*Introduction company*

*Explanation assignment*

*Description data*

*Analysis*

*Building the tool*

*Other project (On-Time)*

*Conclusion / Further research*

# Table of contents

Heineken 3

The assignment 4

Events 4

Requirements of the tool 4

Description of the data 6

The dataset 6

Advantages of building your own dataset 6

Disadvantages of building your own dataset 6

Available data 7

# Heineken

Heineken International is the largest Dutch brewing company, and does not need an introduction for most people. Heineken is the third largest brewer in the world, with over 180 million hectolitres of beer produced per year. Once the British – South African brewing company SABMiller (known in the Netherlands for Grolsch and Bavaria) has been taken over by the world’s largest brewer AB InBev it will be the second largest brewing company.

Since the foundation in 1864, Heineken has grown to produce more than 250 different beers that are served in 192 countries. In addition to of course their own brand, they are known in the Netherlands for Amstel and Brand beer, as well as e.g. Jillz (apple cider) and Desperados (tequila flavoured beer). As a sponsor of the UEFA Champions League and organiser of the Holland Heineken House during the Olympic Games they have gained a lot of international publicity.

Heineken employs approximately 76.000 people internationally, owning more than 160 breweries in more than 70 countries. With the emergence of data analytics Heineken would like to know how they could improve different aspects of their supply chain. One of these aspects will be discussed in this paper.

*More information about the units and the breweries*

# The assignment

Heineken has a lot of customers all over the world, and every customer is different. Their brands are served in many different types of restaurants, pubs and clubs, and their products are also available in supermarkets. Because these supply chains differ a lot from each other, they are part of different branches of the company.

Since the demand of supermarkets is quite consistent and therefore relatively easy to predict, the other branch is much more interesting in terms of data analysis. The consumption depends on how many customers visited the restaurant, pub or club, and not everybody drinks the same amount of beverages. During weekends and on holidays it is a lot busier, which influences the size of the order Heineken receives. To optimize the supply chain it is important for Heineken to be able to predict how much they should have in stock in order to supply each of their customers on time.

## Events

Apart from the regular orders, customers can place additional orders for events they organise. As the organiser of an event you want to make sure that you order enough, so you will not be out of stock during the event and miss out on possible revenue. Therefore, in the case of an event the customer arranges with Heineken thatproducts that are not consumed can bereturned. If the amount of returned products is smaller than 15 per cent of the original order there will be no additional costs for the customer. However, when the amount of returned products is larger than 15 per cent, the customer has to pay for each extra product that he/she sends back to Heineken.

When products are returned there will also be costs for Heineken. Everything has to be collected, shipped back to thedistribution centre, sorted and put back in the warehouse. Therefore Heineken also benefits from having as few products as possible returned.

Currently the organisers determine the amount of products they order based on consumption during previous events, and if there are no previous events they have to guess how much they need for their event. They are assisted by one of Heineken’s account managers, who have a lot of experience with similar events, but the decisions are not data-driven. Therefore, Heineken would like to support their account managers by creating a tool to predict the consumption on events.

## Requirements of the tool

Since the account managers have to use the tool eventually, they have to be able to quickly understand the functionality of the tool. Therefore, the dashboard has to be clear and intuitive. The users have to see at a glance where the input is, and what they have to give as input. Because the tool is going to be used to predict consumption and create orders based on those predictions, the users have to be able to convert or use the output as an order. What this means in practice will be explained in the chapter ‘Building the tool’.

Account managers do not have a lot of time when they want to make a prediction (they do not want to make the customer wait for a long time when they are creating an order), so it is also important that the tool is fast. Once all of the input fields are filled in, it should not take long before the output is generated. It is also possible that a customer changes his / her mind during the meeting with the account manager, in which case it has to be easy to compute new predictions for the event.

The third requirement is quite an obvious one, but nevertheless a very important one. Since the tool is going to be used to create orders, it directly influences the number of products supplied to the events. This means that the tool should always give an advice consisting of enough products, to prevent the organisers of the events from running out of stock. In order to achieve the goal of reducing the number of products that are sent back to Heineken, the tool should not predict a consumption that is too high either. Reliability is therefore another important requirement.

We can conclude that the main requirements of the tool are usability, performance in terms of speed and reliability. We will take these requirements into account when building the tool, and I will elaborate further on how I am going to do this in the chapter on ‘Building the tool’.

# Description of the data

In this chapter I will not only give a description of the data, but also explain how the data is stored and how I had to obtain the data.

## The dataset

Since there wasn’t a dataset already available, I had to create my own dataset. There were some advantages, but of course also some disadvantages to this. I will shortly discuss those advantages and disadvantages.

### Advantages of building your own dataset

* *You are able to decide immediately which features you do, and which features you do not want to include in the dataset*

Instead of having to filter the data before you can start you analysis, you can decide beforehand whether you want to include features in the dataset or not. This saves time and prevents you from having to save a lot of different versions of the dataset (with different combinations of features).

* *You are able to determine the formatting of data*

When you enter the data into the dataset, you can make sure that it is in the correct format. For example, you can enter dates in the date format you prefer, instead of having to perform a transformation on that column afterwards.

* *You know exactly what the dataset looks like*

While this may seem obvious, it is still a big advantage. Instead of having to get familiar with the data first, you already know exactly what is in the dataset. This also gives a better overview of the possibilities for further analysis.

### Disadvantages of building your own dataset

* *It takes a lot of time to build your own dataset*

If you have a ready-to-use dataset, you can immediately start exploring the data and make a method of approach based on the data that is provided. If you have to build your own dataset this is not the case. It takes time to decide what kinds of data you want / need, it takes time to gather the data and it takes time to enter the data into your dataset.

* *You have to decide what data you want to use before you start the analysis*

When you are analysing the data and you decide that you want to include certain features after all, you cannot just take them from the original dataset, since the features were left out in the first place. Therefore it is very important to determine the types of analysis you want to perform, so you can make sure all the data you need is added to the dataset.

* *There is a higher chance that mistakes are made*

Even though you can check the data you put into the dataset, a mistake is made easily. It requires a lot of carefulness to ensure that the data in the dataset is correct and reliable.

## Available data