# ST2187 Business analytics, applied modelling and prediction coursework

## Student number: 220656407

# Executive Summary

The main aim of this analysis is to examine the data about orders from the store and its details. This information could be useful for the company in identifying its weaknesses and making decisions based on it. To conduct this research, it is essential to compare various factors.

The dataset presented here provides information about orders for a variety of products delivered to different locations all over the world. Due to the lack of information about the sellers, we can assume that all of the products are sold from a single "store". The dataset includes information about different categories of products, its shipping costs, profit, delivery terms together with market segment and locations.

Also, a large dataset includes some outliers, which might be interesting to investigate further. While the overall trend of sales and profits for the store is positive, it could have been even higher.

Various subcategories of products, in some of the cases have a negative profit depending on country or even total negative profits. This is not what the company expected, so we need to take a closer look to identify the reasons for these negative profits and make suggestions about changes which could improve the situation.

Additionally, an important factor for the company's international operations is the cost of shipping goods. Some outlying countries are particularly expensive for shipping, so it might be beneficial to consider better supply chains to reduce costs.

In summary, the comprehensive analysis presented below provides insights into sales trends, profit drivers, shipping costs, and future market trends. The identified outliers and trends can be used to make strategic recommendations, such as optimizing shipping operations, refining discount strategies, and changing the product portfolio to maintain profitability and growth in the long run.

**Table of contents:**

**Overview**

To start with, we are interested in analysis of the biggest clusters as here we have division by categories. Wondering to see the generalized data.
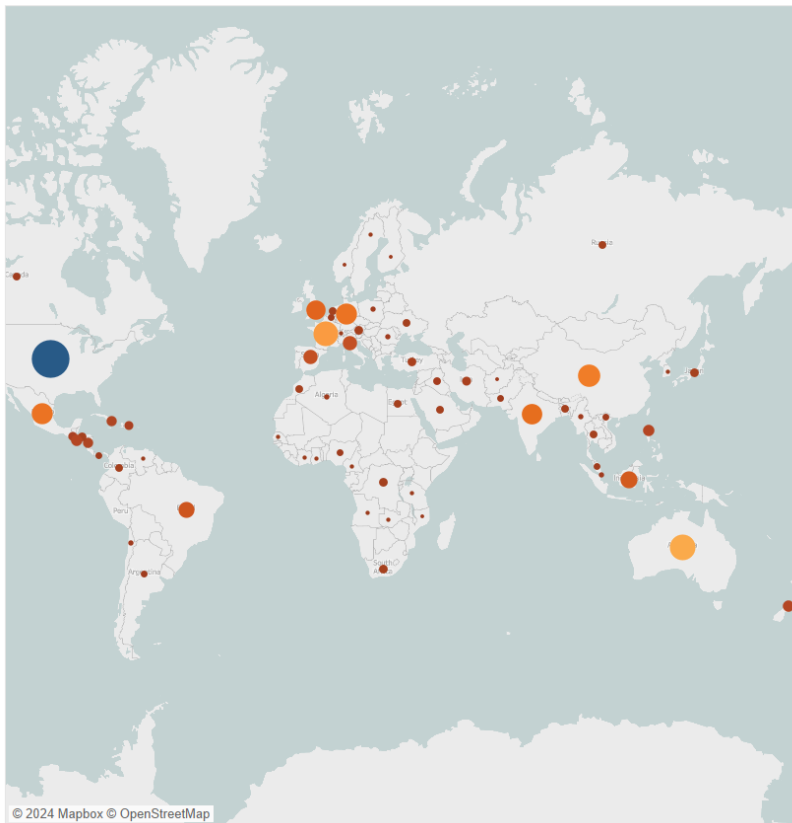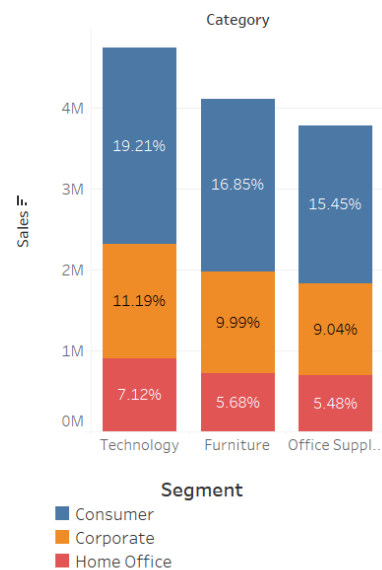
figure 1.1



Figure 1.1

While the goods are delivered to many countries all over the world, it is necessary to find out the hugest buyers to draw any conclusions. Here by the size and color of the markers we could notice that most goods are bought in particular countries. As the store is selling goods internationally, then total sales per country less than 20000 are miserable and here countries like that are not displayed for better perception.

Also, according to this reason there are many demographic and economical factors that could affect sales in various countries. Sales and quantities of products per each country could be explained by the size of population of the country together with the Gross Domestic Product per Capita of the particular country, although there are many other factors these two might be the most apprehensible. Thus, as the population of not all countries may order from the store frequently, having divergent

consumption powers, *managers may think of narrowing the range of countries to which the store delivers its products.*

Figure 1.2
Having three different segments we may see that the distribution of goods by each is nearly the same having slightly varying percentages of sales inside of the category, this may give us an idea about the same demand for goods from each segment. However, while the proportions are nearly the same, the most popular segment is "Consumer".

**During the year**

As we are looking for some outstanding data, we also may look for a trend to see the whole data at a time
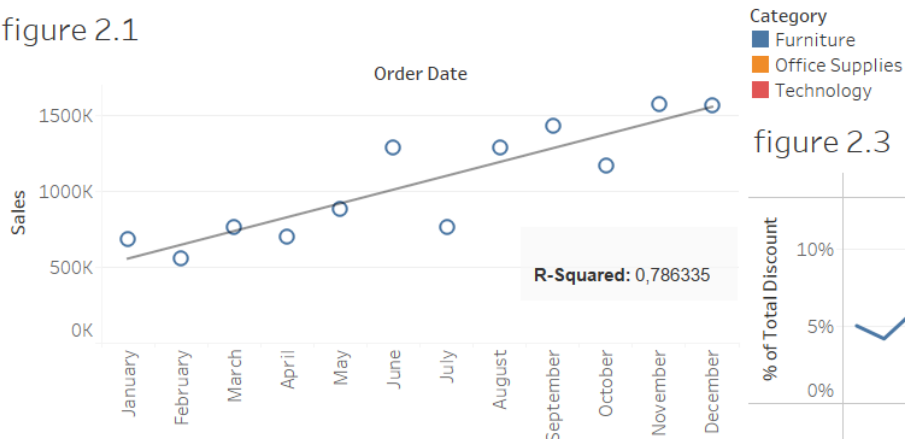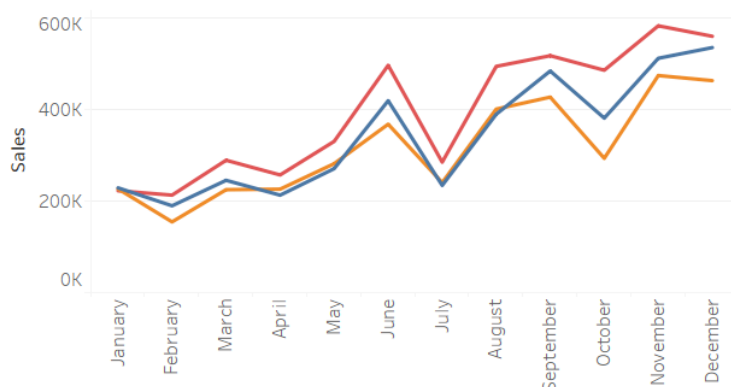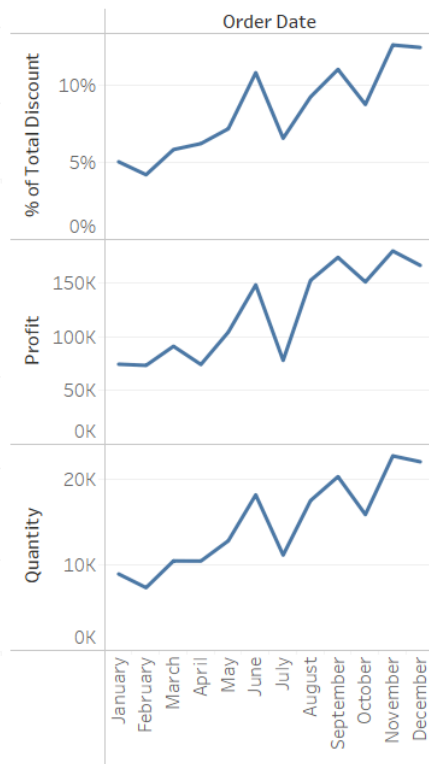


Figure 2.1

The figure shows the trend line that depends on the month of the year compared to the sales during this month. In marketing and consumer behavior the period of the year is often taken into account, on this widget we may see some of the outlying months while the whole trend of sales is upwarding. Having an R-Squared correlation index equal to 0,786335, it is clear that there is a strong correlation and relation between the month of the year and the amount of sales.

Figure 2.2

While the trend is still seen and the pattern is nearly the same for all segments, here are some months that show the peak values. July becomes the month with lowest sales while December has the highest number. It might be caused by the outer

factors in the world such as holidays, which marketologists and sellers are trying to take into account.

Figure 2.3

With the same trend as before it might be interesting to analyze and figure out the causal relationship. Sales, quantity and profit usually have the huge relation to each other, here we are interested to figure out the impact of them: the discount, profit and quantity have nearly the same graphs, so, it might be concluded that if the fluctuations in discount percent are relatively low, as in January to May period, then there would be lower drop in profit if discount goes up. On the other hand, if the market experiences larger differences in discount, then consumers' behavior might change bringing less profit and being more reliant on discounts as in July and October examples sales go down extensively.

*Thus, it is important for marketologists to manipulate discounts cautiously not to influence further sales and profit.*

# Profit and outliers

While the sales bring us a lot of information about the trend and market sentiment, profit probably is the most vital and important index for sellers and their decisions
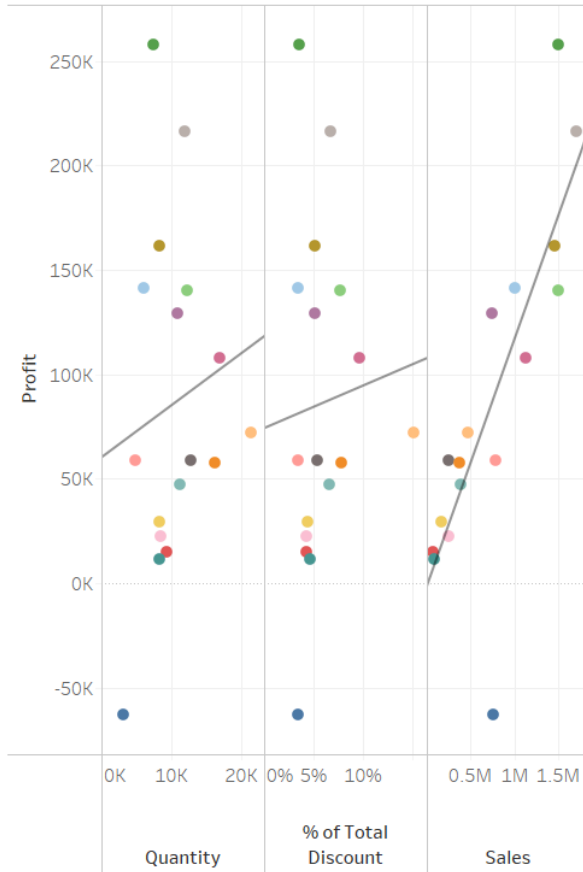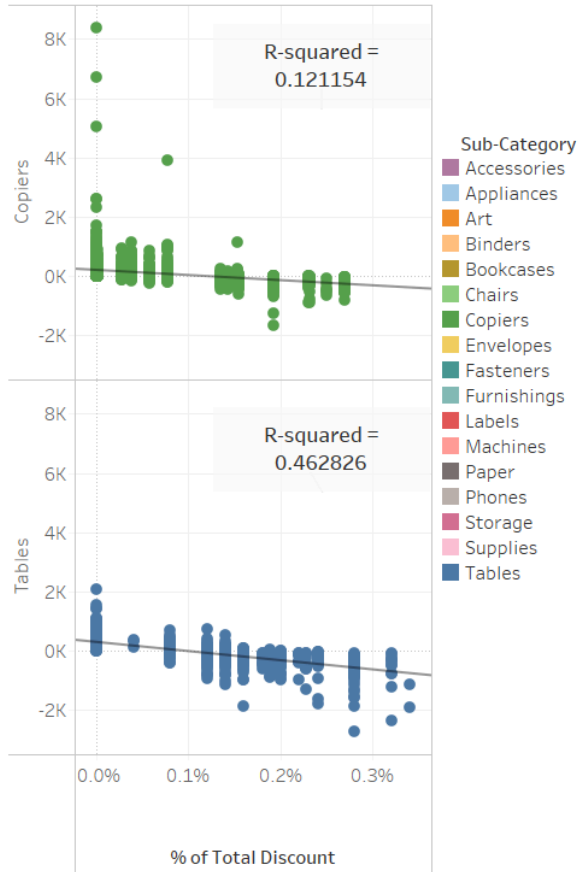


Figure 3.1

The figure shows relations between profit, quantity, discount and sales depending on the subcategory of the goods. There is a relation between all of them, however, while discount and quantity seem to affect profit less, sales have the highest impact as the trend line is steeper. Here we could notice some of the outliers which affect all the data and need a more precise view.

Figure 3.2

Outliers with the highest and lowest profit are "Copiers" and "Tables" subcategories, respectively. While both categories have relatively low quantities of goods sold there is a huge difference in profit, which might be caused by the discount effect. The correlation coefficient between profit and discount for "Copiers" is low being equal to

0,121154 while for "Tables" R-squared is 0,462826. Thus, sales and profit of "Tables" are more affected by the discount percent while "Copiers" are less. The difference in profit might also be caused by the distribution of sales depending on the profit. On the widget for "Copiers" it is seen that the major amount of them is sold by the initial price made by the seller or with a small discount. From the widget for "Tables", the largest number of them are sold with relatively high discounts and a really small amount is sold at the initial price which causes negative profit. The distribution of sales by the discount here plays a key role, and has a strong impact on profits.

*Thus, it is really important for sellers to try selling a higher quantity of goods by initial price or at least with a lower discount.*

**Shipping cost**

It might be really helpful for the company to understand where the cost is the highest to create better supply chains and find out the weak spots.
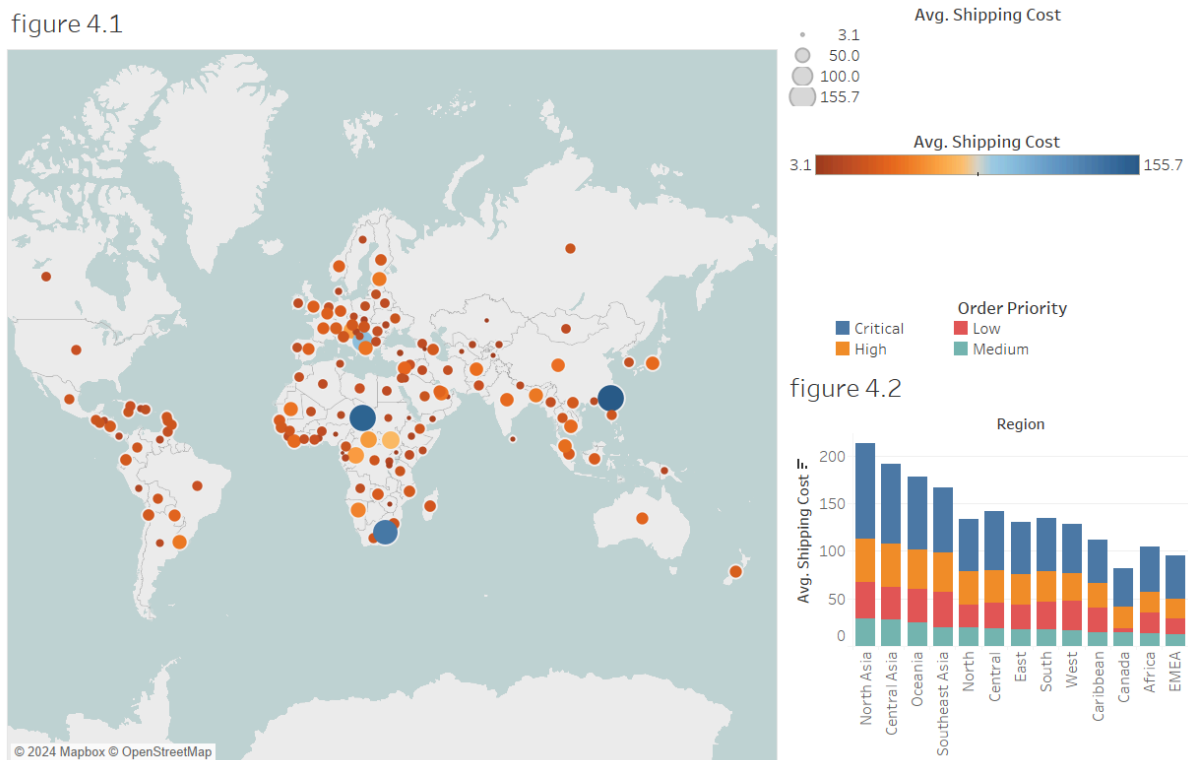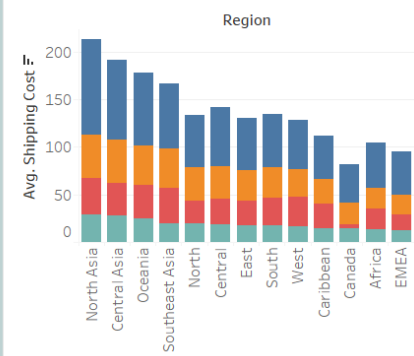


Figure 4.1

The widget shows average shipping cost depending on the country. It is very important to use average measurements as the sum measurement would not give us any information being affected by the quantity of deliveries. Here the countries with bigger markers of cooler colors have the highest average shipping cost. Probably being situated on the most distant places from the department where the orders arrive initially and then distributed by the countries. As it could be seen most of the countries with the highest average cost of shipping are situated near to the Indian Ocean or in the center of the mainland. *So, in a situation like this it might be helpful to make an additional shipping department somewhere in this area.*

Figure 4.2

There is a dependence between average shipping cost, order priority and ship mode. "Order priority" and "Ship mode" have a strong relation, thus, it is better to consider only one of them. Widget shows the dependence between shipping cost, order

priority and region, while it is already known that in some of the regions shipping cost is higher than in others: shipping to the North Asia region is the most expensive one, shipping to EMEA is the cheapest. There also might be a slight difference in relation of shipping cost to the order priority depending on the region. It could be seen here that the cost of shipping is the highest for the critical and high priorities which may be driven by the high resource usage and urgency. According to the cost of shipping for low and medium priority orders, the low priority orders shipping cost should probably be lower in real world condition as orders might be sent in any of the ships having place for it. However, here the cost of shipping for low priority orders is higher than for the medium, probably due to the smaller number of low prioritized orders, thus, the average cost per goods goes up. *In this situation the possible solution is excluding the "low priority" category and assigning all of its orders to medium one, thus, making the shipping cheaper for the company.*

# Forecasting

It is easy to calculate the profit at the moment, however, any company is interested in further predictions and its possible decrease in profit. Also, forecasting might be helpful in terms of the decision making process.
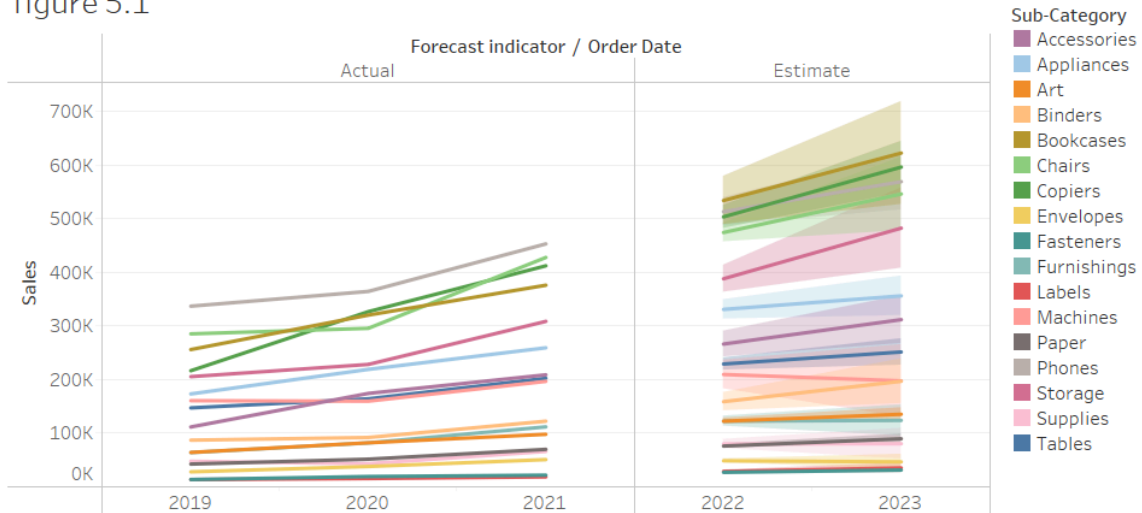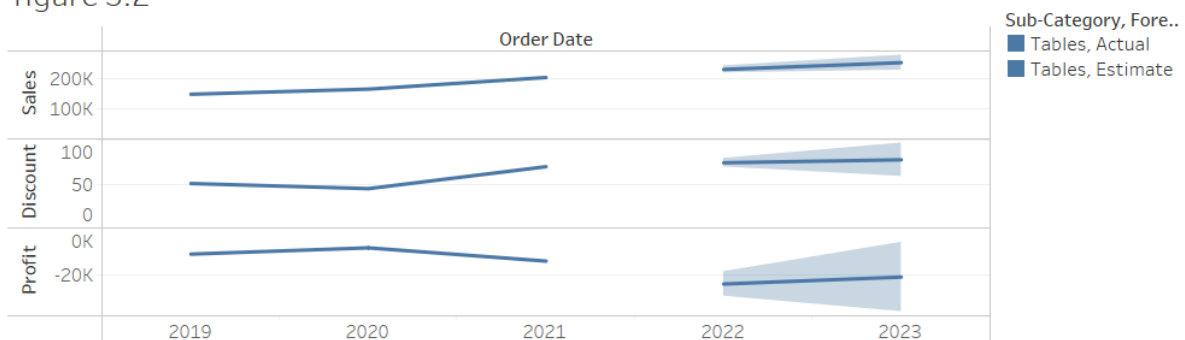


figure 5.1



figure 5.2

Figure 5.1

The widget presents predictive data about the subcategories sales. The trend of the past years should remain the same for most of the subcategories, however, for the less profitable ones it would be better to analyze other factors.

Figure 5.2

For the least profitable subcategory - "Tables" it is interesting to conduct a better analysis. Here the dependence of discount to profit was always relatively high, looking at the forecast, this trend remains the same. According to the experience of the previous years "Tables" was the most unprofitable category with the negative earnings, from the graph it is clear that the same relation is going to stay for the next

years, lowering the total profit for the company, thus, *it might be an idea to stop selling this products in future and concentrate the attention on better opportunities or try the best promotion together with sales with lower discount.*

**Number of words: 1458**

*excluding an executive summary, table of contents, and the title page