|  |
| --- |
| Business Case |
| “Life & Garden” - ADS-A 34 – Group B  Anjela Melkonyan, Ivan Germanov, |



Anjela Melkonyan Ivan Germanov

Stijn Groenen Bart Verhagen

Ries van Geffen Redzhep Molaahmed

**Revisions**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Status** | **Date** | **Changes** |
| 0.1 | Concept | 10-10-2018 | Initial start. |
| 0.2 | Concept | 12-10-2018 | Added introduction, current situation, business goal, business requirements, scope and KPIs. |
| 0.3 | Concept | 19-10-2018 | Changed current situation, business goal, business requirements, scope and KPIs. Added approach and assumptions. |
| 1.0 | Concept | 08-11-2018 | Added names to cover, created version control, added exploration results, ethics and law, improved risks’ section. Changed KPIs. Improved Business Requirements section, removed risks from Executive summary, added conclusion. |

**Distribution**

This document is sent to:

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date sent** | **Name** | **Function** |
| 0.1 | 11-10-2018 | Niek Schmitz | Teacher |
| 0.3 | 19-10-2018 | Rob Prop | Client |
| 0.3 | 26-10-2018 | Niek Schmitz | Teacher |

#### Executive summary

Nowadays, a lot of organizations are opening up their hands to Big Data, unlocking its power and adopting its ways of working. Data science can empower management and officers to make better decisions, direct actions based on trends, challenge the employees to adopt the best practices and identify and refine target audiences. These are just a few of the benefits, a data-driven organization can gain.

Knowing that, a company with a web-shop in the sector of Life & Garden, has asked Informa, which operates in the area of Business Intelligence and Data Management for an advice. They have collected a lot of data over the past few years and want to use it to steer the organization in a beneficial way.

The purpose of this business case is to justify the undertaking of adopting data analytics as a means to improve Informa’s client’s market share and revenue.

To begin the operation of the project, the team has assumed that the company operates in the Netherlands and Belgium. To understand the dataset provided – some descriptive questions were asked. The questions can be found in [Appendix A](#_Appendix_A_–). Thus, answering them with Exploratory Data Analysis will help the understanding of the big amount of data provided and the visualization of the findings.

To explain and visualize why and how data can help the organization, a data analysis report will be conducted.

After adopting data analytics, the following benefits will be seen:

* opportunity to recognize the problems in a quantifiable manner
* identify unnoticed problems, which hinder the company’s progress
* aid the company’s decision-making process and long-term strategy

Each one of them is elaborated on further in the document with specific examples and explanations.

To see whether the results of the project can become opportunity for the company and to measure if the goal is met, the following KPIs will be used:

* Growth in revenue
* Growth in market share
* Customer retention

Going in depth of the data analysing, we have calculated some risks in the field of Data and Privacy. We will explain them in depth in the Risks section of the document.

During this project different types of analysis will be performed. Via their application the company will have the opportunity to recognize the problems it is facing in a quantifiable manner, identify new, previously unnoticed problems, improve its decision-making process and long-term strategy while utilizing the data.

Table of Contents

[Introduction 4](#_Toc529460282)

[Current Situation 4](#_Toc529460283)

[Business Goal 4](#_Toc529460284)

[Scope 5](#_Toc529460285)

[Approach 6](#_Toc529460286)

[Initial Research 6](#_Toc529460287)

[Exploratory Data Analysis 6](#_Toc529460288)

[Machine learning model 6](#_Toc529460289)

[Business Requirements 7](#_Toc529460290)

[Benefits 8](#_Toc529460291)

[Introduction 8](#_Toc529460292)

[Exploratory data analysis 8](#_Toc529460293)

[Areas of application 8](#_Toc529460294)

[Forecasting demand 9](#_Toc529460295)

[Understanding customer retention/customer churn 10](#_Toc529460296)

[Targeting recommendations 10](#_Toc529460297)

[KPIs and metrics 12](#_Toc529460298)

[KP1 12](#_Toc529460299)

[KP2 12](#_Toc529460300)

[KP3 13](#_Toc529460301)

[Risks 14](#_Toc529460302)

[Not enough data from internal dataset 15](#_Toc529460303)

[Data is not clean or is unorganized 15](#_Toc529460304)

[Insufficient data storage 16](#_Toc529460305)

[Data privacy 16](#_Toc529460306)

[Not being able to track market share due to lack of information 17](#_Toc529460307)

[Project will not be finished on time 17](#_Toc529460308)

[Constraints and Restraints 18](#_Toc529460309)

[Law and Ethics 18](#_Toc529460310)

[Limitations 18](#_Toc529460311)

[Prospects 18](#_Toc529460312)

[Assumptions 19](#_Toc529460313)

[Conclusion 20](#_Toc529460314)

[Appendix A – Research Questions 21](#_Toc529460315)

[Works Cited 24](#_Toc529460316)

# Introduction

The company Informa is knowledge centre for the Netherlands and Belgium in the area of ​​Business Intelligence and Data Management. Their consultants make analyses of existing environments, advise on sizing and migration and implement data models and data warehouses. They also play a key role in advising, implementing and training in Business Intelligence solutions.

One of their clients is a company with a web-shop in the field of Gardening & Lifestyle, where individuals can find professional and personal advice for home and garden in their store. There is a large assortment of gift and atmospheric items. But the main focus is on everything what is needed to set up and maintain a garden. Since a few years, web-shops have become a standard.

## Current Situation

Informa’s client is seeing quite a steady order amount across the last years, while their competitors see around 20% year-growth. Their business and revenue are not dropping down, but they are still interested in improvement.

As a means to achieving that goal, the company is interested in utilizing data, which they have collected a sufficient amount of during the past few years. The company wants to know how this data can help them improve their business to see growth in revenue and market share.

## Business Goal

The goal of the project is to capture the benefits of using Big Data and use them to increase the business’ revenue and market share. In addition to that, the company will also better understand customer churn. All of this will give the retailer a sustainable competitive advantage on the market.

In addition, a successful ending of the project will be to prove the company’s management that becoming a data-driven organization and encouraging data collection and analysis can steer the organization in a direction from which their market share can even further increase.

# Scope

The scope of this project will consist of 3 main categories:

1. Company research, discovering the benefits of undertaking the project
2. Analysis of the provided dataset and Data quality improvements
3. Development of Machine learning algorithm and reporting its findings

For the provided dataset, a selected number of angles, but at least two, will be further investigated.

The plan and what will be done step by step towards the Business goal is the following:

1. Research the field in which the company is working
2. Explore the benefits, which can be extracted from data according to the business field
3. Perform an Exploratory data analysis to understand the dataset provided
4. Use algorithms learn during the course to identify the problems the company might be having, as seen through the collected data
5. Report the findings
6. Define why certain things are happening and drill down to the root-cause
7. Create a machine learning model that will help the company reach its business goals
8. Prove the client that investment in data is worth it

# Approach

### Initial Research

The company, described in the introduction, has a web-shop in the market “Lifestyle and Gardening”. To successfully determine the problems and introduce solutions, the field of market must be properly researched. From that research will be determined relevant point of views to approach the problems as well as correct KPIs and metrics to evaluate the successful result of the project.

### Exploratory Data Analysis

The data, collected from orders and customers during the years, will be explored in detail with machine learning algorithms. For that purpose, a few descriptive questions were asked [(Appendix A)](#_Appendix_A_–). All finding will be reported in a Data Analysis Report, where the management of the company can easily understand the problems and see possible solutions.

### Machine learning model

To provide solution to the reported problems in the previous step, machine learning algorithms will be used to create a model. The areas of application of the model will be the following:

* Forecasting demand
* Understanding customer retention/customer churn
* Targeting recommendations

# Business Requirements

Rob Prop from Informa has given us a number of business requirements. With this we can indicate what his client wants to achieve with the desired situation.

|  |  |  |  |
| --- | --- | --- | --- |
| Category | ID | Business Requirement | Subscription |
| Data Driven Organization | BR1 | The client wants information to steer the organization wherefrom they can benefit with focus on correlations. | The client has a lot of data from day-to-day business but doesn’t use its advantage. |
| Company improvement as a whole | BR2 | The client wants the data analysis to touch all departments of the business. | The client wants the data to be used to benefit the whole company, not only a specific department. |
| Competitors | BR3 | The client wants to keep up with the competition or, even better, build up a lead. | The client tries to find out the opportunities they can benefit from by using the collected data. |
| Data Ethics | BR4 | The client wants to comply with the regulations in order to gain trust by his customers. | The client does not know how to deal with competitive information. |

# Benefits

## Introduction

The purpose of this introduction is to outline the benefits of applying data analytics. Firstly, they are explained in a more abstract way in this introductory subsection, whereas the more concrete application of data analytics is presented in the following subsections.

## Exploratory data analysis

To get a good impression of the dataset we created a list of questions. We answered these questions based on the results we got from the dataset. These results can be found in Appendix A.

## Areas of application

It is a well-known fact that successful web-based retailer services have been utilizing big data for a long time to drive growth. Some of the main areas they have applied data analytics to, are also the areas in which data analytics will be applied to in Informa’s client’s company. They are:

* Forecasting demand
* Understanding customer retention/customer churn
* Targeting recommendations

Even though some of these areas are overlapping at certain points, they can be separated into distinct categories. For instance, even though understanding customer retention can overlap with targeting recommendations and even their end goal can be the same (e.g. to increase revenue or to optimize customer retention), they are not one and the same thing for the reason that the derived knowledge from them differs. Where they overlap is that they both try to understand customer behaviour. However, the result of targeting recommendations is to market products to a customer on the basis of the customer’s behaviour. The result of trying to understand customer retention is the derived knowledge of the conditions under which the customer is likely to return or not (based on several criteria, such as rating score). Therefore, it is seen that they differ, even though what they are both trying to understand is customer behaviour.

Finally, how are problems going to be recognized in a quantifiable manner? When business problems are concerned, specifics matter. Thus, giving concrete examples is a proper way to demonstrate (a part of) the available opportunities.

### Forecasting demand

One instance is to forecast the demand of a given set of products. Forecasting demand can be done upon the basis of many criteria – seasonal changes, place of residence, popularity of given category of products, demographics (esp. male and female, as males drive around 70% of gardening purchase decisions (Figure 1)).

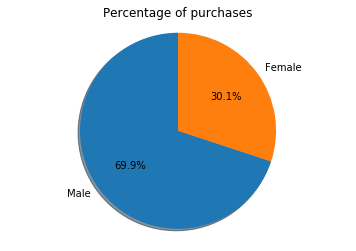


Figure 1

For the purposes of this example, forecasting demand will be based upon seasonal changes. In that case, the company will not have to predict demand based (only) on their instincts, but instead they will utilize the data they have collected in the past years to predict future demand based on facts and figures (e.g., which articles were sold most in the winter/summer season, what was the exact number of sales, where did they sell most, what is the customer profile most likely to buy a certain product, etc.).

Put into more abstract terms, by correctly assessing and learning from what has happened in the past, the company will adapt for the future. Put into more concrete terms, by analysing and learning from the trends of consumer decisions during the different seasons of the past, the company will be able to, through data analytics, more accurately determine the demand of certain products for the future, thus maximizing revenue and market share.

### Understanding customer retention/customer churn

In a business, it is way more difficult to acquire new customers in terms of expenses in comparison to optimizing customer retention rates and minimizing customer churn rates. Therefore, understanding why customers walk away or stay is key to reducing an organization’s expenses and increasing revenue.

According to research, it costs 5 times as much to attract a new customer than to keep an existing one (Lee Resources 2010). It is also the case that the probability of selling to a prospective customer is estimated at around 5-20%, whereas selling to an existing one is 60-70% (Farris, 2009). In addition to that, existing customers are 50% more likely to try new products while also spending 31% more in comparison to new customers (Saleh, n.d.). What is more, increasing customer retention rates just by 5% increases profits by 25% to 95% (Reichheld, 2018)! Those are some of the reasons why 70% of companies agree that it is cheaper to retain than to acquire a customer (Moth, 2013).

Now that we have established the value of retaining customers, how do we set out to optimize it by the usage of data analytics? First off, reducing customer churn starts with pinpointing the reasons why a customer decides to discontinue purchasing from the web shop and determining the probability of that customer leaving. Only after that will a churn management solution be possible to develop on top of the model in order to address the underlying causes for the churn. For instance, if customers receiving a late delivery are more likely to churn than customers who receive their delivery on time, that could be an indicator to inspect and improve the logistics department. The same also applies to users being more likely to leave after giving a bad rating to a product, which could be an indicator of the product’s quality. By account of the aforementioned examples, it can be deduced that by predicting the likelihood of customer churn, the company can identify the problem areas which sabotage the desired customer retention. Subsequently, by being cognizant of this information, the company can deploy a solution to ameliorate a problem area.

### Targeting recommendations

Most, if not all, successful web shops have started utilizing recommendation systems (engines) which provide personalized recommendations to their customers in order to increase user interaction and enrich shopping opportunities.

There are numerous benefits to deploying recommendation systems. First off, one of the main benefits of deploying a recommendation engine is improving cart value. A company such as yours has an inventory of thousands of items. Thus, it becomes apparent that there needs to be a means of filtering to suggest new products that a customer is likely to buy, which will result in increased purchases (and cart value).

Secondly, once deployed, a recommendation system gets better with time at predicting the likelihood a customer will like, then purchase, a product. The reason for that is that as more data accrues, the recommendation system will continuously learn to calibrate a customer’s preferences in a more and more accurate way.

Finally, having a recommendation system in place will improve customer engagement and satisfaction. As a user’s experience is made more relevant to them on a website, that results in increased engagement and satisfaction on their part. Increased satisfaction is likely to contribute to a higher customer retention rate and even revenue. Increased engagement is important, because it translates into customers staying longer on the website, which results into them viewing more potential offers, thus making them more likely to buy. In addition to that, a user’s time and attention are finite. Having them engaged on the web shop’s platform for longer means they are less likely to be on a competitor’s platform.

KPIs and metrics

In order to determine whether the given dataset provides one or more opportunities wherefrom Informa’s client can benefit we have drawn up a couple of KPIs which can be measured

In the table below an overview of success factors and corresponding KPIs is shown.

|  |  |  |
| --- | --- | --- |
| ID | Key success factor | Key Performance Indicator |
| K1 | Growth in revenue | There is a correlation between growth in revenue and a factor to be determined. |
| K2 | Growth in market share | There is a correlation between growth in market share and a factor to be determined. |
| K3 | Customer retention/churn | There is a correlation between customer retention and a factor to be determined. |

K1

To determine whether potential opportunity can be a success Informa’s client must track their revenue every quarter.

**Indicator**

All records from the dataset show a correlation between growth in revenue and a factor to be determined.

**Target**

The factor to be determined must have a positive effect of at least 10% on growth in revenue.

**Actions**

To measure growth in revenue, the current quarter’s revenue should be higher than the second last quarter.

K2

To determine whether potential opportunity can be a success Informa’s client must track growth in market share.

**Indicator**

All records from the dataset show a correlation between growth in market share and a factor to be determined.

**Target**

The factor to be determined must have a positive effect of at least 10% on growth in market share.

**Actions**

To measure growth in market share:

* The company’s total revenue will be calculated. All publicly-traded companies must release quarterly or annual financial statements. These statements will include a record of all of the firm’s sales, and may also include an itemized explanation of sales of specific product or service types within the footnotes of the financial statements.
* The total market sales (total amount of sales or revenue the entire market is acquiring) will be gathered. The market sales amount may be found through industry trade associations or publicly-available research reports.
* The target company’s total revenue will be divided by the entire industry’s total market sales.

K3

To determine whether potential opportunity can be a success Informa’s client must track customer retention/churn, which are two sides of the same coin. Churn rаte is cаlculatеd by dividing thе number оf custоmer cаncellations within a time period by the number of active custоmers аt the start оf that periоd. In this case, customer churn is non-contractual (customers are free to buy or not at any time, thus churn event is not explicitly observed). What needs to be done in such a case is to define a clear churn event timestamp. Frequently, this is done by finding a certain threshold for a period of inactivity and using it as a definition for the churn event.

**Indicator**

All records from the dataset show a correlation between customer retention and a factor to be determined.

**Target**

The factor to be determined must have a positive effect of at least 15% on customer retention.

**Actions**

To measure growth in customer retention, the period of time between two moments of buying products will be calculated. Regular customers come back each quarter so we focus on this period.

# Risks

Using data analytics there are pros and cons. Below a list is shown of what we believe to be potential risks, organized in a table. Keep in mind that the data utilized for the machine learning models does not only come from the provided dataset by the company but can also come from external datasets.

|  |  |  |  |
| --- | --- | --- | --- |
| Risk | Low | Medium | High |
| Not enough data from internal dataset | **** |  |  |
| Data is not clean or is unstructured |  | **** |  |
| Insufficient data storage | **** |  |  |
| Data privacy |  |  | **** |
| Not being able to track market share due to lack of information |  |  | **** |
| Project will not finished on time |  | **** |  |

Identifying risks and planning for how they will be handled prior to undertaking a project or even just assessing its hypothetical business value is key.

The aforementioned risks in the table will be further explained in more detail in the following sections. Then, for each risk, a risk mitigation subsection is included, in which a strategy for overcoming that particular risk is described.

There are 4 basic risk management categories – avoid, mitigate, transfer and accept.

Generally, avoiding a risk is the best solution as preventing it from happening bears the smallest risk to hurting the project. However, avoiding a risk is not always possible.

Mitigating a risk is less preferable than avoiding it, but more preferable than the other 2 options. Mitigation means taking a set of actions which will cause the risk to do as little harm to the project as possible.

Transferring a risk is also a viable alternative. This implies delegating the risk to a third party, usually by paying them. An example of transferring risk is insurance.

Accepting a risk is the least desirable solution. However, when all alternatives fail, there is nothing left but to accept a risk as is and make the best of the situation.

## Not enough data from internal dataset

#### Risk level - Low

#### Risk explanation

One of the associated risks is if there is not enough data. The risk for that case is low, because the client has provided us with a sufficient amount of data in their dataset. However, the risk still exists, because a big proportion of the records have certain attributes missing – mostly the optional attributes such as rating and reason for return.

#### Risk management

Mitigate  
The plan to overcome the risk is by mitigation. This means that in the project, there will be mostly a focus on machine learning models which do not need those missing attributes. However, there will still be machine learning models which will take them into account. In those cases, the risk has to be accepted that the accuracy of those models may be negatively impacted as there may be too few data to train the models.

## Data is not clean or is unorganized

#### Risk level - Medium

#### Risk explanation

Another possible risk is if data is not clean, meaning data that contains errors or inaccuracies, which will lead to faulty conclusions. Alternatively, the data may be unorganized and unstructured, as a survey has pointed out, about 80% of the data created in the world is unstructured (Schneider, 2017). The better the dataset’s data, the better the machine learning model. This risk may not only apply to the provided dataset but could also apply to datasets from external sources.

#### Risk management

Mitigate  
This risk can also be overcome through mitigation. The way the unclean data is going to be handled is in 2 ways. First, by choosing data sources that are as clean as possible. Second, by actually cleaning the data. It is a fact that data analysts spend between 50% and 70% of their time cleaning data (Press, 2016).

## Insufficient data storage

#### Risk level - Low

#### Risk explanation

Clearly, data storage is of the essence when it comes to utilizing big data. As machine learning models get deployed to bolster the organization, more and more data will be collected, which results in the models becoming more accurate. Since big data encompasses data sets of great magnitude, traditional data storage methods are outdated and inappropriate for big data.

#### Risk management

Transfer  
Since 80% of the data is unstructured, as pointed out previously, it is inefficient to store it in traditional relational databases as they need structured data. The need for technologies such as Apache Hadoop, NoSQL, etc. arises, therefore this risk needs to be delegated to experts in the field of data storage, who are skilled in such technologies.

It is vital to point out that this risk grows bigger once the company starts utilizing big data more and more efficiently as that growth in efficiency is usually accompanied with more new data. Not only more new data, but also more varying data, encompassing more facets, attributes of the customer, and even of previously untracked phenomena such as site activity, time spent on the site, site clicks, etc.

## Data privacy

#### Risk level - High

#### Risk explanation

Data breaches, discriminatory algorithms and identity reverse engineering are only a few of the embarrassment’s companies suffer if they do not pay enough attention to addressing the issue of data privacy. Companies must also be acting in accordance with data privacy laws and regulations. Otherwise, a company may suffer consequences such as losing customers as their trust in the company has deteriorated. An organization may also be the subject of lawsuits.

#### Risk management

Mitigate and Transfer  
Ensuring transparency is one way to mitigate the effects of possessing people’s personal data. That means disclosing what it is that is being collected to consumers and educating people as a whole. This encompasses explaining the advantages of analytics for the consumer in terms of better services, marketing, etc. to foster a fair value exchange in order that people are happy to provide their data.

In addition to that, rigorous governance systems should be deployed, preferably delegated to subject experts. Experts well-versed in the relevant data privacy laws should also be contacted to ascertain that laws are strictly followed.

Furthermore, intermediary third-parties can give out trust accreditations to help the company demonstrate that data privacy is of vital importance to it and is reliably taken care of.

## Not being able to track market share due to lack of information

#### Risk level - High

#### Risk explanation

Information about the current situation of market share is missing which makes it difficult to track whether this KPI.

#### Risk management

Accept  
We have to accept that there is a little we can do about this, but we still can gather current information to track the current market share.

## Project will not be finished on time

#### Risk level - Medium

#### Risk explanation

Project will not be finished due to lack of organisation in the project group.

#### Risk management

Avoid  
Assign a project manager who will keep track of what has to be done, by who and how it has been executed.

# Constraints and Restraints

## Law and Ethics

Emerging technologies in the field of data science promise to improve human communication, health, reproduction and cognition. Ultimately, in that process, these technologies may transform the human condition itself.

One of the biggest challenges within the legal-ethical governance of these developments is that each of these technologies radically challenges foundational concepts of legal and ethical systems of thought, such as responsibility, dignity, autonomy, freedom and privacy. These central values and principles will be taken in consideration before applying the data analysis.

## Limitations

The project will be completed for 17 weeks, which is followed by limitations. The knowledge and understanding of the business of the company will be limited to the few facts, provided by Rob Prop, Manager Professional Services of Informa.

Analysis will be limited: a greater depth of understanding and evaluation can only occur with utilisation of other resources such as comparisons with similar businesses, more data provided or the existence of more time.

The machine learning algorithm will not be improved a 100% by reason of time and knowledge limitations.

## Prospects

At this point the company has a steady amount of orders as its competitors see a 20% growth annually. To keep up with the competitors, the stakeholders should consider adopting data in every way but especially to the day-to-day business tasks and decisions.

Via the application of descriptive and diagnostic data analytics the company will have the opportunity to recognize the problems it is facing in a quantifiable manner or even identify new, previously unnoticed problems. Finding solutions for these problems will not be based on instinct and “gut feeling”, but rather on facts derived from the data. In addition, with the benefits of predictive and prescriptive data analytics, the company will also have the opportunity to aid its decision-making process and long-term strategy while utilizing their data, possibly in addition to external data as well.

Data science is the solution to remove all of the unknowns from the business investment, marketing expense, and future plans (including forecasting and prioritizing new marketing resource, expertise, and experiments). In addition, combined with machine learning and intelligent algorithms it can tell you the strengths and weaknesses of your website.

## Assumptions

To clarify the process and focus on the exploratory analysis, the following assumptions were made:

1. The data analytics will be done with use of Pandas, Python 3. Algorithms will not be explained in detail as assumed that only the final result matters.
2. Informa’s client operates in Belgium and The Netherlands.

# Conclusion

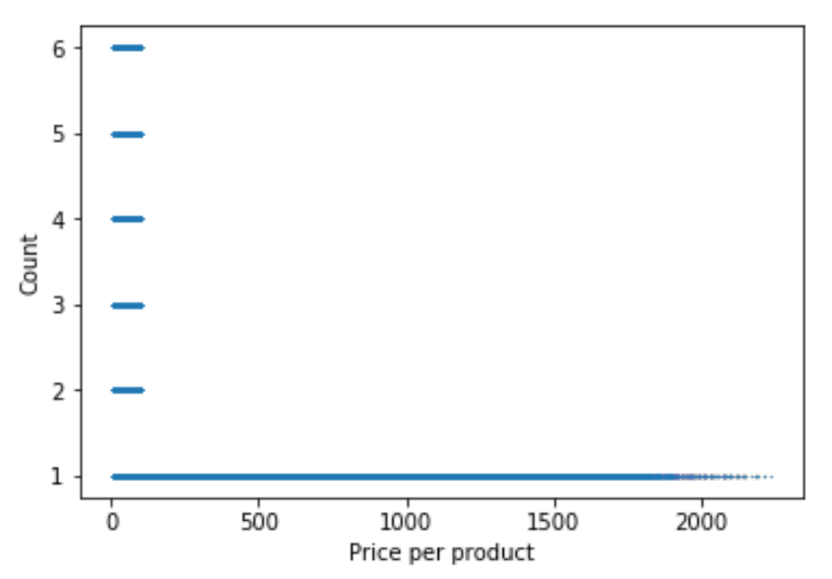
During this project, the company and its field of business will be closely researched to discover all the benefits and possibilities of undertaking the project. In addition, a detailed analysis of the provided dataset will be conducted, to better visualize the findings. In the end, a machine learning model will be developed. The areas of application of the algorithm will be:

* + Forecasting demand
  + Understanding customer retention/customer churn
  + Targeting recommendations

The benefits for adopting data analytics are numerous. Firstly, via the utilization of descriptive and diagnostic data analytics the company will have the opportunity to recognize the problems it is facing in a quantifiable manner or even identify new, previously unnoticed problems, which nevertheless hinder the company’s progress. The realization for the existence of those problems will not be based on instinct and “gut feeling”, but rather on facts derived from the data. Secondly, with the benefits of predictive and prescriptive data analytics, the company will also have the opportunity to aid its decision-making process and long-term strategy while utilizing their data, possibly in addition to external data as well.

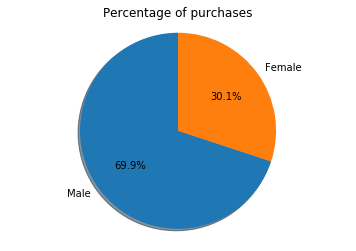
# Appendix A – Research Questions

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Question | | Answer |
| Data Driven Organization | | What is the average time between the date of purchase and the date of delivery? | The average time between different purchases is 3.45 days. |
|  | | What is the average time difference between the expected delivery and the actual delivery date? | The average time of delivery is 0.75 days longer than expected. |
|  | | Is there a correlation between the price of the product and the quantity of the product people order? | No there is no correlation between the price of the product and the quantity of the product people order. |
|  | | What is the gender split? | 69,9% Male 30,1% Female See figure 1. |
|  | | What is the age range of customers? | The youngest customer is 20 years old, and the oldest person is 87 years old. |
|  | | What is the distribution in age groups? | See figure 2 for reference. |
|  | | Which category has the most revenue? | The category with the most revenue is “Garden hand tools” |
|  | | Is there a correlation between the rating and the price? | We found no correlation between the rating and the price. See figure 4. |

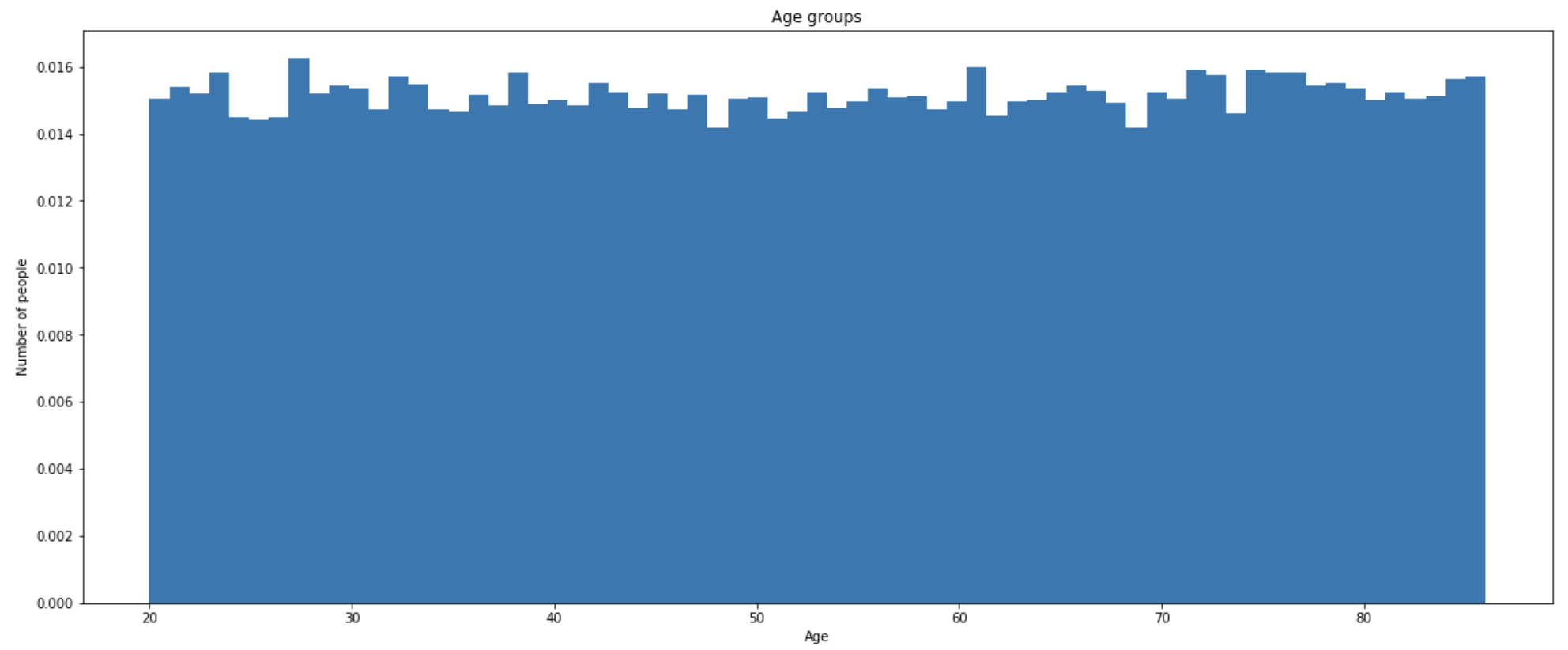


|  |  |
| --- | --- |
| Quantity | Average price |
| 1 | €549,13 |
| 2 | €56,21 |
| 3 | €56,13 |
| 4 | €56,18 |
| 5 | €56,25 |
| 6 | €56,24 |

**Figure 1**In the figure its visible that the average price of a product is way higher when 1 product is bought. If the quantity becomes higher the average price drops to about €56, -. So, there is no correlation between price and quantity.

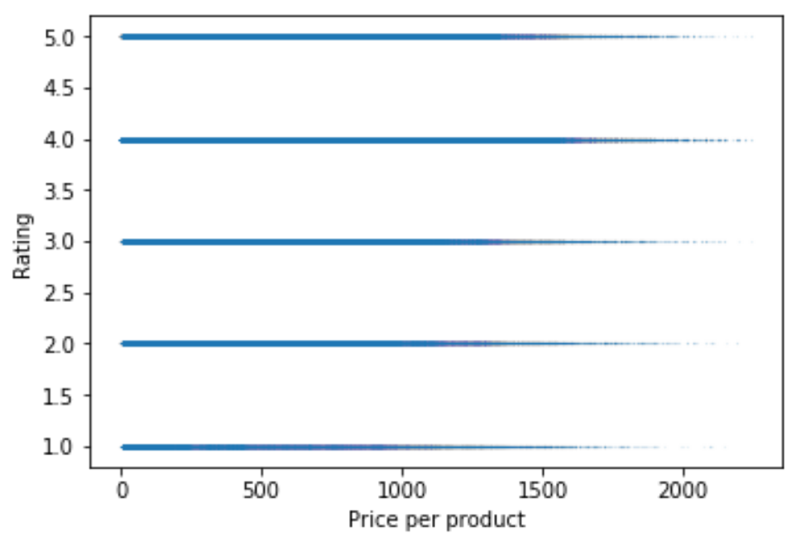


**Figure 2**In the chart above it is shown that 69.9% of the customers are men. 30.1% are women.

****

**Figure 3**In the chart above, the percentage of customers per age is shown.

|  |  |
| --- | --- |
| Rating | Average price |
| 1 | €439,83 |
| 2 | €444,48 |
| 3 | €442,75 |
| 4 | €441,74 |
| 5 | €441,54 |



**Figure 4**It is shown in the figure that the average price is about the same for every rating.

# Works Cited

Farris, P. (2009). *Marketing metrics: 50 metrics every executive should master.* Upper Saddle River (N.J.): Wharton School Publ.

Moth, D. (2013, August 28). *Almost a quarter of businesses don’t carry out any relationship marketing: report*. Retrieved from Econsultancy: https://econsultancy.com/almost-a-quarter-of-businesses-don-t-carry-out-any-relationship-marketing-report/

Press, G. (2016, March 23). *Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says*. Retrieved from Forbes: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/

Reichheld, F. (2018, September 14). *Prescription for Cutting Costs*. Retrieved from Bain & Company: https://www.bain.com/insights/prescription-for-cutting-costs-bain-brief/

Saleh, K. (n.d.). *Customer Acquisition Vs. Retention Costs – Statistics And Trends*. Retrieved from Invesp: https://www.invespcro.com/blog/customer-acquisition-retention/

Schneider, C. (2017, May 16). *Biggest data challenges might not even know*. Retrieved from IBM: https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/