ADS-ABusiness Proposal

korte regel

Mai Linh Luong  
Mitchel Kuijpers  
Nikita Gavrilov  
Huy Nguyen  
Jak Benev  
24 September 2020  
V2.0  
Status: Final

# Management Summary

For this project we are going to analyze the dataset of a pharmacy. It is important that we can tell the pharmacist where the money is based on our analytics and where the pharmacy can maybe perhaps save money. The pharmacy wants to make the health situation in Belgium still possible, this can only be done by cutting down the costs in health medication but doing so without affecting its quality. So in other words, to keep the patients as healthy as possible.

The objective for this project is to give insight to the pharmacist. To do so, we are going to analyse the dataset. Our goal for the pharmacist is to be prepared for the particular “illness season”, so not only that the pharmacist can be prepared for the amount of medications but also to be able to improve their services for their patients during that season. To do so, analyzing the most sold medicine we can understand what illness appears the most in the region by the season  
of the year.

The business requirement for this project is mainly focused on decreasing the cost of the pharmacist but also on the customer satisfaction since this is an important factor in the revenue. Business requirement is not something a system must do but rather something that the business needs to do or have in order to stay in business. In this case for the pharmacy, they want to see an increase in their revenue and decrease the unnecessary amount of cost.

The key performance indicators (KPI’s) are measurable values that show how effectively Informa is achieving key business objectives.

The scope for this project is to investigate different topics connected to the data set and examine the different interesting observations that we come across during the process. Furthermore, it portrays the working process and clarifies why the project group is working in a certain way.

The duration of the project is 18 weeks in total. To make the milestones more clear, we’ve set milestones that are identified as phase A, phase B and Phase C.

In the last chapter, the risk log of this project will be described. This provides a means of recording the identified risks, the and the necessary management actions that has to be taken during the project. It gives us the indication of how likely a certain risk would happen.

# Table of Contents

[**Management Summary**](#_u5tr1ul730li) **1**

[**Version History**](#_nh6u8xktfknd) **3**

[**Introduction**](#_xod4t1ojcmyi) **4**

[**1.Project Statement Summary**](#_cxbc6nq1cg4g) **5**

[1.1.Client’s Background](#_h606iig868kc) 5

[1.2.Current Situation](#_8s2evw4q0g69) 5

[1.3.Business Understanding](#_2kha5rpe2db5) 5

[**2.Business Requirements**](#_ecwhmmekl3gu) **6**

[**3.KPI’s**](#_cg4y1giii4j5) **7**

[**4.Scope**](#_xh6r3l476v0z) **9**

[**5.Data Understanding**](#_b2kcve9ns1no) **11**

[5.1.Genders](#_y9fdbxheacpw) 11

[5.2.Seasons](#_yecalwk8a9qj) 12

[5.3.Prices](#_d1rq3gfasbx0) 14

[5.4.Medicine Products](#_y9fdbxheacpw) 16

[5.5.Clinics](#_d80kbqhfso6x) 18

[**5. Planning**](#_sjd3cxrm18pa) **20**

[**6. Risk Log**](#_jqg8646esfnv) **22**

[Impact](#_wfigd3sthv0v) 22

[**7. Conclusion**](#_t26bayfq1tqd) **24**

[**8. Appendices**](#_s52mmeqb2k2q) **25**

# 

# Version History

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Description** | **Name** |
| V0.1 | 24-09-2020 | * Setting up the document | Mitchel Kuijpers  Nikita Gavrilov  Mai Linh Luong |
| V0.2 | 07-10-2020 | * 1.1.Client’s Background | Mitchel Kuijpers |
| V0.3 | 07-10-2020 | * 1.3 Business understanding * 5. Planning | Mai Linh Luong |
| V0.4 | 08-10-2020 | * 1.Project Statement Summary (Introduction) * 3.KPI’s | Mitchel Kuijpers |
| V0.5 | 25-10-2020 | * Introduction | Mai Linh Luong |
| V0.6 | 25-10-2020 | * 3.KPI’s - Updated with new Kpi’s | Mitchel Kuijpers |
| V0.7 | 29-10-2020 | * Processing the feedback on chapter 2 * Processing the feedback for management summary | Mai Linh Luong |
| V0.8 | 29-10-2020 | * Feedback KPI’s chapter 3 processed | Mitchel Kuijpers |
| V0.9 | 01-11-2020 | * Chapter 5.2.Seasons added | Mitchel Kuijpers |
| V1.0 | 02-11-2020 | * Chapter 5.5 Clinics added | Mai Linh Luong |
| V1.1 | 02-11-2020 | * Chapter 5.4 Medicine Products added | Nikita Gavrilov |
| V1.2 | 02-11-2020 | * Chapter 5.1 Genders added | Jak Benev |
| V1.3 | 04-11-2020 | * Chapter 5.3 Prices added | Nguyen |
| V2.0 | 03-12-2020 | * Last Check on document | Mai Linh Luong |

# Introduction

For this project it is expected that we can show where the money is in the data. Therefore questions like, where the organizations improve or how to increase the upsell could be answered when analyzing the data. Comparisons could also be useful, so we could show which pharmacists are underachieving compared to others and why.

The objective for this project is to give insight to the pharmacist. To do so, we are going to analyse the dataset. Our goal for the pharmacist is to be prepared for the particular “illness season”, so not only that the pharmacist can be prepared for the amount of medications but also to be able to improve their services for their patients during that season. To do so, analyzing the most sold medicine we can understand what illness appears the most in the region by the season  
of the year.

# 

# 1.Project Statement Summary

The Project Statement Summary gives the reader an overall summary of the project. This chapter will describe the clients background, so where do they come from and what do they do, the current situation the client is in. So why is the project being done and the business understanding, where is explained what is being done during the project.

## 1.1.Client’s Background

Informa is active in the Benelux as a specialist in data related challenges. Originally they were specialized in data(base) management, but over time Informa developed itself in other data related issues like data security, data integration and replication, data governance and business intelligence. The last couple of years they invested a lot of effort in the areas of data governance and data science.

## 1.2.Current Situation

Our customer is working for a significant number of pharmacies in Belgium. Their business consists of two major areas. Their first major activity is acting as a representative for the pharmacists in dealing with insurance institutions and pharmacy manufacturers and distributors. The other major activity is advising pharmacists on how to optimize their business.

The pharmacy company expects that project group can show where the money is in the data. Where could the organisations improve, for example by making agreements and predictions in purchasing goods and how to increase the upsell.

## 1.3.Business Understanding

Before analyzing the data, it is important to properly understand the structure of the business that we are going to investigate. When understanding the structure of the business better, it allows us to make a better starting point of our data analysis and explain different trends we find in the data regarding the business requirements.

We are investigating a Belgium pharmacy that actually sells the medications to their patients. This means that the revenue of the pharmacy relies on the amount of the medications. There are different pharmacies in different regions that all have an unique code for the medications.

# 2.Business Requirements

Business requirement is not something a system must do but rather something that the business needs to do or have in order to stay in business. In this case for the pharmacy, they want to see an increase in their revenue and decrease the unnecessary amount of cost.

**The business requirement for the pharmacy is therefore:**

1. Increase customer service by using the data effectively with the opportunity that the amount of patients will be increased
2. Reduce unnecessary processed orders by the end of the next quarter with the opportunity to decrease the amount of cost
3. Increase order accuracy by using the data effectively

# 

# 3.KPI’s

The key performance indicators (KPI’s) are measurable values that show how effectively Informa is achieving key business objectives.

The KPI’s listed in the table below are based on the business requirements that have been stated in the previous chapter. With the help of the following KPI’s based on seasonal, prices, contribution, gender and age data, we are able to get indications about the customer satisfaction, orders that are being processed and increase the order accuracy with the help of efficiently using the data. With the help of these KPI’s we can improve and help Inform increase their revenue and decrease the unnecessary amount of cost of products.

The most important indicators we are going to keep track of are the following:

|  |  |
| --- | --- |
| **Nr.** | **Key Performance Indicators** |
| **1.Seasonal** | |
| **1.1** | What is the average price of products per season? |
| **1.2** | What season has had no growth in sales compared to the previous season? |
| **1.3** | What are the total sales per season? |
| **2.Prices & Contribution** | |
| **2.1** | What is the average contribution per month? |
| **2.2** | What customer has contributed the most? |
| **2.3** | What products have the highest average price? |
| **3.Gender & Age** | |
| **3.1** | What is the average age of customers? |
| **3.2** | What gender has paid the most for products? |
| **3.3** | What gender bought the most products? |
| **4.Regional** | |
| **4.1** | What region has contributed the most? |
| **4.2** | What region has the most orders? |
| **4.3** | What is the most sold product per region? |

With the help of the described kpi’s we can help Informa make better use of the data they gather. These kpi’s could also help with improving the organisations Informa works with and increase the upsell of products.

# 

# 4.Scope

The scope for this project is to investigate different topics connected to the data set and examine the different interesting observations that we come across during the process. Furthermore, it portrays the working process and clarifies why the project group is working in a certain way.

**4.1.Functional areas in scope**

The main scope of the project is to optimize preparation for the illness season. For that reason We will look at certain criteria such as Sales of the products, Delivery data and delivery time and last but not least is the Location. The data which is not related to drugs such as the fee for the rest house will not be included in the scope of this project.

**Sales**, the most important key performance indicator of a business is how much revenue and profit/loss it gains so dealing with sales and using data to improve upon sales is mandatory.

**Delivery data and Delivery time**, can help the organization understand when people purchase the medicine the most. Thus, the organization can predict and be ready for the high demand for a certain medicine

**Location**, can help the organization understand where people purchase the medicine the most. Thus, the organization can understand where it should increase the stock and provide medicine to people without any delays. Moreover, in a bigger picture, it can show the problematic regions for a certain disease.

**4.2.Organizational scope**

The project team consists of 6 people and one mentor which helps the project team when necessary and also helps the group to keep on track with the project. Also, there is our customer who evaluates our work.

The estimated man hour of this project is 336 hours. Considering that a person’s fully scheduled day adds up to 6 hours, the team expected to have at least 4 hours available every week for the working on the project.

**4.3.Technical scope**

The project group uses several working environments for cleaning, analyzing and visualizing the data.

The team chose the following programming environments and tools:

Libre office - a powerful tool for cleaning and handling missing data

Anaconda navigator and Jupiter notebook - our main working environment for the project, which help of python and pandas we are going to collect, clean, munging, analyze, interpret and visualize data

Tableau - a good tool for visualisation of data and will be used as a validation of our investigation

**4.4.Project deliverables**

The project deliverables are divided in tree parts.

Part A: Business Proposal supported by Exploratory Data Analysis (Descriptive/Diagnostic Analysis).

Part B and part C: Data Analysis Report (Predictive/Prescriptive Analytics) supported by Data Quality improvements, Machine Learning results, and ethical considerations.

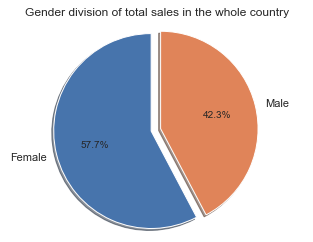
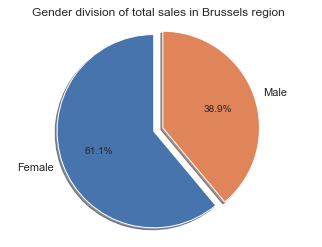
# 5.Data Understanding

In this chapter, we will talk about the foundings in the dataset. The dataset that was given by the customer was relatively large. Therefore it was crucial to clean the dataset before we wanted to work with it. The full data analysis can be found on the following link: [Exploratory Data Analysis](https://docs.google.com/document/d/1jZtyaYeidgT7hUBY51QIil3SKBxppnPkw7sUx8w1nzQ/edit?usp=sharing)

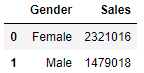
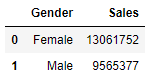
## 5.1.Genders

To start things off we’ll be looking into total sales according to gender division.

Initially we will look into the second biggest region in sales - Brussels (the capital) and the whole country.



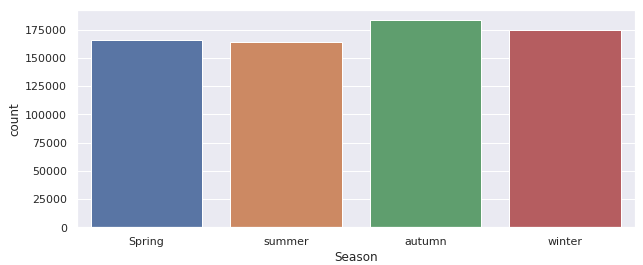
A small difference of 3.4% towards equal division, however something to keep in mind as we’ll be looking into top clinics in sales in a few pages and the capital has one of the biggest. The following is the total number of sales in the whole country and in brussels, accordingly:



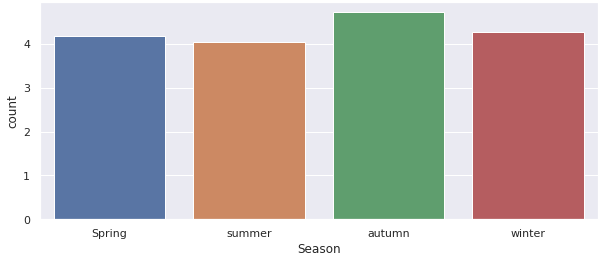
## 5.2.Seasons

In this chapter we will take a look at the seasonal data that have been explored in the exploratory data analysis (EDA). The season data has been segmented in the four different seasons summer, winter, spring and autumn. Before any analysis could be done per season we had to clean the data and merge the delivery date and time. This made it possible to see in what season the date is in.

During the analysis we found out that the seasons have about as much sales compared to the others, however, the season autumn did have the most sales in the region of Antwerp.

  
Image: *Seasonal sales in region 22 (Antwerp)*

The analysis has shown that all the regions have nearly the same sales percentages per season. All regions show that the most medication is bought during the season of autumn. When we look at the overall sales we can see that the autumn season is still the season that has the most products sold as seen in the image below:

  
Image: *Seasonal sales in all regions total*

However, the total sales per region do vary. The products sold in region 22 are between 15.000 and 19.000 per season, where region 30 has sales between 27.000 and 32.000 per region. This could show that in some regions more products are being sold or less or more pharmacies are available. This can be seen in the images below where the left image shows the number of products sold in region 30 and the right image shows the sales in region 22.

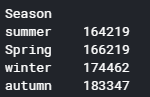
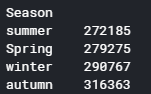


Image: *Sales per season region 30 (Left) and region 22 (right)*

## 

## 5.3.Prices

In this part, we will explore closely the difference between the product prices explored in the Exploratory Data Analysis (EDA). The data for the product has been divided into 3 main types: Honorarium, Pills and Tablet. After cleaning up the whole dataset, we divided the types of the product depending on whether it name contains pill, tablet or honorarium.

During the analysis, we found that the Honorarium type had the lowest selling price compared to the other medication types.

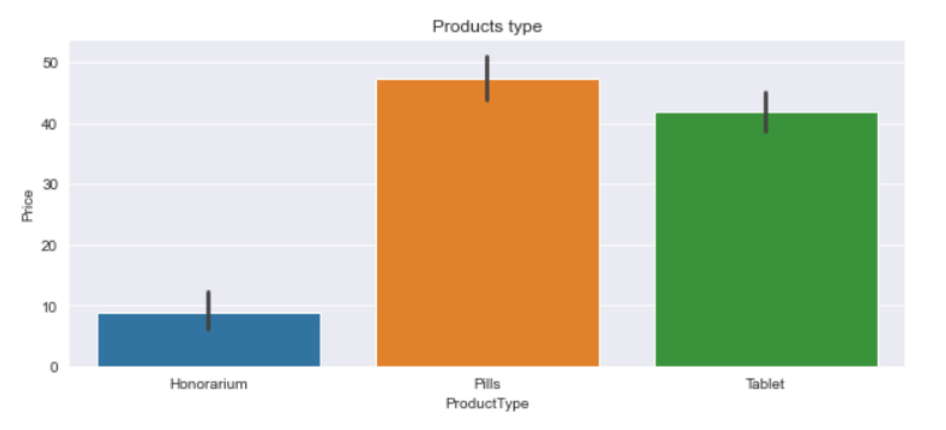


Image: *Product price comparison in region 20*

The analysis has shown that the pill and the tablet products both have their price in between 40 and 50.Moreover, the most expensive medication is the pill which have the price of over 45. When we look at the price in detail for each products sold as seen in the image below:

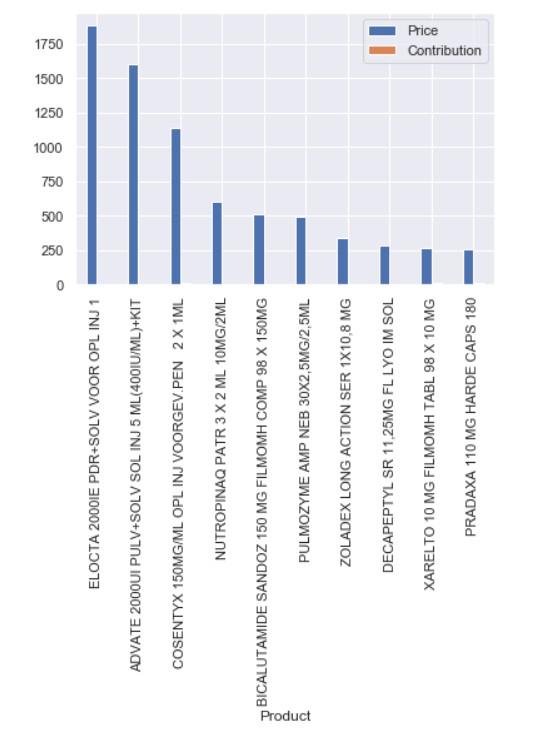


Image: *10 most expensive products in region 20*

The chart below noted on which product has the highest income. The ten most expensive products have a price range between 250 and 1750.

## 5.4.Medicine Products

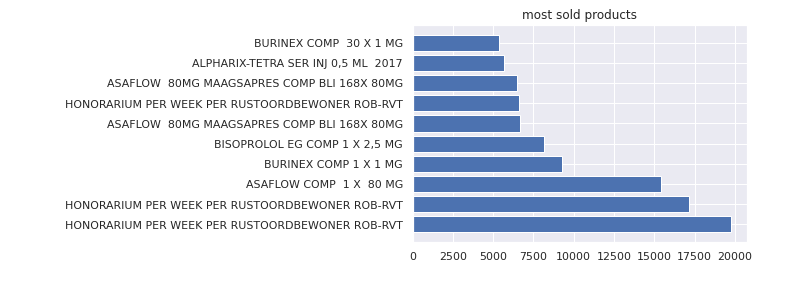
In this chapter, we want to see what are the most sold products. For that purpose, we compared large regions which are Leuven and Liege.

If you look at both of the graphs you can clearly see that the most sold medicine by far is “Honorarium per week per rustoord bewoner”, however, it is not particularly a medicine. “Honorarium” is a fee per week for the rest home. Thus, it is not in the scope of this chapter.

In the region of Leuven, the most sold medicines are Bisoprolol for heart diseases, Asaflow (in different variations)which is also preventing some cardiovascular diseases, Burinex is a medicine used to treat swelling and high blood pressure, Alphrix-tetra medicine which is used against flu.

If we compare the first graph with the second one we can see some similarities. Asaflow(in different variations), Alphrix-tetra and Burinex are also in the list. Asaflow and Burinex are the most sold medicines in that region by far. On the other hand in that region we can see drugs that not listed in the graph one such as Quetiapine Sandoz which is used to treat schizophrenia and bipolar disorder, Coversyl which is used to treat high blood pressure and heart failure, and last but not least Paracetamol which is used to treat pain and fever.

To conclude we can see that the most popular medicines in both regions are used to treat cardiovascular diseases such as Bisoprolol, Asaflow, Burinex and Coversyl.



*Figure 1: Postcode 30 (Leuven)*

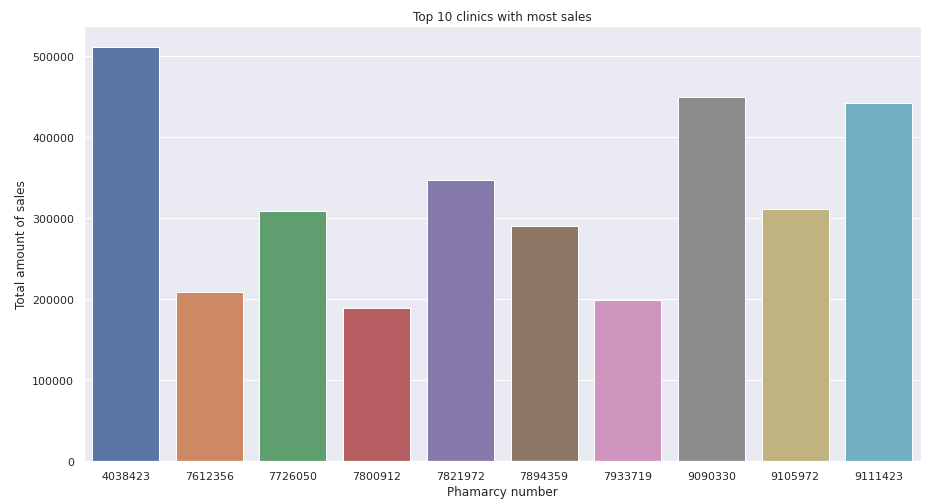
## 

*Figure 2: Postcode 40 (Liege)*.

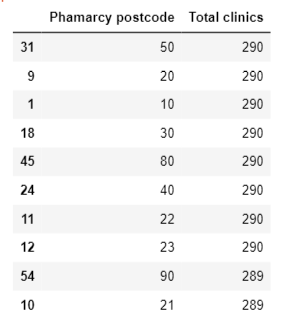
## 

## 5.5.Clinics

In this chapter the most popular clinics will be discussed. This tells us which clinics sell the most products and gives us a better indication for the sales between the clinics. If we take a look at figure 1, where the top 10 clinics are plotted you can see that pharmacy number 4038423 in region 20 has the most sales. As you can see on the table, most of the clinics are from region 10. Region 10 is Brussel so the capital of the country. That explains why there are 3 clinics from region 10 in the top 10 clinics with the most sales. *Table 1: Top 10 clinics with most sales*

*Figure 1: Top 10 clinics with most sales*

But what is also interesting to see is, how many clinics are there in each region. This also gives us a better indication of the amount of sales in each region. For example, if a region only has one clinic, then it would be logical that this clinic made it to the top 10 list.

As we can see in table 2 where the top 10 region with the most clinics is shown, we can see that region 50 also has the most clinics, but didn’t make it to the top 10 clinics with the most sold products. This could be because of the good division between the clinics where the amount of sales between could be equally divided. 

But comparing table 2 with table 1, we can see that region 20 is also at the top with the most clinics in its region. It seems that region 20 is from Antwerp, the biggest city in Belgium with the highest population. This explains the amount of sales in clinic 4038423.

*Table 2: Top 10 region with the most clinics*

# 

# 

# 

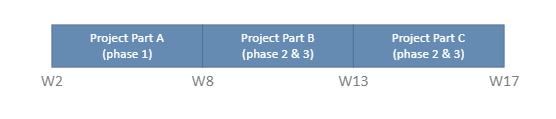
# 

# 

# 

# 5. Planning

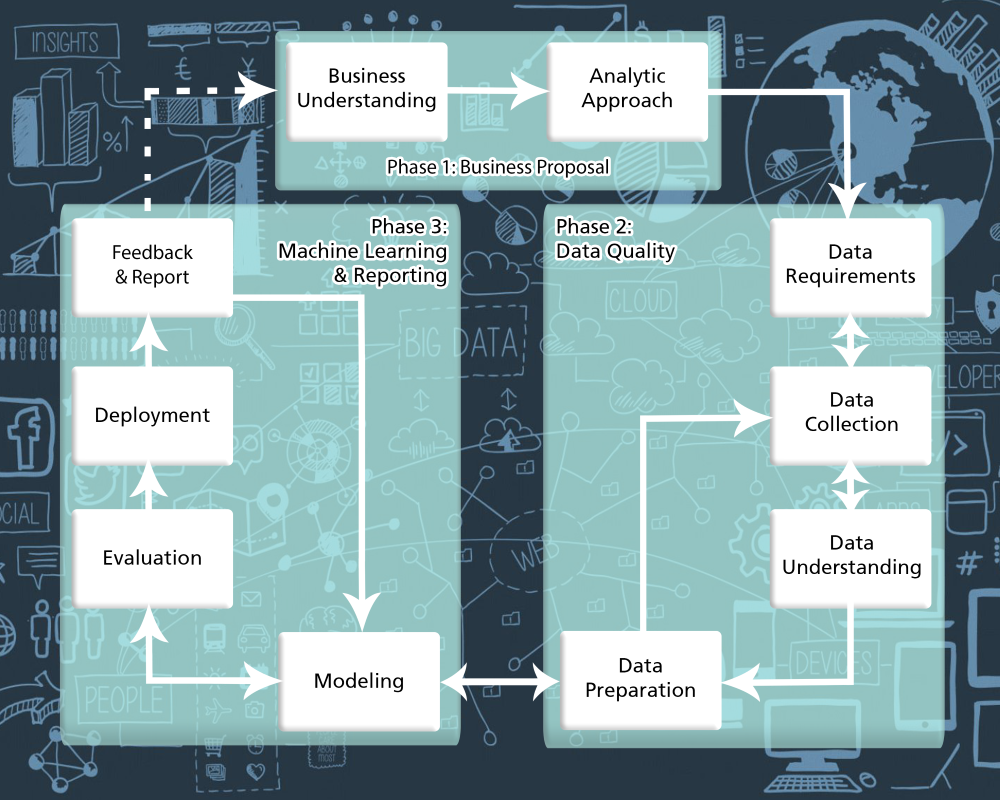
In this chapter, we are going to explain the project planning for the upcoming weeks this semester and how to achieve our goal. It gives us more insight of the structure of the semester but it also helps us to see when the deadlines are. We are using sprints that are represented on Canvas (see image 1).



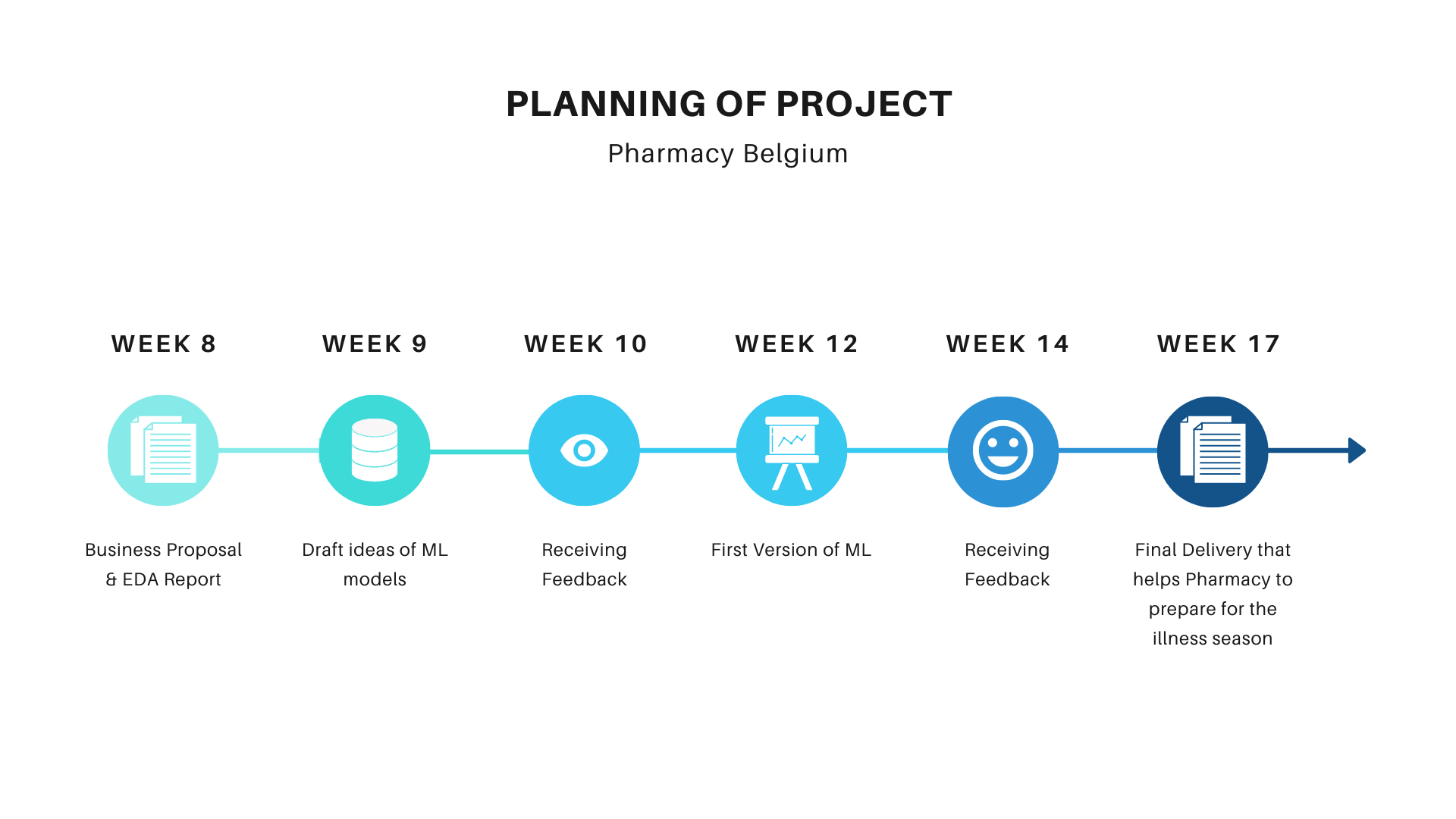
*(Image 1: Sprint)*

So from week 2 until week 8, this is the first part (part A) of the project. Also known as phase 1. This is where the business proposal and the exploratory data analysis report will be delivered.

The next sprint is week 8 until week 13. This is the second part of the project (project part B), also known as phase 2 and phase 3. The reason why we are doing phase 2 and 3 in the same sprint is because in the second iterations, the product should be improved compared to the first one. This is the same for the next and final sprint of the semester which is project part C. This is the final sprint where we have the desired quality of the product ready for the customer.

  
*(Image 2: Phases of the project)*

As you can see on image 3, where the deliverables of this project is explicitly explained the first 8 weeks will be spent on the business proposal and the EDA (exploratory data analysis) that is then going to be written in a report. After 8 weeks, the time will be spent on designing and thinking about potential ML (machine learning) models. This will eventually help us to achieve our goal, which is to help pharmacy to be prepared for the particular illness season.

  
(*Image 3: Planning of Project)*

# 6. Risk Log

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Description** | **Impact** | **Likelihood** | **Control measurement** | **Score** |
| **1** | The product is different from customer expectations | 4 | 2 | The features of the project and customer expectations from the project are described in the business Proposal document. | **H** |
| **2** | Scope creep | 3 | 2 | The scope of the project is defined, documented and controlled inthe Business Proposal document. | **M** |
| **3** | Coronavirus | 4 | 5 | The global situation with coronavirus has to be studied frequently. To ensure the safety of customer and group members the group will obey the requirements of the local authorities | **C** |
| **4** | Inaccurate or/and inconsistent data | 2 | 3 | The team tries to manage the transformation and work on the quality without falling behind schedule. | **M** |
| **5** | Poor project planning | 1 | 3 | Weekly group meetings in MS Teams and in R10 where potential delays of the deliverables are discussed | **L** |

Below you can see the legend of this risk log. In the table above are the numbers 1 to 5 and scores are used and referred to the numbers and score in the table below.

## **Impact**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Likelihood** | **Insignificant (1)** | **Minor (2)** | **Moderate (3)** | **Major (4)** | **Extreme** |
| **Almost certain (5)** | **M** | **M** | **H** | **C** | **C** |
| **Likely (4)** | **M** | **M** | **H** | **C** | **C** |
| **Possible (3)** | **L** | **M** | **M** | **H** | **H** |
| **Unlikely (2)** | **L** | **L** | **M** | **H** | **H** |
| **Rare (1)** | **L** | **L** | **M** | **M** | **M** |

# 

# 7. Conclusion

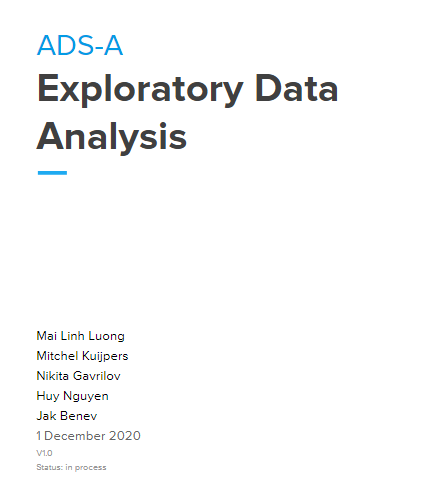
The primary goal of this project is to prove how the pharmacy could make profit or lose money for buying and selling their products. Try to find the highest profit margin for the pharmacy client. The pharmacy needs to make the health situation in Belgium still possible, and this can only be achieved by reducing the cost of medicines. They still want to keep medicines accessible and keep patients as safe as possible. But if we can find anything that's safe for the patients, it's a good way to find something for the project.

The final destination of this project is to provide the pharmacist with insight. We're going to evaluate the dataset to do this. Our goal for the pharmacist is to be prepared for a specific "illness season" so that not only can the pharmacist be prepared for the amount of medicines, but they can also enhance their care for their patients during that season. To do so, by evaluating the most sold medicine, we can understand what disease is most prevalent in the area by the season of the year.

### 

# 

# 8. Appendices

  
File 1: Exploratory Data Analysis Document