta version of data analytics "Data Studio" by Google.

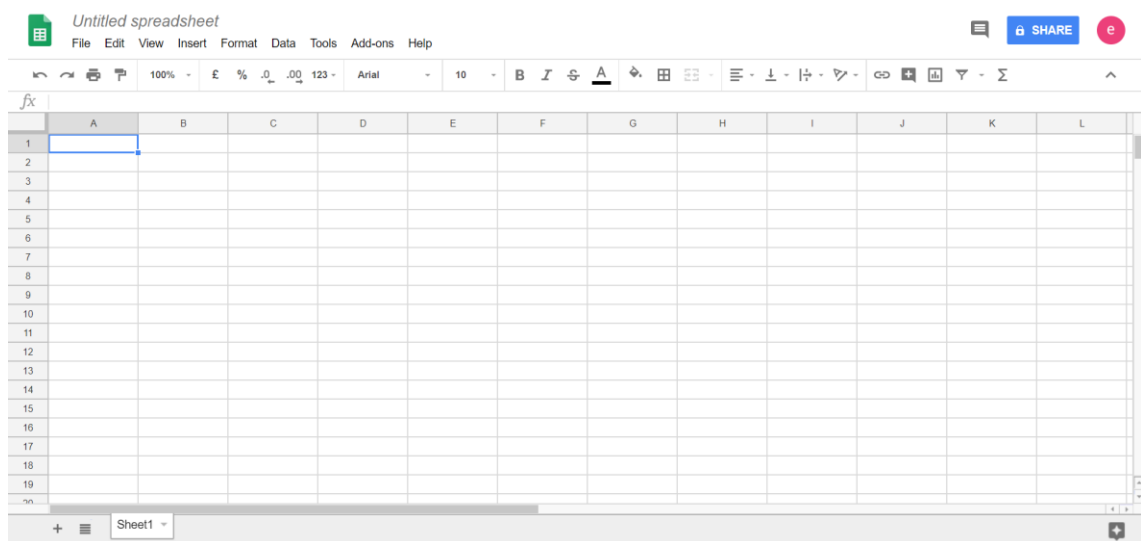


Figure 4. Google Spreadsheet

5.3.4 Data Studio

Data Studio is part of Google Drive family. The Data Studio is still in beta version but it can be used by the consumers. Data Studio is a software that allows its users to create reports and visual dashboards to analyze data on real time. Data studio can be as well easily connected to other sources like google sheet, Adword or databases like Bigquery or My Sql and Excel through Google Spreadsheet.

Since Data Studio is part of Google Drive family. It makes it easy to share, modify, and add tasks.



Figure 5. Data Studio

5.3.5 Excel

Excel is a software that allows users to create spreadsheets based on a column and row format. Due to its ability and flexibility to manipulate data, Excel has become one of the most popular and widely used software in the world. Excel is part of Microsoft Office family.

The columns are labeled with letters, and the rows are labeled with numbers. In this chess-table type of format that allows its users to make calculation, create pivot tables, build charts, and has a programming language called VBA (Visual Basic for Applications). One of the most important features of Excel is the creation of charts, this allows its users to transform data into visual information.

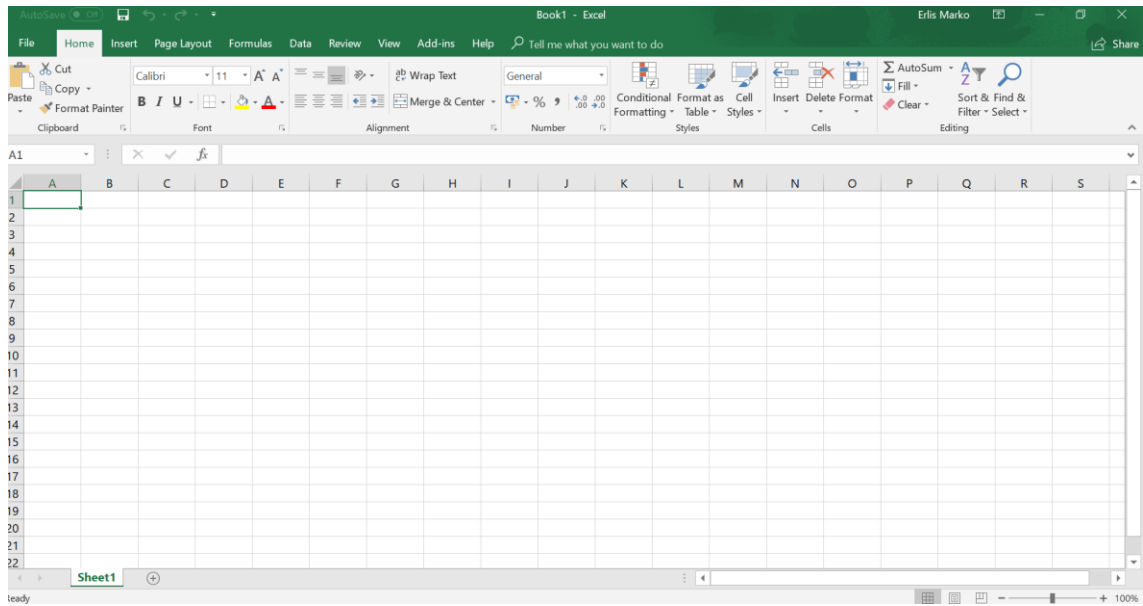


Figure 6. Excel

VBA or Visual Basic for application, is a very powerful programming language that allows its users to build their own functions or applications and adapt these applications to their systems.

Another good feature of Excel is the ability of the software to analyze and display data to the users. Some of the most notables features of Excel Data Analysis are sort, filter, conditional formatting, charts, tables, pivot table, what-if analysis, solver. Through these tools almost all the data entered in the software can be analyzed and understood.

5.3.6 Power BI

Power BI (Business Intelligence) is a cloud-based tool that permits individuals and companies to analyze and visualize raw data in a form of reports, charts, maps and dashboards, the software creates also reports. The software is developed by Microsoft and allows the users to access the information from any Windows, IOS, or Android device connected to the internet.

Through Power BI we get access to a wide range of data sources like: SQL, Excel, Access, SharePoint, Oracle Database, JSON, Text/CSV, SAP/HANA database, Amazon

Redshift, Azure SQL database, Google Analytics, Smartsheet, Odata Feed and R script. Merge the data from different sources, clean the data so we can have only the necessary information, and break it down into different columns and rows so that it can be comprehensible to reader. After all the steps the users will have to create the necessary dashboards and reports based on their needs. The information can be shared online through Azure – Microsoft cloud if necessary (Figure.7). [11]

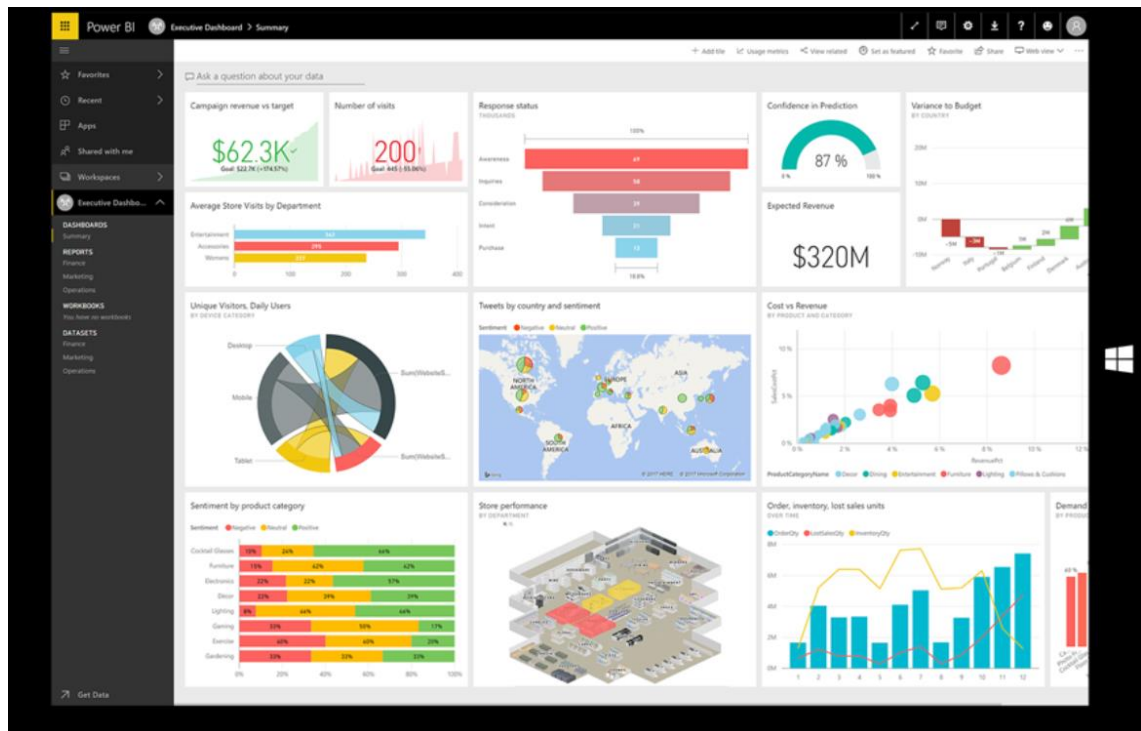


Figure 7. Power BI Dashboards [10]

6 Roche project

After the design phase, and after most of the options mention above were handled two software got closer to the goal and requirements of the company, Excel-Power BI and Google Spreadsheet and Data Studio. One of the reasons for choosing these software were because the company was familiar with Excel and Google Drive, the other reason was because they fulfilled most of the requirements set by the company at the beginning of the project.

6.1 Roche Google Spreadsheet

Since the company has about five teams, and each team has around six employees. In total there are 30 sales employees and every employee has about 2-3 products. To keep it simple one small group of employees and their information were used to create a beta version. This was a precaution that was taken to reduce the amount of data, since the amount of data used in the reality is larger.

First, some of the company's information was used and inserted in the software to make a tryout with the personnel. By creating groups and ranges with a team's information regarding their members and products.

The second phase was to create drop down boxes by using Google Spreadsheet script, the code can be found below (Appendix 1). The idea of the drop-down boxes was simple, each worker will choose his/her team, after they have chosen the team they will choose their name, then the product. GI account will come based on the product, based on the name Activity will come, and WBS will come as a result of the team in the end based on the WBS and GI account Product number will come. An example is part of the get data script (Appendix 1). For every product each employee will have to fill also the comment box to explain in detail how the money is used and select the city where they were. Next after the employees have filled up the details they will have to add the amount of money they are spending for each product. This amount will be compared to the amount each employee has set in the beginning of the year (Figure.8). A send data button is located close to month table to send the data to sheet 2.

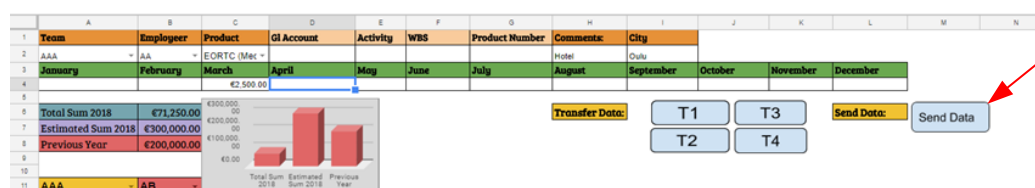


Figure 8. Front page, where the data is entered

The third phase was to create a database environment for supervisors and employees to check and compare the data available (Figure. 9). All the data entered can be found there and modified or cancelled until a certain date.

Team	Employer	Product	GI Account	Activity	WBS	Product Number	Comments	January	February	March	April	May	June	July	August
AAA	AA	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Hotel	€1,000.00							
AAA	AB	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Hotel		€2,000.00						
AAA	AC	Preceptoshp (Medical)	418116027	HCP registration fees pr	M145-AU1.1.05	MA_1454101003	Conference			€2,500.00					
BBB	BB	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Taxi				€50.00				
BBB	BA	ESMO (Medical)	418116003	Travel & Accommodation	M145-AU1.1.04	ME_1454106003	Hotel					€5,000.00			
BBB	BC	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Hotel			€5,000.00					
CCC	CC	ASCO (Medical)	418129030	Prom. giveaways / tax d	M145-AU1.1.01	ME_1454106003	Flight				€5,000.00				
CCC	CA	ESMO (Medical)	418116003	Travel & Accommodation	M145-AU1.1.04	ME_1454106003	Flight/Hotel						€1,200.00		
CCC	CB	Preceptoshp (Medical)	418116027	HCP registration fees pr	M145-AU1.1.05	MA_1454101003	Conference								
AAA	AA	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Hotel	€1,000.00							
AAA	AB	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Hotel		€2,000.00						
AAA	AC	Preceptoshp (Medical)	418116027	HCP registration fees pr	M145-AU1.1.05	MA_1454101003	Conference			€2,500.00					
BBB	BB	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Taxi				€50.00				
BBB	BA	ESMO (Medical)	418116003	Travel & Accommodation	M145-AU1.1.04	ME_1454106003	Hotel					€5,000.00			
BBB	BC	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Hotel			€5,000.00					
CCC	CC	ASCO (Medical)	418129030	Prom. giveaways / tax d	M145-AU1.1.01	ME_1454106003	Flight				€5,000.00				
CCC	CA	ESMO (Medical)	418116003	Travel & Accommodation	M145-AU1.1.04	ME_1454106003	Flight/Hotel						€1,200.00		
CCC	CB	Preceptoshp (Medical)	418116027	HCP registration fees pr	M145-AU1.1.05	MA_1454101003	Conference								
AAA	AA	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Hotel	€1,000.00							
AAA	AB	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Hotel		€2,000.00						
AAA	AC	Preceptoshp (Medical)	418116027	HCP registration fees pr	M145-AU1.1.05	MA_1454101003	Conference			€2,500.00					
BBB	BB	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Taxi				€50.00				
BBB	BA	ESMO (Medical)	418116003	Travel & Accommodation	M145-AU1.1.04	ME_1454106003	Hotel					€5,000.00			
BBB	BC	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Hotel			€5,000.00					
CCC	CC	ASCO (Medical)	418129030	Prom. giveaways / tax d	M145-AU1.1.01	ME_1454106003	Flight				€5,000.00				
CCC	CA	ESMO (Medical)	418116003	Travel & Accommodation	M145-AU1.1.04	ME_1454106003	Flight/Hotel						€1,200.00		
CCC	CB	Preceptoshp (Medical)	418116027	HCP registration fees pr	M145-AU1.1.05	MA_1454101003	Conference								
BBB	BB	ESMO (Medical)	418116003	Travel & Accommodation	M145-AU1.1.04	ME_1454106003	Hotel	€1,000.00							
BBB	BA	EDRTC (Medical)	418129031	Prom. giveaways / tax d	M145-AU1.1.04	MA_1454101003	Hotel		€5,000.00						

Figure 9. The total amount of data entered is on Sheet 2

The fourth phase was to create visual dashboards for the supervisors. The dashboards were supposed to be selective, meaning that the supervisor can select an employee and verify the budget (Figure. 10). One figure 10 team or employee can be chosen and verified. Other dashboards were created to present all the teams and all the employees.

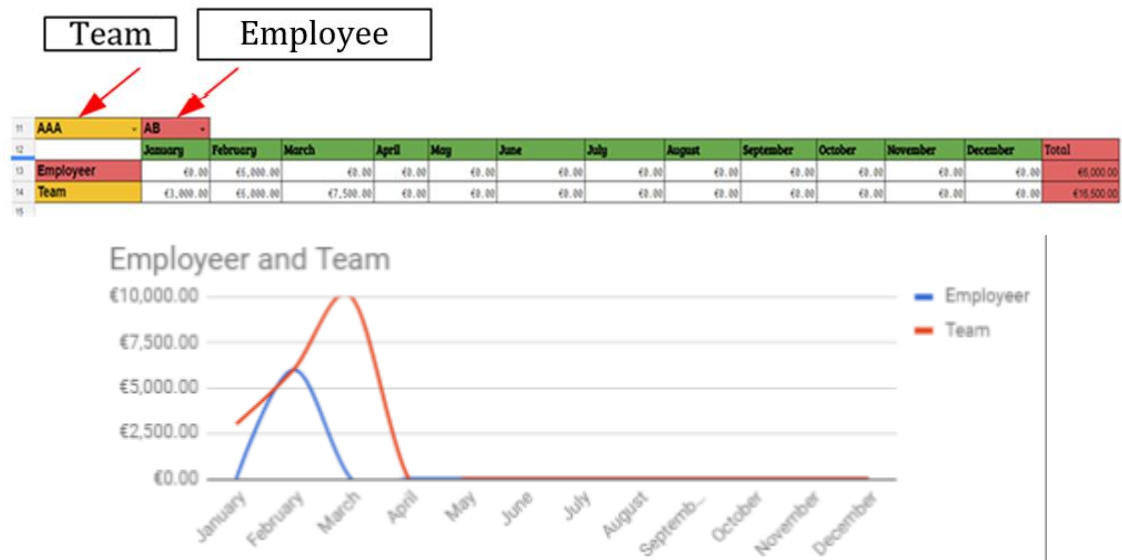


Figure 10. Selective Tables and Graph

To make it easier some tables were created so the supervisor can see the estimated budget which was created at the beginning of the year and the present budget (Figure.11 & Figure.13).

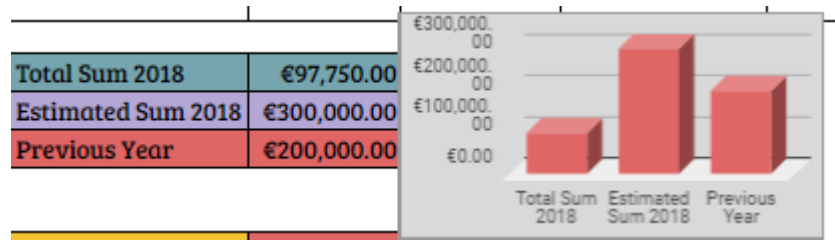


Figure 11

The fifth phase was to create other important features. A quarter button was created to send the quarters of the year to another special page and compare it to the previous year and the estimated budget (Figure12 and Figure 13, 14). The script for transferring the data can be found at appendix 1.

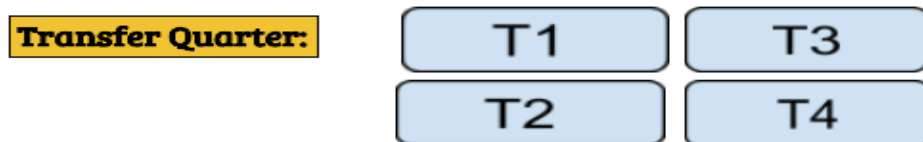


Figure 12. Quarters buttons

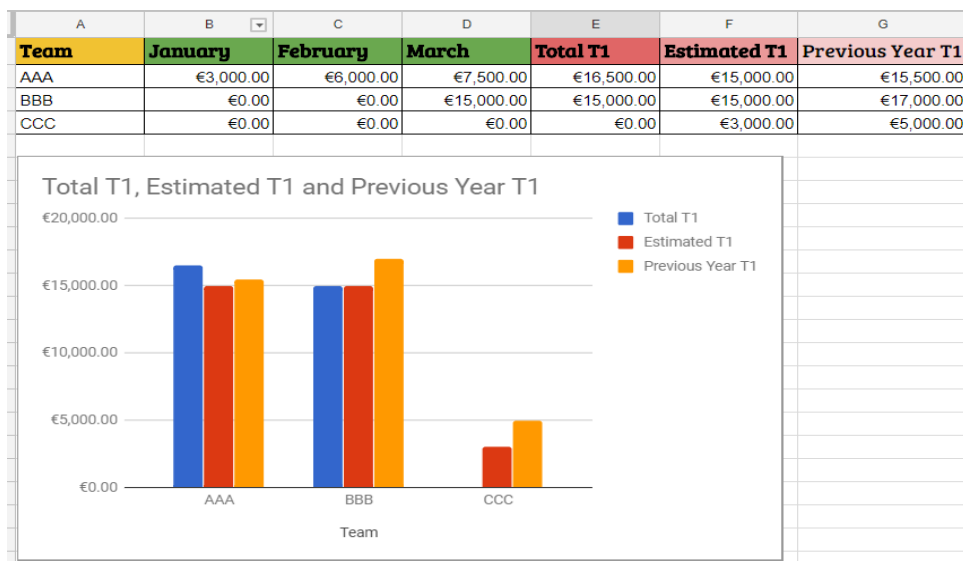


Figure 13. Comparing Teams for the first Quarter

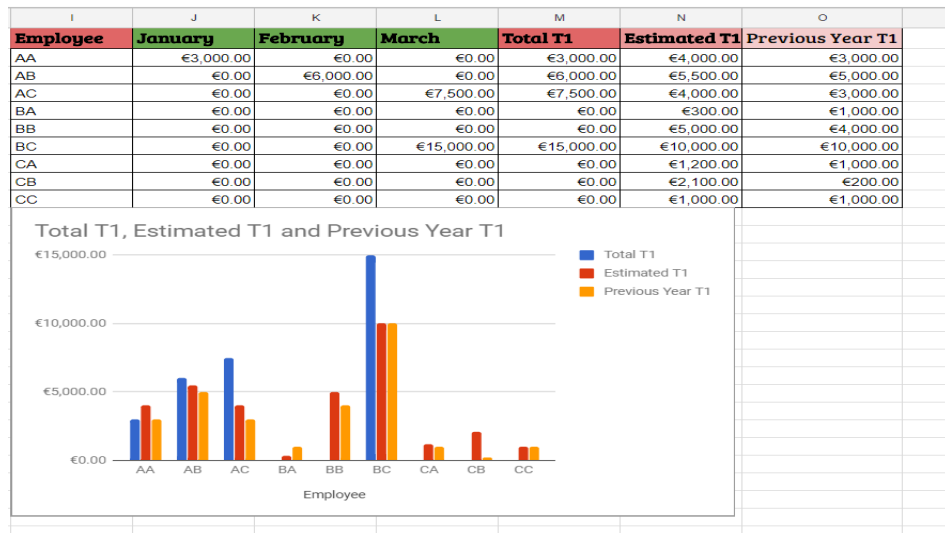


Figure 14. Comparing employees for the first quarter

An email page was created for supervisors to send different email to different people with different data (Figure.15). The email script is as an attachment in appendix 1. On the page we can choose the team or the employee and the right amount will come and the email will be sent by pressing “Sent Email”.

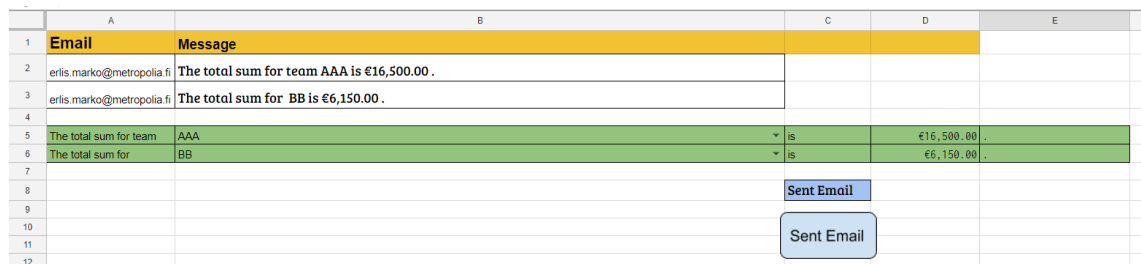


Figure 15. Email sheet, were emails can be sent to multiple users in the same time with different data.

On Figure 16 two email come in the same time to my email address with different message.

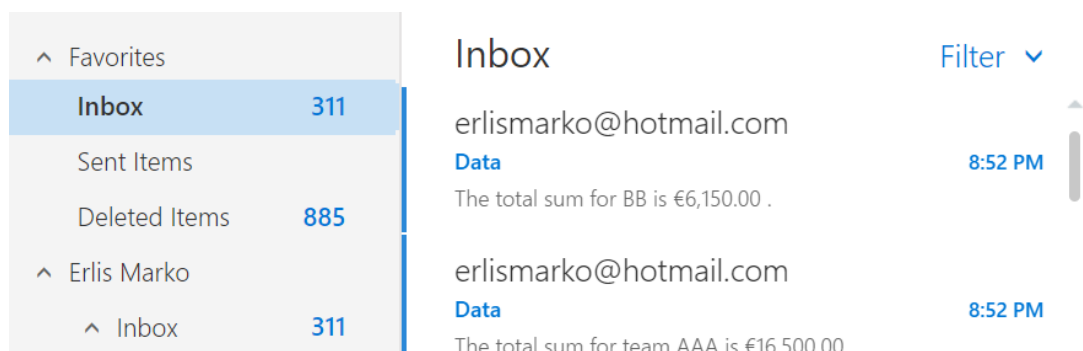


Figure 16. Example, emails with different data arrive in the same time.

Other pages were created to subtract GI account and the total, Product and the total, and another page was created for the product number with the monthly total. These pages were used in the future by another member of the company.

7 Data Analysis in Industry

Industry in general is a very competitive environment, facing competitions from local and international business. Most of the companies try to use all the available tools to improve their products, get more information and make better decisions before the competitors. Since most of the industries have a global network of suppliers and clients they create a lot of data. Keeping up and managing large amount of data is a lot of work and analysing. Business Intelligence can be used in all sectors of the industry, from the manufacturing to logistics to customer service, and can help its client to see in real time the data and the problems.

Business Intelligence tools are not the only tools available in the market. Larger companies have created their own Business Intelligence system adapted to their system and products. Different manufacturing companies are creating Business Intelligence software to follow their own products, and their international partners, which production line is more problematic; where is the product, which orders are overdue and can give a day to day accurate report of most of the activities in the company.

Other companies have created services for their clients to improve their products and help them to see in real time their products and predict the failure. A very good example of this is EKE-electronics' Smart Vision.

7.1 EKE –Smart VisionTM

Eke-electronics is one of the leading companies in providing smart solution for train manufactures and operators. The company delivers solutions and products for train automation, system integration, on board communication between coaches and with line controller and safety systems. The systems can be applied to high speed and commuter trains, metros, trams and to passenger coaches and locomotives.

EKE Smart Vision is a Multifleet Remote Condition Monitoring System or RCMS. The system is able to collect and analyse data in real time related to a certain fleet and help

its users to have a visual view of the fleet. The system improves maintenance and operation. The system is very easy to maintain and apply, and since the system is modular it allows different people to access different information based on the needs. Some of the application that can be implemented are:

- Driving assistance
- Train on map
- Predictive analytics
- Online monitoring
- Condition monitoring

Driving assistance

Driving assistance application allows the driver to be instructed on how to apply a mode which will optimize the energy consumption on the train. Driving assistance allows the drivers and the line controllers to get access to information like momentary speed, distance, time and deviation from schedule. Through driving assistance, the line controller is able to send messages and instruction to the drivers display. [12;13]

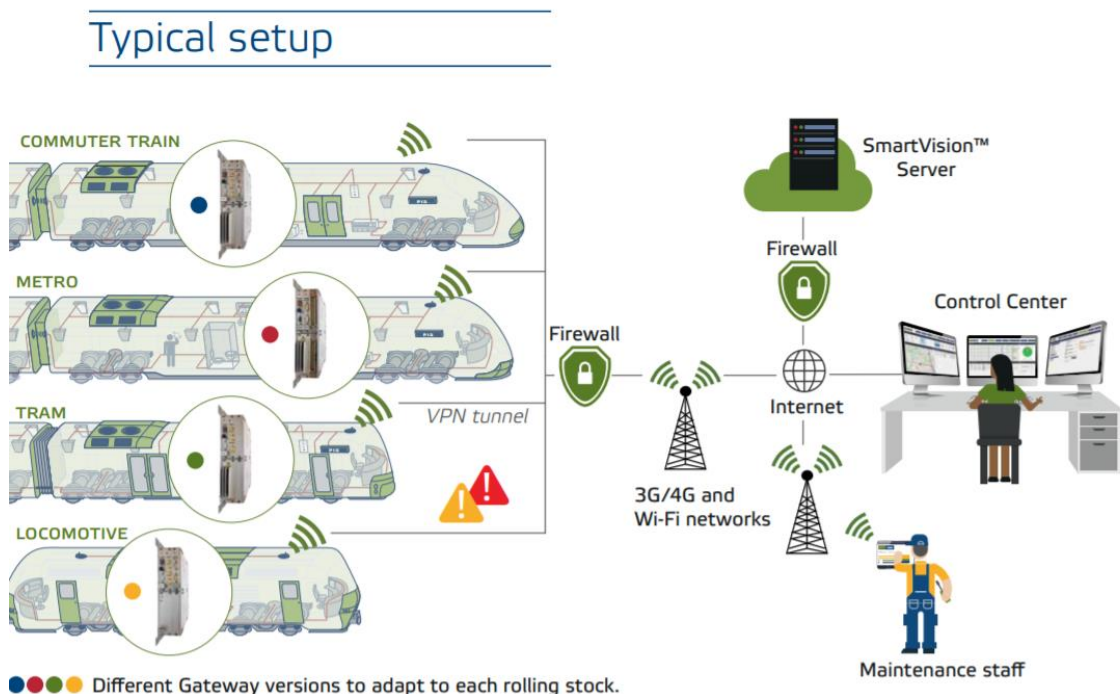


Figure. 17 Driving assistance

Train on map

Train on the map permits the line controller to have a visual view of the entire fleet on the monitor. Through this application the line controller can follow the trains and their delays. Important alarms and the situation of the train can be displayed on the map, information can be accessed by clicking on the train number. [12; 13]

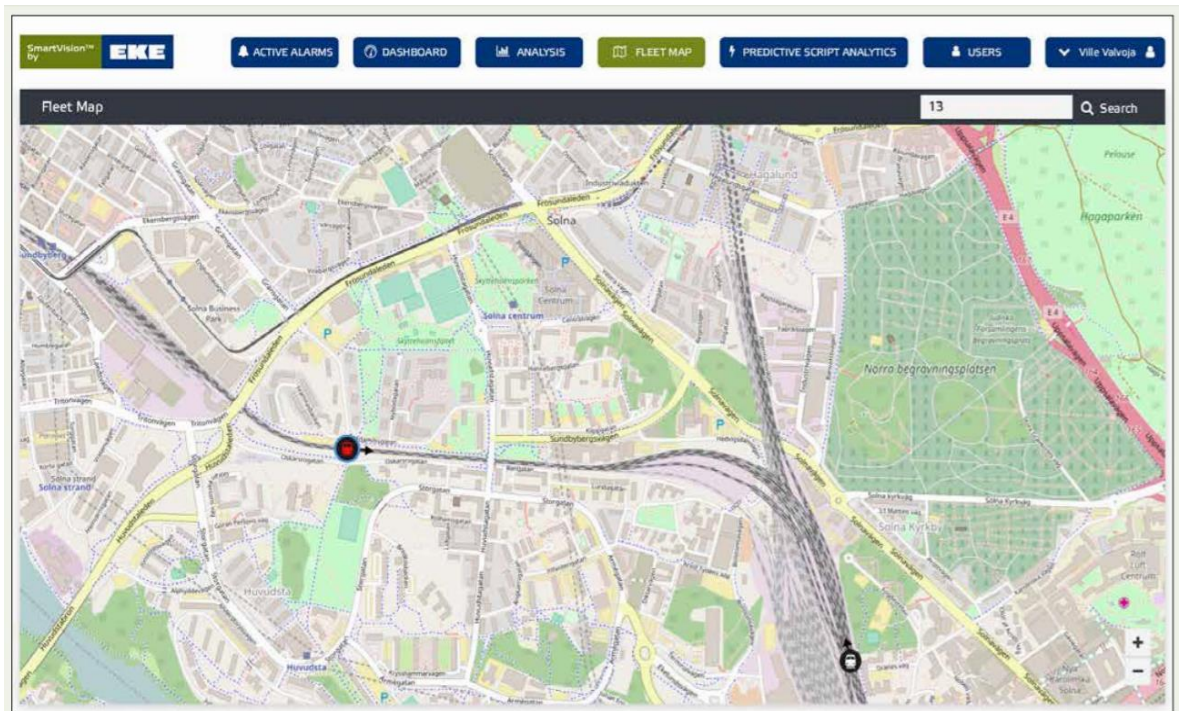


Figure 18. Train location on the map

Predictive analytics

Predictive analytics function is created to help maintenance have a better understanding of the trains. The software uses algorithms and mathematical calculation based on the previous data to give an estimation of component failure. When the maintenance team has this knowledge, it allows them to be prepared and replace the part before their failure. [12, 13]

Online monitoring

Online monitoring allows line controller, maintenance and operator to be updated on real time about the situation of the fleet. The online monitoring gives an overview of the whole fleet. The online dashboard shows which trains are on service, which trains are on maintenance, and which trains will need maintenance so that the teams can be prepared. The software gives a view as well of the most maintenance-needed systems, such as doors, HVAC and brakes. [12; 13]

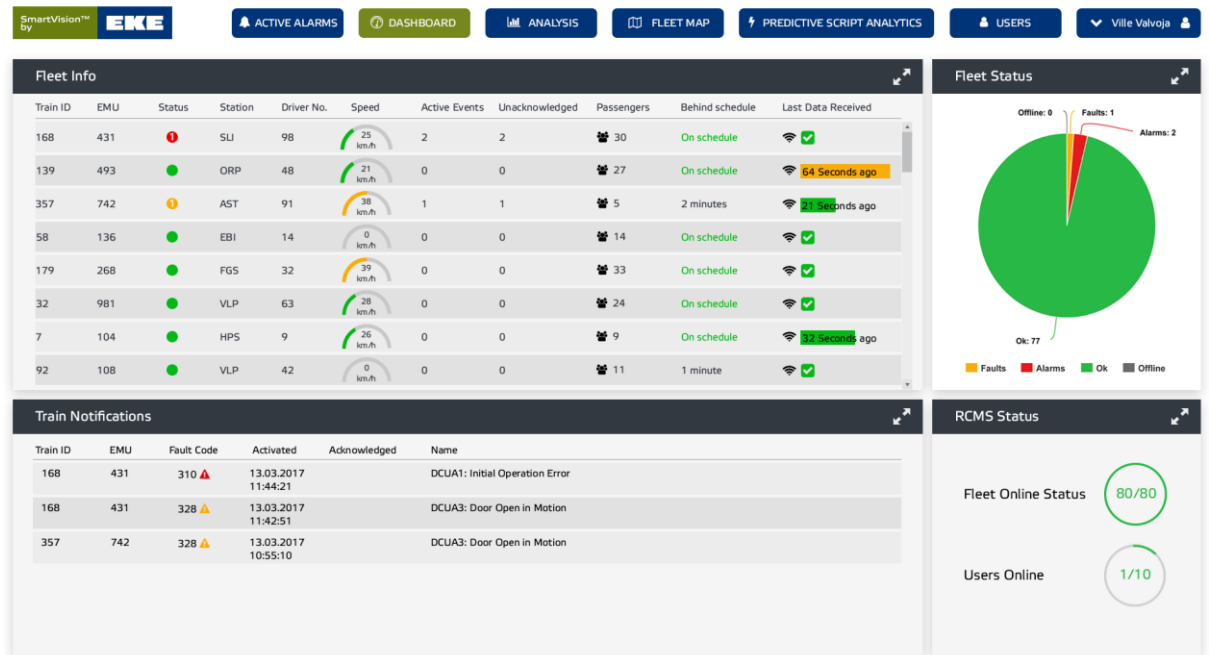


Figure 19. Online monitoring of a fleet

Condition monitoring

Condition monitoring allows the operators, maintenance and line controller to decide what data they will like to save and use in the future. The system allows for supplementary sensors to be added to the system in case they are not included on the fleet. [12, 13]

How it works

Data collection happens through Gateways modular systems. The Gateways can be designed to collect the data based on the required by the clients. The Gateways can be placed in every coach of the train and they can communicate between each other through Ethernet, WTB, MVB, CAN and serial links. The system gathers data from any existing onboard systems that already exist on the trains and send it to the wayside.

The data can be send either to the local server of the company or to the cloud. The system is very safe, and firewalls are applied to limit the access of the software and servers by other users. [12; 13]

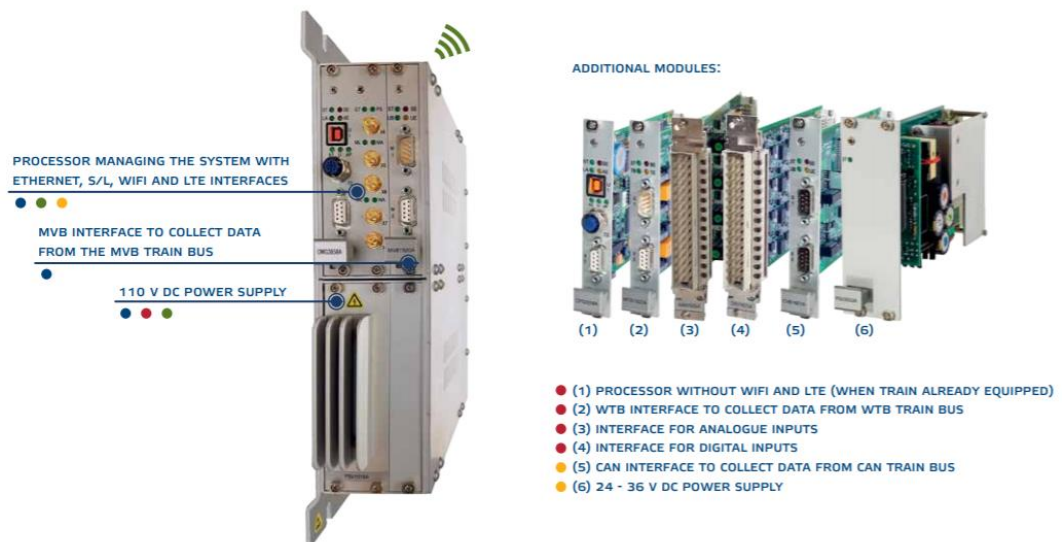


Figure 20. Typical modular used in train

8 Conclusion

The objective of this project was to create an environment that can be easily used and managed by the company, for this reason different software option were used to select the most suitable one. The environment was important to the company

and should be very accurate since most of the data related to their products and employees were uploaded there.

After trying with different software, Google Spreadsheet was chosen as a perfect fit since the company was familiar with the Google Drive platform and the information can be accessed from every online computer.

The results in the end were satisfactory, and the company started to use the online environment that was built for them. A list with pictures from all the sheets can be found at appendix 2.

References

- [1] <ftp://ftp.ingv.it/pub/manuela.sbarra/Data%20Mining%20Practical%20Machine%20Learning%20Tools%20and%20Techniques%20-%20WEKA.pdf>
Data Mining: Practical Machine Learning Tools and Techniques, Second Edition
- [2] http://www.spumonte.com/Files/PDFs/_PC%20Books_/Microsoft%20Office/Office%202010/Microsoft%20Access%202010%20All-In-One%20for%20Dummies.pdf
Microsoft Access 2010 All-In-One for Dummies
- [3] Practical Data Analysis, Hector Cuesta
- [4] Real-time analytics. Technique to Analyze and Visualize Streaming Data. By Byron Ellis.
- [5] <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3234&context=cais>, by Solomon Negash
- [6] Successful Business Intelligence, Second Edition: Unlock the Value of BI. By Cindy Howson
- [7] <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3234&context=cais>
- [8] <https://www.tableau.com/products/online/request-trial>
- [9] <https://www.benlcollins.com/dashboard/google-data-studio-tips/>
- [10] <https://powerbi.microsoft.com/en-us/features/>
- [11] Pro Power BI Desktop By Adam Aspin
- [12] <https://www.railway-technology.com/contractors/computer/eke-electronics/>
- [13] https://www.eke-electronics.com/images/eke/Brochure/EKE_Trainnet_Brochure.pdf

Appendix 1

Google spreadsheet code

Drop Down Box

```
function Drop(Size, source){  
  var rule = SpreadsheetApp.newDataValidation().requireValueInRange(source,  
  true).build();  
  Size.setDataValidation(rule);  
}
```

```
function onEdit (){  
  var aCell = SpreadsheetApp.getActiveSheet().getActiveCell();  
  var aColumn = aCell.getColumn();  
  if (aColumn == 1 && SpreadsheetApp.getActiveSheet()){  
    var Size = SpreadsheetApp.getActiveSheet().getRange(aCell.getRow(), aColumn + 1);  
    var source = SpreadsheetApp.getActiveSpreadsheet().getRangeByName(aCell.get-  
    Value());  
    Drop(Size, source);  
  
    }  
}
```

Get Data

```
function myFunction() {  
  var app = SpreadsheetApp;  
  var activeSheet = app.getActiveSpreadsheet().getActiveSheet();  
  var workingCell = activeSheet.getRange(2, 3).getValue();  
    var ss = SpreadsheetApp.getActiveSpreadsheet();  
    var sheet1 = ss.getSheetByName("Sheet1");  
    var sheet2 = ss.getSheetByName("Sheet2");
```

```
if (workingCell == "ASCO (Medical)"){\n    activeSheet.getRange(2, 4).setValue("418129030") && activeSheet.getRange(2,\n5).setValue("Prom. giveaways / tax allowable (Com. only)")&& activeSheet.getRange(2,\n6).setValue("M145.AVA1.1.014")&& activeSheet.getRange(2, 7).set-\nValue("ME_1454106003");\n\n}\nelse if(workingCell == "EORTC (██████████)"){\n    activeSheet.getRange(2, 4).setValue("██████████") && activeSheet.getRange(2,\n5).setValue("Prom. giveaways / tax disallowable (Com. only)")&& ac-\nactiveSheet.getRange(2, 6).setValue("██████████")&& activeSheet.getRange(2,\n7).setValue("██████████");\n\n}\nelse if(workingCell == "ESMO (Medical)"){\n    activeSheet.getRange(2, 4).setValue("██████████") && activeSheet.getRange(2,\n5).setValue("Travel & Accommodation product specific - on label (Com. only)")&& ac-\nactiveSheet.getRange(2, 6).setValue("██████████")&& activeSheet.getRange(2,\n7).setValue("██████████");\n\n}\nelse if(workingCell == "Preceptoship (Medical)"){\n    activeSheet.getRange(2, 4).setValue("██████████") && activeSheet.getRange(2,\n5).setValue("HCP registration fees product specific- on label (Com. only)")&& ac-\nactiveSheet.getRange(2, 6).setValue("██████████")&& activeSheet.getRange(2,\n7).setValue("██████████");\n\n}\nelse if(workingCell == "WCGC (██████████)"){\n    activeSheet.getRange(2, 4).setValue("██████████") && activeSheet.getRange(2,\n5).setValue("other special costs product specific on-label (██████████)")&& ac-\nactiveSheet.getRange(2, 6).setValue("██████████")&& activeSheet.getRange(2,\n7).setValue("██████████");\n\n}\n\nelse{\n    activeSheet.getRange(2, 3).setValue("Choose one option");\n}
```

```
}  
var app = SpreadsheetApp;  
var activeSheet = app.getActiveSpreadsheet().getActiveSheet();  
  
sheet1.getRange("A2:H2").copyTo(sheet2.getRange(sheet2.getLastRow()+1,1,1,7),  
{contentsOnly:true})  
  
sheet1.getRange("A4:L4").copyTo(sheet2.getRange(sheet2.getLastRow()+0,9,1,7),  
{contentsOnly:true})  
  
sheet1.getRange("I2").copyTo(sheet2.getRange(sheet2.getLastRow()+0,21,1,1), {con-  
tentsOnly:true})  
  
activeSheet.getRange("A4:L4").clearContent();  
activeSheet.getRange("A2:i2").clearContent();  
  
}
```

Transfer Data

```
function TransferData() {  
  
var ss = SpreadsheetApp.getActiveSpreadsheet();  
var sheet1 = ss.getSheetByName("Sheet1");  
var sheet2 = ss.getSheetByName("T1");  
  
sheet1.getRange("A18:D20").copyTo(sheet2.getRange(sheet2.getLastRow()+1,1,1,1),  
{contentsOnly:true});  
sheet1.getRange("A26:D34").copyTo(sheet2.getRange(sheet2.getLastRow()-2,9,1,1),  
{contentsOnly:true});  
}
```

```
}
```

Send email

```
function sendEmails() {
```

```
    var sheet = SpreadsheetApp.getActiveSheet();
```

```
    var startingRow = 2;
```

```
    var RowNumber = 2;
```

```
    var dataRange = sheet.getRange(startingRow, 1, RowNumber, 2)
```

```
    var data = dataRange.getValues();
```

```
    for (i in data) {
```

```
        var row = data[i];
```

```
        var emailAddress = row[0];
```

```
        var message = row[1];
```

```
        var subject = "Data";
```

```
        MailApp.sendEmail(emailAddress, subject, message);
```

```
    }
```

```
}
```

Appendix 2

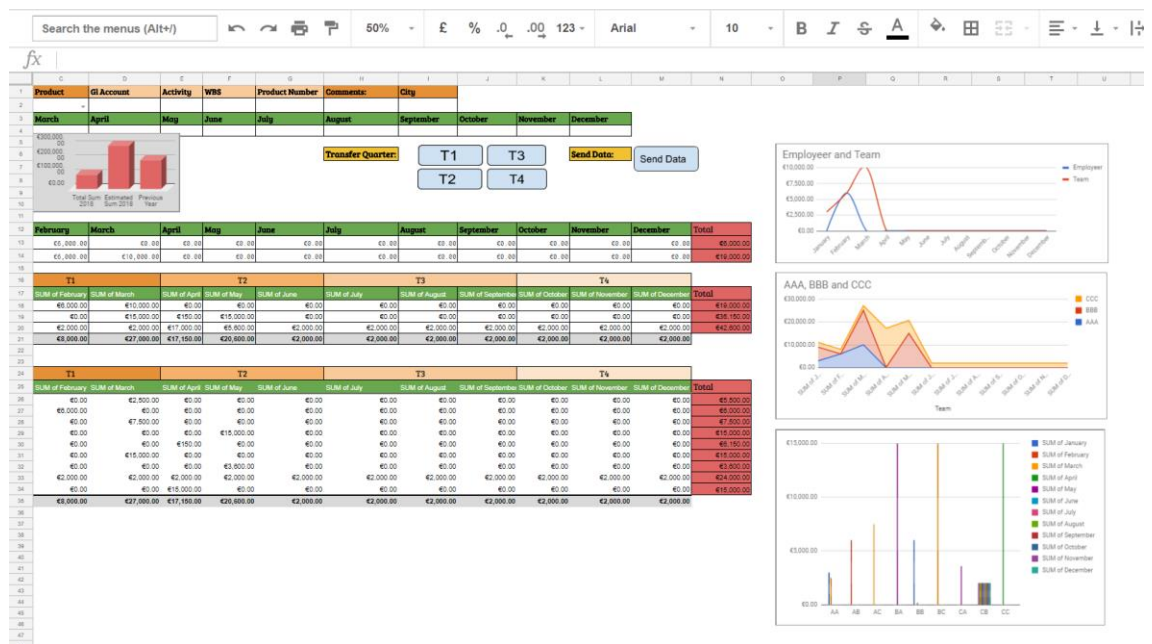


Figure 21. Front sheet

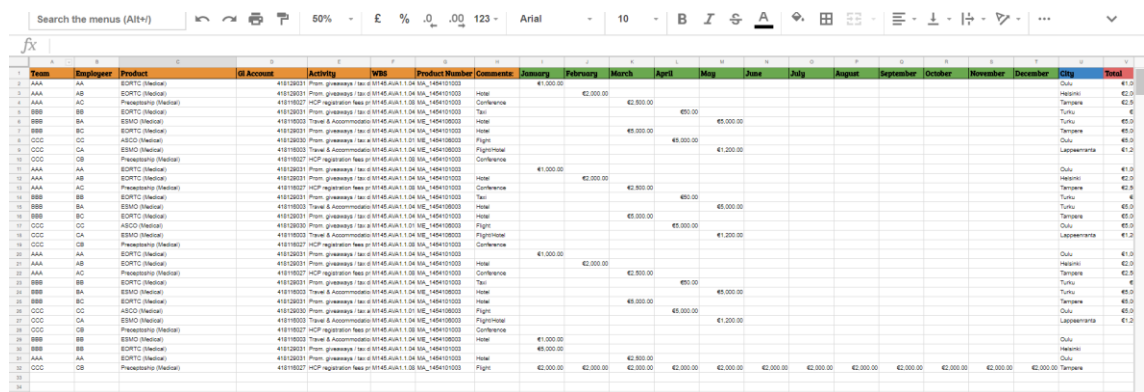


Figure 22. Data collecting sheet

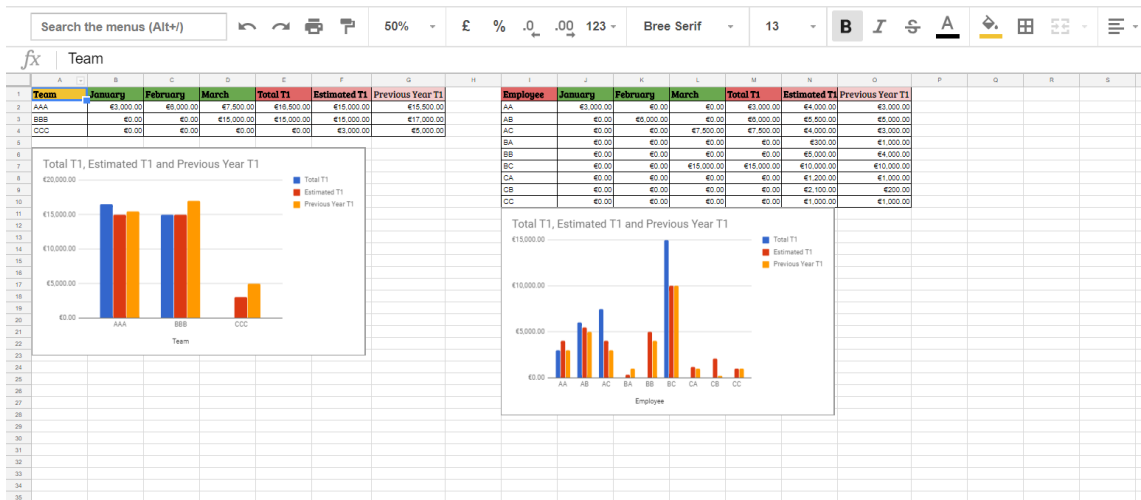


Figure 23. Comparing tables for the first quarter

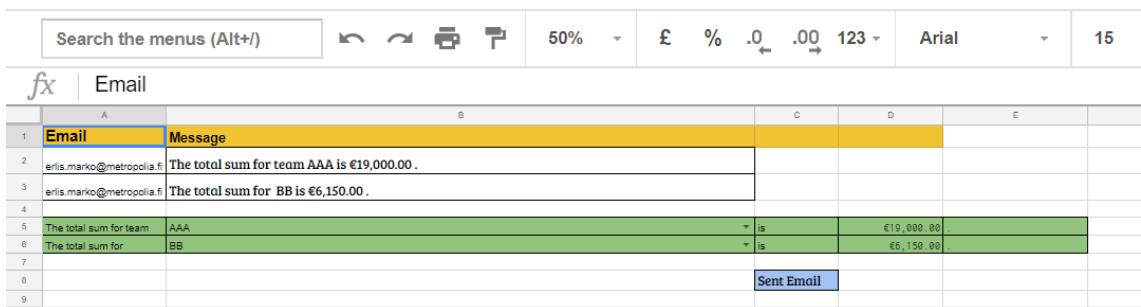


Figure 24. Send email sheet