Date: 27 October 2019

BUSINESS PROPOSAL

GARDEN & LIFESTYLE

DataTeam

Purple I

Members: Nadya Cheperkova
Mihai Alexandru Tudor
Max Post
Niek Sleddens
Rob Hendrikx

DataTeam

CONTENT

VERSION INDICATION

Version	Date	Editor	Change
0.1	16/10/2019	Niek	Initial layout Added content for Project Statement, Client, Team, Current Situation, Problem Description, Project deliverables and non-deliverables, constraints, risk
0.2	23/10/2019	Max	Changed Version history to correct format Edited Changed layout
0.3	23/10/2019	Nadya	Change completely the whole layout Change completely the content for Project Statement, Client, Team, Current Situation, Problem Description, Project deliverables and non-deliverables, constraints, risk
0.4	24/10/2019	Mihai & Niek	Review and make some changes on current situation and problem description
0.5	24/10/2019	Rob & Max	Analysis the data and looking for correlations, generates some graphs
0.6	24/10/2019	Max	Generates graphs
0.7	25/10/2019	Nadya & Mihai	Write Project Goal and KPI's and metrics
0.8	25/10/2019	Rob & Max	Generates more graphs focused on delivery time
0.9	25/10/2019	Niek	Generates graph for showing the reasons of returning the items
0.10	25/10/2019	Nadya & Mihai	Create Phasing and Gantt Chart
0.11	26/10/2019	Nadya	Reporting EDA
0.12	2/10/2019	Mihai	Write more KPI's and metrics
0.13	27/10/2019	Rob	Create a graph which shows which products are sold the most for the last year
0.14	27/10/2019	Everyone	Reviewed

PROJECT STATEMENT

The purpose of this document is to describe the process of exploring, cleaning and analyzing the provided data in order to produce predictive analysis that will create value for the company. Firstly, the formal client for whom this predictive analysis is being created, is described. Secondly, the people involved in the project are introduced and their roles and responsibilities regarding the final product. Moreover, the project is well described by clearly defining the current situation, the problem description and the goal. The deliverables that the team is going to provide at the end of the project are presented as well as the non-deliverables. Additionally, the project constraints and risks are determined and well explained, followed by the phasing which is illustrated with the needed explanation and Gantt Chart. Finally, at the end of the document, the KPI's and Metrics are introduced measuring success based on the specific business goals and targets, supported with the possible benefits, the findings and the data sources and methods.

FORMAL CLIENT

Informa is a leading international events, intelligence and scholarly research group. Their purpose is to champion the specialist. Through hundreds of powerful brands, they work with businesses and professionals in specialist markets, providing the connections, intelligence and opportunities that help customers grow, do business, make breakthroughs and take better informed decisions.

Informa's leading brands and expert teams operate events and exhibitions, deliver intelligence-based products and data-driven services, convene communities in person and digitally and provide access to cutting-edge research for specialist customer communities worldwide. Each of their five operating divisions (Taylor & Francis, Informa Intelligence, Informa Markets, Informa Connect and Informa Tech) has a distinct focus. Informa's Taylor & Francis business commissions, curates, produces and publishes scholarly research and reference-led content in specialist subject areas, advancing research and enabling knowledge to be discovered and shared. Informa Intelligence provides specialist digital information and data products, consultancy and research services to over 25,000 businesses worldwide, helping companies make better informed decisions, spot opportunities and gain competitive edge. Informa Markets creates platforms for specialist markets to connect, trade, innovate and grow. Informa Connect aims to provide expert content which you cannot Google and access to specialist networks and communities, through major branded annual events and specialist digital communities and content services. Informa Tech aims to inspire the global technology community to design, build and run a better digital world through providing research, media, training and events that inform, educate and connect businesses and professionals working in technology.

TEAM

The project team consists of five young and motivated specialists who are going to work together on this project in order to achieve its objectives. It consists of the project leader, project members, and other members who are not directly involved with management but are responsible for monitoring the work. This team consists of people from different technical and business areas with different knowledge but with the required skill set to carry out the work for the project. The structure and characteristics of a project team usually vary, but the project leader's role remains constant.

Project leader

The team leader is Nadya Cheperkova who is a young female specialized in Data Science.

Contact Information
Nadya Cheperkova
Email: n.cheperkova@student.fontys.nl

Team members

Contact Information
Max Post
Email: m.post@student.fontys.nl

Contact Information
Niek Sleddens
Email: n.sleddens@student.fontys.nl

Contact Information
Rob Hendrikx
Email: rob.hendrikx@student.fontys.nl

Contact Information
Mihai Tudor
Email: m.tudor@student.fontys.nl

CURRENT SITUATION

One of Informa's customers, a web shop focused on 'Garden & Lifestyle', has reached the point where the data volume is getting bigger, more complex and more versatile. They require a machine learning /deep learning solution which should be applicable to any similar data format. The web shop's IT department has comprised a dataset containing all the information that the web shop collects.

Informa asked our team to look into the given dataset and return with a solution supported by correlations and give insights. Based on those findings they can decide how their organization can benefit from Data Science and how to expand their partnership.

PROBLEM DESCRIPTION

The company has received a request from a customer with a running web shop, which focuses on products related to gardening and garden maintenance. The web shop owner requested help with improving the sales of the shop.
By looking at the data it can be noticed that the number of orders is quite steady over the last years. Moreover, a significant number of orders are being returned due to delivery delays. This cost the company a lot of money and customers.

PROBLEM JUSTIFICATION

We started our exploration of data by checking the revenue per year in order to be aware of how serious the problem is. From the graph below (Figure 1), it can be noticed that there is not much change in the revenue over the last five years, even the last year it went down compared to the previous. Moreover, the second graph (Figure 2) shows the revenue per month over the last five years, as expected the highest revenue is in the summer months. From these graphs our team concluded that there is need for a new approach in order to improve the customer satisfaction and sales. We dived deeper into the data and found an interesting insight. The third graph (Figure 3) illustrates the percentage of the people who returned their orders grouped by reasons. The majority (32%) of the customers did not give a concrete reason but if we exclude it, 3.9% remains a significant number. This is

the number of people who returned their orders because the delivery took too long. With the help of Machine Learning, our team will highlight how the delivery time can be improved so more satisfied customers which will lead to increase of sales.

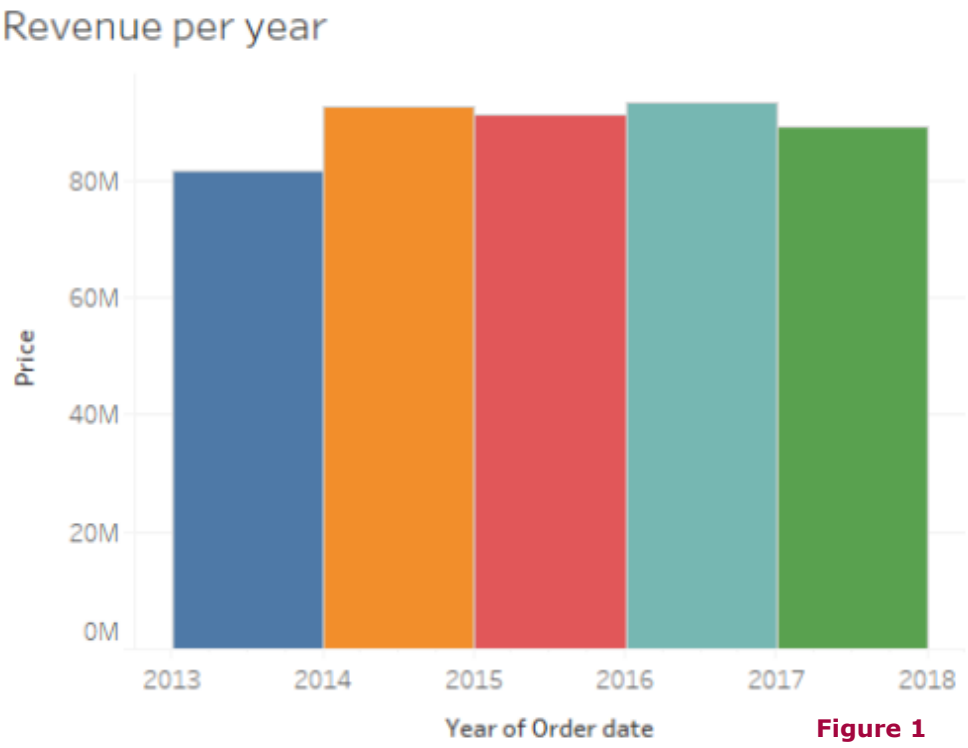


Figure 1

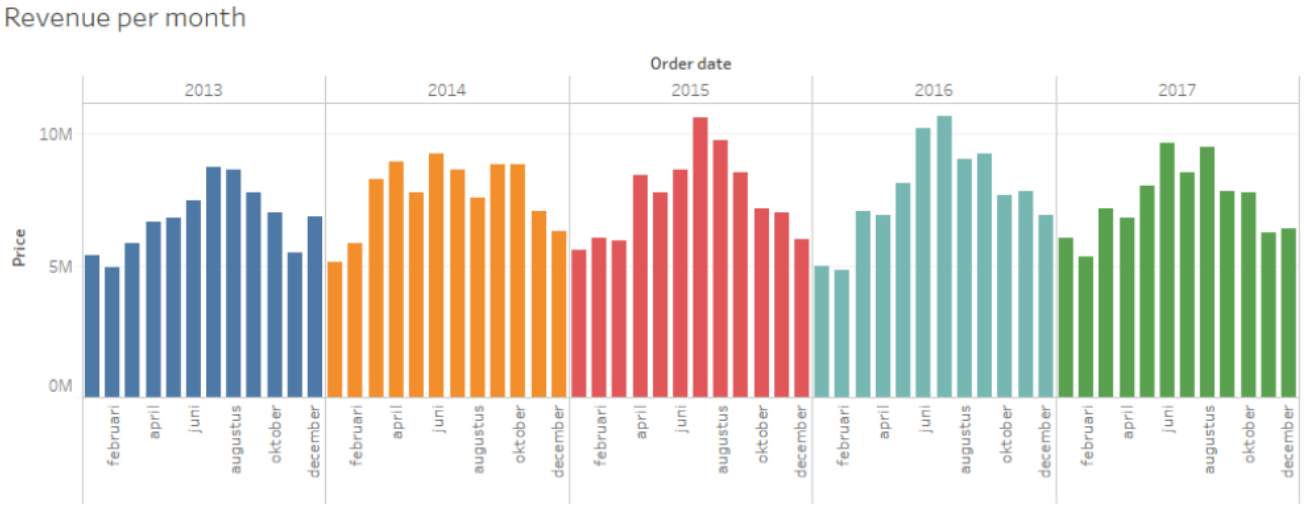


Figure 2

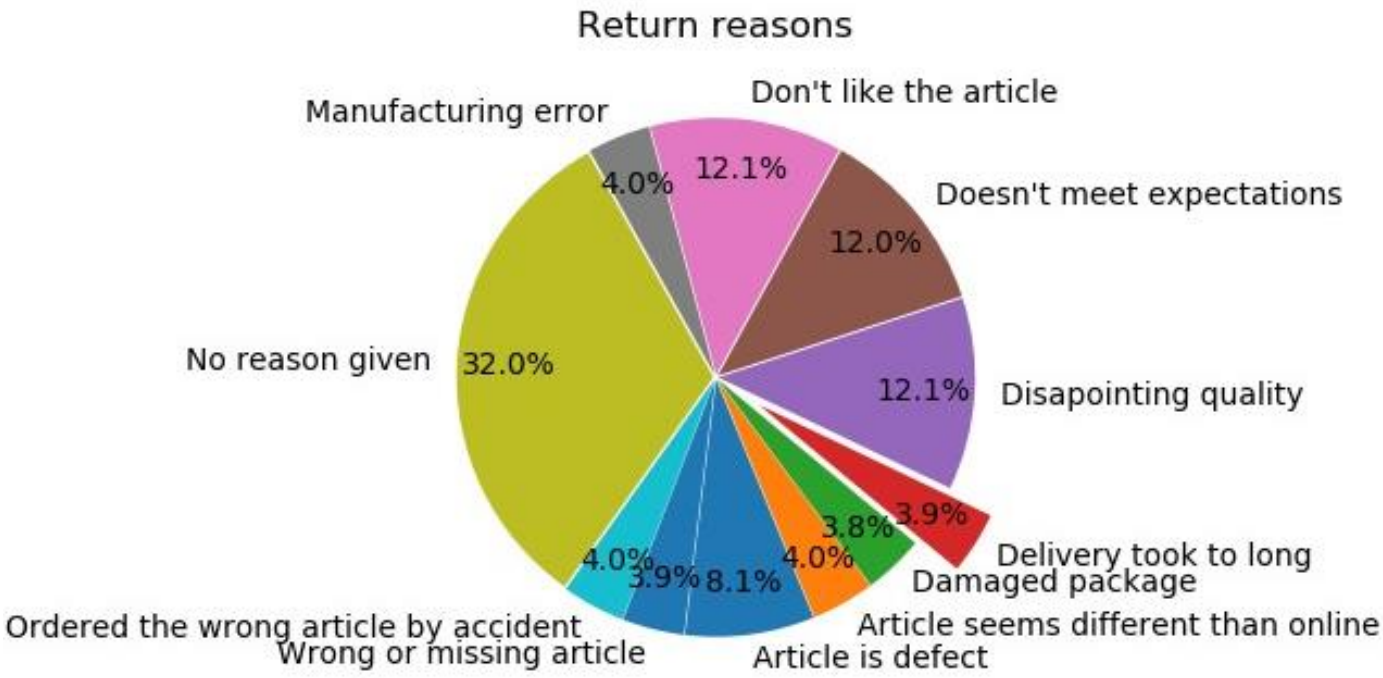


Figure 3

Moreover, we noticed that there are some products which have a really low number of sales (Figure 4) and we can combine the complaints about long delivery with the low sales of the selected products to improve the customer satisfaction. As a compensation for a longer than expected delivery time, the customers can be given a discount for these selected items (only on the ones that match the customer's profile based on his previous orders of the last year). Figure 5 shows that the typical customer of this web shop is represented by men and Figure 6 presents the total sold and total returned products over the last five years, this information can be useful later in modelling.

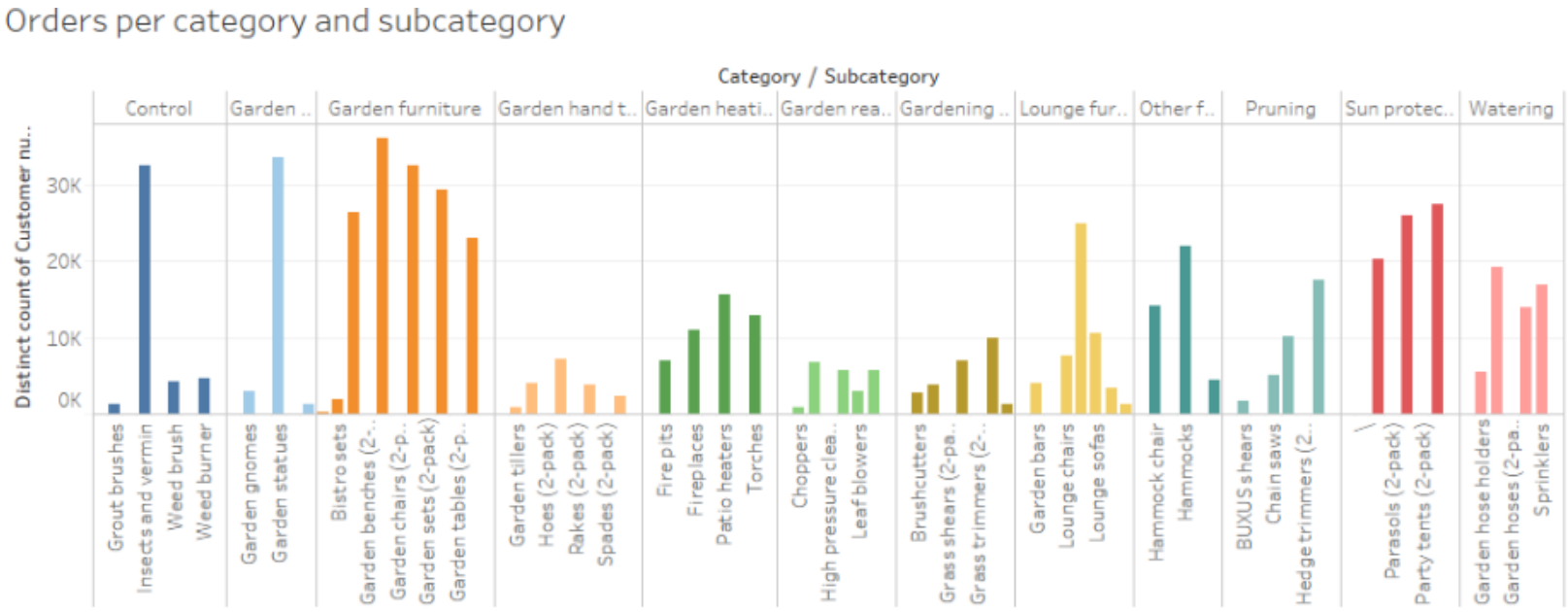


Figure 4

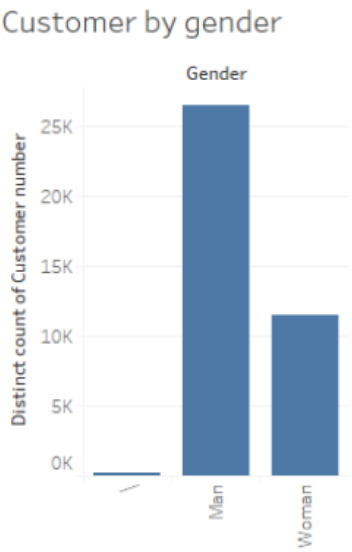


Figure 5

Figure 6

Furthermore, Graph 7 shows historically which products have been sold the most in the last year. This data can be used for improving the accuracy and scale of sales forecasts. By forecasting the demand for products, the stock will not run out and the delivery time will not increase.

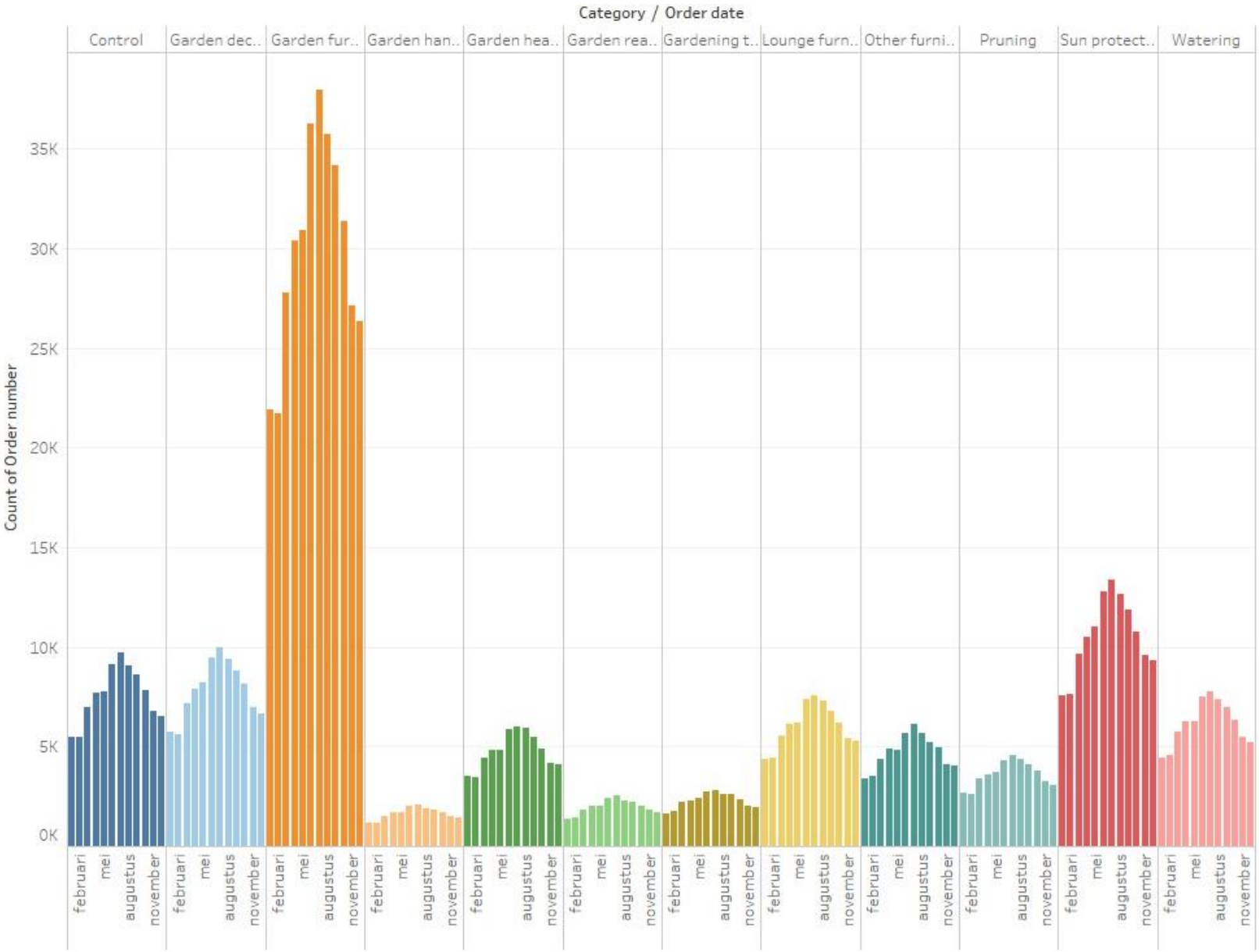

Category / Order date
Control
Garden dec..
Garden fur..
Garden han..
Garden hea..
Garden rea..
Gardening t..
Lounge furn..
Other furni..
Pruning
Sun protect..
Watering
Count of Order number
35K
30K
25K
20K
15K
10K
5K
0K
februari
mei
augustus
november

Figure 7

Finally, the figure below (Graph 8) illustrates the number of orders per region in the Netherlands. At this stage of the analysis it is not possible to draw any conclusions from it, but we can use it later for improving the delivery time when we know in which region people tend to buy more. The company can decide to invest in warehouses in these regions. In the next iteration, we will look deeper into it and create clusters and outliers.

Figure 8

PROJECT GOAL

The web shop's Sales Department, Marketing Department and Logistics Department can benefit from our final conclusions for the sake of improving the delivery time of the orders. Therefore, the goal of this project is to find different correlations and insights for improving the customer satisfaction, the sales and logistics with the main focus on the expected delivery time. The result of that can help the decision-makers of Informa to take further steps in creating the final solution for the improvement of their customer's web shop.

We noticed that one of the reasons of complaints is the actual delivery time being later than the expected delivery time by a large margin. Therefore, it is one of the significant reasons of why people are returning their orders. The following areas are the ones that can benefit from our findings:

- **Customer Service**
  The customer satisfaction can be improved by compensating for the orders that were not delivered on time. Machine Learning algorithms can be used to create a personal profile based on the customer's previous orders. Once that is created, the information can be used to increase the sales of low selling items by giving a discount on these items (selected items) for their next purchase.

- **Sales**
  With the help of Machine Learning algorithm, historical selling can be used for improving the accuracy and scale of sales forecasts. Being able to give a buying history can accurately predict the future buying levels. Forecasting demand for products can improve the availability in the stock and the delivery time can be reduced.

- **Logistic Department**
  The Logistics Department can improve its Management System with the use of Machine Learning. Based on Machine Learning algorithms, the customer of the web shop will be able to see the real-time changes of the estimated delivery time as well as how many products are available in the stock. This can be beneficial for the organization because the customer will be able to monitor, in real-time, the time of their delivery and the availability in the stock.

PROJECT DELIVERABLES AND NON-DELIVERABLES

Deliverables:

- ***Business Proposal*** - Describes the process of exploring, cleaning and analysing the provided data in order to produce predictive analysis that will create value for the company. The project is well described by clearly defining the current situation, the problem description, and the project goal. The project team and the management plan are identified as well as the work process.

- ***Data Analysis Report*** – Presents the results supported by Data Quality improvements.

- ***Ethical considerations (Ethics Report)*** - Determines if the final findings are within the principles of ethical and legal behaviour. The team looks at the current and upcoming laws (like GDPR) and checks if the project is compliant and at the ethical side, using the Five Ethics Principles.

Non-deliverables

- A solution in the form of application
- Database
- Multiple applications

PROJECT CONSTRAINTS

Constraint 1: Time

The project must be completed before the 12th of January 2020.

Constraint 2: Scope

The final product will be a Data Analysis report supported by Machine Learning results and Data Quality improvement suggestions, not an actual system/application for improvement.

Constraint 3: Resources

In terms of given resources, we can use only the provided dataset.

Constraint 4: Programming Language

The programming language that is going to be used is Python.

Constraint 5: Tools

The tools that are going to be used are Excel and Tableau.

Constraint 6: Language

English will be the main language.

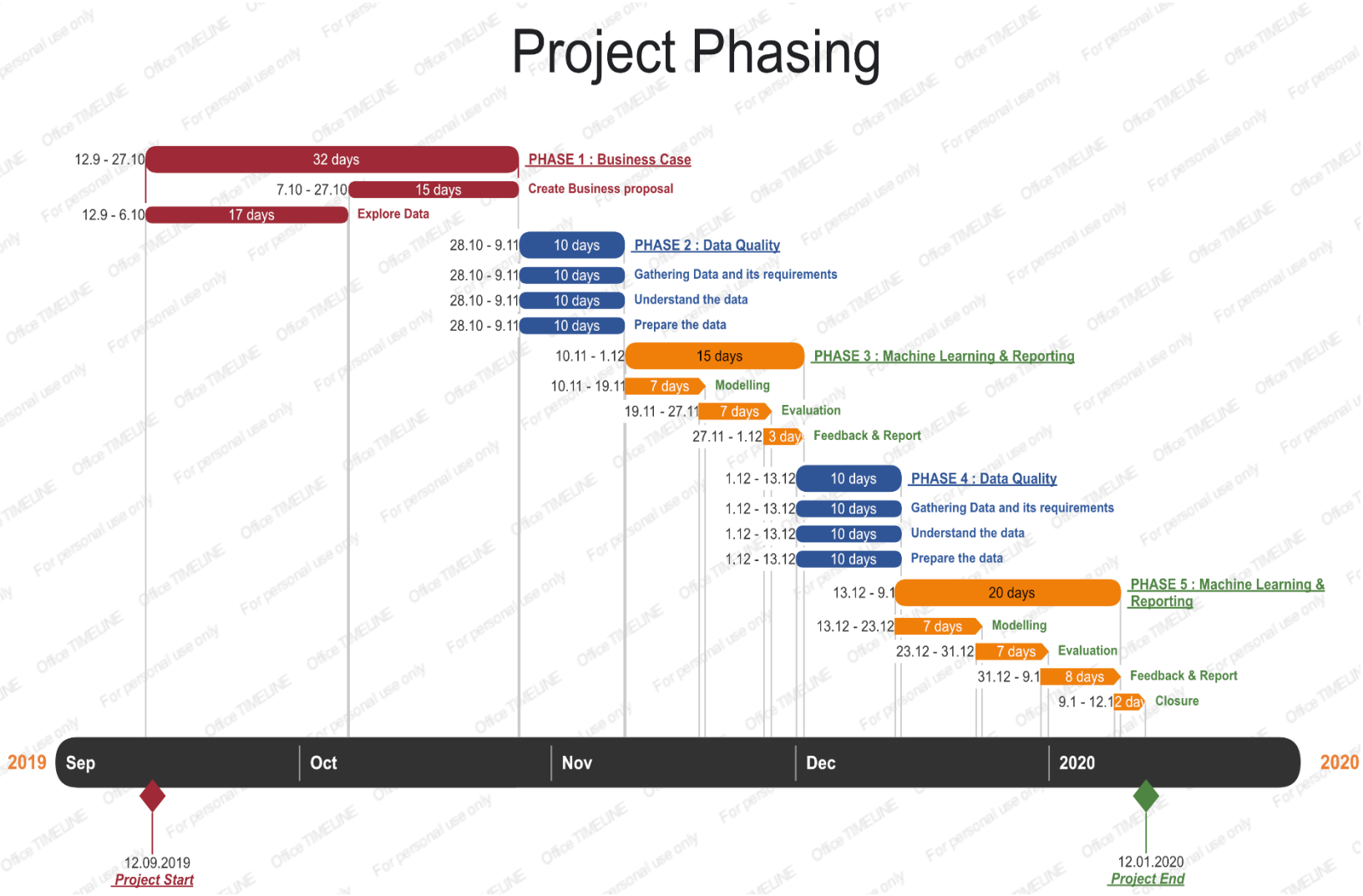
PROJECT RISKS

Risk	Probability	Impact	Mitigation
The client asks for changes in the project requirements	**Medium**	**Medium**	Ask for a regular client contribution in revising the functional requirements of every iteration and invite him to a retrospective meeting
The resources are limited or imprecise	**Low**	**High**	Ask the company for more datasets or more accurate and up to date data
A team member is unable to continue the work with the team.	**Low**	**Medium**	Every two weeks ask in advance for providing an update on the presence and in the case of accidental absence redo the work division to match with the current resources during the sprint
A team member does not deliver his task on time	**Medium**	**Medium**	Estimating the workload, try to divide the work in such a way that it is not dependent and make strict working rules
Insufficient data results	**Low**	**High**	Finish the project few weeks before the deadline in order to have time to check the results and improve them
The estimated time is inaccurate	**Low**	**Medium**	The good planning and conducting Backlog Refinement once in each sprint for having in advance insight for forthcoming phases

DataTeam

PROJECT PHASING

Gantt Chart



Phase 1: Business case

During this first phase, the project objectives are business understanding and analytic approach.

Activity: Explore data

Tasks for the activity are:

- Load the dataset and try to understand it
- Clean the data
- Look for insights and correlations
- Create graphs supporting the analysis

Estimated working time: 3 weeks

Activity: Create a business proposal

Tasks for the activity are:

- Divide the tasks among the team members
- Define the current situation, the problem description, and the desired result
- Identify the deliverables and non-deliverables
- Determine the risks and the project constraints
- Illustrate the phasing with the needed explanation
- Agree upon KIPI's and metrics

Estimated working hours: 3 weeks

Estimated duration is from 12-09-2019 until 27-10-2019

Deliverables for Milestone 1:

- Business proposal with exploratory data analysis

Phase 2: Data Quality

During this second phase, the project objectives are data requirements, data collection, data understanding and data preparation

Activity: Gather data and its requirements

Tasks for the activity are:

- Check the requirements for data
- Agree on specific requirements
- Research on publicly available data related to the project

Estimated working hours: 1 week

Activity: Understand the data

Tasks for the activity are:

- Try to understand the data
- Clean the data

Estimated working hours: 1 week

Activity: Prepare the data

Tasks for the activity are:

- Find correlations and insights
- Create graphs

Estimated working hours: 1 week

Estimated duration is from 28-10-2019 until 09-11-2019

Deliverables for Milestone 2:

- Data Quality

Phase 3: Machine Learning & Reporting

During this third phase, the project objectives are modelling, evaluation, deployment and feedback & report

Activity: Modelling

Tasks for the activity are:

- Create a trial version of the machine learning model

Estimated working hours: 1 week

Activity: Deployment

Tasks for the activity are:

- Commit the models
- Test the models
- Build the models

Estimated working hours: 1 week

Activity: Feedback & Report

Tasks for the activity are:

- Receive feedback
- Reflect on the feedback
- Report the process with the feedback

Estimated working hours: 1 week

Estimated duration is from 10-11-2019 until 01-12-2019

Deliverables for Milestone 3:

- Data Quality
- Machine Learning models
- Report

Phase 4: Improve Data Quality
Phase 5: Improve Machine Learning & Reporting

At this phase, Phase 4 and Phase 5 are repeated for the sake of improvement

Estimated duration is from 02-12-2019 until 09-01-2020

Deliverables for Milestone 4 & 5:

- Final Data Quality
- Final Machine Learning models
- Final Report

Phase 6: Closure

This is the final phase in which the final product is going to be created

Activity: Present the final product in the form of a presentation

Tasks for the activity are:

- Create a presentation

Estimated duration is from 10-01-2020 until 12-01-2020

Deliverables for Milestone 6:

- Final presentation

KPI'S AND METRICS

Customer satisfaction

Increase the average of all the ratings given by customers by 1 point (the scale being from 1 to 5) over the next 12 months.

Number of sales of low selling items

Increase sales of item sub-categories which had an order amount below 5000 over the last 12 months for the next 12 months.

Average estimated delivery time per item

Reduce the average estimated delivery time per item of the last 12 months by 20% for the next 12 months for items that have an average actual delivery time greater than the average of the estimated delivery time.

Stock

Increase the availability of products in stock by 20% over the course of the next 12 months.