**ASSIGNMENT 2 FRONT SHEET**

|  |  |  |  |
| --- | --- | --- | --- |
| **Qualification** | **BTEC Level 5 HND Diploma in Computing** | | |
| **Unit number and title** | Unit 14: Business Intelligence | | |
| **Submission date** | 03/03/2020 | **Date Received 1st submission** |  |
| **Re-submission Date** |  | **Date Received 2nd submission** |  |
| **Student Name** | Vo Trong Nghia | **Student ID** | GCD18468 |
| **Class** | GCD0606 | **Assessor name** | Srikanth Raju Kandukuri |
| **Student declaration**  I certify that the assignment submission is entirely my own work and I fully understand the consequences of plagiarism. I understand that making a false declaration is a form of malpractice. | | | |
|  |  | **Student’s signature** | Nghia |

**Grading grid**

|  |  |  |  |
| --- | --- | --- | --- |
| P3 | P4 | M3 | D3 |
|  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **❒ Summative Feedback: ❒ Resubmission Feedback:** | | |
| **Grade:** | **Assessor Signature:** | **Date:** |
| **IV Signature:** | | |

Table of Contents

[1. Business intelligence 6](#_Toc34129458)

[1. BI concept 6](#_Toc34129459)

[2. BI techniques 7](#_Toc34129460)

[3. BI tools 11](#_Toc34129461)

[2. Tableau report 15](#_Toc34129462)

[3. Report review and evaluation 22](#_Toc34129463)

[References 23](#_Toc34129464)

Table of Figures

[Figure 1 - Data warehouse 7](#_Toc34129465)

[Figure 2 – Analytics 8](#_Toc34129466)

[Figure 3 - Data mining 9](#_Toc34129467)

[Figure 4 - Data visualization 10](#_Toc34129468)

[Figure 5 – OLAP 11](#_Toc34129469)

[Figure 6 – Tableau 12](#_Toc34129470)

[Figure 7 – Sisense 12](#_Toc34129471)

[Figure 8 - Weka 13](#_Toc34129472)

[Figure 9 - Jupyter notebook 14](#_Toc34129473)

[Figure 10 - R studio 14](#_Toc34129474)

[Figure 11 - List of orders 15](#_Toc34129475)

[Figure 12 - Order Breakdown 16](#_Toc34129476)

[Figure 13 - Sales targets 16](#_Toc34129477)

[Figure 14 - Monthly sales 17](#_Toc34129478)

[Figure 15 - Monthly sales and targets 17](#_Toc34129479)

[Figure 16 - Sales difference for furniture 18](#_Toc34129480)

[Figure 17 - Sales difference for office supplies 19](#_Toc34129481)

[Figure 18 - Sales difference for technology 19](#_Toc34129482)

[Figure 19 - Map 20](#_Toc34129483)

[Figure 20 - Profit and sales 21](#_Toc34129484)

[Figure 21 - Dashboard 22](#_Toc34129485)

Table of Tables

**No table of figures entries found.**

# 1. Business intelligence

## 1. BI concept

Business intelligence is a process of applying specialized techniques and principles of finding data pattern, meaningful information and data mining in order to support organizational purposes and decision making in business. The word business intelligence emerged in the 1940s-1950s when people needed to have a way of analysis in making decisions effectively. Over the years, the technology has developed in many areas and aspect due to the growing in global data and information as well as the capacity of data storage in lots of database system and computers. Since the evolution of the Internet, data has been a valuable resource for business and that is how many companies and organizations gain their understanding of markets and customers. Business intelligence greatly enhance organizational performance, business statistics and productivity. Lots of companies apply business intelligence to have a closer look to a five-time increase [customer experience analysis](https://callminer.com/blog/what-is-customer-experience-analysis/) and the boost of decision making process. The completion of the process will delivery services to customers and we can then decide what to do next in the future based on how the market and feedbacks is going to be. It also helps organizations achieve significant goals. The traditional process requires input in order to produce output. To preserve competitive, any managers and organizations must get how to analyze data pattern and utilize the information that is given, and we know that business intelligence can come in handy in this kinds of situation. With the application of business intelligence, we can gain some insights from a bunch of available data to deliver reliable, actionable, and almost real-time inputs for decision making. Nowadays, big data and data science are usually mentioned in lots of newspapers and blogs that we can think of how the benefits that it can give us in many ways.

The data that is collected into one consistent data storage can be used for the process of data analysis which is the techniques of applying ETL that is extracting, transforming and loading to a centralized database system. However, data can be in a form that we sometimes cannot store all of them which means that there are 3 main types of data we need to deal with including structured data, semi-structured data and unstructured data. All of these types need to convert into a type for storing easily and readable. In addition, business intelligence can consist of other types of process which are management, operation and supporting helping to improve business model and performance.

## 2. BI techniques

* **Data warehouse:**

Data warehouse is one of the most important aspects in business intelligence which stores all sorts of data into one consistent data structure that can then retrieve and query from a specific database. Then we can do some data analysis and study on data which is the next step after finishing querying and aggregating. Since the expanding of the Internet and more information is exchanging across the world every day, we have to find a way to store all those information, extract them and applying data mining techniques. The more data we have, the more valuable information we can get for our business decision making. By storing lots of information that we can use to predict for future data and decisions, visualizing data in order to discover pattern, making business report and doing some research.

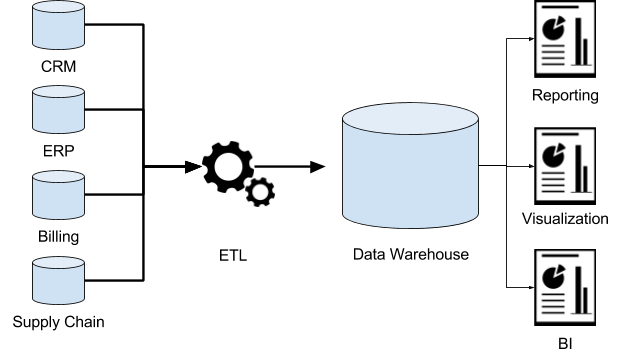


Figure 1 - Data warehouse

(dremio.com, n.d.)

* **Analytics:**

Analytics is one of the most useful techniques which includes the study of available data in order to gain meaningful insights and how the change of data. This is a popular analysis technique since it helps analysts gain deeply understanding of available data and combine as well as utilize ultimate value with data-driven skills and business decisions. For example, we can apply analytics in a technology company that is selling computers and digital cameras which helps to understand the market segmentation and customers’ needs in order to produce appropriate products that can increase company revenue. In addition, Statistical analysis and other general techniques is applied for the process of understanding data and pattern which helps to improve productive models. It uses the mathematic foundations to estimate the correlation, relationships and reliability of the observations in terms of features and target. Some of the most useful concepts and techniques are ROC - AUC, features scaling, data distribution analysis, principal component analysis, confidence interval, P value, data leakage, etc.

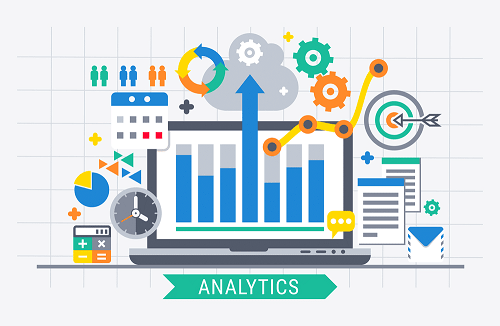


Figure 2 – Analytics

(timhieuvewordpresswebsite.webflow.io, n.d.)

* **Data mining**

Data mining is a process of finding data pattern that can help to make a more effective prediction model. It can includes data extraction from raw data and external resources, data visualization, features engineering, data encoding, features selection and machine learning algorithms. By applying some key techniques for data analysis we can then move to the next step of data mining is to try to build predictive model that can give out some useful information and prediction in the future. Depending on the types of problems we are trying to solve that we can choose to classify any new data or using regression algorithms to predict figures. Data mining enhance the process of decision making because its reliability and possibility when it comes to predict data which can take times in order to discover.

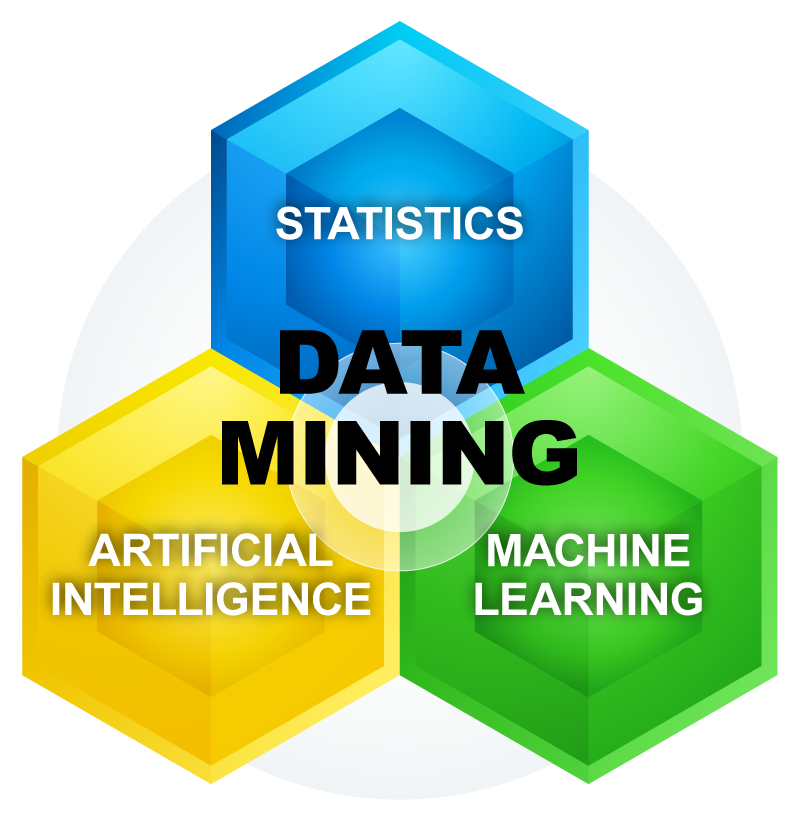


Figure 3 - Data mining

(mfahri.web.id, n.d.)

* **Data visualization**

Data visualization is part of statistics analysis which show how data looks in different ways. We can use some types of charts and tables in order to gain useful insights from given data. It can be bar chart, distribution chart, scatter plot, line plot, bubble chart, confusion matrix, pairwise matrix, map, heat map, pie chart, etc. We can see its benefits in any aspects that involves the use of data. For example, the normal distribution chart tell how the data is distributed across the x-axis which can describe as the probability of each case. Thus, we can predict that if a certain type of market segment can gain any benefits from customers and how likely someone will continue using the products in the future.



Figure 4 - Data visualization

(nordiccoder.com, n.d.)

* **Models and prediction**

It’s a kind of technique that is used to predict data which means can tell us the probability that something can happened or a certain value that we can gain from new data. Building predictive models involves various statistics techniques and machine learning algorithms. It can be a classification problem which predict whether it belongs to a specific label or if we want to predict a continuing value such as house pricing or stock market.

* **Online analytical processing**

Online analytical processing which is abbreviated as OLAP, is a way of applying data visualization in different views when retrieving from database. It can be described as a model of a cube that involves multi-dimensional structures that can help data analysts and scientists view from separate angles and directions. It is also best known for the OLAP-cubes due to the way it shows in multidimensional data. The cubes illustrate dimensions on the cube edges such as sales, company revenue, customer name, product type, category, date, transaction, etc. The values in the cube describe figures and measurements like number of customers, amount of money, target revenue for each month, etc. we can control the way it looks by performing some of the operations like slicing, roll-up, drill-down, pivot and dicing.



Figure 5 – OLAP

(educba.com, n.d.)

## 3. BI tools

* **Tableau**

It’s a tool that need to be included in business intelligence packs that will help both technical and non-technical analysts to easily use for data visualization and analysis. Tableau can be one of the essential application that everyone should learn to use at least once. Also, it supports many database systems and some of popular file types such as excel or comma-separated values.

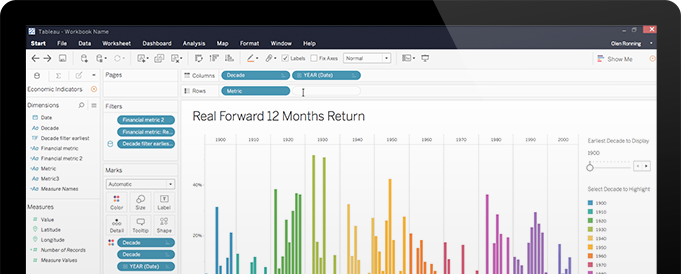


Figure 6 – Tableau

(tableau.com, n.d.)

* **Sisense**

Sisense provides appropriate tools as well as user-friendly design for non-technical users to use. It can help to gather data from various sources that could be from Internet or uploading data to the cloud. It allows users to do data processing, visualization, analysis and sharing data among colleagues since it run on a much faster platform and technology.

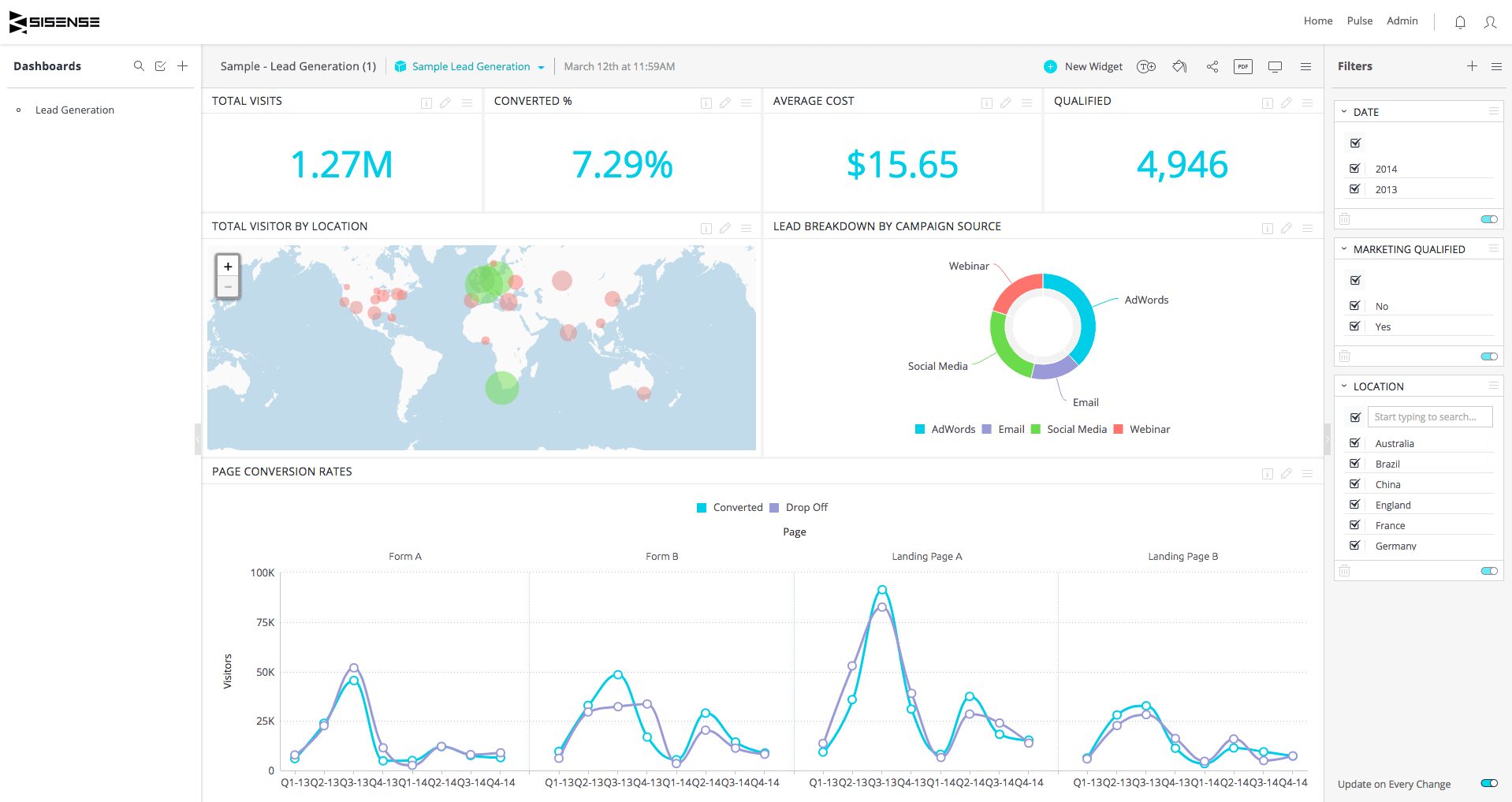


Figure 7 – Sisense

(invoiceberry.com, n.d.)

* **Weka**

According to cs.waikato.ac.nz Weka is tried and tested open source machine learning software that can be accessed through a graphical user interface, standard terminal applications, or a Java API. It is widely used for teaching, research, and industrial applications, contains a plethora of built-in tools for standard machine learning tasks, and additionally gives transparent access to well-known toolboxes such as [scikit-learn](https://markahall.blogspot.co.nz/2015/06/cpython-integration-in-weka.html" \t "_blank), [R](https://markahall.blogspot.com/2012/07/r-integration-in-weka.html), and [Deeplearning4j](https://deeplearning.cms.waikato.ac.nz/) (cs.waikato.ac.nz, n.d.). Therefore, it’s also a tool for researchers who don’t have much technical knowledge to use other complicated tools or software that can cause confusion.

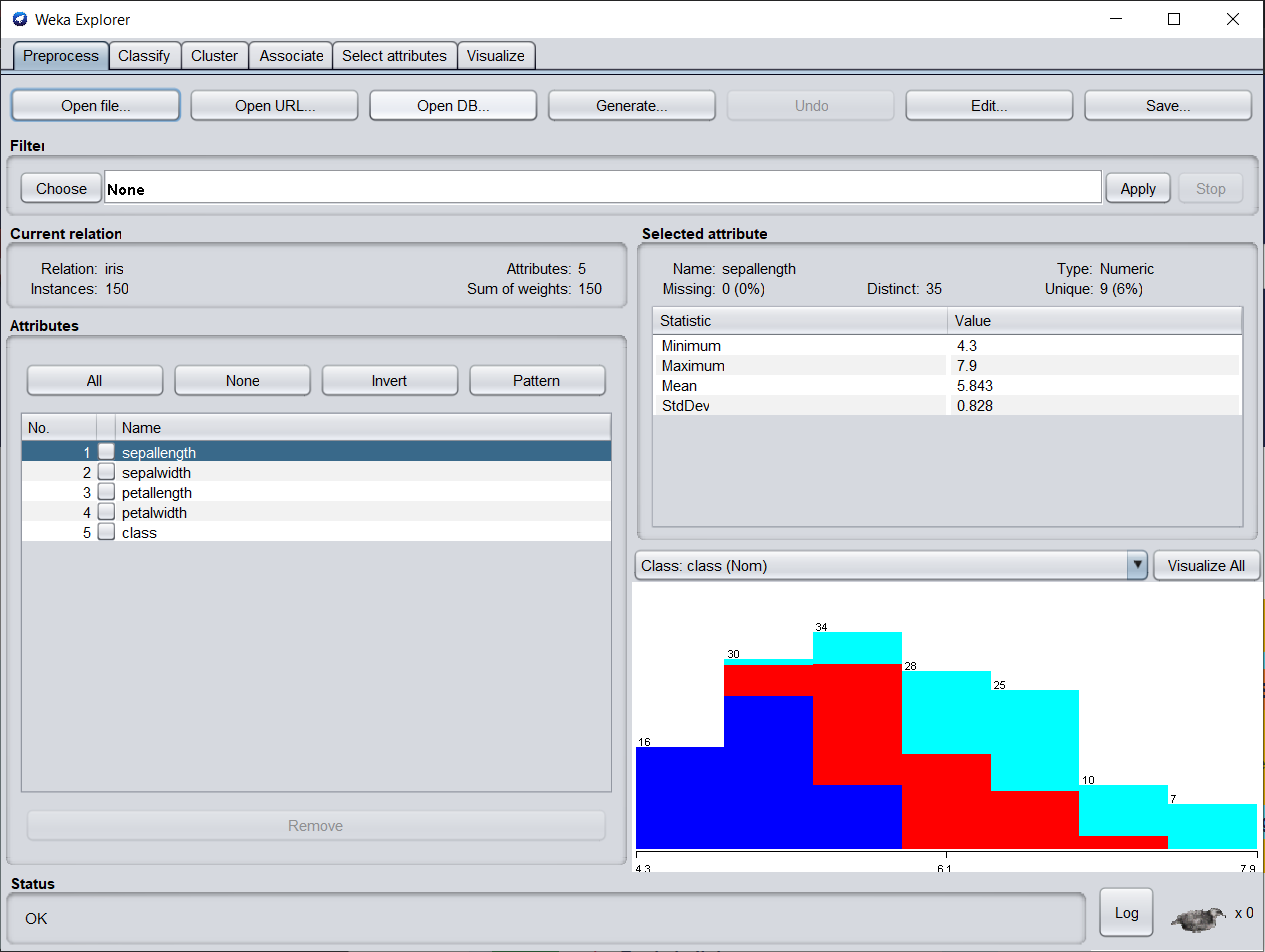


Figure 8 - Weka

* **Jupyter Notebook**

Jupyter notebook is one of the best tools for data science that helps to perform data analysis, visualization, data cleaning and building machine learning models. It utilizes Python as one of the most popular programming languages recently which has user-friendly syntax and easy to learn for beginners. However, in order to use this one, it’s important to have an understanding of programming and coding which aims for technical users.

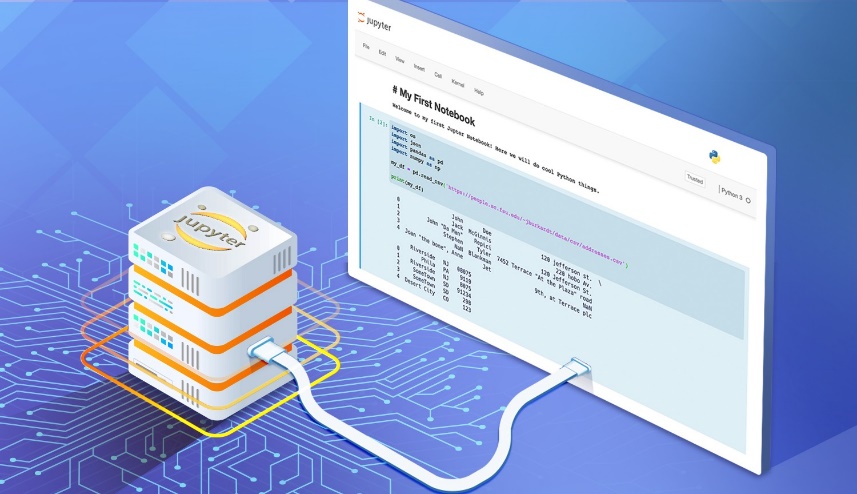


Figure 9 - Jupyter notebook

(hackingandslacking.com, n.d.)

* **R studio**

Just similar to Jupyter Notebook, R studio is usually used by many data scientists which utilizes R language for writing code, however, it’s less coding than Python which is for people less familiar with programming.

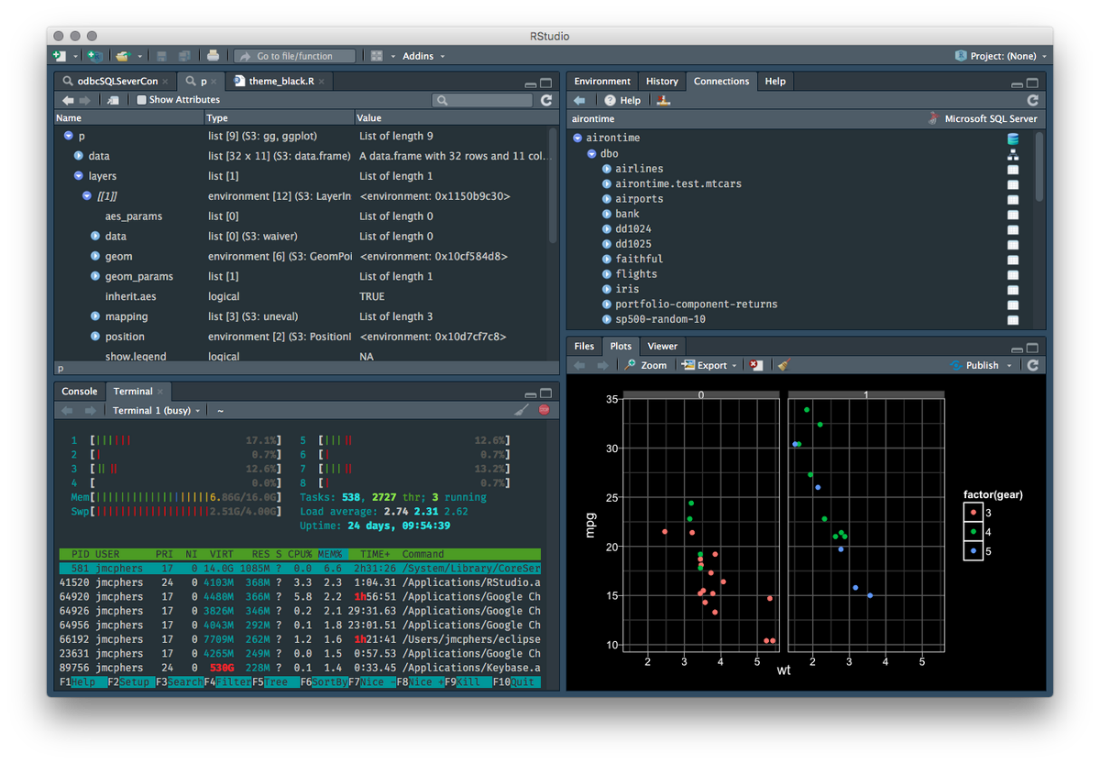


Figure 10 - R studio

# 2. Tableau report

In this section, we will look at the data of a company that sells 3 types of products which are furniture, office supplies and technology. This file contains 3 sheets:

* List of orders that is all the orders from customers that the company has

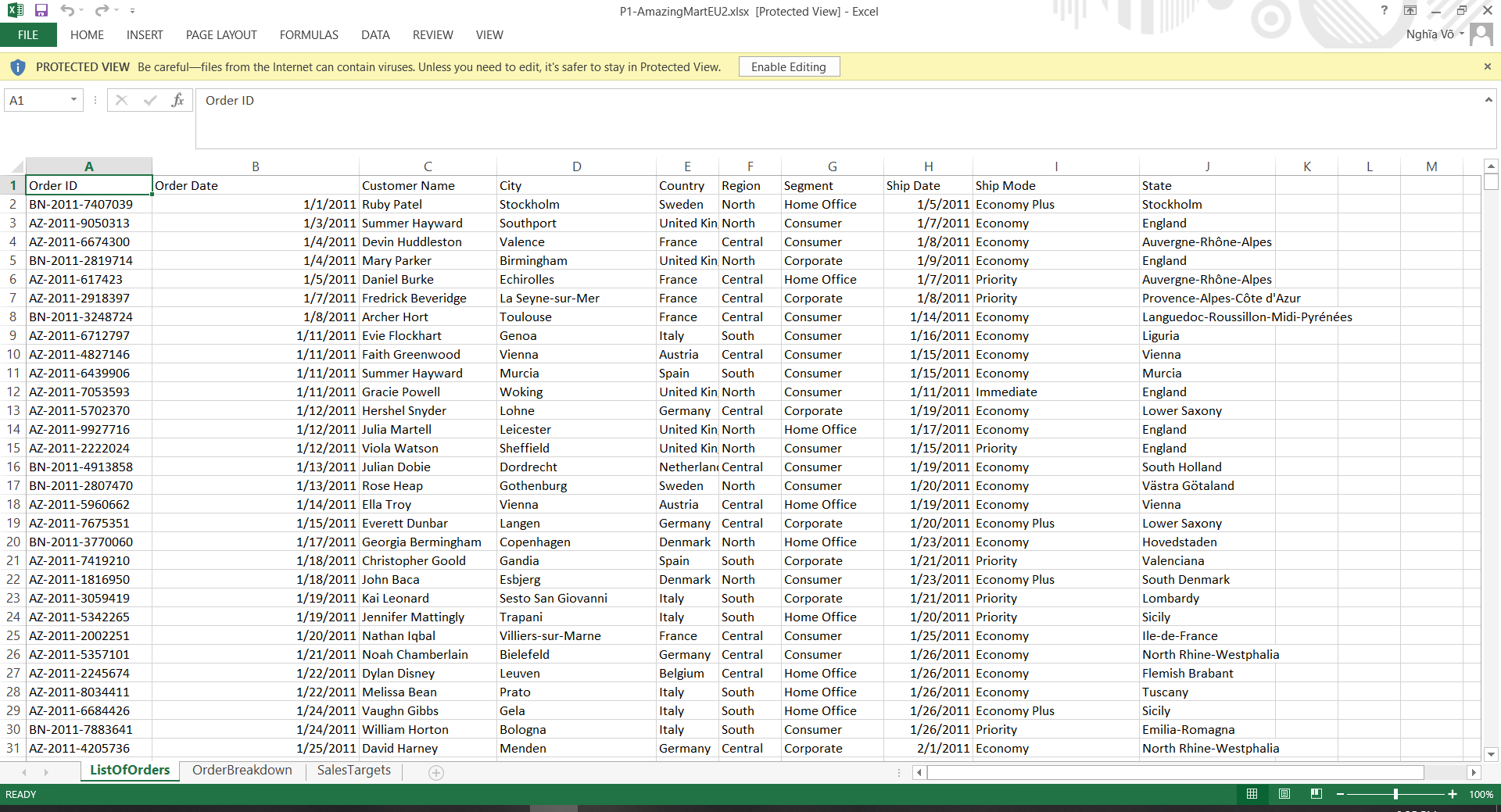


Figure 11 - List of orders

* Order break down which is the details of each order

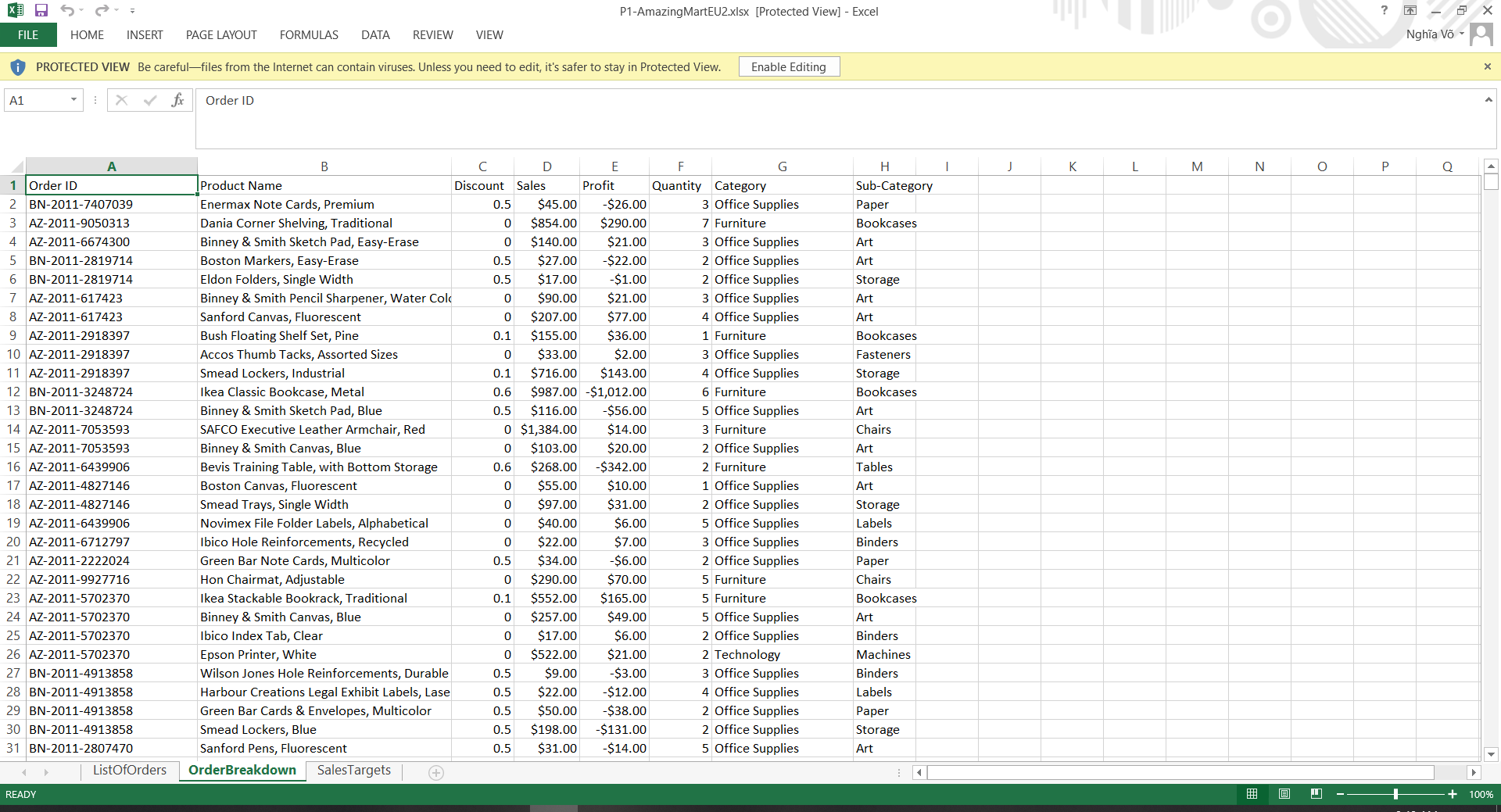


Figure 12 - Order Breakdown

* Sales targets which are the goals for sales at the end of each month.

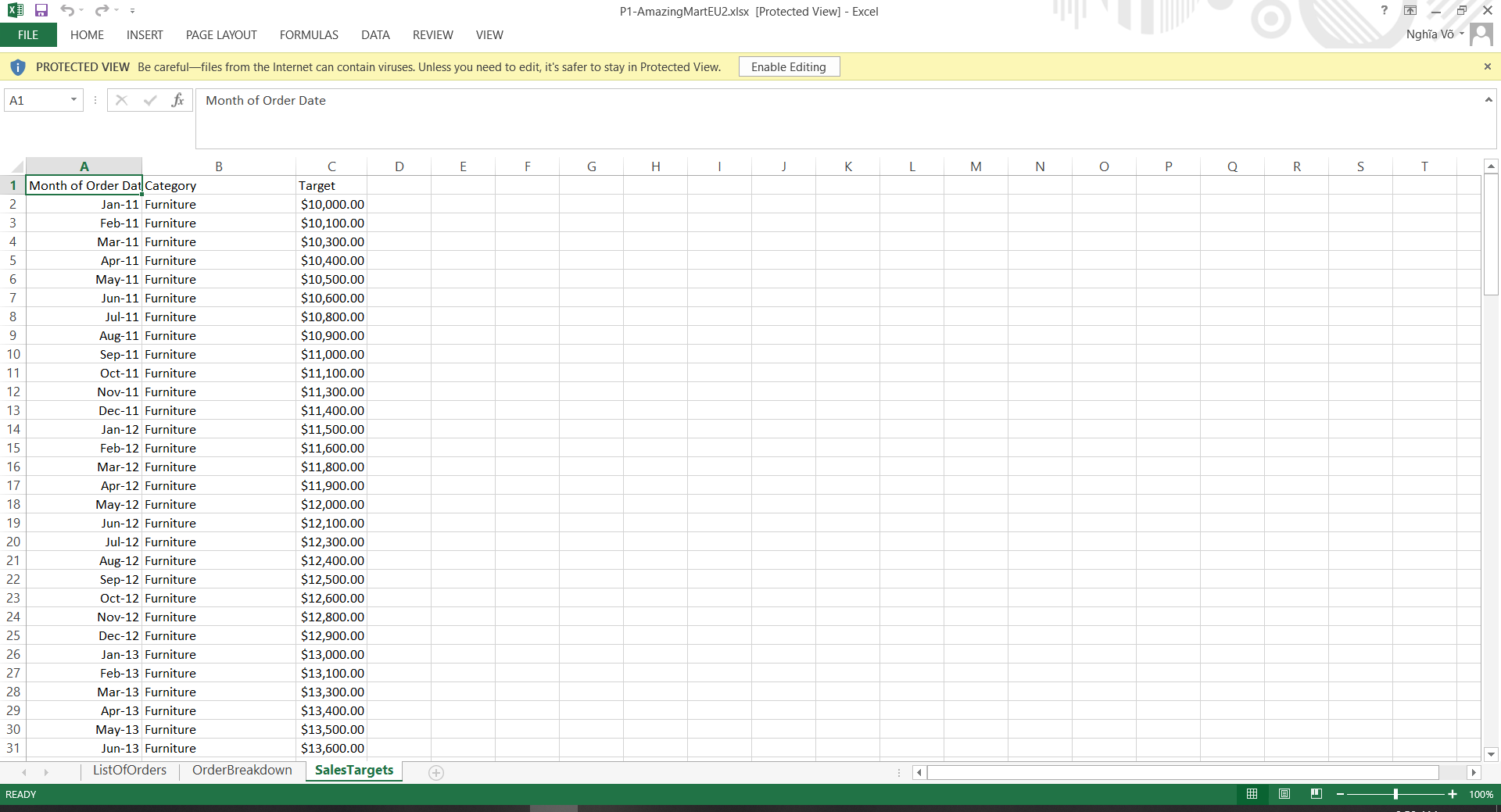


Figure 13 - Sales targets

In order to know how the company is doing its job of selling products, we can look at the sales for each month and then compare all 3 types of products. Bar charts for these illustrates as below:

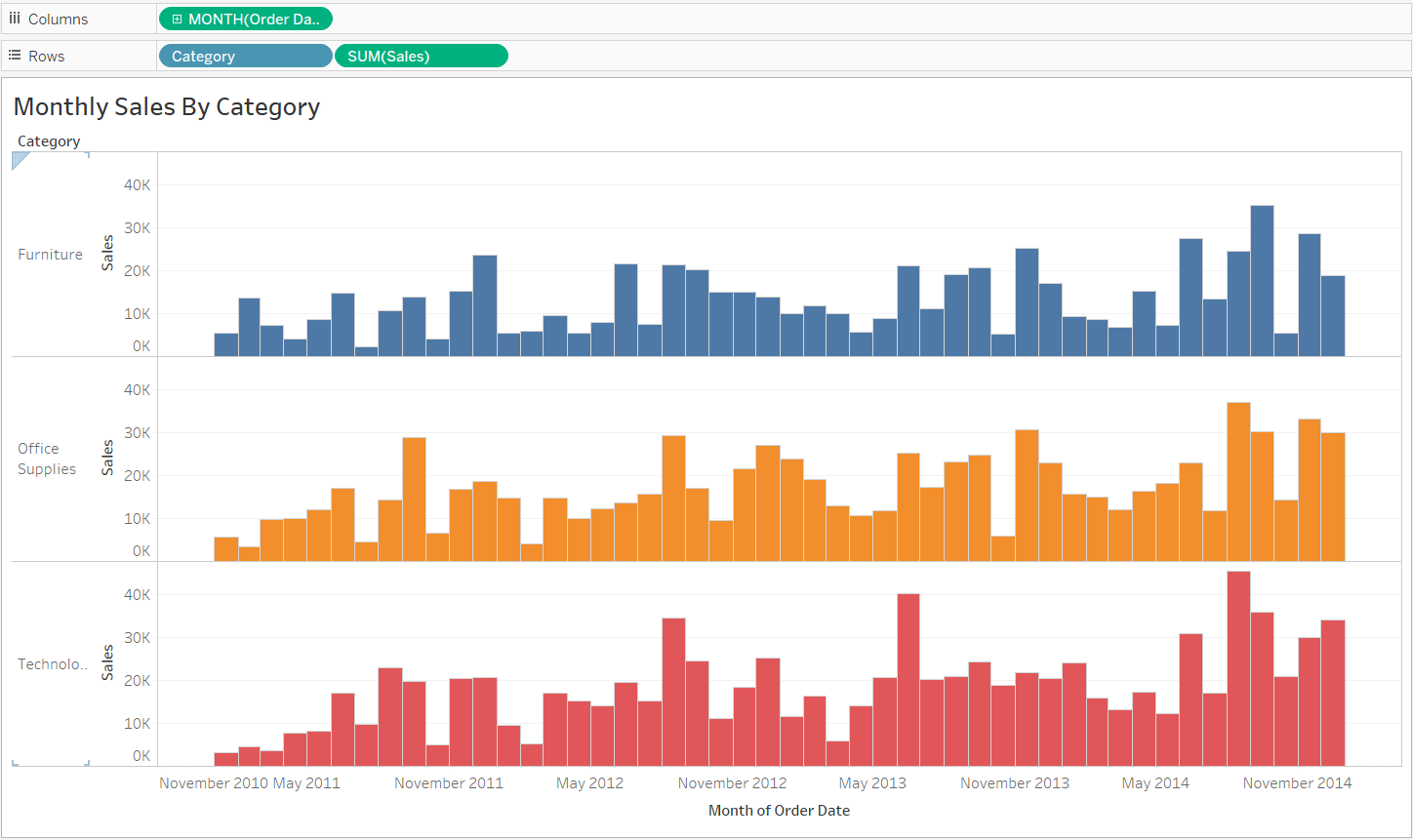


Figure 14 - Monthly sales

However, what we want to know is if the sales of a month meets or greater than the target sales that the company has been setting. The grey area chart behind is the target sales.

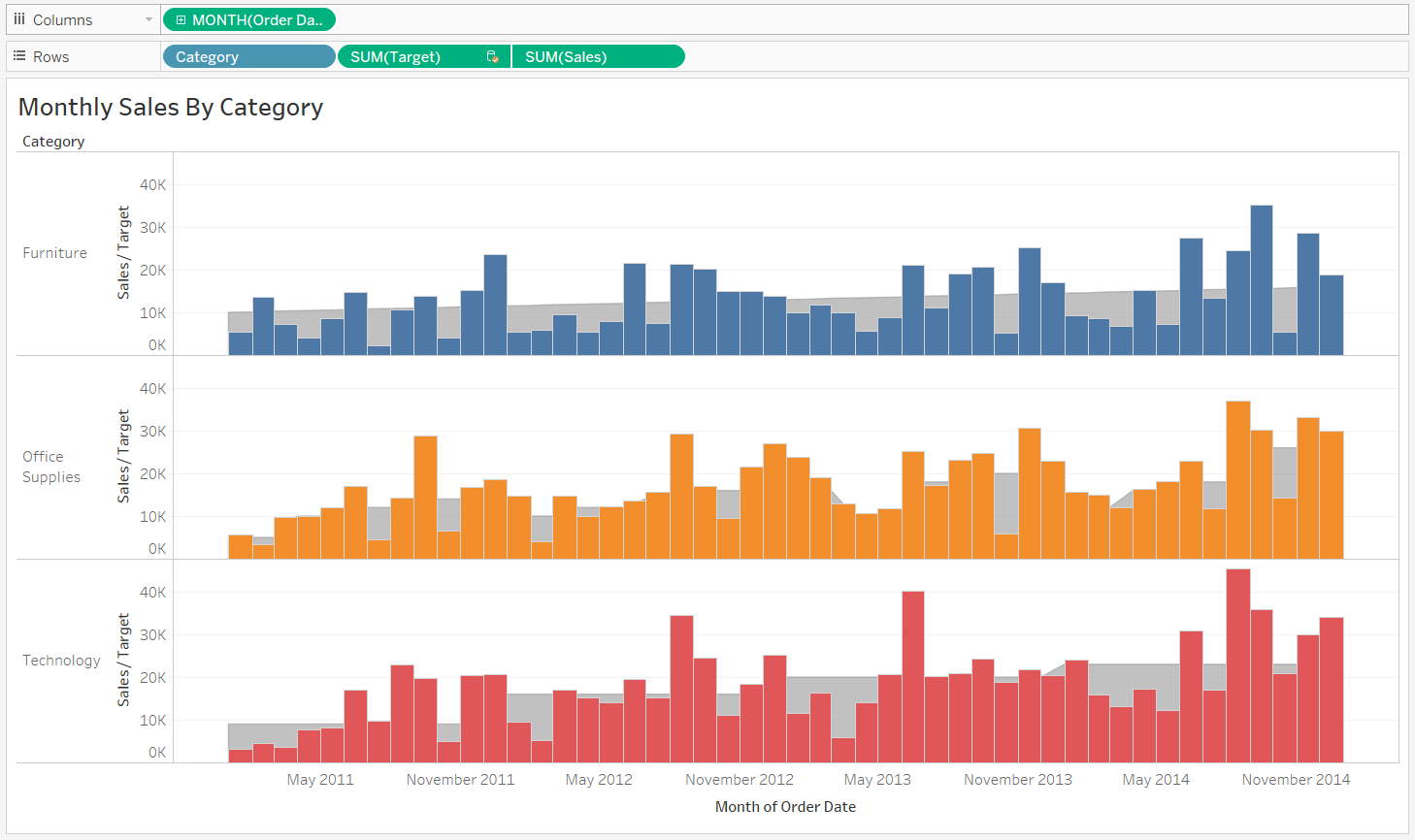


Figure 15 - Monthly sales and targets

From these charts, we want to explore which month the company didn’t meet the required targets for sales of a specific product category because it’s important to know if the company is doing great job in any areas or not.

This is the sales over the past few years and target sales for furniture:

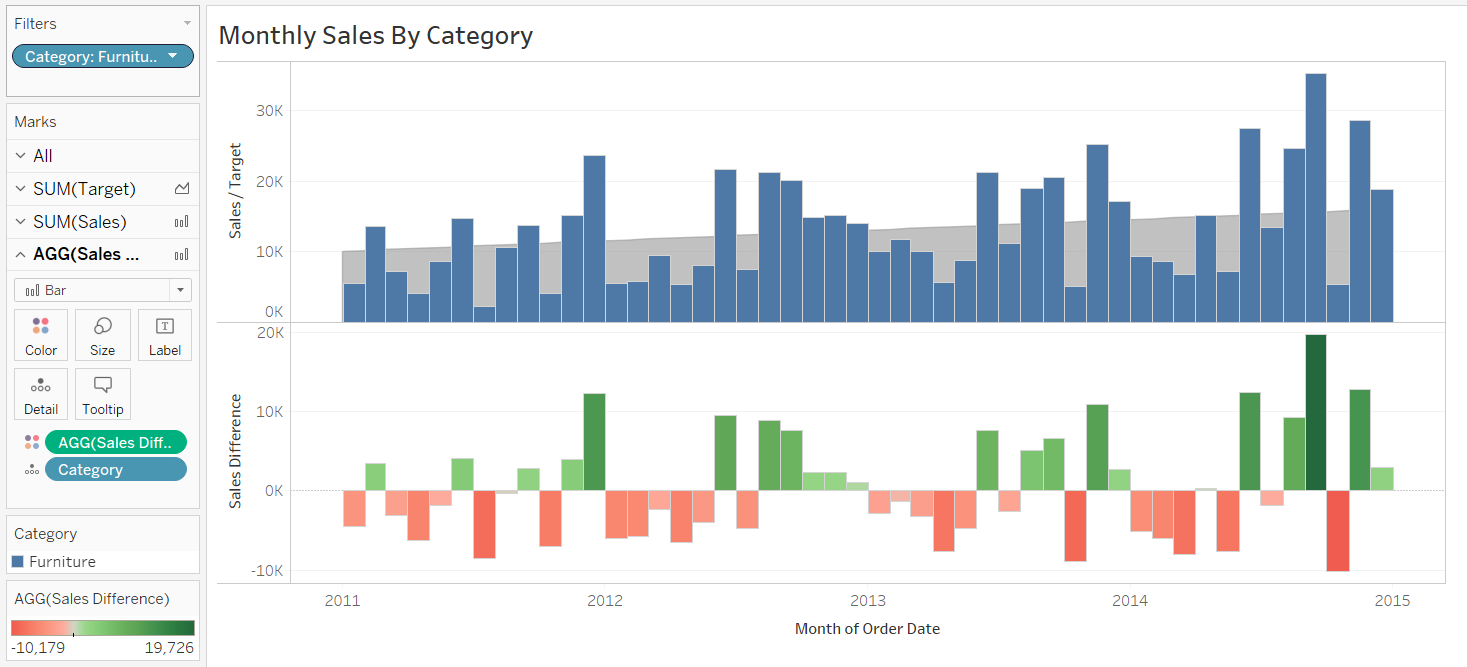


Figure 16 - Sales difference for furniture

We can see that most of the months shown in the bar charts above are below the sales target which tells us that there’s a problem in products belonging to furniture. From this, we can try to come up with some solutions that can help to increase the number of furniture products sold to customers.

This is the sales over the past few years and target sales for office supplies

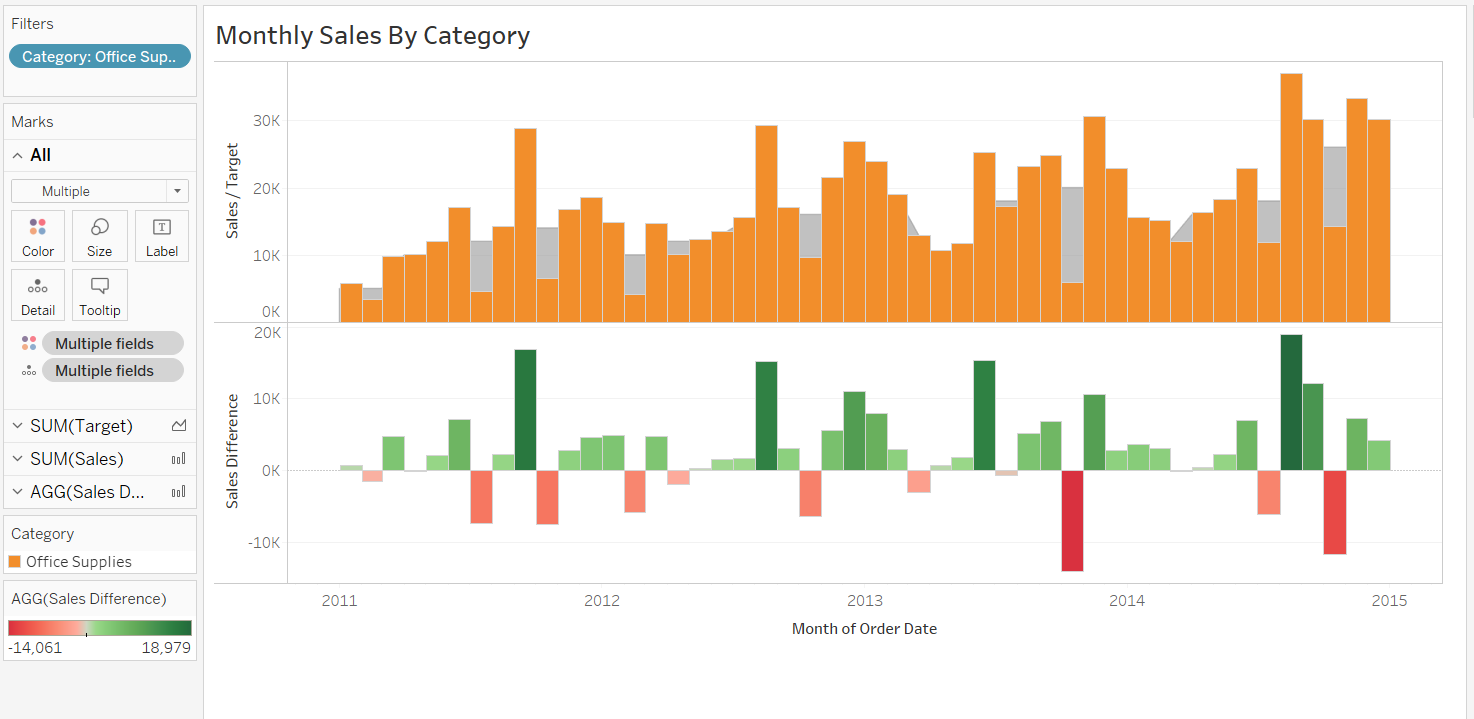


Figure 17 - Sales difference for office supplies

From the charts above we can see that the sales team is doing a great job in selling office supplies which means that it’s the main source of the company’s profit. This can help us understand what we are good at and make appropriate decision in the future.

This is the sales over the past few years and target sales for technology

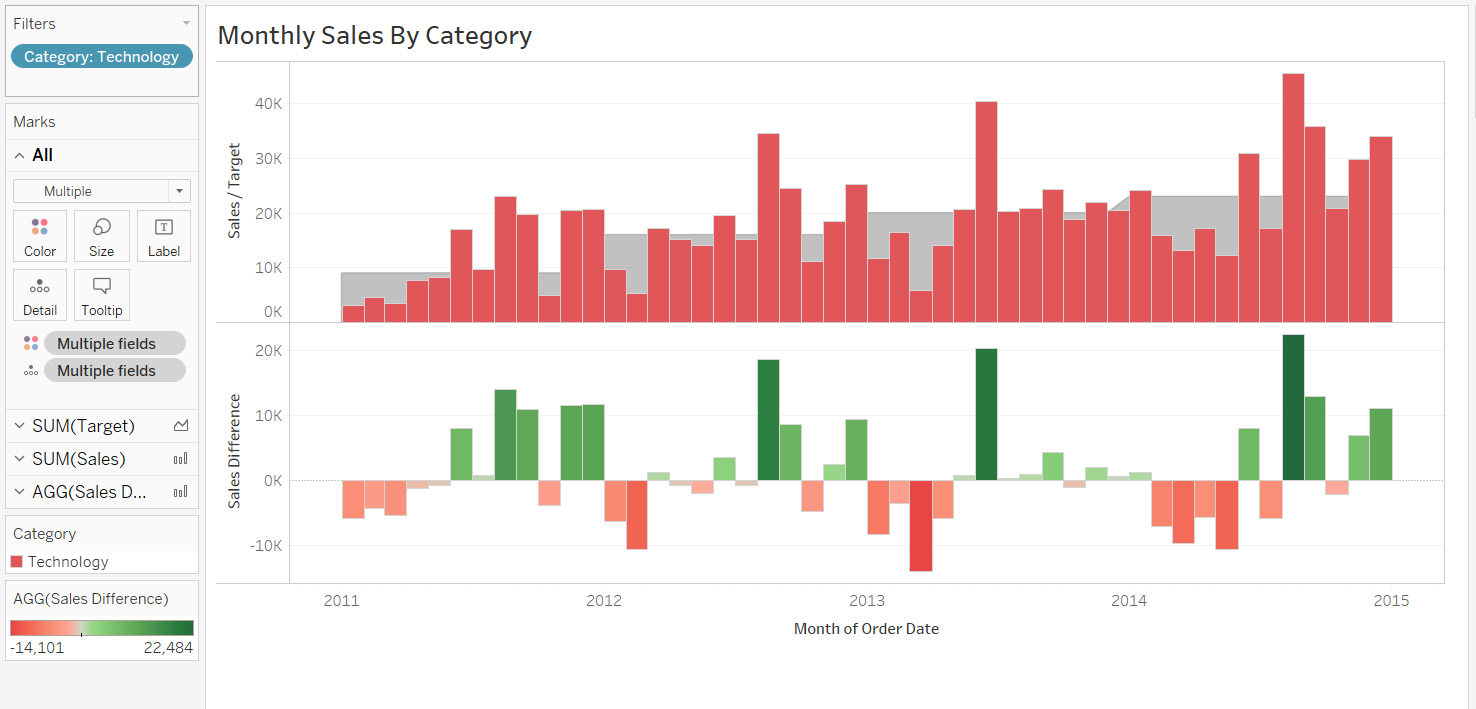


Figure 18 - Sales difference for technology

We can see that there are months that the company sold technology products with a number that is greater than the sales target but also there are months that below the target. The sales for technology is not as good as that of office supplies but also not bad as that of furniture so we can see there’s another way to improve marketing skills and increase number of sold product in this area.

We will move on to the next sheet which is the map of profit that describes how the company is gaining benefits from selling products in European countries.

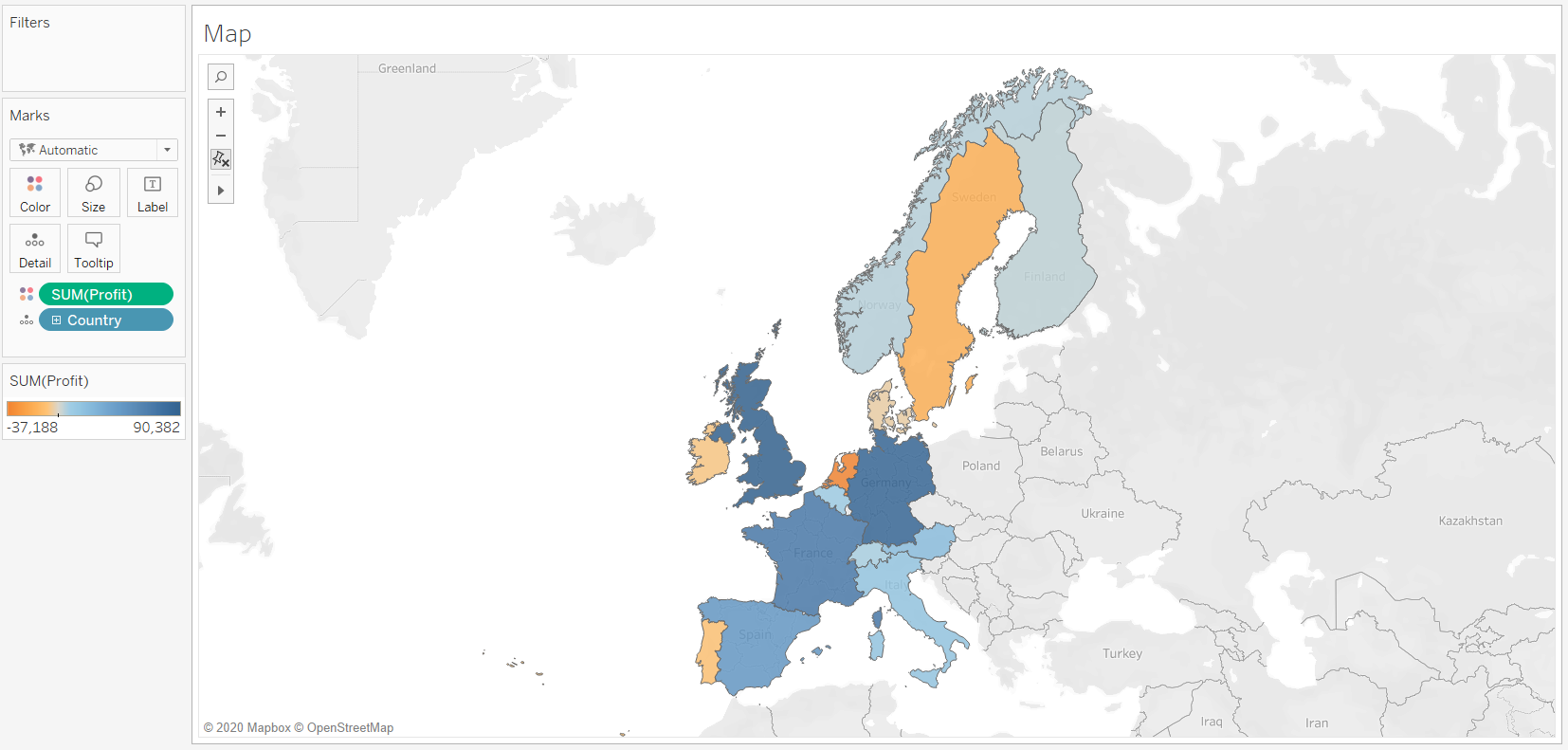


Figure 19 - Map

It shows that most of our benefits come from UK, France, Spain and Germany. Other countries like Italy, Switzerland, Belgium, Austria, Norway and Finland that have average sales can be improved in the future. In addition, we can see that the company products is not quite popular in Sweden, Ireland, Denmark, Netherlands and Portugal which can be a main source of causing sales below the targets. It’s important to try to come up with solutions that can help to improve our productivity and product marketing.

In order to know how much the company has gained, we can plot the profit and the sales as well as compare them together with profit margin. From the chart below, we can compare the how the sales is changing in term of profits. The following figures shows the profit margin in color:

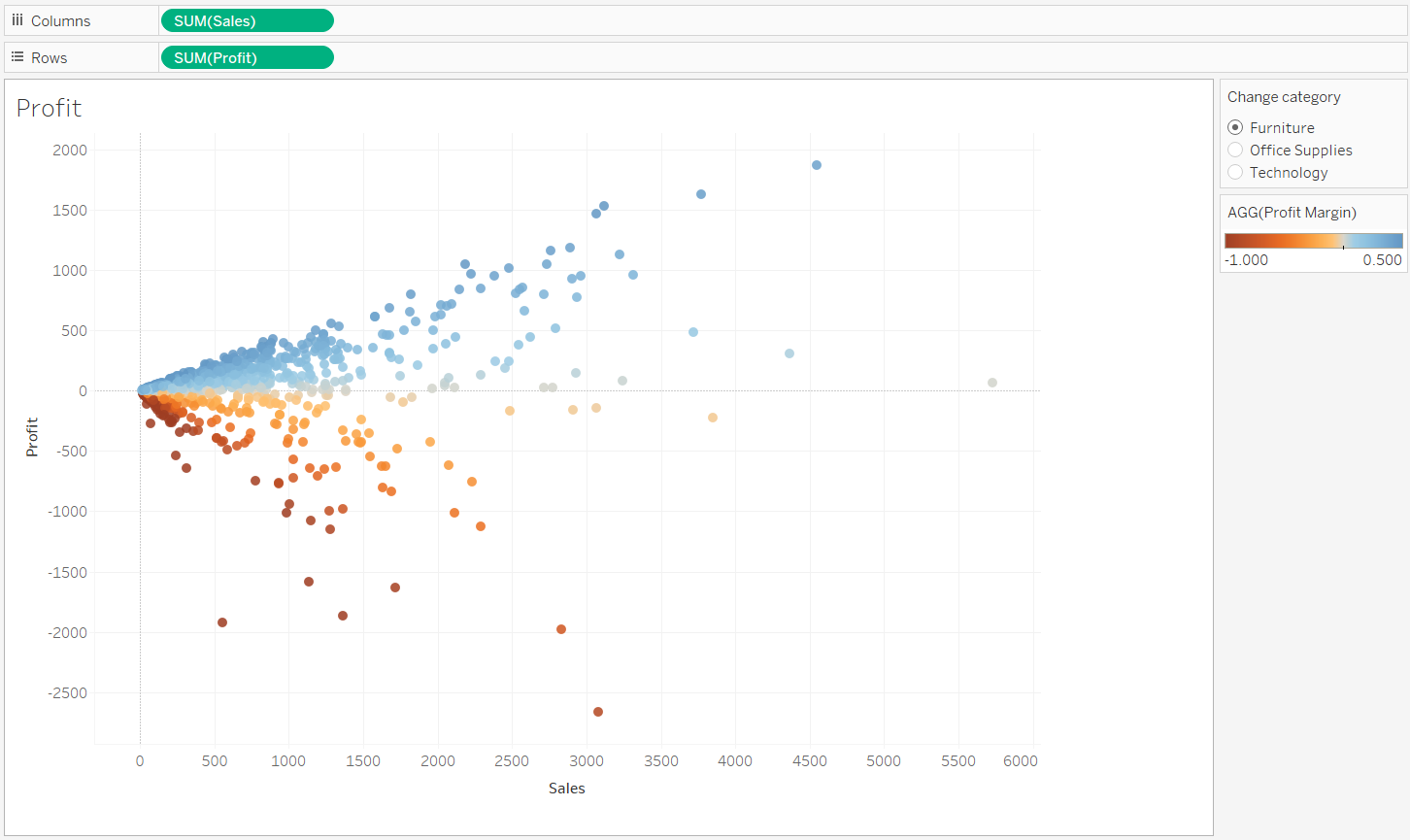


Figure 20 - Profit and sales

# 3. Report review and evaluation

The dashboard of all the charts can be shown as below:

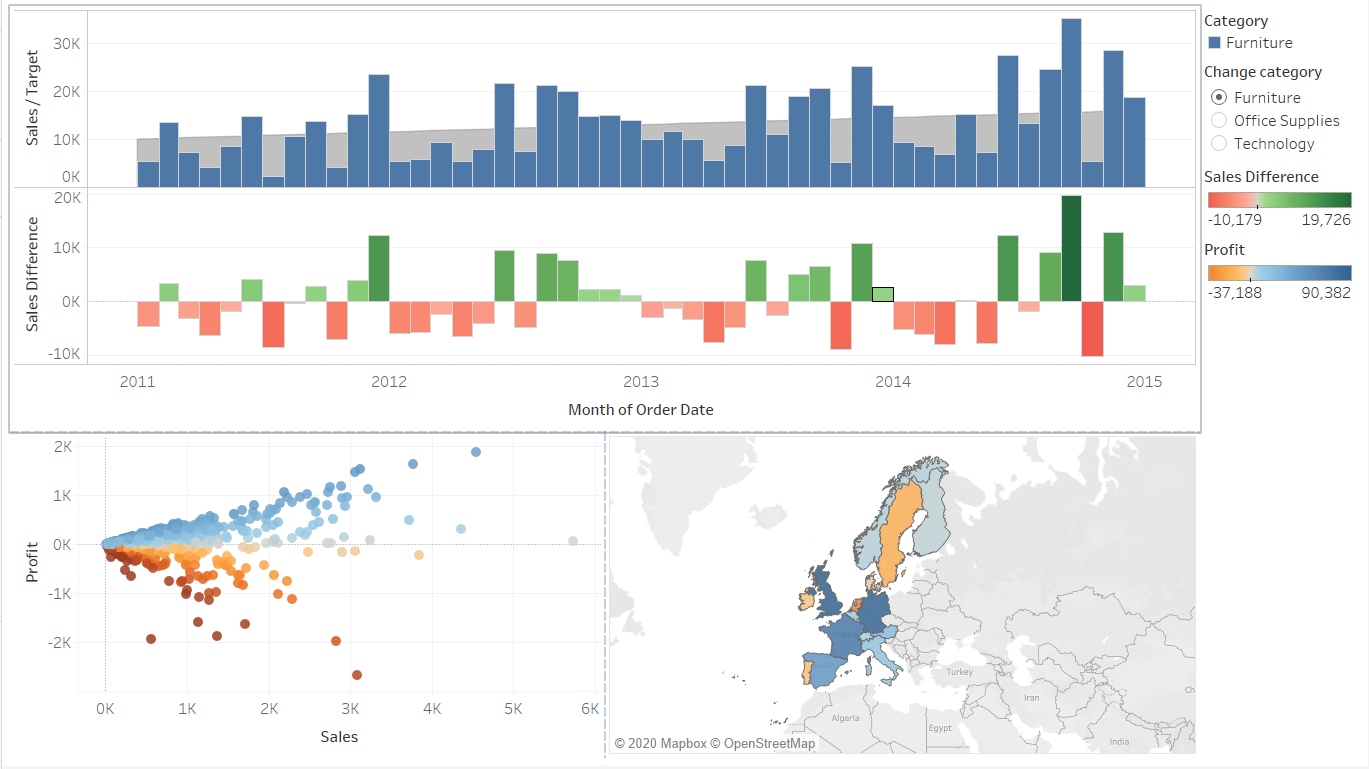


Figure 21 - Dashboard

In general, the report has shown one of the most important aspects in business is the sales for each month of a product type. We can get to know which month the company sell all products and which is not. The map show how we sell products across countries and we can compare the relationship between profit and sales. The user and business requirements that we have achieved can be shown as below:

* Design a reliable and comprehensive charts and reports.
* Analyze the sales, profit and sales target for each month.
* Compare sales of every month in order to come up with an approach helping to increase revenue.
* Know how well the company is doing its job on each type of product.
* From the profit margin we can define appropriate price for each product.
* Analyze sales of countries by map.
* Solutions for furniture products to increase sales.
* Improve marketing for advertising to more customers depending on country.

The dashboard can be improved with some features by enhance the appearance of the charts and user interface that help reporter make the report more concise and clean. It’s crucial to have a design that is familiar to people and readable so that they can see how the data looks immediately. Sometimes, the report need to be simple which means not causing people confused and having difficult time in order to fully understand the charts. Some of the good points that can be added:

* Create a responsive design as well as how we can visualize it in different devices.
* We can compare the sales for each state and city.
* By looking at the subcategory we can detect the problem in selling a certain product category.
* Compare the product quantity for each month.
* We can analyze how the ship mode can affect the business process and customer experience.
* Add more filters and granularity level for charts and data types for more precise looks.
* Make a story that analyze the impact, causing and suggestion.

# References

(n.d.). Retrieved from dremio.com: https://www.dremio.com/what-is-a-data-warehouse/

(n.d.). Retrieved from timhieuvewordpresswebsite.webflow.io: http://timhieuvewordpresswebsite.webflow.io/posts/cai-dat-google-analytics-cho-wordpress-nhu-the-nao-dung-cach

(n.d.). Retrieved from mfahri.web.id: https://mfahri.web.id/data-mining-php/

(n.d.). Retrieved from nordiccoder.com: https://nordiccoder.com/blog/data-visualization-tools/

(n.d.). Retrieved from educba.com: https://www.educba.com/olap-tools/

(n.d.). Retrieved from tableau.com: https://tableau.com

(n.d.). Retrieved from invoiceberry.com: https://www.invoiceberry.com/blog/enhance-small-business-in-the-digital-age/sisense-small-business-in-the-digital-age/

(n.d.). Retrieved from cs.waikato.ac.nz: https://www.cs.waikato.ac.nz/ml/weka/

(n.d.). Retrieved from hackingandslacking.com: https://hackingandslacking.com/running-jupyter-notebooks-on-a-ubuntu-server-b7b2cb88a986