Anne-Sophie Josée

635725

DATUM

Portfolio Data science analytics skills

Data Driven Decision

Title: Portfolio data science analytics skills

Author: Anne-Sophie Josée

Studentnumber: 636725

E-mail:  [a.Josee@student.han.nl](mailto:a.Josee@student.han.nl)

Date: xxxx

Place: Arnhem 

Educational program: Minor Data Driven Decision Making in Business

Client: Hogeschool van Arnhem en Nijmegen

Content

[1. Business case 3](#_Toc124339448)

[1.1 Introduction 3](#_Toc124339449)

[1.2 General information about dashboards 3](#_Toc124339450)

[1.2.1 What are dashboards 3](#_Toc124339451)

[1.2.2 What benefits does a company experience using dashboards 3](#_Toc124339452)

[1.2.3 Dashboards within supply chain 3](#_Toc124339453)

[1.3 Tools to make dashboards 4](#_Toc124339454)

[1.3.1 Tools to make a dashboard 4](#_Toc124339455)

[1.3.2 General guide lines for dashboards 4](#_Toc124339456)

[1.4 CRISP-DM model 5](#_Toc124339457)

[2. Data science and techniques and tools 8](#_Toc124339458)

[2.1 The dataset 8](#_Toc124339459)

[2.2 Excel as a dashboard tool 8](#_Toc124339460)

[2.3 Power BI as a dashboard tool 8](#_Toc124339461)

[2.4 Airtable as a dashboard tool 8](#_Toc124339462)

[3. Visualization and justification 9](#_Toc124339463)

[3.1 Visualization and justification Excel dashboard 9](#_Toc124339464)

[3.2 Visualization and justification Power BI dashboard 12](#_Toc124339465)

[3.3 Visualization and justification Airtable dashboard 16](#_Toc124339466)

[5. Applying CRISP-DM to business case 19](#_Toc124339467)

[5.1 Project process steps linked to CRISP-DM 19](#_Toc124339468)

[5.1.1 Github link to deploy project 19](#_Toc124339469)

[6. Code / Script quality 20](#_Toc124339470)

[Literature 21](#_Toc124339471)

# 1. Business case

## 1.1 Introduction

This report is about increasing dashboard making skills. The goal is to use multiple tools to get more experience. I study logistics management, so I expected to get in touch with dashboards quite a lot. This is why I want to strengthen my experience.​

The tools that will be used are Excel, Power BI and Airtable.

## 1.2 General information about dashboards

### 1.2.1 What are dashboards

A dashboard is an information controlling tool. It visualizes key performance indicators (KPI), analyses and displays the status of a company, department or a specific process (Tazmi, 2022). KPI’s are indicators that are used to monitor the results, but also to analyse and improve them (Beek, n.d.).

Dashboards can be easily adjusted to the specific needs, they are quite flexible. Often they are connected to all kind of data sets, but the dashboards shows these in tables, graphs and meters. A dashboard is an efficient way of analysing because it is based on real time monitoring.

There is a certain difference between dashboards and reports. Reports are often used to display something with a narrow focus (Tableau, n.d.). They have the function provide a deep-dive view and therefore concentrate on only a single event or item. Dashboards are used to create a high-level view of broad amounts of data, they are made to answer just one single question.

### 1.2.2 What benefits does a company experience using dashboards

The most companies want to follow it’s KPI’s and statistics, to monitor and analyse these can cost a lot of time and resources. Investing in a creating a dashboard can prevent this (Tazmi, 2022). A dashboard processes raw data and turns it into easily understandable tables, graphs and meters. The users of the dashboard know at a glance where to find the most import statistics and how to read them.

Another benefit is that dashboards can be used for the monthly report. The users can communicate about the information directly without hours of preparation and analyses.

In conclusion they give the user the change to make better informed data-driven decisions. As a user you get a better real time understanding of the happenings around you.

### 1.2.3 Dashboards within supply chain

In many companies the supply chain is a crucial part of their business. If there is an inefficiency, a bad structure or vision it will have consequences for the whole company (Calzon, 2022). Therefore an supply chain dashboard to keep an eye on the KPI’s can be very helpful.

Below some examples form possible subjects for a supply chain dashboard are named:

* Cash-to-cash time cycle: Time period in which the suppliers have to be paid until money from your customer is received.
* Freight bill accuracy: How effective the freight bills are processed.
* Perfect order rate: The ability to correctly deliver orders.
* Inventory turnover: The number of times the inventory has been sold.
* Warehousing costs: Cost distribution, for example order picking, storage, shipping etc.

## 1.3 Tools to make dashboards

### 1.3.1 Tools to make a dashboard

To make a dashboard a dashboarding tool is used. This is a data-management software that can display the complex data (Ozuysal, 2022). There are a lot of tools that can be used, there are some things that make a dashboard tool sufficient.

**Easy set up**

It is very crucial to that a tool is quick and easy when it comes to setting things up. The base of a dashboard is very important and it shouldn’t be hard to get a good set up.

**Quick access**

If a dashboards needs a lot of loading time after starting it up, it is not efficient to use it. It might be due to slow loading time or it cannot handle heavy databases, but you should use a tool that can assist you with data visualization directly when you need it.

**Access on the go**

This depends on your working environment and agreements, but this could be an important criteria. If you for example work a lot from home, it would be nice that you can easily access the dashboard from that location.

### 1.3.2 General guide lines for dashboards

When making a dashboard there are a couple of things to pay attention to, these are discussed below.

**Define the goal and audience**

If you know which problem the dashboard should solve, it is much easier to create an usable dashboard if you understand the goal of it (Tazmi, 2022). If you understand the goal, the chances are you use the right statistics.

Knowing the needs of the users makes sure that the dashboards matches their general goals.

**Choosing the right visualizations**

There are a lot of visualizations to choose form, for example tables, graphs, maps etc. For each kind of data, of course you should choose the one which represents the data the best (Tableau, n.d.). Pie charts for example can become overwhelming rather quickly, when they include too much information.

Also, use interactive elements if this is needed. It gives people the opportunity to dive deeper into the data if needed.

**Easy design**

In this case ‘less is more’ comes to its right. People tend to get distracted when a lot is going on. Therefore it is better to only use a couple of colours, one font style and a minimal amount of pictures.

**Placing the statistics**

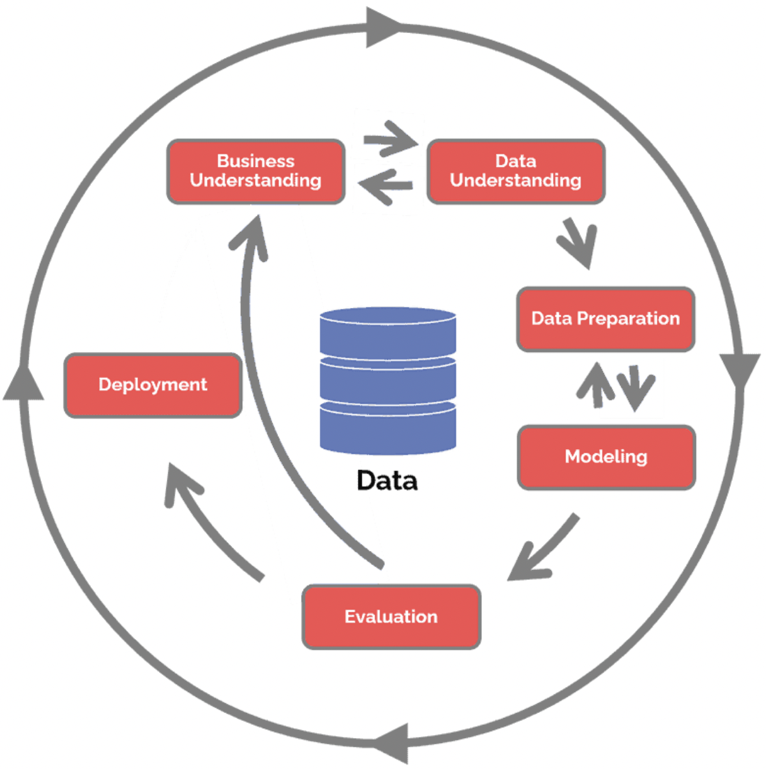
To make your dashboard easy readable, it is best to structure the information from left to right. The most important data should be placed in the upper left corner.

## 1.4 CRISP-DM model

For the execution of a data science project a special model has been developed, called the CRISP-DM model. This stands for Cross Industry Standard Process for Data Mining (Hotz, 2023). The model consist of six phases:

* Business understanding: What does the business need?
* Data understanding: What data do we have and or need? Is it clean?
* Data preparation: How should de data be organized for modelling?
* Modeling: What modelling techniques should be applied?
* Evaluation: Which model meets the business objectives best?
* Deployment: How can stakeholders access the results?

These six phases form a path that can lead to a successful data driven decision. The model is shown in picture 1.4 below.



*Picture 1.4: CRISP-DM model (Hotz)*

**Business Understanding**

The first step of the model is business understanding, the focus lays on understanding the objectives and requirements. This helps giving context to the goal and to the data.

**Data understanding**

The next stap is data understanding. This step is meant to help know what can be expected and achieved from the data. It drives to collect, analyse and identify data sets, this is crucial part of the project. It defines how viable and trustworthy results can be.

**Data preparation**

The third step is data preparation, involving extractions, transformations and load processes (ETL). The data is being prepared for modelling. Standardizing data is an example of what is often done in this step, data preparation is often the most time consuming step.

**Modeling**

The step after the data is prepared, is modelling. This step is responsible for the results of the project and often seen as the most exciting work. If the previous steps are done correctly, this is also the shortest phase of the project. Selecting modelling techniques, generating a test design, building a model and assessing the model are done in these phase.

**Evaluation**

The fifth set is evaluation. Whereas modelling focuses on the more technical side of the project, this step looks more broadly. The needs of the business are taking into account to decide which model fits best and what should be done next. If results are wrong, the methodology of the CRISP-DM model permits the review to go back to the first step in order to understand why results are wrong.

**Deployment**

The final step of the model is deployment. In this step the results are presented in a useful and understandable way.

# 2. Data science and techniques and tools

## 2.1 The dataset

The dataset that will be used is called Juicy Fruit Corporation. This is dataset is supplied by school for an data visualization assignment, in an xlsx Excel format. However, my part of the teamwork was to search for correlations so I never really got to make the visualizations. This dataset has just the kind of information I will get in touch with during my professional career so is very suitable.

The dataset contains company information from the following departments:

* Purchasing
* Production
* Sales
* Warehouse
* Distributor
* SCM (inventory)

## 2.2 Excel as a dashboard tool

The first tool where a dashboard will be made in is Excel. This tool is chosen because Excel is a tool which is often used in companies during the daily work operations. It is in the Microsoft package that a lot of companies have and therefore it can easily be used to create a dashboard.

My reason to choose Excel is because I want to learn how to make graphs interactive in this program, this is my goal for this particular tool. I already knew the basics but wanted to broaden my experience and give the viewer the option to add variables to a graph to deepen their analysis.

## 2.3 Power BI as a dashboard tool

The other tool is Power BI, this is also owned by Microsoft but not in the standard package.

My reason to choose Power BI is because it is a much used tool in business, but I have never used it. I hoped that at this minor it would be used, but unfortunately this was not the case. Therefore I want to create my own experience.

## 2.4 Airtable as a dashboard tool

The last tool that will be used is calles Airtable, learned about this tool through other students from the minor.

The reason the use this tool was to get familiar with it, in case I will use it in the future. So purely to broaden my experience with dashboard tools overall.

# 3. Visualization and justification

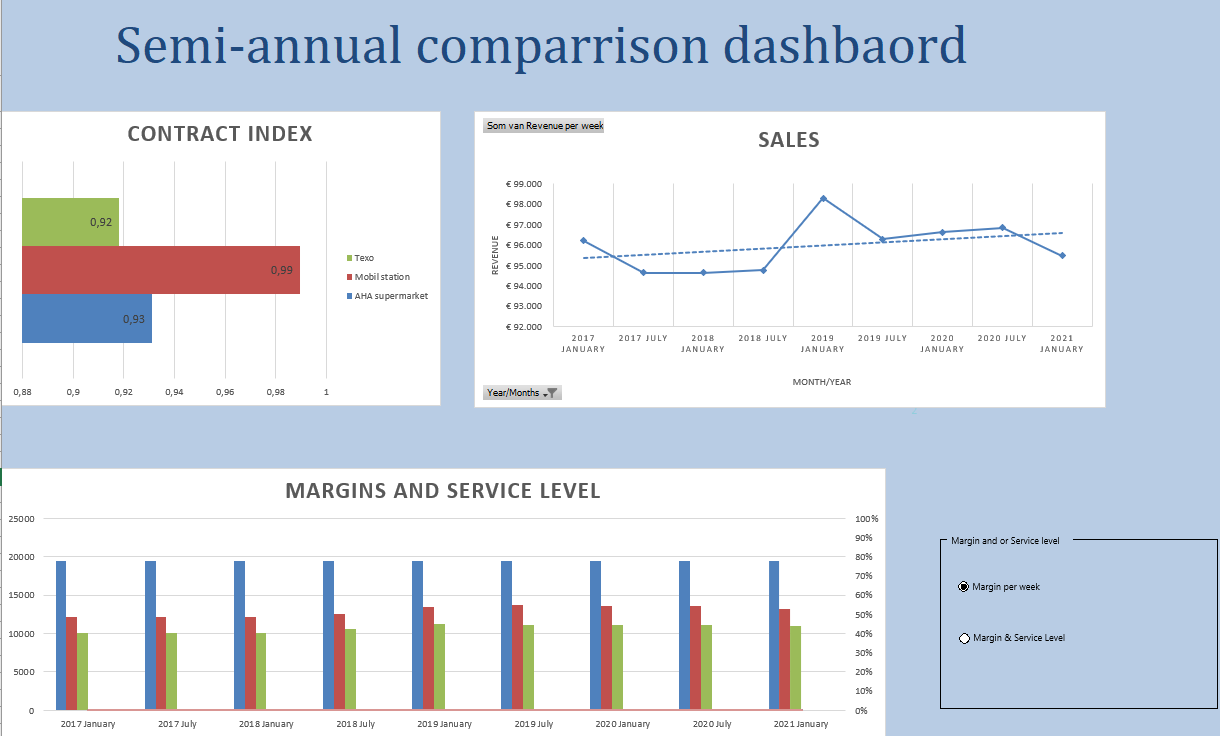
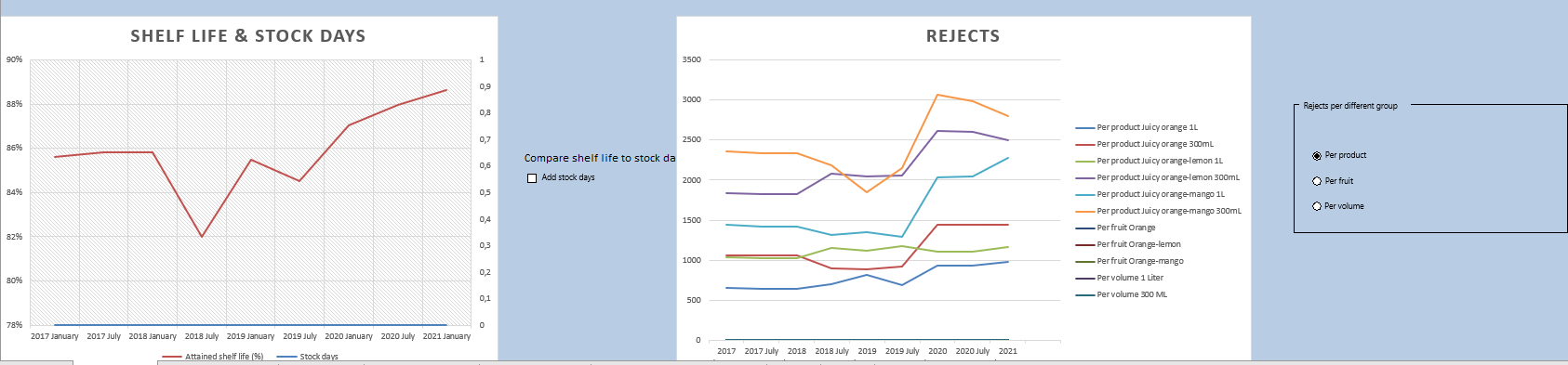
## 3.1 Visualization and justification Excel dashboard

**Description**

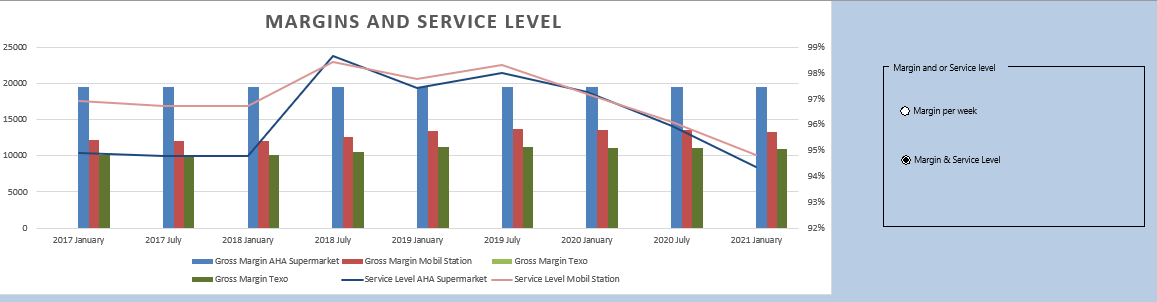
In the pictures 3.1.1 and 3.1.2 below an overview of the dashboard made in Excel can be seen. On the top left and right the two very important indicators for a company are shown. The contract index indicates the periodic adjustments to the long-term contract prices based on the current delivered quality of goods and service. Also the sales over the past years are shown, with an trend line added to show if overall the sales are increasing or decreasing. The contract index and sales are shown here because according to the described theory the most important data should be placed on the left top, and work from left to right.

Then the margins are shown. This indicates per customer which margins are earned.

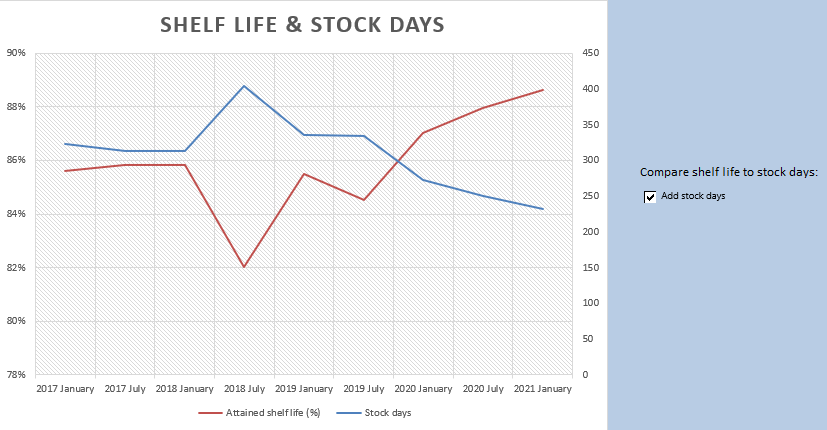
In picture 3.1.2 the data has a more logistical vision. First the shelf life of products is shown, this shows how fresh the products are that the company delivers to the customers. On the right the rejects from the product line are shown.

*Picture 3.1.1: Top half dashboard Excel (Own work)*

*Picture 3.1.2: Bottom half dashboard Excel (Own work)*

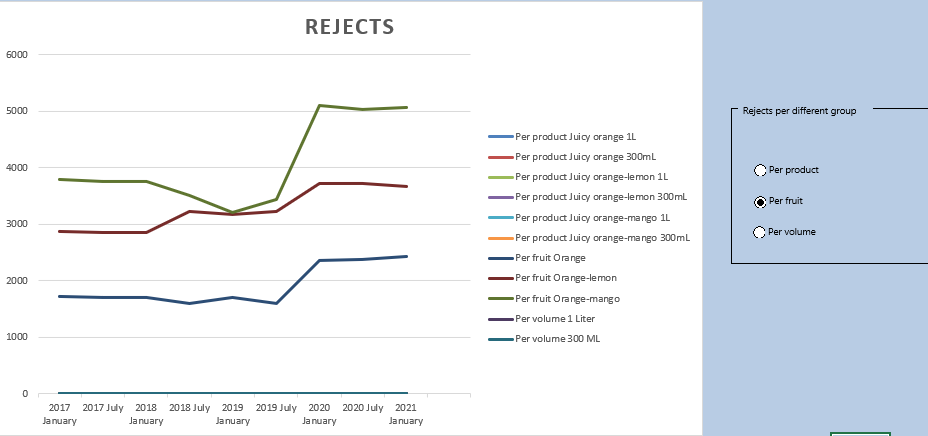
The goal was to make interactive graphs. As you can see in the pictures above, there are select and choice boxes. For the margins for example, you have the option to show margins in comparison to the offered service level. Then it is easy to see what the company earns from a customers to witch offered service level and if maybe the service level should be raised. When selecting the other option, the graph changes into what can be seen in picture 3.1.3.

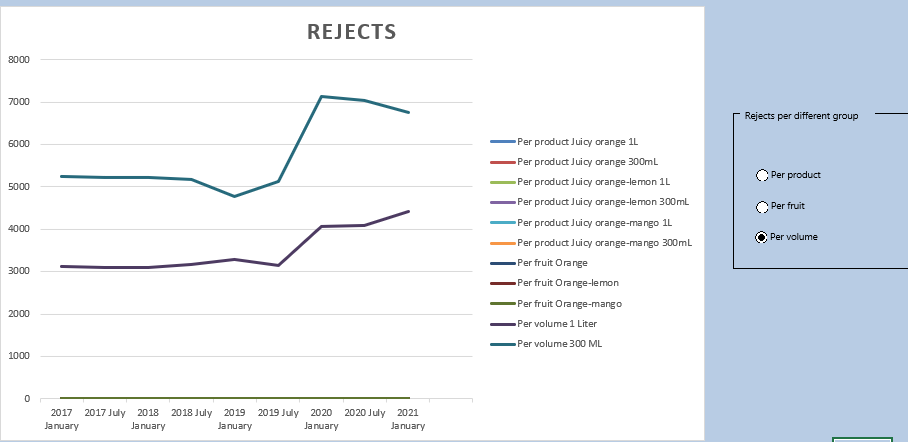
*Picture 3.1.3: Interactive graph margins and service level Excel (Own work)*

Also the logistical graphs are interactive to give to viewer an option to get a more deeper view on the data. If you check the box ‘Add stock days’, another line graph appears, this is shown in picture 3.1.4. The viewer can see the development of the shelf life over the years and what this means for the stock days for the company. As you can see are these variables directly influenced by each other, there is some kind of correlation.

*Picture 3.1.4: Interactive graph shelf life & Stock life Excel (Own work)*

For the last graph, there are three options for the viewer. He can see the rejects per product, per fruit or per volume. Because of this he can more easily track down where the problem lays, since he can see the rejects at multiple categories. Pictures 3.1.5 and 3.1.6 show these categories.



*Picture 3.1.5: Interactive graph rejects per fruit Excel (Own work)*

*Picture 3.1.6: Interactive graph rejects per volume Excel (Own work)*

**Reflection**

My goal for Excel as a dashboard was to make the graphs interactive, I am quite satisfied on how I did this. At first I struggled with making another choice selection, because the buttons linked with the ones from the margins graph causing that graph to change as well. After reading some articles online I found out that you can make a group of selection buttons, which solved my problem.

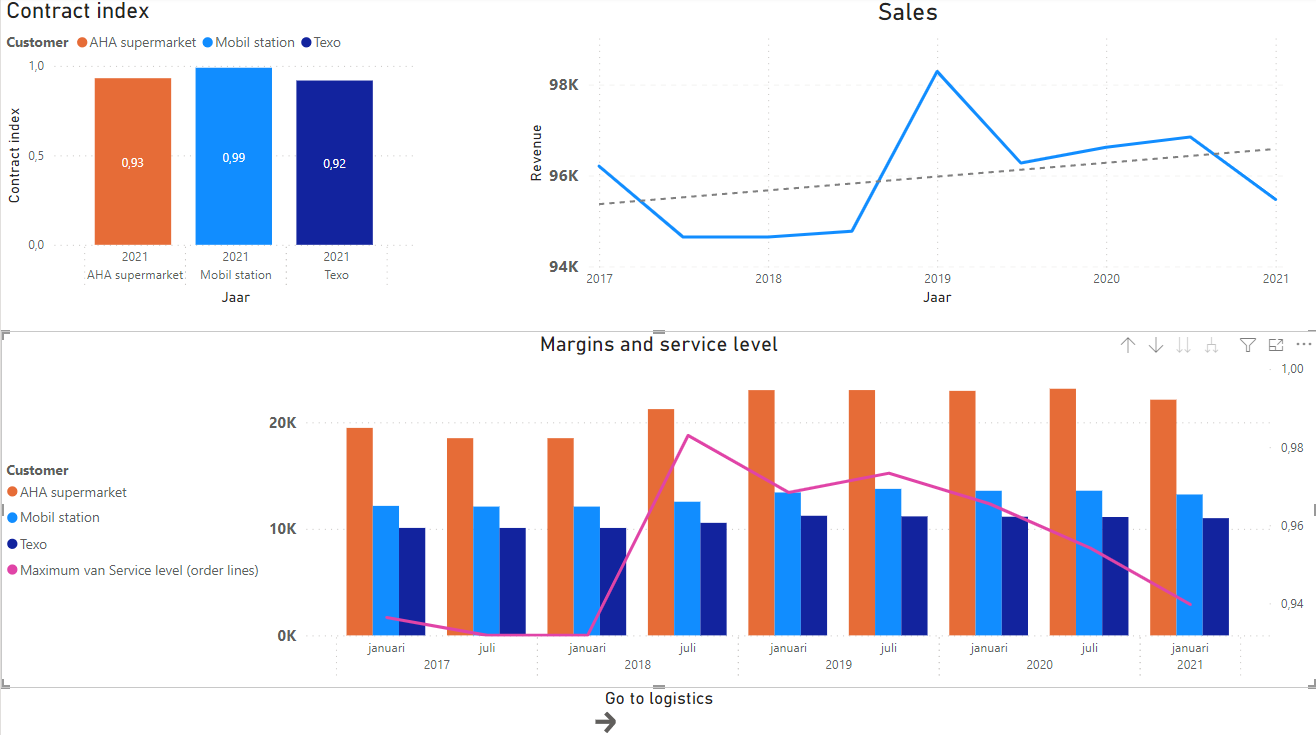
For my personally there is one particular downside to my graphs. The legends are very cluttered and not how I wanted it to look. In the ideal situation the legends would change with the line graphs, but now the legend include the lines of all choices.

I know have a better understanding of how to make graphs interactive in Excel, this can help to make data more understandable. I also learned that Excel is not very flexible in particular things and sometimes it is quite hard to adjust graphs to your liking.

## 3.2 Visualization and justification Power BI dashboard

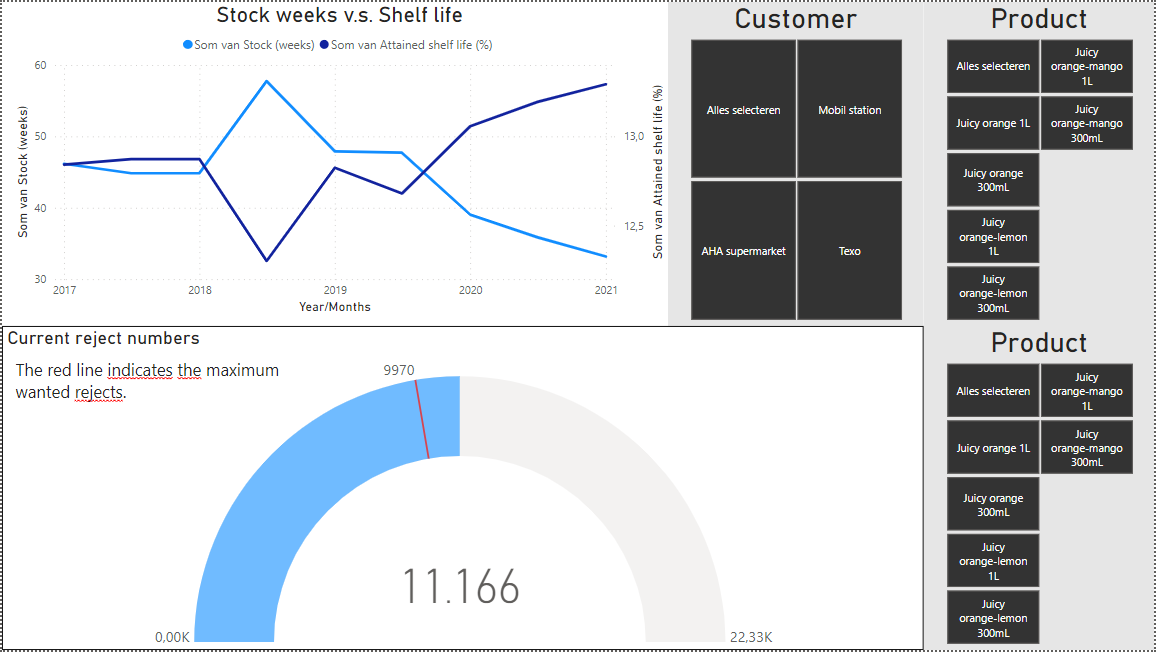
**Description**

In the pictures 3.2.1 and 3.2.2 the dashboard made in Power BI is showed. On the top left and right, again the most important indicators.

Below this the margins and service levels are shown. I was not able to add an selection option in Power BI, therefore the margins and service levels are always shown together. I also was not able to make multiple line graphs in this graph, I only could add one line. I did make a button that when it is clicked on, you automatically go to the logistical graphs.

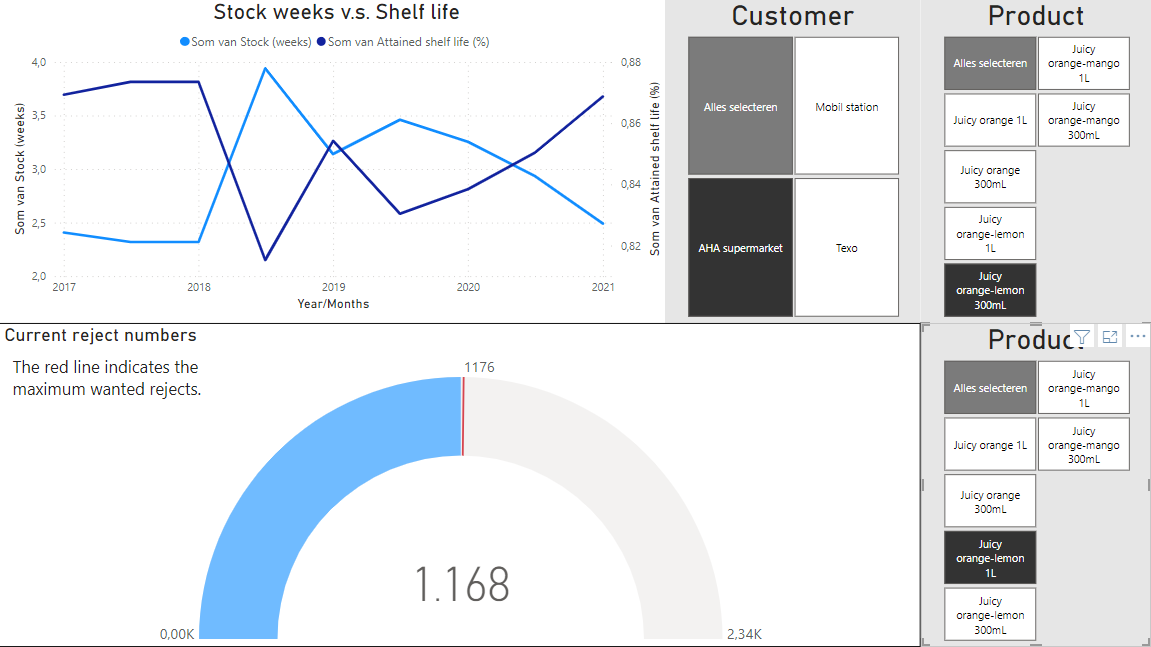
*Picture 3.2.1: First page dashboard Power BI (Own work)*

In picture 3.2.2 the logistical page is showed. First the stock weeks versus the shelf life is showed, also in this case I wasn’t able to add a function which manually adds a line. Instead of that, I decided to add slicers, one for the customer and one for the products. In this way the viewer has the option to compare customers and or products. This gives the viewer a change to see were the most improvement is needed.

Under this the current rejects numbers are shown, because I was not able to make an similar graph as in Excel I decided to make a KPI meter of it. To make this meter I struggled quite a lot, also because at first I wanted to have the same slicer work on it. Because the data for the graphs came from different tabs in my database, I could not succeed. I decided to manually make a different table, copying together the data I needed and using the average rejection numbers as the maximum of rejects. Then I added a slicer to make the graph interactive.

*Picture 3.2.2: Second page dashboard Power BI (Own work)*

In picture 3.2.3 you can see how the graphs look different when slicers are applied.

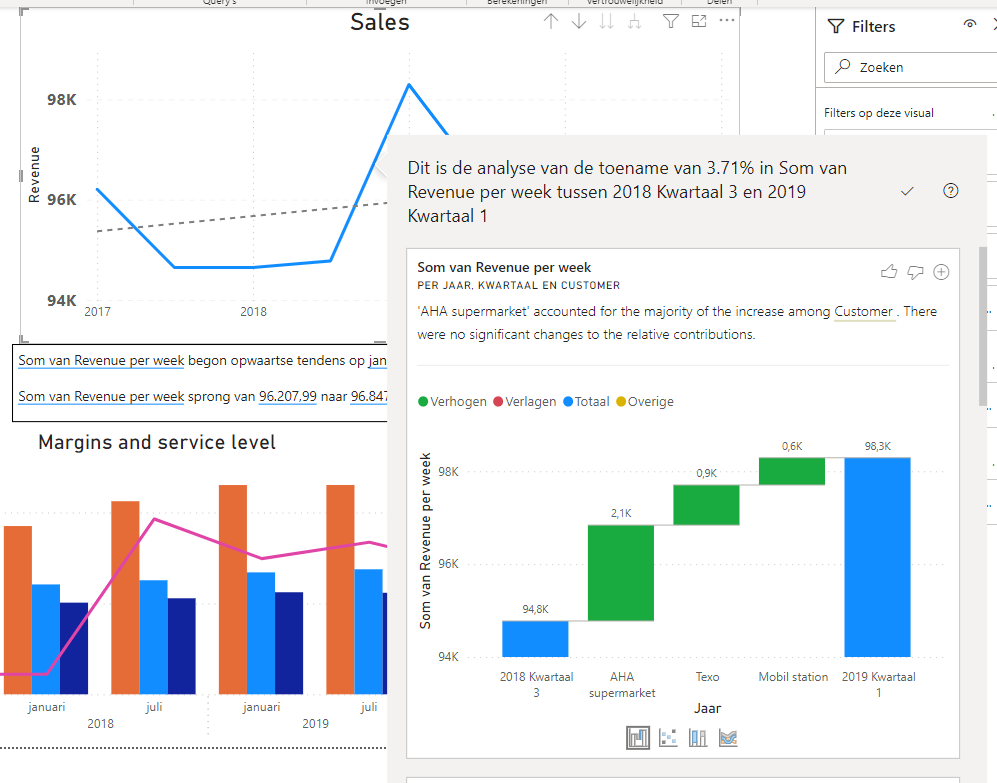


*Picture 3.2.3: First page dashboard Power BI (Own work)*

**Reflection**

My main goal for working with Power BI was getting to know the application. I really wanted this because the company where I will start my intern soon, asked me if I was familiar with Power BI. I achieved this goal and I think I learned the very basics of the tools, but I think there are so many options that I don’t know of and have to learn about from somebody else.

I found out that the tool can automatically make summaries of a graph, describing the most important changes in for example a line graph. It is also able to make an analysis explaining a particular point in the graph. An example of this is shown in picture 3.2.4 Because of these options you notice that the tools is much more focused on data visualization and understanding then a tool like Excel.

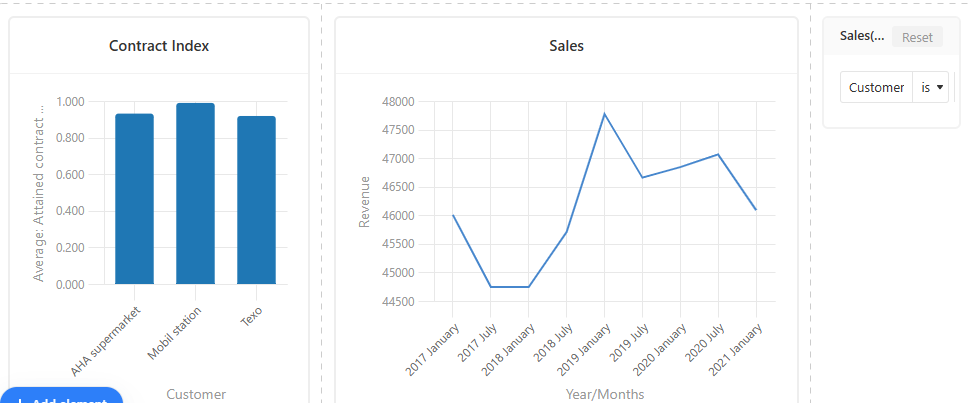
*Picture 3.2.4: Analysis option Power BI (Own work)*

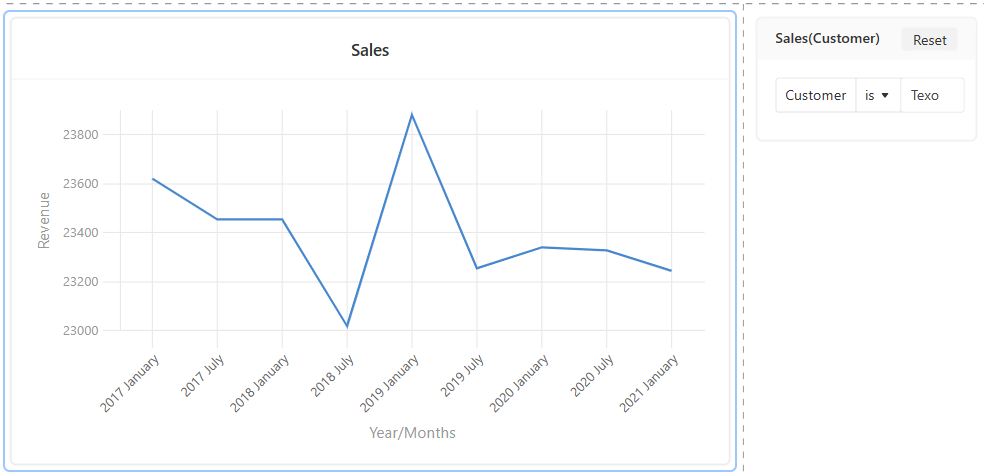
Something that I struggled with was wanting to use data from different tabs in the Excel database. You will have to manually make relationships between the tabs, but I did not succeed in this. Being able to do this will give you more options, so this something I want to learn in the near future.

## 3.3 Visualization and justification Airtable dashboard

**Description**

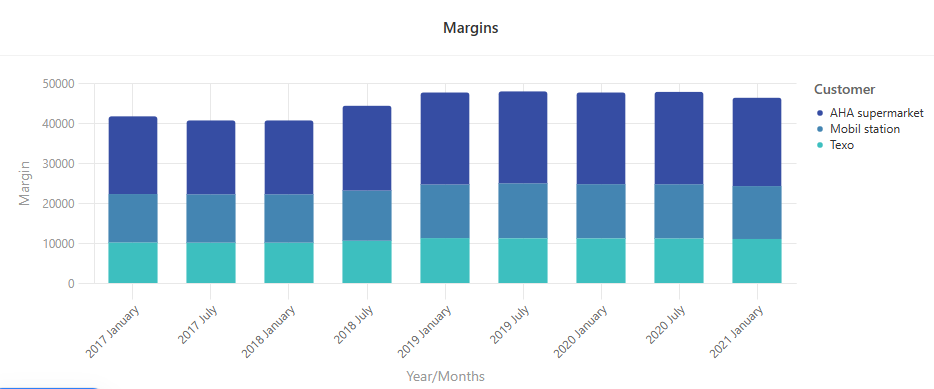
In the pictures 3.3.1 and 3.3.3 the dashboard made in Airtable can be seen. The first pictures shows to most important indicators. Something I struggled with was getting the columns for the first graph in different colours, but as you can see I was not able to do this. Also a trend line in the sales graph was not possible. I did add a filter for the sales to see the revenue per customer so that there was at least one extra dimension. An example of the applied filter can be seen in picture 3.3.2.



*Picture 3.3.1: Top half dashboard Airtable (Own work)*

*Picture 3.3.2: Applied filter dashboard Airtable (Own work)*

Below this the margins are shown. I was not able to add lines to a bar chart, so I could not make an second y-axis to compare the service level with the margins. Therefore I decided to make an stacked graph where you can easily see which customer provided the largest margins.



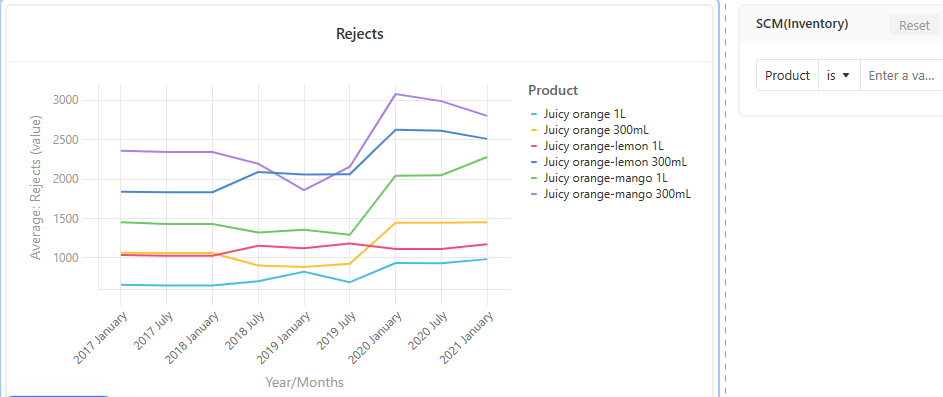
*Picture 3.3.3: Margins graph dashboard Airtable (Own work)*

I also was unable to select more variables in one graph, so also for the stock weeks versus shelflife I experienced some problems. I decided to place the graphs next to each other, but this makes the comparing harder. This can be seen in picture 3.3.4.



*Picture 3.3.4: Margins graph dashboard Airtable (Own work)*

For the rejects I could only add the products to a graph, not the volumes or fruits. Also to this graph I added a filter to select a particular product.



*Picture 3.3.5: Rejects graph dashboard Airtable (Own work)*

**Reflection**

I am quite disappointed in Airtable the options are very limited. I could only choose between a line graph, bar chart, pie chart, donut chart or chatter plot. When choosing one of these, I could only add 1 variable so it is hard to compare multiple variables to see if there is a certain correlation.

The program is easy to understand and does a lot of things automatically, but even though I think Airtable can only be used for very simple data.

# 5. Applying CRISP-DM to business case

## 5.1 Project process steps linked to CRISP-DM

Some steps that have been taken in the process of this project can be linked to the steps of CRISP-DM model.

In the case of this project there was not a business to understand. Instead of that, information about dashboards was gathered. Usually at business understanding the requirements and goals become clear, with the dashboard understanding requirements and guidelines for a sufficient dashboard were established.

Data understanding was very minimal, because the data set was already known and divided into the right departments and columns.

Data preparation happened later, the CRISP-DM model allows to go back to the circle, to repeat or do a step later. First modeling in the form of making graphs in the dashboard tools was done. For a particular graph the data needed to be adjusted and copied together, this was when the data preparation took place.

Evaluation was done by reflecting what was learned from making the dashboards. Deployment is discussed in the next subgraph.

### 5.1.1 Github link to deploy project

Deployment is the last step of the CRISP-DM model. The results are being presented in a suitable way.

This project is meant to be shared online so that it could perhaps help people from all around the world. To make this possible, this report is posted on Github. Below a linked to the report is placed:

<https://github.com/Anne-Sophie31/Portfolio.git>

# 6. Code / Script quality

This chapter does not apply to the goal of this report. I want to focus on making dashboards and getting familiar with the tools that can be used for this.

I expect to get in touch with dashboards in my professional career, but not with anything like the data cleaning part. Therefore I decided not to use code / scripts in my learning curve.

# Literature

Beek, D. van. (n.d.). *KPI: Key Performance Indicators*. <https://www.passionned.nl/strategie/pm/kpi/>

Calzon, B. (2022, 1st September). *The Top 20 Supply Chain Metrics & KPIs For Your Dashboards*. BI Blog | Data Visualization & Analytics Blog | datapine. <https://www.datapine.com/blog/supply-chain-metrics-and-kpis/>

Hotz, N. (2023, 8 januari). *What is CRISP DM?* Data Science Process Alliance. <https://www.datascience-pm.com/crisp-dm-2/>

Ozuysal, J. (2022, 21st September). (*Best 10 Free Dashboard Tools in 2022 [In Depth Guide]*. (z.d.). <https://www.datapad.io/blog/free-dashboard-tools>

Tableau, (n.d.). *What is a dashboard? A complete overview*. (z.d.). Tableau. <https://www.tableau.com/learn/articles/dashboards/what-is>

Tazmi, K. (2022, 13th September). *Wat is een data dashboard? Uitleg, meerwaarde & tips*. Decidas. <https://decidas.nl/wat-is-een-data-dashboard/>