*Introduction company*

*Explanation assignment*

*Order processing*

*Description data*

*Analysis*

*Building the tool*

*Other project (On-Time)*

*Conclusion / Further research*

# Table of contents

Table of contents 2

1. Heineken 3

1.1 Breweries 3

1.2 Distribution units 3

2. The assignment 4

2.1 Events 4

2.2 Requirements of the tool 4

3. Order processing 6

3.1 Placing an order 6

3.1.1 The website 6

3.1.2 The call centre 6

3.1.3 The account managers 6

3.2 SAP 6

3.3 Order picking 7

3.4 Delivering the order 7

4. Description of the data 9

4.1 The dataset 9

4.1.1 Advantages of building your own dataset 9

4.1.2 Disadvantages of building your own dataset 9

4.2 Available data 10

4.2.1 Event forms 10

4.2.2 Order details in SAP 10

4.3 Usability of the data 11

4.4 Building the database 12

4.4.1 Factors 12

4.4.2 Product categories 14

4.4.3 The assortment table 16

4.5 Filling the database 16

5. Analysis 17

5.1 Influence of different factors 17

5.1.1 Expected number of visitors 17

5.1.2 Type of event 19

6. Building the tool 20

7. On-Time project 21

8. Conclusions and further research 22

Appendix 23

# 1. 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.

## 1.1 Breweries

Text

## 1.2 Distribution units

Text

# 2. 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.

## 2.1 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.

## 2.2 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 we will elaborate further on how we are going to do this in the chapter on ‘Building the tool’.

# 3. Order processing

There are a lot of steps between the customer placing an order and receiving that order. In order to understand the data and the importance of the tool in the order process we will explain the different steps of placing and processing orders.

## 3.1 Placing an order

There are three ways for customers to place an order: on the website, by calling the Heineken call centre (located in Houten, the Netherlands) or through their account manager (in the case of an order for an event). Each method is slightly different:

### 3.1.1 The website

When a customer visits the website, he / she can browse through the entire assortment that Heineken offers to their customers. The customers can log in with their own customer ID, and place their order. This order is then sent to Heineken and entered into SAP (see chapter X for more information on SAP).

### 3.1.2 The call centre

Customers can also place their order by telephone. The call centre in Houten receives every call, and processes the order. The employees in the call centre do not only take customer calls, they also make calls themselves. A lot of regular customers, such as restaurants and cafés, place their order on the same day every week. A call centre employee calls them one day before they want to receive the products and their order is entered into SAP.

### 3.1.3 The account managers

As mentioned before, the account managers assist the customers in creating the orders for events. They sit together with the customer and discuss which products the customer wants to order. Once they have decided what the order is going to be, the account manager fills in an event form with the details of the customer, the event and the order. The order is then entered into SAP.

## 3.2 SAP

SAP is the order processing system used by Heineken. In SAP all of the information about the stock is stored, as well as the information on the orders. As explained in the previous sections, (almost) every Heineken employee uses SAP. Since all of the supply chain information is stored here, it is a very useful source for data analysis. For example, the historical stock levels can be used to reduce future stock levels to save money. The historical order data can be used to predict future orders, as is the case in this project.

## 3.3 Order picking

In every distribution unit there is a large warehouse where the products are stored. Each different product has its own place on a shelf or on the floor, and has its own barcode. When an order has to be picked, the warehouse employees scan the order barcode to receive a list of products. They collect the products and scan the product barcode.



*Source:* [*http://www.ed.nl/economie/helmonds-heineken-voortaan-uit-oss-1.2139824*](http://www.ed.nl/economie/helmonds-heineken-voortaan-uit-oss-1.2139824)

This process minimizes the chances of someone accidentally forgetting a part of the order, since it is not possible to finish the order without scanning every product (unless, of course, a product is out of stock). Once they have finished an order, their hand terminal gives instructions on where to deliver the cart with the products. This way the truck drivers know where to find the carts with the orders they have to deliver the next morning.

## 3.4 Delivering the order

Each morning the truck drivers arrive at the distribution unit and receive a list with the customers they have to visit that day. They connect their truck to the dock they have been assigned, where the carts with their orders are placed. Before loading everything in their truck, they have to double-check every order, to make sure the order pickers did not forget something.



*Source: http://www.archiproducts.com/en/products/77787/loading-dock-kdr-kopron.html*

Once they have checked everything, they load the carts in their truck and start their route. Each customer has specified a time window, within which the products have to be delivered. The truck driver has to make sure that he / she delivers the order at the right time; otherwise the customer will likely file a complaint. More information about this can be found in chapter X (On-Time project).

# 4. Description of the data

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

## 4.1 The dataset

Since there wasn’t a dataset already available, we had to create our own dataset. There were some advantages, but of course also some disadvantages to this. Those advantages and disadvantages will be discussed shortly.

### 4.1.1 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.

### 4.1.2 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.

## 4.2 Available data

To create a dataset, we received the event forms of every event for the Amsterdam distribution unit in 2014. The event forms are digital files, although the account managers print an event form when they are meeting with a customer. Using these event forms we could look up the order in SAP to find the products that were actually delivered, and the products that were sent back.

### 4.2.1 Event forms

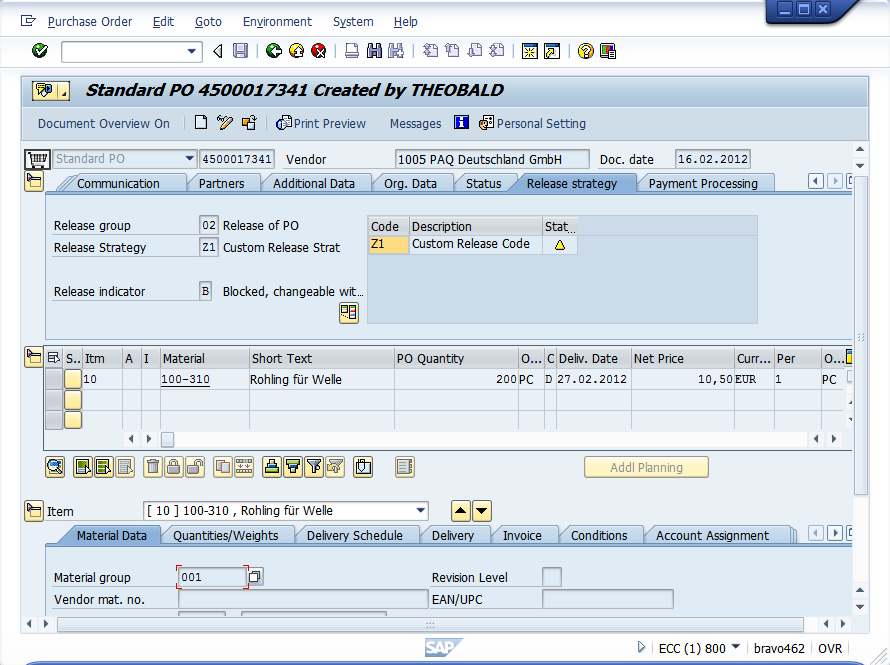
There were event forms for approximately 250 events, varying from big dance festivals to sporting events, and from rock concerts to parties in cafés. An event form contains all of the necessary information regarding the order for the event. Some important elements on the event form are:

* *Name of the event:* Naturally, the name of the event is included on the event form.
* *Ship-to number:* The identification number of the customer that placed the order.
* *Expected number of visitors:* The organiser and the account manager estimate the number of visitors of the event.
* *The order:* The details of the order are on a separate tab.

For an example of an event form, see appendix X.

### 4.2.2 Order details in SAP

For each event we looked up the order details in SAP, using the ship-to numbers on the event forms. In SAP every order has a separate number, which makes it easy to differentiate between different orders placed by the same customer. Since deliveries and returns are both entered as orders, there is a column indicating whether an order is a delivery or a return.



*Source: http://help.theobald-software.com/ERPConnect-Services-EN/default.aspx?pageid=bcs-purchase-order-release-in-sap*

When looking at a separate order, there are several important columns that we can use:

* *Product code:* Each product has its own unique product number, used to identify each product.
* *Amount ordered:* The amount of the product that has been ordered.
* *Product description:* A description of the product, including the brand, type and volume (e.g. Heineken cans 6x4x33cL)

This information will all be included in the dataset, because in order to predict the consumption at future events, we have to know how much has been consumed at previous events.

Unfortunately it is not possible to directly export an order from SAP, so we had to manually copy the contents of each order into the dataset. In the section on ‘Filling the database’ (chapter X) we will explain how we automated this as much as possible, to save time and to make it easier for future users to add new events to the event database.

## 4.3 Usability of the data

There was data available for approximately 250 events, but not every event was suitable to be included in the dataset. There are several reasons for an event not to be included in the dataset:

* *The event did not have a return order*

For some events such as fairs, especially events organised by cafés, there was no return order in SAP. When cafés organise an event and they still have products left after the event, they keep it and sell it later. Because of this it is not possible to know the exact consumption at the event, and therefore we cannot include these events in the dataset.

* *There is missing information*

On some event forms not all fields are filled in. Some information is crucial, such as the number of expected visitors at the event. If this information is not present, the event cannot be used and is therefore not included in the dataset.

* *The order is only half an order*

This category of orders is not completely useless, but needs to be treated differently. Not every customer orders everything at Heineken. Organisers may choose to order their beer at Heineken, but their wine and their soft drinks at another supplier. Since we cannot be sure that these customers order all products of one category (a category is e.g. beer, wine, etc.), the data of these events needs to be handled with care. These events can probably be used for analysis on consumption per category, but definitely not to predict the total consumption.

## 4.4 Building the database

Because of availability and usability, we chose to build the database in Excel. There was no server available to run the database on, so the tool and the database have to be ‘portable’. Therefore an Excel file with different tables is used.

With the usable events, we could start building the dataset. However, before entering the data of the events in a database, we first had to determine what kind of additional information we wanted to collect. Since some of the events are very different from each other, we cannot build a prediction model based on just the expected number of customers. Therefore a number of factors had to be specified, to be able to differentiate between different events.

### 4.4.1 Factors

* *Expected number of visitors*

This information is already available on the event forms, but it is still an important one to include in the list of factors. The number of visitors has a large influence on the total consumption at the event, so this is one of the key factors.

* *Type of event*

To determine the different types of events, and assign a type to each individual event, we made use of the experience of the account managers and other people involved in the event service. We did not want to create too little event categories, to prevent events that are too different from each other to fall in the same category, but we did not want to create too many categories either, because the information gain would be too little. Eventually we came up with the following five categories:

* *Village festivals:* Local festivities and fairs
* *House:* House- and dance parties and festivals (e.g. Amsterdam Open Air)
* *Jazz:* Jazz concerts and festivals (e.g. North Sea Jazz festival)
* *Rock:* Rock concerts and festivals (e.g. Oranjerock)
* *Sport:* Sporting events (e.g. Nederland – Wales)
* *Indoor / outdoor*

Another factor that has been included in the dataset is whether an event is indoors or outdoors. An event being indoors or outdoors has two important consequences:

* The influence of the weather is significantly less for an indoor event. For outdoor events the weather may influence the consumption pattern of visitors, because people prefer other drinks when it is cold than when it is warm. For indoor events this does not matter that much, because the temperature can be regulated and is usually approximately the same for all indoor events.
* The expected number of visitors is usually more accurate for indoor events. Most of the times you have to buy tickets for indoor events, while this is not always the case for events that are outdoors. Referring back to the previous consequence, the probability of visitors not showing up due to the weather is smaller for indoor events than for outdoor events.
* *Temperature in degrees Celsius*

The temperature on the day of the event is also added to the database. If an event lasts more than one day, the average temperature over the days is taken. As mentioned in the previous section, the temperature influences the consumption pattern of visitors. When it is warm, people may drink more water than when it is cold, while people may drink more red wine when it is cold than when it is warm.

* *Precipitation in millimetres*

The last factor that has been included is the precipitation (usually in the form of rainfall) in millimetres. It is possible that whether it is dry or it is raining influences the consumption behaviour of visitors. This will be further investigated in the next chapter ‘Analysis’.

In addition to the factors, there was another part of the database that had to be constructed. In SAP, the product description and the amounts of products are available, but this amount does not say anything about the product category and the number of litres. Therefore another table, containing the product category and the number of litres per product needs to be added to the database.

### 4.4.2 Product categories

To be able to predict more accurately than just the total consumption at events, the products are divided into different product categories. Since Heineken does not only supply beer but also many other products such as wine and soft drinks, there are six main categories, consisting of several subcategories:

* *Beer:*

Heineken is known for its beer, so naturally beer is a main category. However, next to their lager, Heineken supplies a lot of different types of beer:

* Cider
* Malt beer
* Lager
* Radler (beer and sparkling lemonade)
* Rosé beer
* Specialty beer
* Beach beer (e.g. Desperados, tequila flavoured beer)
* Wheat beer
* *Soft drinks:*

Heineken also supplies a lot of different types of soft drinks. Vrumona, a producer of soft drinks, is part of Heineken and produces a lot of the soft drinks that Heineken supplies. Apart from the brands that Vrumona produces, there are also a lot of ‘external’ brands in the assortment. The soft drinks subtypes in the database are:

* Bitter lemon
* Cassis
* Chocolate milk
* Coke
* Energy drinks
* Fernandes (a soft drink of Surinamese origin)
* Yoghurt drinks
* Ginger ale
* Ice tea
* Seven up
* Orange soda
* Tonic
* Fruit juice
* *Liquors:*

In the liquors category all the strong liquors are included. These are not produced by Heineken, but they are supplied to customers:

* Apfelkorn
* Cognac
* Gin
* Jenever
* Liqueur
* Mixed drinks
* Port
* Rum
* Sherry
* Tequila
* Vodka
* Whisky
* *Water:*

Water is a relatively simple main category, since there are not a lot of variations on water. It is important as a separate main category nevertheless, because it is non-alcoholic and significantly different from soft drinks. The three subcategories are:

* Uncarbonated water
* Sparkling water
* Vitamin water
* *Wine:*

Like the liquors, the wine Heineken supplies is not produced by Heineken itself, but customers can still order a large variety of wines from the following subcategories:

* Mulled wine
* Sparkling wine
* Red wine
* Rosé
* Vermouth
* White wine
* *Others:*

The final main category consists of other products, like glasses, plastic cups and coasters. These products are not used in the prediction model, but they are included so they can be used for ordering in the tool.

### 4.4.3 The assortment table

In the assortment table, information on all of the products is stored. The table contains a few columns with information to make entering data into the database easier, and a few columns that can be used for analysis and prediction. We will shortly discuss the database elements.

* *Product code:* The unique code that is assigned to a product. When an order is entered into the database, the product code can be looked up in the table to easily add the product to the order in the database.
* *Main category:* The main category the product belongs to.
* *Subcategory:* The subcategory the product belongs to.
* *Description:* A description of the product.
* *Units per package:* The number of units in each package. A crate of beer, for example, contains 24 bottles.
* *Volume per unit in litres:* The volume of one unit of the product. The beer bottles in the previous example contain 0,33 L.
* *Total volume per package:* The total volume of one package of the product, equal to the units per package times the volume per unit.

## 4.5 Filling the database

After determining and specifying the factors and product categories, the database was ready to be filled. For every event form order was looked up in SAP. The expected number of visitors was taken from the event form and filled in the table. The event type was not always filled in on the event form or sometimes the event type was not one of the specified event types in the tool. In that case the account manager responsible for that event was contacted to make sure that the correct event type was entered. This was also the case for the indoor / outdoor factor.

The temperature and the precipitation were obtained from (link X en link X). As mentioned in the section on the factors, the temperature and precipitation were averaged if the event lasted multiple days.

To make the input of a new order easier, a small VBA tool was created in which the user can paste the order copied from SAP. The tool then displays for every product in the order whether it is already included in the assortment table, and it adds the products that are included in the assortment table to the order in the database.

This is of course less efficient than exporting an order in one click, but in this case the possibilities were limited. This way the number of manual actions is minimized, so the database can be filled as soon as possible.

# 5. Analysis

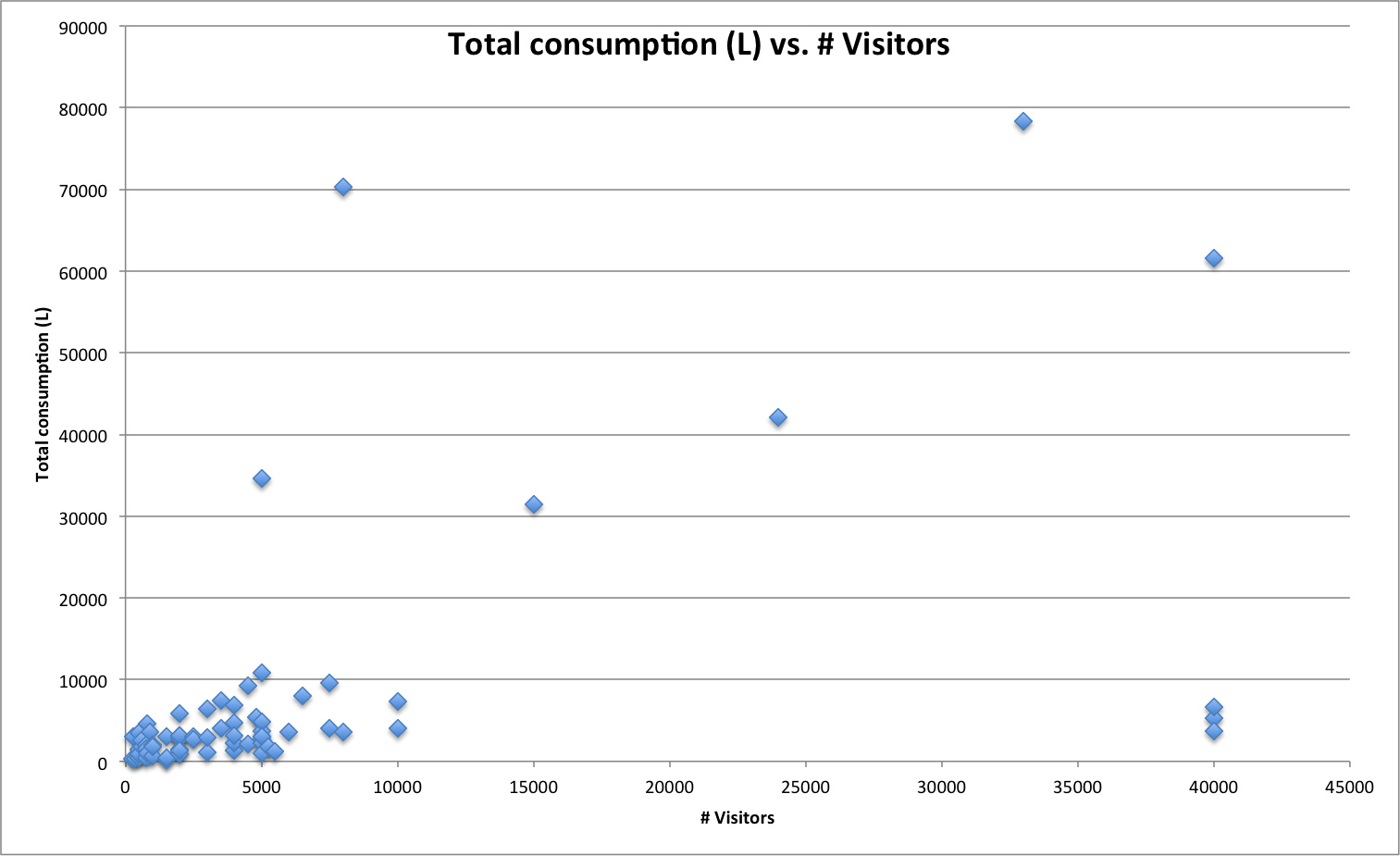
Before we can make a prediction model, we first have to analyse the influence of the different factors on the consumption. In the next sections we will discuss the influence of the different factors, and take a closer look at combinations of factors. Because the consumption of drinks is sensitive to trends, we will also take look at the development of the consumption of different categories over time.

## 5.1 Influence of different factors

First we will take a look at the influence of the individual factors on the total consumption, and the consumption per main category. The results of this analysis can be used as input for a regression model for example. After analysing the individual factors, we will try combining factors to see whether they are correlated with the consumption.

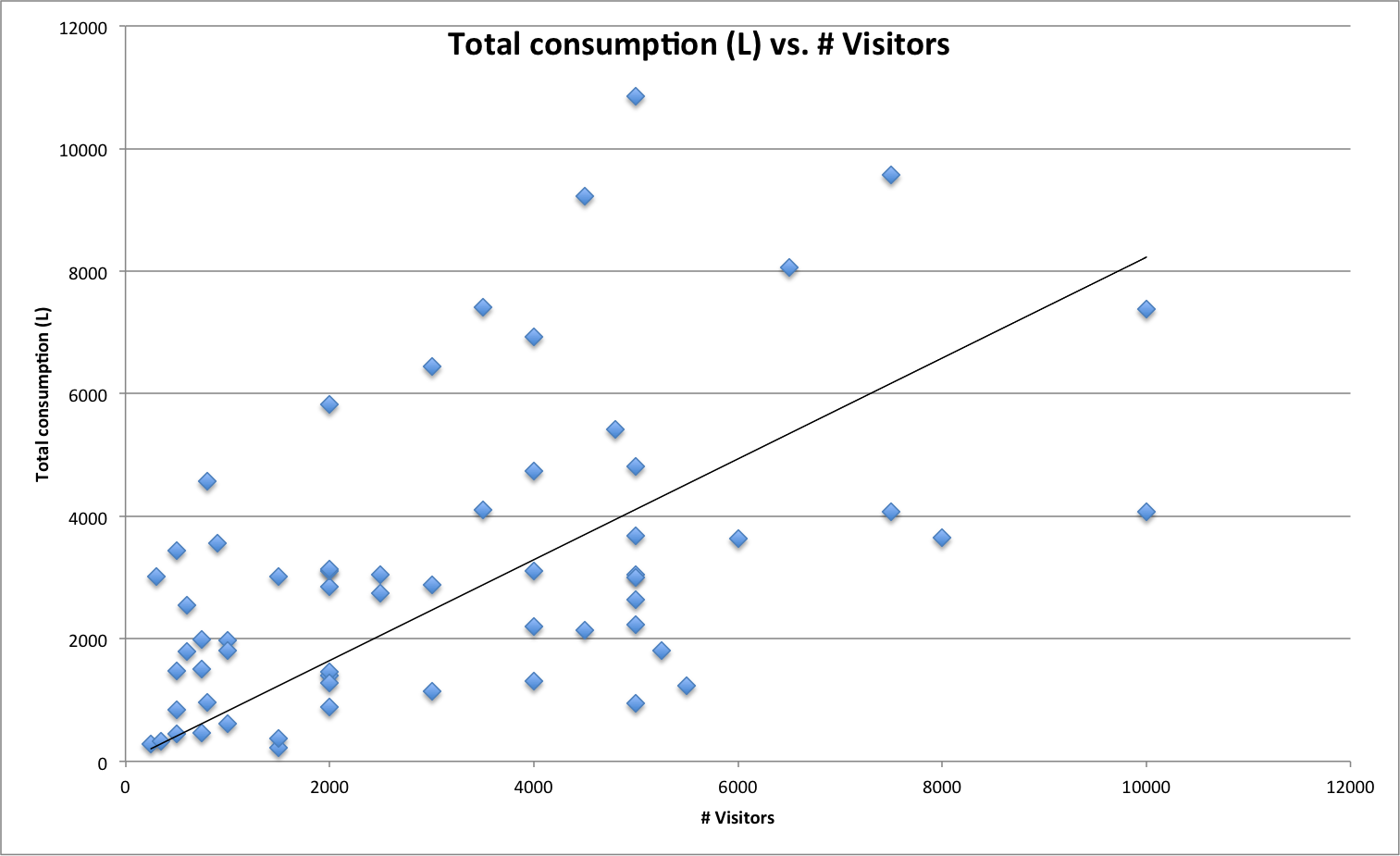
### 5.1.1 Expected number of visitors

The first factor that we will look at is the expected number of visitors. We will plot the total consumption in litres against the number of visitors to see what the influence on the total consumption is.



*Figure X: The total consumption (L) versus the number of visitors*

As shown in figure X, most of the data points are in the range of 0 to 10000 visitors, and 0 to 10000 litres consumed. There are some outliers with a lot of visitors and relatively little consumption and there are outliers with relatively few visitors but a lot of consumption. It might be interesting to take a closer look at the region where most of the data points are (figure X).

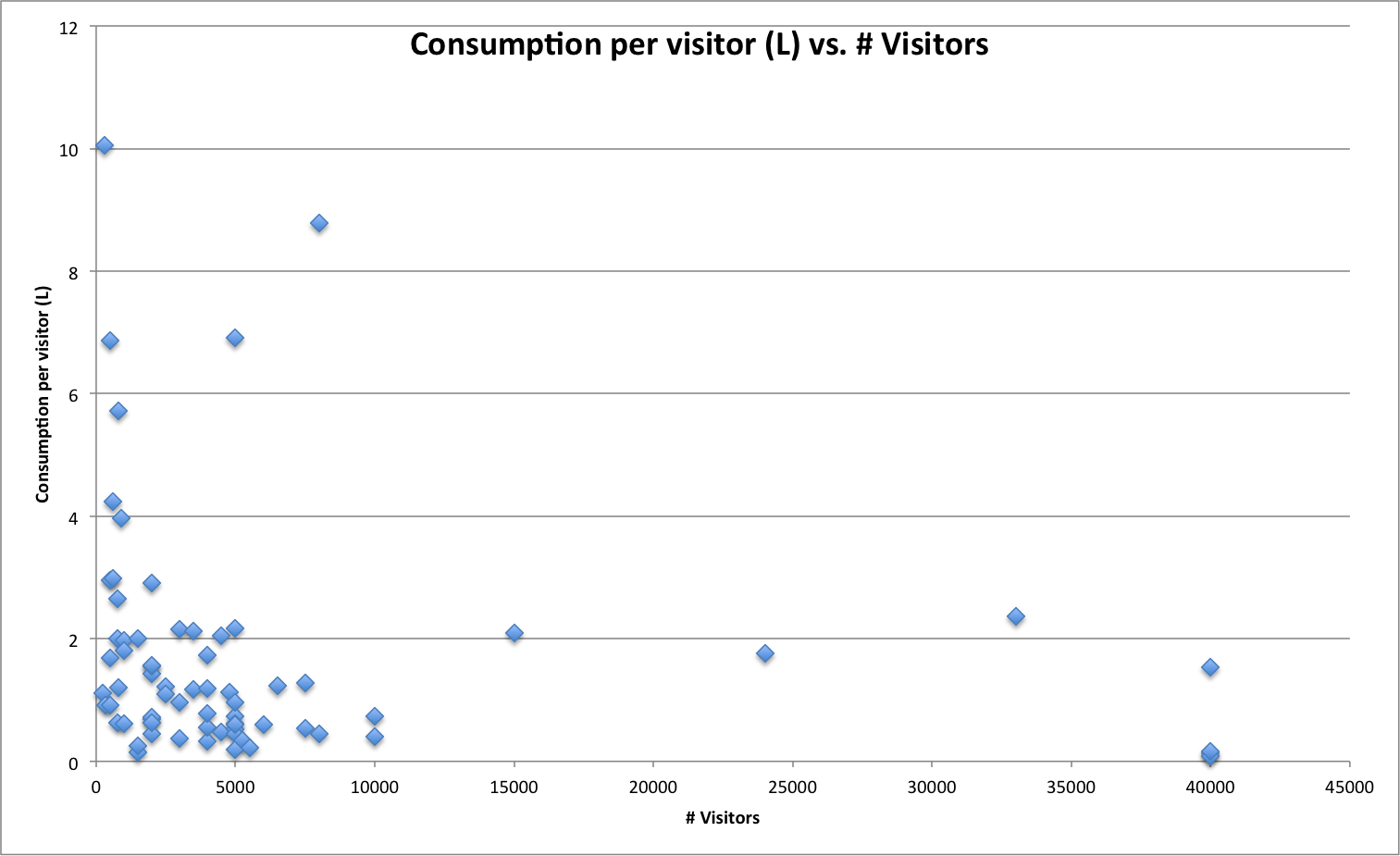


*Figure X: The total consumption (L) versus the number of visitors (zoomed in) with trend line*

There seems to be a linear trend, but the total consumption and the number of visitors do not seem to be very highly correlated. When we compute the correlation we get a value of 0,52. This confirms our expectations and tells us that the total consumption and the number of visitors are moderately correlated.

Something else that is interesting to look at is the average consumption per visitor against the number of visitors. That way we can determine if people consume more on average at bigger events, or at smaller events. The results are displayed in figure X.

There does not seem to be a correlation between the number of visitors and the average consumption per visitor. For some smaller events there is a relatively high average consumption per visitor, with sometimes up to 10 litres consumed per visitor. This is not very likely and probably an error in the data. For large events the average consumption per visitor is not significantly higher than for small events. The correlation between the average consumption per visitor and the number of visitors is -0,16 and therefore tells us that there is no significant linear correlation.



*Figure X: Average consumption per visitor (L) versus the number of visitors*

### 5.1.2 Type of event

The next factor to be analysed is the type of event. To see if the type of event has a strong influence on the consumption, we will take a look at the average consumption per visitor for the different types of events.

# 6. Building the tool

Text

# 7. On-Time project

Text

# 8. Conclusions and further research

Text

# Appendix

Text