AEKI REPORT

Pol - Dimas - Juan - Gabriel - Monia

2024-01-08

Introduction

This is a comprehensive report of a thorough investigation and evaluation of the historical data provided by AEKI, containing data about the operations of AEKI supermarket during its best years, from 2014 to 2017. The main objective of this analysis is to discover insights from the supermarket's past performance, aiming to identify successful strategies that could potentially serve as a guide for improving their current business performance.

AEKI Supermarket finds itself in a crisis of significant downturn in performance over recent years. Therefore, the analysis will be vital in finding their way back to success.

Throughout this report, a extensive examination of datasets provided by AEKI build up of our analysis. It must be noted that the accuracy and reliability of our conclusions are directly linked to the authenticity and reliability of the data provided by AEKI themselves.

The following sections of this report aim to uncover key insights derived from exploration of AEKI's historical data, highlighting the factors that contributed to their past success and providing possible recommendations of actions to guide the company towards its previous glory.

Understanding the Structure of the datasets

Check the integrity and consistency of the data sets

Orders dataset

There are 10000 observations for 18 variables. We have 3 different classes of variables: Character (11), POSIXct (2) and Numerical (5)

```
# Summary statistics
summary(AEKI_Orders)
```

By summarizing all the dataframe we can appreciate that every variables seems to have logical observations except Sales, Profit and Discount. Although the maximum Quantity varies considerably from the mean, we consider it to have logical observations because they are plausible.

The following code was used to identify missing values from the AEKI_Orders dataset:

```
# Checking for missing values
colSums(is.na(AEKI_Orders))
```

Thanks to this we can appreciate that there aren't any NA values throughout the dataframe.

Products dataset

We have 1847 observations and 4 variables of only 1 class (Character)

< br>

```r
# Summary statistics
summary(AEKI_Products)
```

```r
# Checking for missing values
colSums(is.na(AEKI_Products))
```

Thanks to this we can appreciate that there aren't any NA values throughout the dataframe.

Demographics dataset

This dataset has 52 observations for 8 variables, smaller than the other two. There are 2 variable classes: Numerical (4) and Character (4).

```r
# Summary statistics
summary(AEKI_Demographics)
```

From the coded summary, we can see that all variables seem to have logical observations. The only observation that could be an error when computing is Density=11707 for District of Columbia, but through a deeper investigation of the data we found out it is CORRECT.

```r
# Checking for missing values
colSums(is.na(AEKI_Demographics))
```

Thanks to this we can appreciate that there aren't any NA values throughout the data frame.

Data Cleaning and Preparation

From the "Understanding the Structure of the datasets" section above, we identified that AEKI_Orders is the likely has outliers in the Sales, Profit and Discount variables.

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
##     0.002    17.280    54.490   230.735   209.947 22638.480
```

We can see that the maximum is extremely superior to the 3rd quartile, what suggests the existence of outliers.

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
##  -6599.98      1.73      8.67     42.40     29.36 135072.00
```

For profit we have the same situation with the maximum and also with the minimum.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.0000  0.0000  0.2000  0.1564  0.2000  1.6000
```

We can see that the maximum is 1.6, what represents a 160% discount on a product. That is not plausible as everything superior to 100% would involve giving money to customers when purchasing a product.

IQR method for outlier detection

```r
Sales_IQR <- IQR(AEKI_Orders$Sales)
Sales_lower <- quantile(AEKI_Orders$Sales, 0.25) - 1.5 * Sales_IQR
Sales_upper <- quantile(AEKI_Orders$Sales, 0.75) + 1.5 * Sales_IQR
```

SALES: Lower-end SALES outliers according to IQR method:

	x
25%	-271.7212

Upper-end SALES outliers according to IQR method:

	x
75%	498.9488

```r
Profit_IQR <- IQR(AEKI_Orders$Profit)
Profit_lower <- quantile(AEKI_Orders$Profit, 0.25) - 1.5 * Profit_IQR
Profit_upper <- quantile(AEKI_Orders$Profit, 0.75) + 1.5 * Profit_IQR
```

PROFIT : Lower-end PROFIT outliers according to IQR method:

	x
25%	-39.726

Upper-end PROFIT outliers according to IQR method:

	x
75%	70.818

```r
Discount_IQR <- IQR(AEKI_Orders$Discount)
Discount_lower <- quantile(AEKI_Orders$Discount, 0.25) - 1.5 * Discount_IQR
Discount_upper <- quantile(AEKI_Orders$Discount, 0.75) + 1.5 * Discount_IQR
```

DISCOUNT: Lower-end DISCOUNT outliers according to IQR method:

	x
25%	-0.3

Upper-end DISCOUNT outliers according to IQR method:

	x
75%	0.5

LIST OF OUTLIERS ACCORDING TO IQR METHOD (Does not necessarily mean to be 100% an outlier) SALES: 875 observations considered outliers

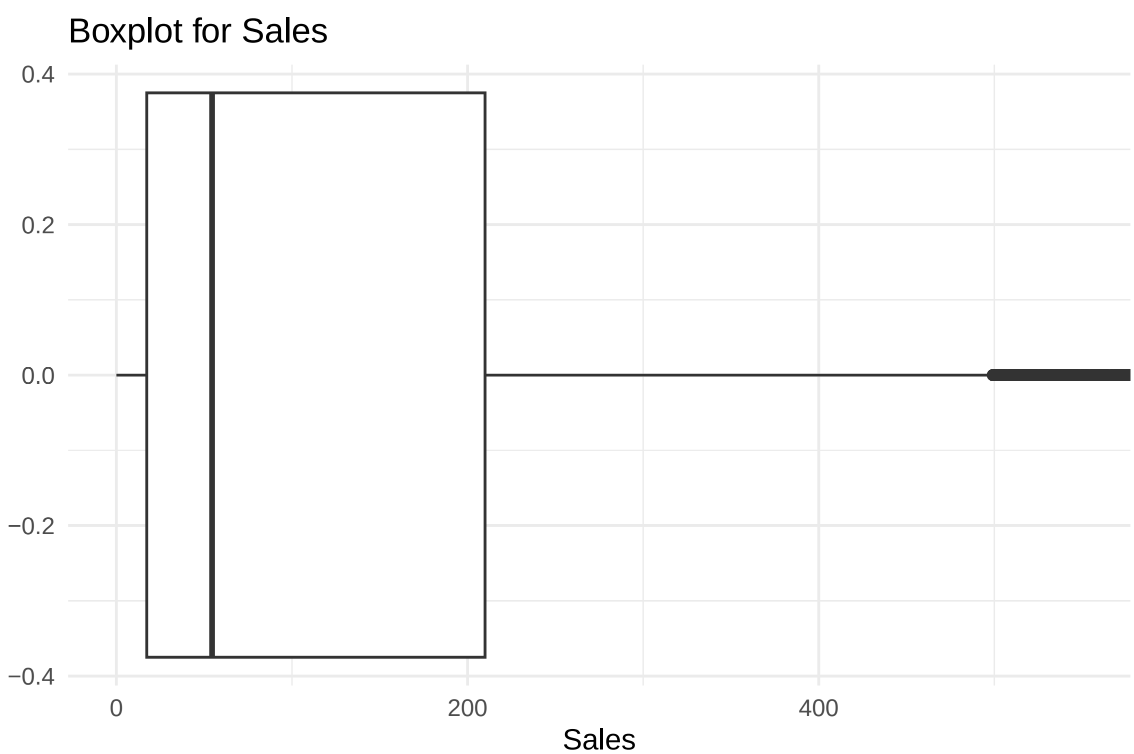
PROFIT: 1400 observations considered outliers

DISCOUNT: 639 observations considered outliers

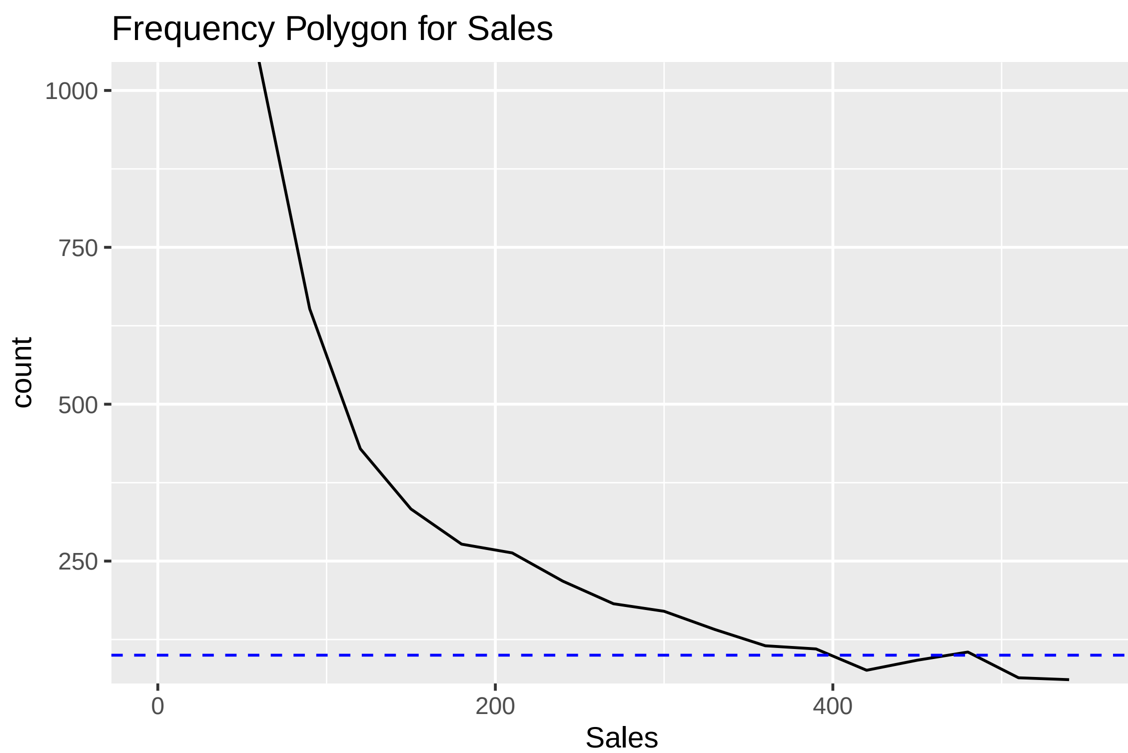
We believe further investigation is required, as the IQR method is likely removing "outliers" which can be justifiable

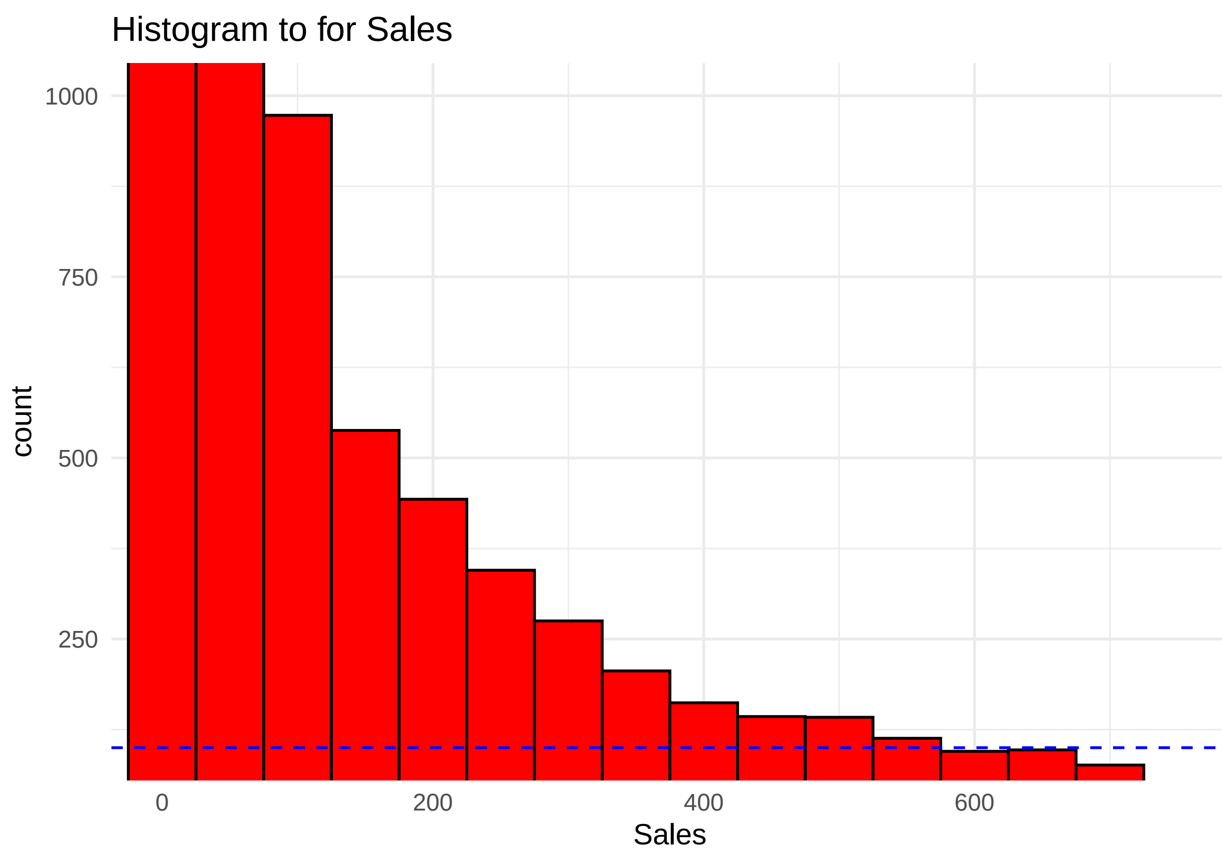
Further investigation into outliers, using visalisation tools.

Visualisation and Analysis of Sales Outliers Variation of plots for better visualisation:



We can appreciate that, approx. from Sales=500 on, there are outliers.

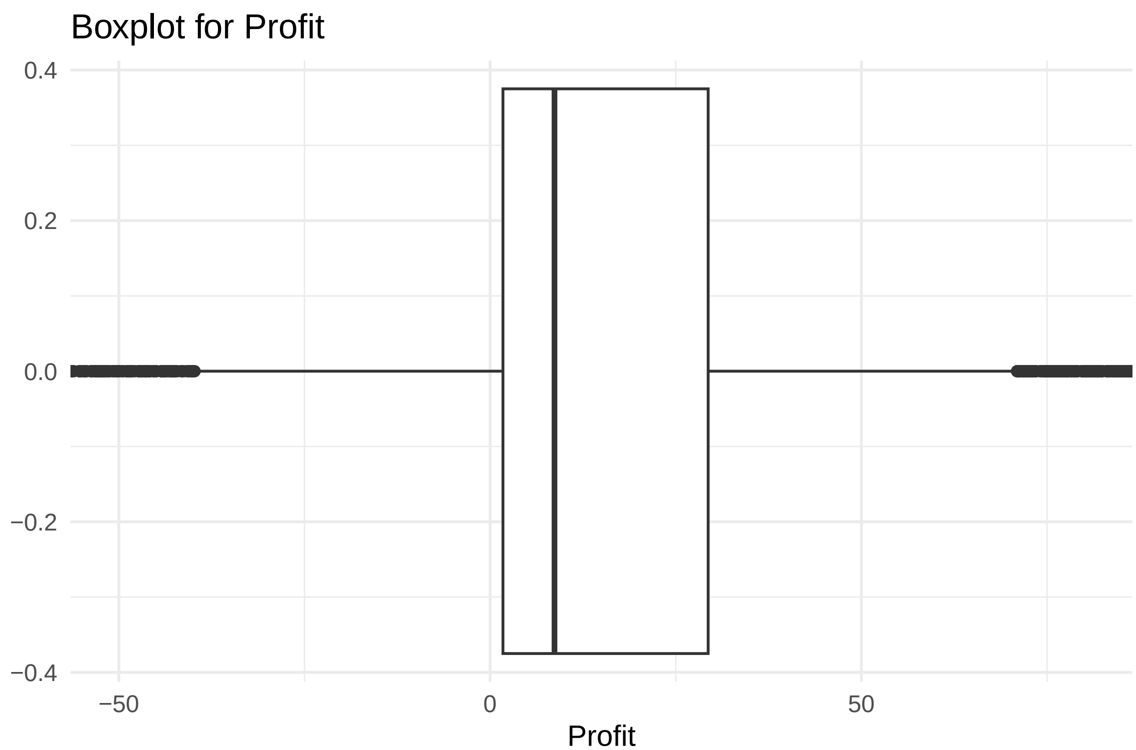




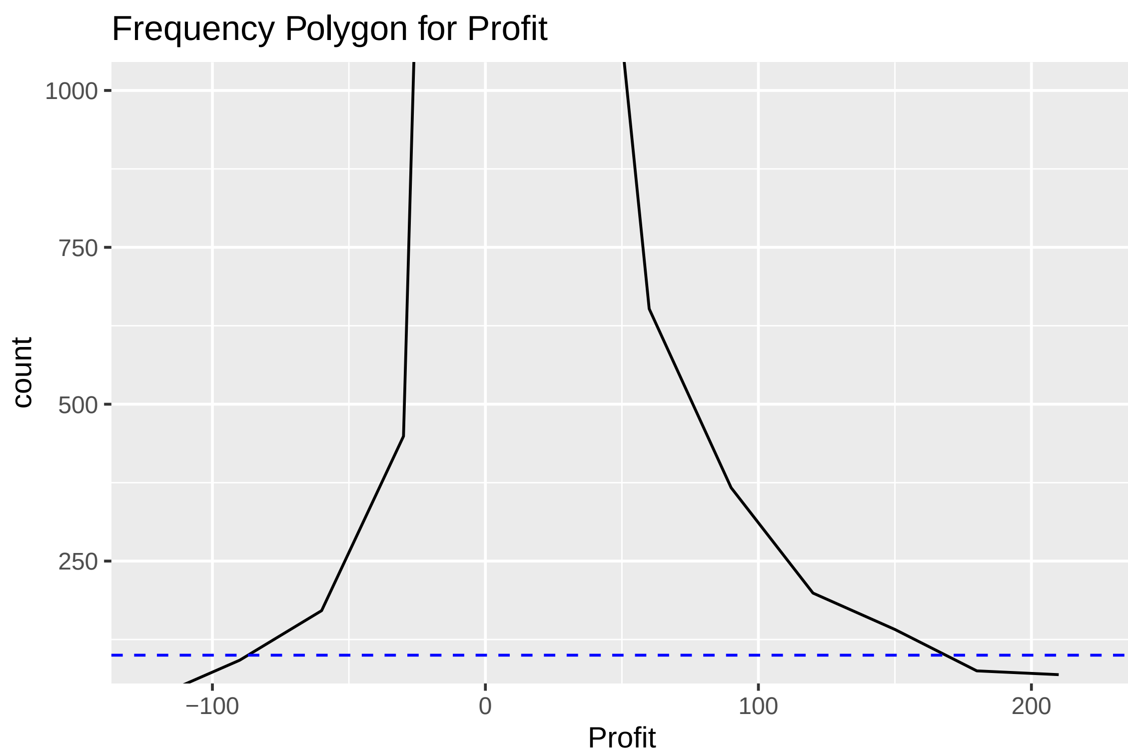
Conclusions from the plots: From the previous two plots, we deduce that the range of Sales 0 to 500 is the correct and accurate one (agreeing with the outlier formula). Thus, the rest of observations are considered outliers.

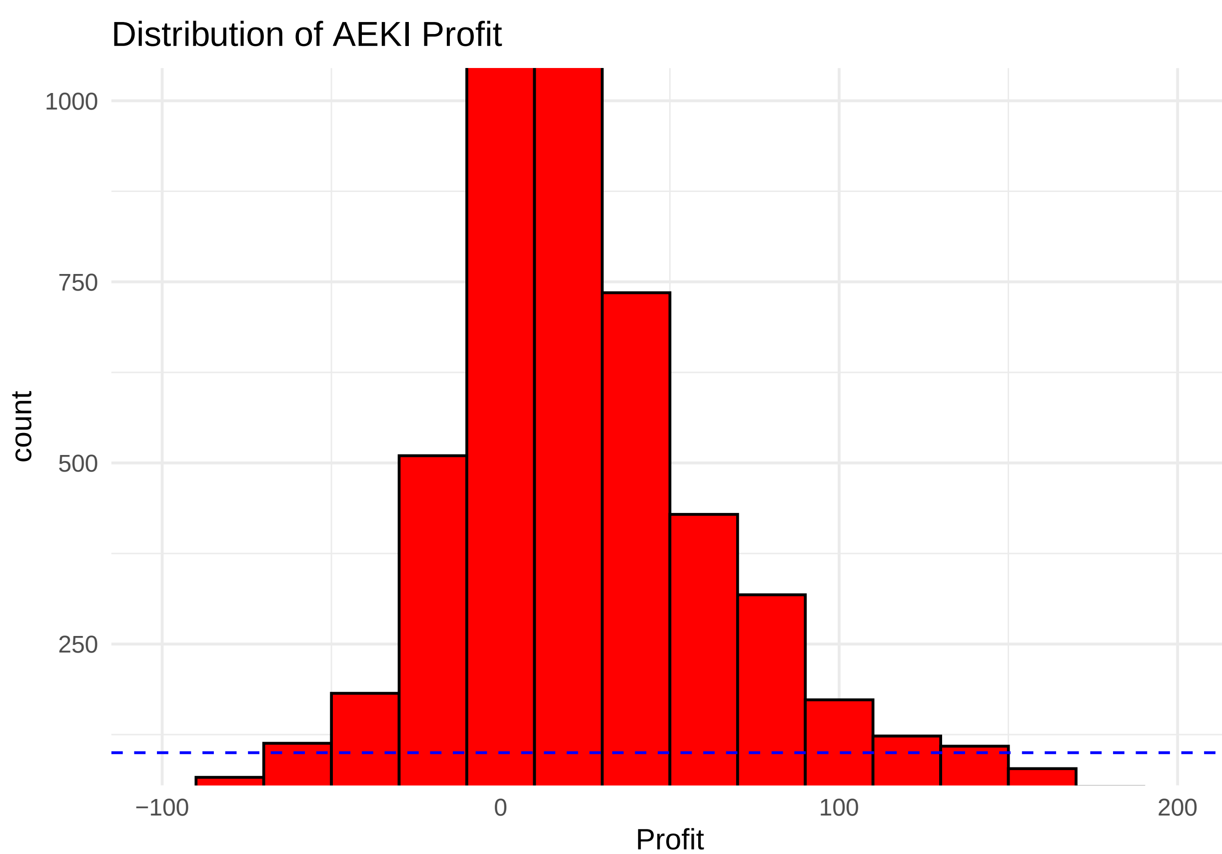
For both the Histogram and the Frequency Polygon, we have agreed that COUNT=100 must be the minimum for Sales to be representative enough. Therefore, every observation below the plotted line will be classified as an outlier.

Visualisation and Analysis of Profit Outliers Variation of plots for better visualisation:



We can appreciate that for -40 approximately and lower there are outliers and for 70 approximately, and higher, too.



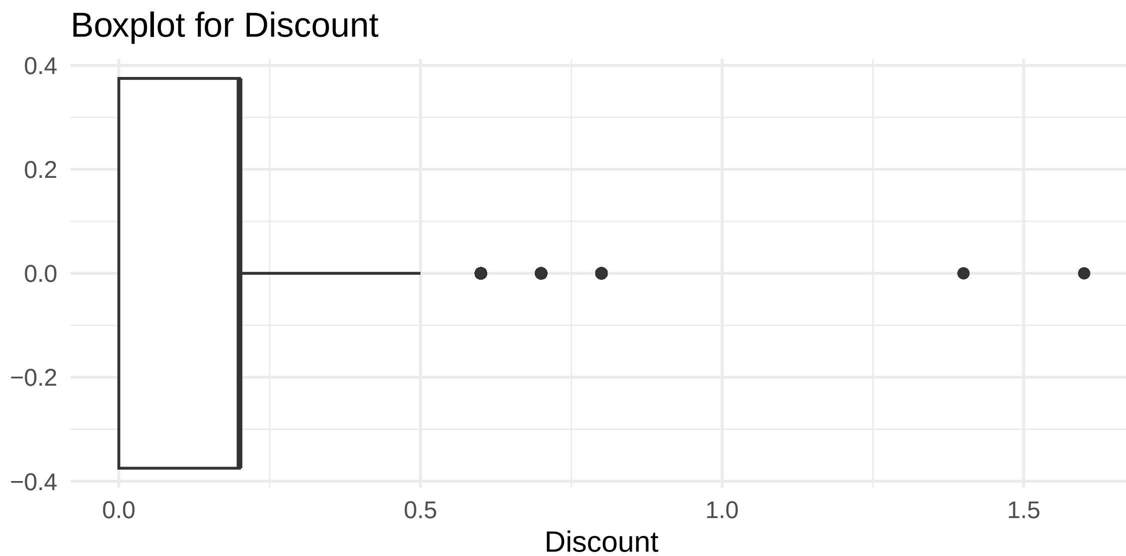


Conclusions from plots:

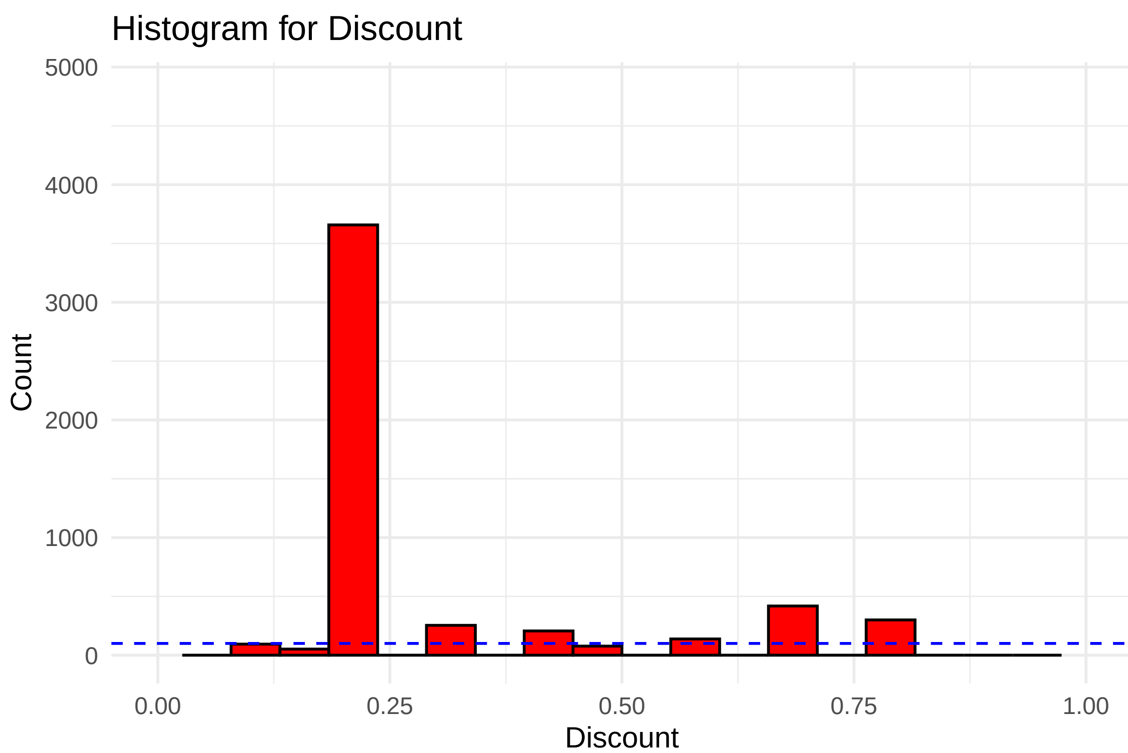
Despite the boxplot, by plotting a histogram or a frequency polygon, we deduce that the range of Profit -75 to 100 is the correct one. Thus, the rest of observations are considered outliers. For both the Histogram and the Frequency Polygon, we have agreed that COUNT=100 must be the minimum for Sales to be representative enough. Therefore, every observation below the plotted line will be classified as an outlier.

Accordingly, as we consider Count>100 to be representative enough, our range for correct values is much wider than the one calculated with the outlier formula (-39.7 , 70.8)

Visualisation and Analysis of Discount Outliers Variation of plots for better visualisation:



According to the boxplot, Discount>0.5 are outliers



Conclusions from plots:

By plotting, we deduce that the range of Discount 0 to 1 is the correct one, as the Count for those Discounts that are mathematically considered outliers (>0.5), follows our condition of Count>100.

Although, according to the outlier formula, every Discount greater than 0.5 is an outlier, we think that every discount equal or lower than 1 should be accepted as they are plausible. Therefore, the only observations that we consider to be outliers are those with Discount>1 (1.6 & 1.4)

```
Clean_AEKI_Orders <- AEKI_Orders %>%
  filter(Sales < 500, Profit > -75, Profit < 100, Discount < 1, Profit <= Sales)%>%
  mutate(Price_Per_Unit= (Sales*(1+Discount)/Quantity),
         Cost_Per_Unit= (Price_Per_Unit-Profit))
```

Cleaning Dataset "AEKI_Orders" from Outliers To create a new dataframe without the outliers we have filtered the problematic variables by the ranges we have previously determined.

In addition, we have also come up to the idea that Profit cannot be bigger than Sales (which equals revenue) as it is not logical. Therefore, we have added this filter: Profit <= Sales.

Also, we have added the variables Price_Per_Unit and Cost_Per_Unit, in order to have more data to work with.

From now on, we will be working on this cleaned dataframe. By cleaning the data we have created 2 new variables and deleted 1647 observations.

Joining Datasets

```
unique_products <- AEKI_Products %>%
  filter(!(duplicated(AEKI_Products$`ID Product`)))
```

Joining AEKI_Orders and AEKI_Products In order to join the tables, we need to remove duplicated Product IDs, so that the join runs without errors.

```
## Orders and products joing
names(Clean_AEKI_Orders)[names(AEKI_Orders) == "Product ID"] <- "ID Product"
Orders_Products <- merge(Clean_AEKI_Orders, AEKI_Products, by = "ID Product")
```

We want to join both dataframes by the Product ID, but they are named different in their dataframes. Therefore, we need to set that "Product ID" = "ID Product", and make the join by that specific variable.

```
Orders_Products <- Orders_Products %>%
  mutate(Category = ifelse(Category == "Office Suplies", "Office Supplies", Category))
```

We used this code to adjust the spelling mistake for the category Office Supplies (One observation was written "Office Suplies")

Financial analysis for AEKI

Preparing dataset to analyse:

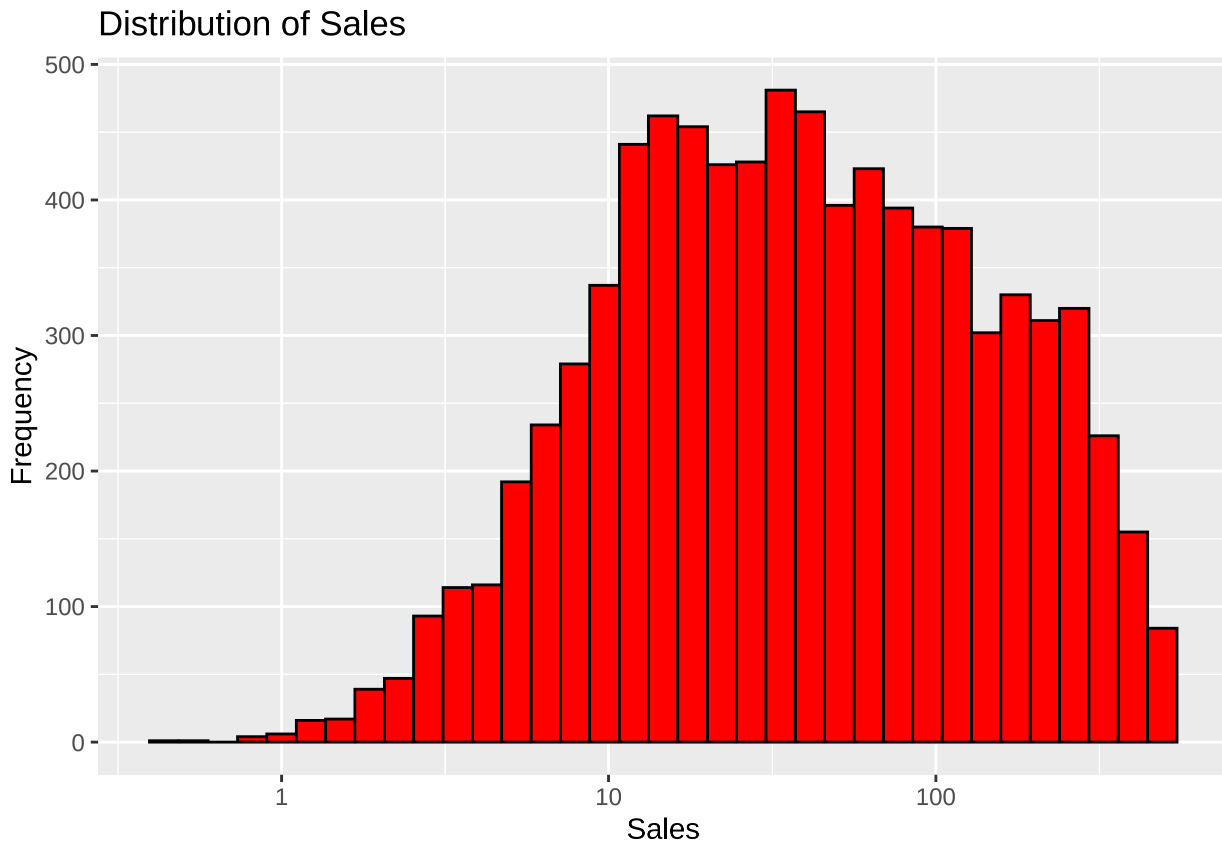
Category breakdown:

```
AEKI_Category_Analysis <- Orders_Products %>%
  na.omit() %>%
  group_by(Category) %>%
  summarise(Total_Sales = sum(Sales), Total_Profit = sum(Profit))%>%
  mutate(Margins= Total_Profit/Total_Sales)
```

Sub-Category breakdown:

```
AEKI_Sub_Category_Analysis <- Orders_Products %>%
  group_by(Category,`Sub-Category`) %>%
  summarise(Total_Sales = sum(Sales), Total_Profit = sum(Profit),
            AVG_Price=(mean(Price_Per_Unit)), AVG_Cost=(mean(Cost_Per_Unit)), count=n(),
            Quantity_Total=(sum(Quantity,na.rm=TRUE)),
            AVG_Profit= AVG_Price-AVG_Cost)%>%
  mutate(Margins= Total_Profit/Total_Sales)
```

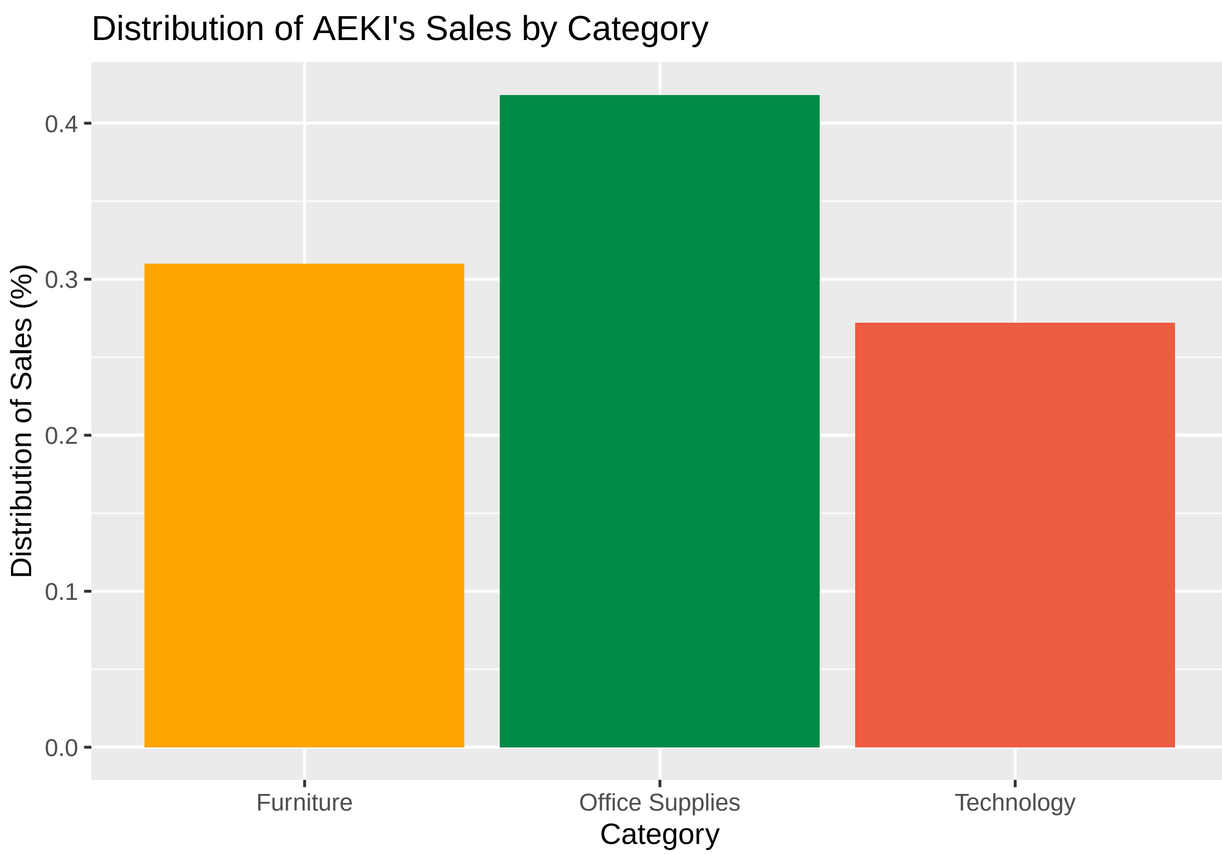
SALES



Left skewed histogram for sales distribution that shows the revenue of every sale. We can see that all sales that end in a loss start out being few and those that end in a profit have been carried out more than 100 times and subsequently begin to increase more in frequency. We can also observe that the sales that are around 36.84 EUR revenue are the ones that are most reputed to repeat more times. On the other hand we have that on average sales in general are concentrated around 80.28 EUR revenue.

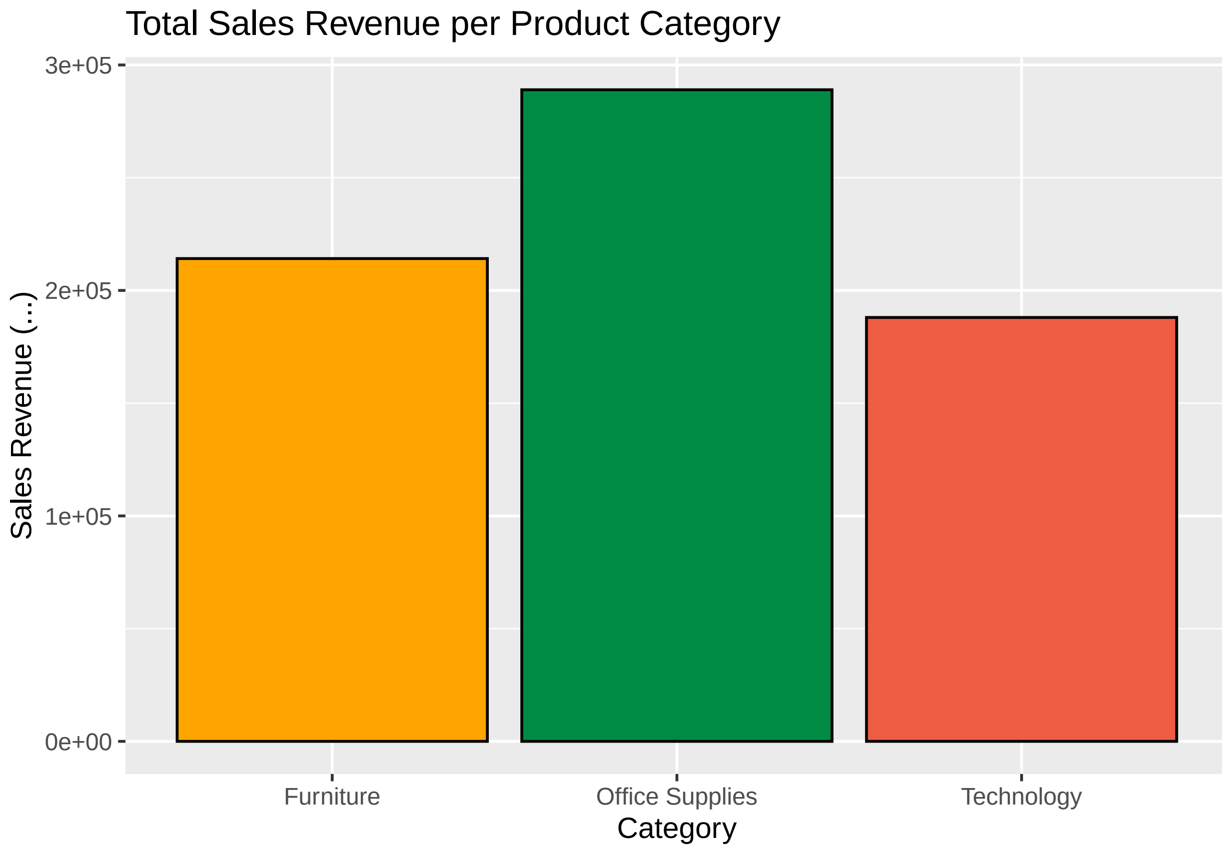
In conclusion we can see that sales ranging from 12 to 60 EUR are the ones that occur most frequently, however, between a range of 10 to 150 EUR the frequency is still very high as well.

Relationship between Sales and Category Graph showing the **percentage** of total sales that each Category represents (distribution of revenue):



We interpret that **Office Supplies is the category that represents the highest percentage of AEKI's Sales**. Office Supplies generate 279,953.4€ which represents a 0.4212 (42.12%) of total Sales. It is followed by Furniture with 0.3064 (30.64%), and Technology with a 0.2723 (27.23%) of the total sales. Therefore, we can agree with our previous analysis and conclude that, in this case for AEKI, the higher number of products sold leads to higher Sales revenue.

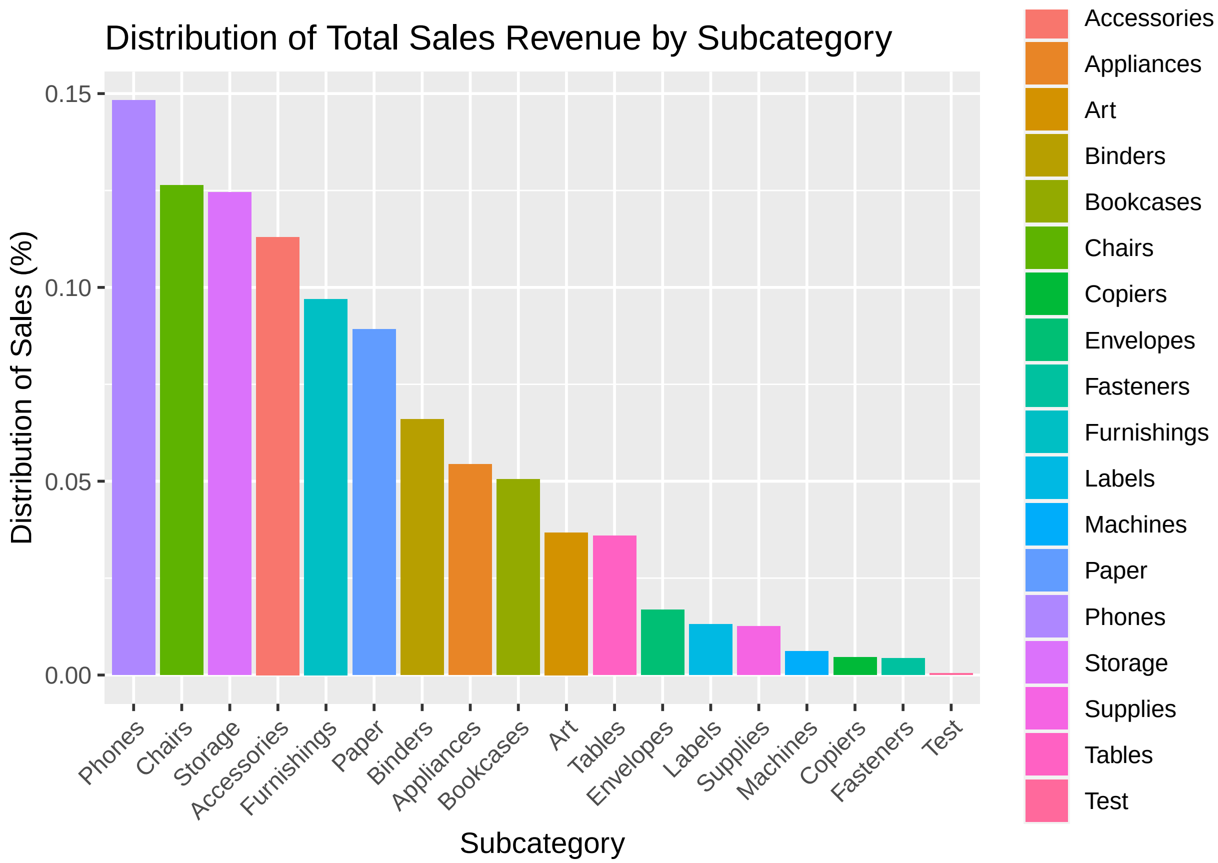
Graph showing the **absolute** of total sales that each Category represents (distribution of revenue):



We can see that the highest revenue earner for Categories is Office Supplies with 279,953.4€ generated.

On the other hand, the Technology Category is the one with lowest Sales with 180,969.9€ generated.

Relationship between Sales and Sub-Category Graph showing the percentage of total sales that each Sub-Category represents (distribution of revenue):



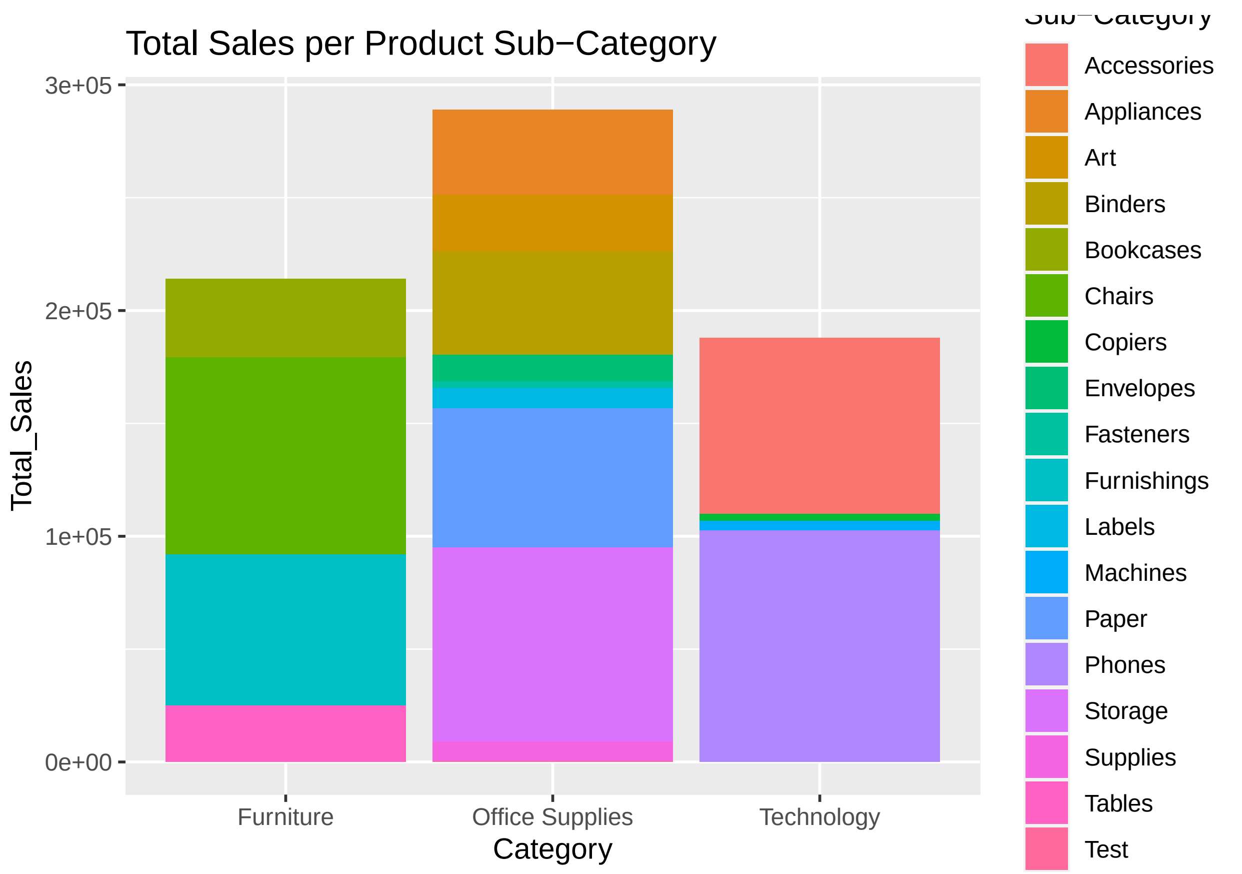
We get that the product **Sub-Category with highest representation of total Sales is Phones.** Phones have generated **98,657.150€** what represents 0.1484 **(14.84%) of total Sales**.

On the other hand, the least important product sub-category, in terms of sales, is Tests with only 306.9€ generated, representing the 0.00046 (0.0005% approx) of total Sales.

In conclusion, as we suggested before, the product subcategories that bring the most money to AEKI are not the most sold one.

Therefore, it has been a good choice to plot the distributions of total sales in order to analyze the product sub-category hierarchy.

Contribution of Sub-Categories in Categories based on Sales



From this graph we can conclude that AEKI'S best sellers per category. In Furniture category: chairs and furnishing. In Office Supplies category: storage, paper, binders, appliances, art. Lastlly, for Technology category: phones and accesories.

Furthermore we can observe the composition of the categories sales based on the subcategories. We mostly see a fair distribution except for a couple of products in tech and office/ We will study if stop selling them.

Sales analysis for AEKI - BREAKDOWN BY YEAR

Preparing dataset to analyse:

Separating dataframes per year:

```
Orders_Products <- Orders_Products %>%
  mutate(Year = year(`Order Date`))
```

We create a new variable including the year, so that it is easier for us to filter the data.

```
AEKI_2014 <- filter(Orders_Products, Year == 2014)
  ##1661 observations
AEKI_2015 <- filter(Orders_Products, Year == 2015)
  ##1738 observations
AEKI_2016 <- filter(Orders_Products, Year == 2016)
```

```
  ##2156 observations
AEKI_2017 <- filter(Orders_Products, Year == 2017)
  ##2798 observations
```

We can see that the year with the highest number of sales was 2017.

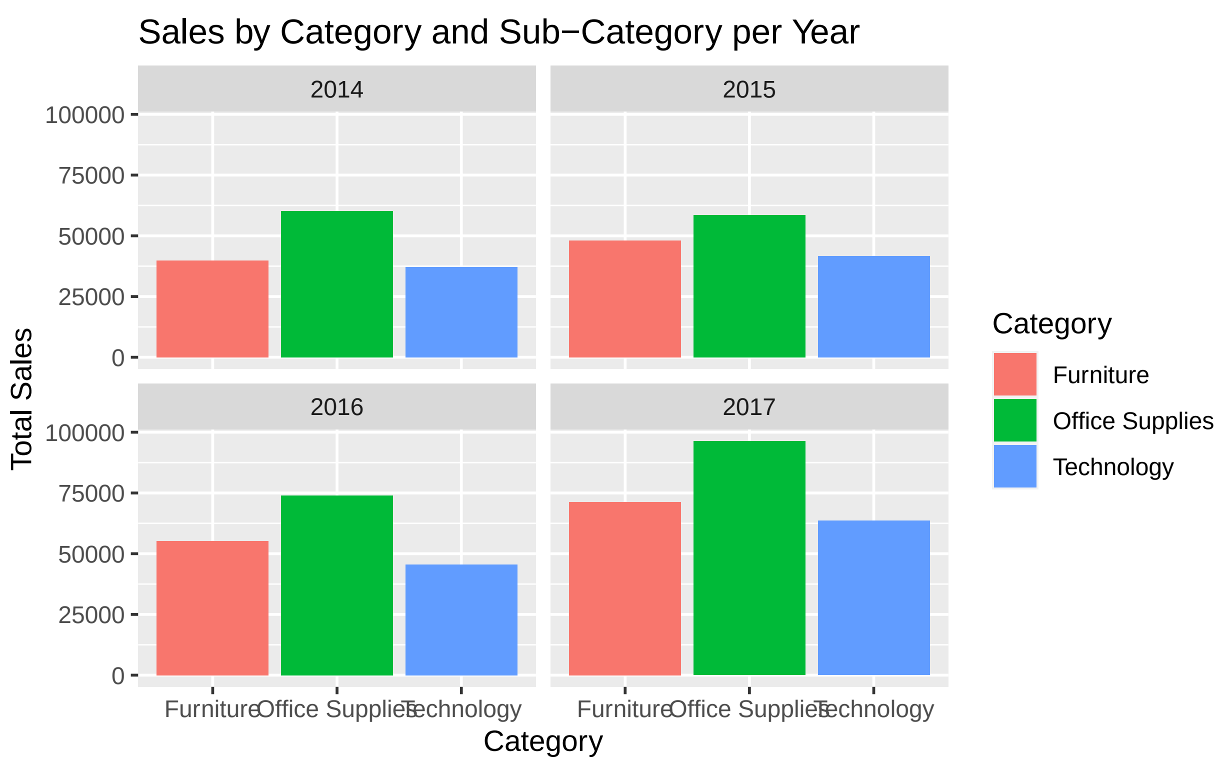
The number of sales presents an increasing trend over years, getting each year more sales than the previous.

However, higher sales doesn't necessarily mean higher revenues or profits, so we will carry out an analysis to figure which was AEKI's best performing year and what did they do well, so that they can now replicate it.

Dataframe for sales and profit by category subcategory and year

```
AEKI_Category_Year_Sales <- Orders_Products %>%
  na.omit() %>%
  group_by(Year, Category, `Sub-Category`) %>%
  summarize(Total_Sales = sum(Sales, na.rm = TRUE),
            Total_Profit = sum(Profit, na.rm = TRUE), "Number of sales"=n())
```

Sales by Category per Year



In general terms, the sales turnover for product categories we can appreciate an increasing trend too. Most likely, this would be caused by the increase in the number of sales. As in the all-year analysis made in the first part of this report, Office Supplies is the category with the highest sales revenue year by year, followed by the Furniture and Technology categories. Between 2014 and 2015, sales turnover for all 3 categories remain more or less the same, with a very scarce growth. However, in 2016 sales rose significantly for all 3 categories, but with a much higher increase for Office Supplies, which kept rising until 2017, becoming the category leader with approximately 25,000€ difference with the second best performing category (Furniture).

Yearly contribution of Sales Data preparation:

```
AEKI_Summary_Per_Year <- Orders_Products %>%
  arrange(Year) %>%
  group_by(Year) %>%
  summarize(
    Total_Sales = sum(Sales, na.rm = TRUE),
    Max_Sales = max(Sales, na.rm = TRUE),
    Min_Sales = min(Sales, na.rm = TRUE),
    Avg_Sales = mean(Sales, na.rm = TRUE),
    Total_Profit = sum(Profit, na.rm = TRUE),
    Max_Profit = max(Profit, na.rm = TRUE),
    Min_Profit = min(Profit, na.rm = TRUE),
    Avg_Profit = mean(Profit, na.rm = TRUE)
  ) %>%
  mutate(
    Sales_Percent_Variation = (Total_Sales - lag(Total_Sales))/lag(Total_Sales) * 100,
    Profit_Percent_Variation = (Total_Profit - lag(Total_Profit))/lag(Total_Profit) * 100
  )
```

Results of preparation:

Year	Total_Sales	Max_Sales	Min_Sales	Avg_Sales	Total_Profit	Max_Profit	Min_Profit	Avg_Profit	Sales_Percent_Variation	Profit_Percent_Variation
2014	137076.8	493.430	0.852	80.02148	20673.40	98.1396	-74.9526	12.06854	NA	NA
2015	148387.6	494.376	0.984	82.80560	24918.10	99.9408	-73.8192	13.90519	8.251458	20.532153
2016	174405.5	496.860	0.836	80.00254	27294.59	99.4896	-73.7061	12.52046	17.533737	9.537234
2017	231254.8	499.584	0.444	80.15766	34356.84	99.9012	-74.7456	11.90878	32.596045	25.874177

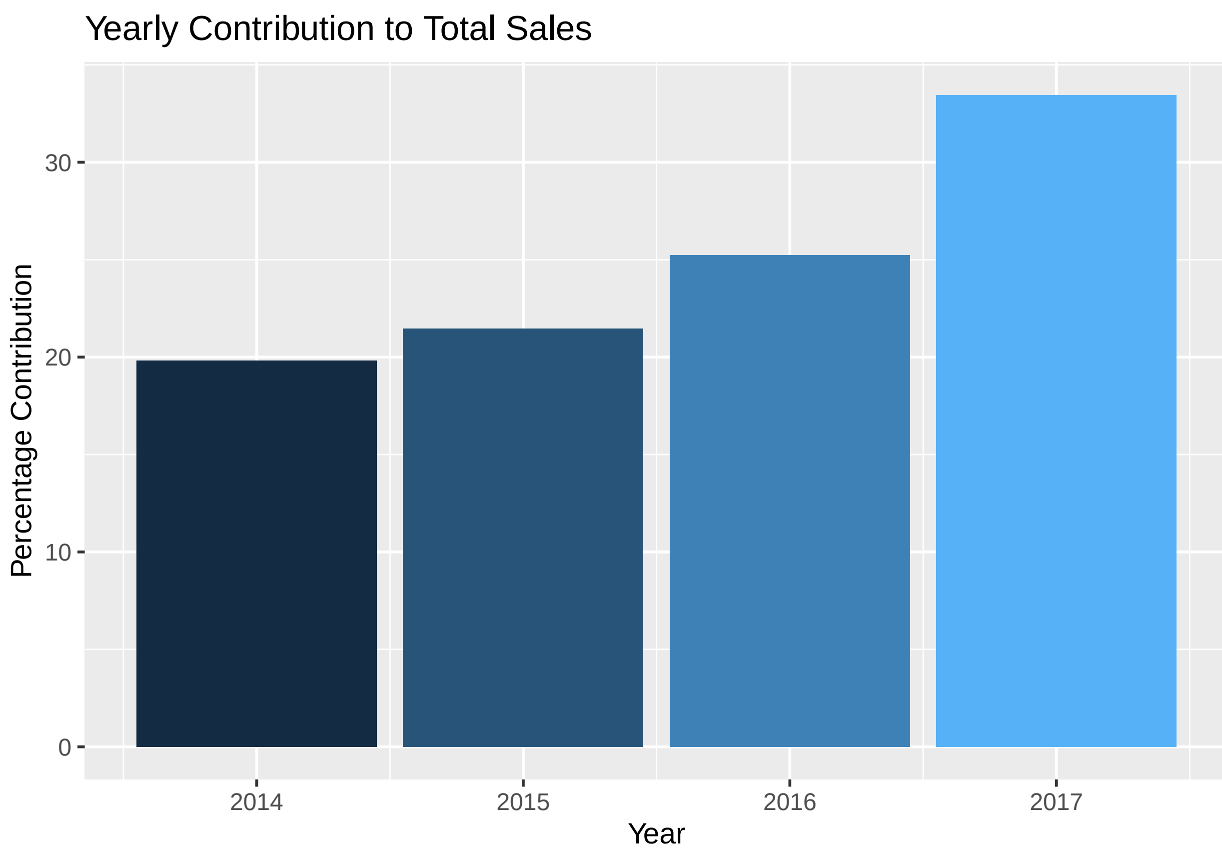
From this new dataframe we can extract that, from 2014 to 2017, the maximum, minimum and average sales revenue and profits per year, are pretty similar.

On the other hand, the only 2 variables showing a clear change over years are total sales and total profit. We can see how both sales and profit increase year by year. 2017 was the year with the biggest growth, with a 30.47% increase in sales and a 22.53% increase in profit, making it AEKI's best financial year.

Percentage of Total Sales per year based on sum of all time sales: Data frame:

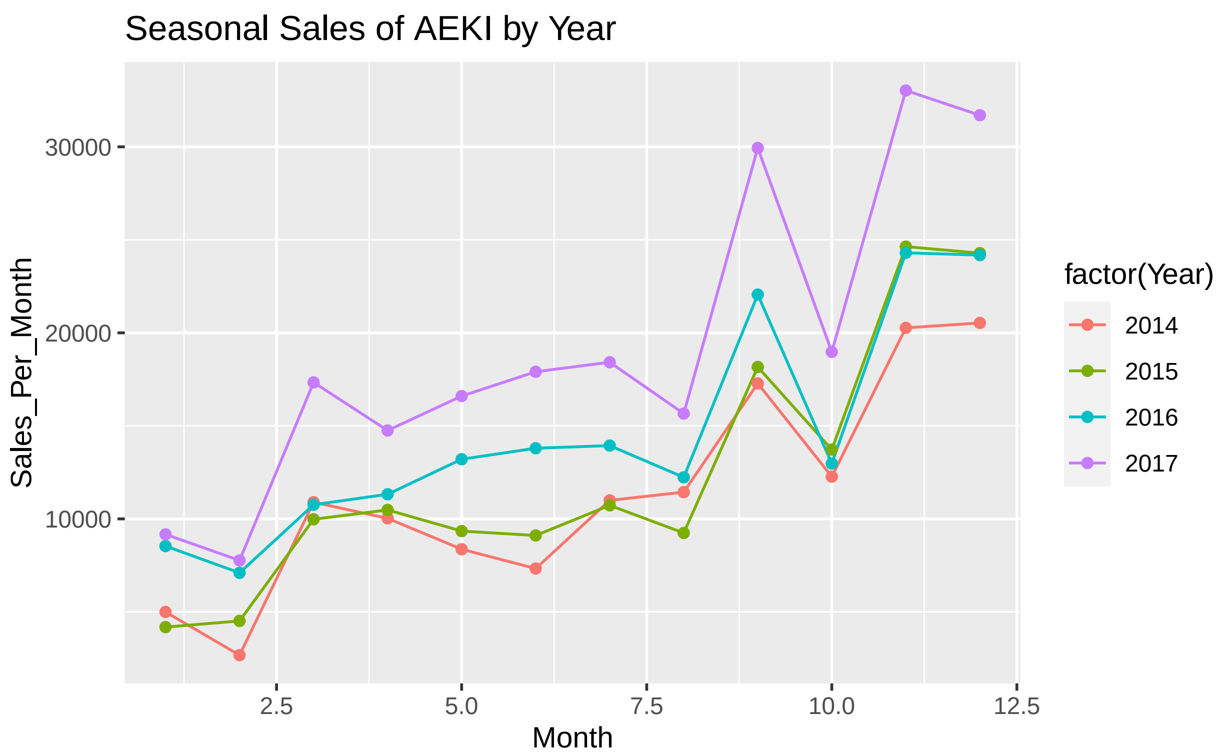
```
AEKI_Yearly_Contribution <- AEKI_Summary_Per_Year %>%
  transmute(Year,
            History_Sales_Contribution=Total_Sales/sum(Total_Sales)*100,
            History_Profit_Contribution=Total_Profit/sum(Total_Profit)*100)
```

We create this data frame to calculate the percentage contribution of each year for both historic total sales and profit.



As expected, 2017 was the year that contributed a higher percentage to AEKI's total sales for the period 2014-2017, with the 33.37% of them. The big jump in sales contribution for the year 2017 could be caused by the increasing demand of phones and technological products, which made AEKI have way more sales.

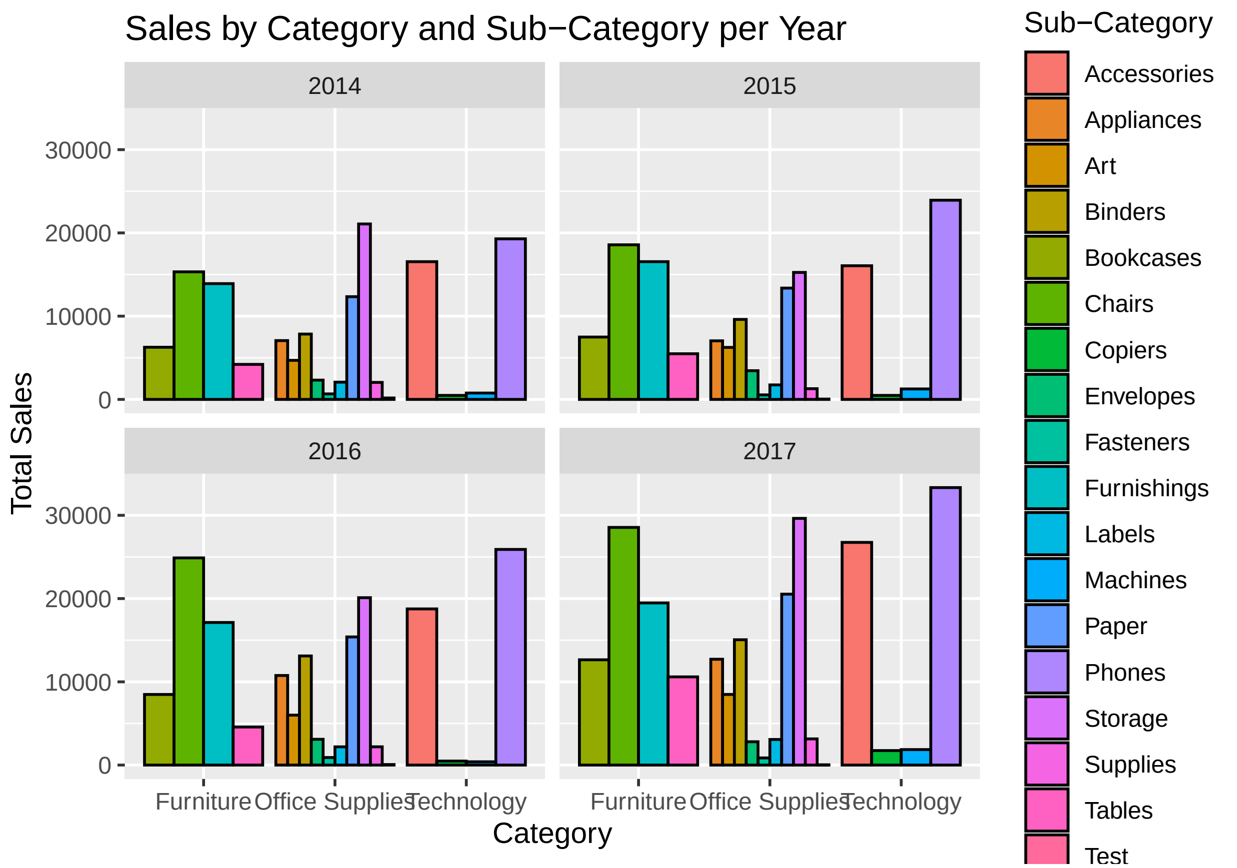
Seasonal and Monthly trends for Sales



We can appreciate that AEKI has more sales as months go by. This is probably caused by seasonal sales or special offers, that tend to take place at the year, such as Black Friday or Christmas. Also, as AEKI sells products that are mainly used by people at work or at school, their peaks in sales matches with the return of summer holidays at the end of the year.

From this data visualization we can make 2 valuable conclusions: AEKI should try to better maintain their number of sales at the beginning of the year and lengthen their high-sales season, and, also, they should fix their big slump in sales in October, which happens every year.

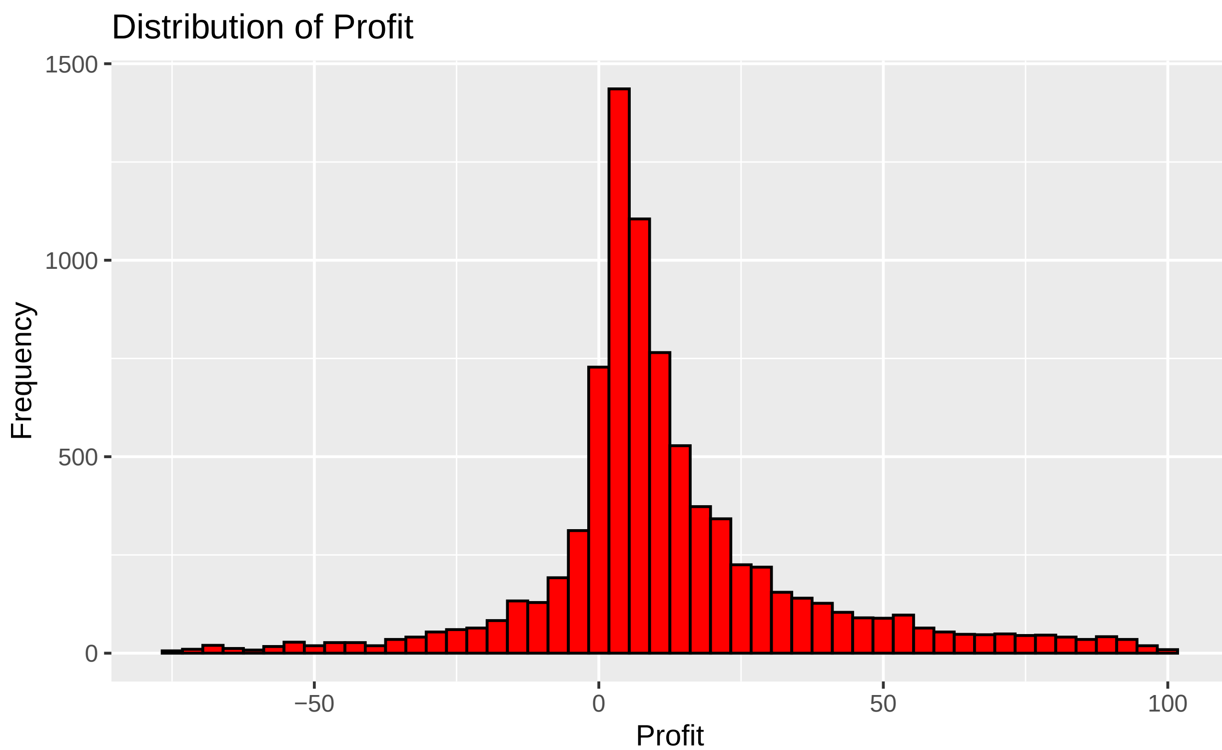
Sales by Category and Sub-Category per Year



Over years, as for categories, subcategories have also kept growing in sales revenue, bably due to the increase in the scale of AEKI's operations year by year. In addition, the most valuable information coming from this graph is the massive increase in Phones' sales revenue in only 3 years. This shows the power of innovation of new technologies and how society adopts these new technological products.

PROFIT

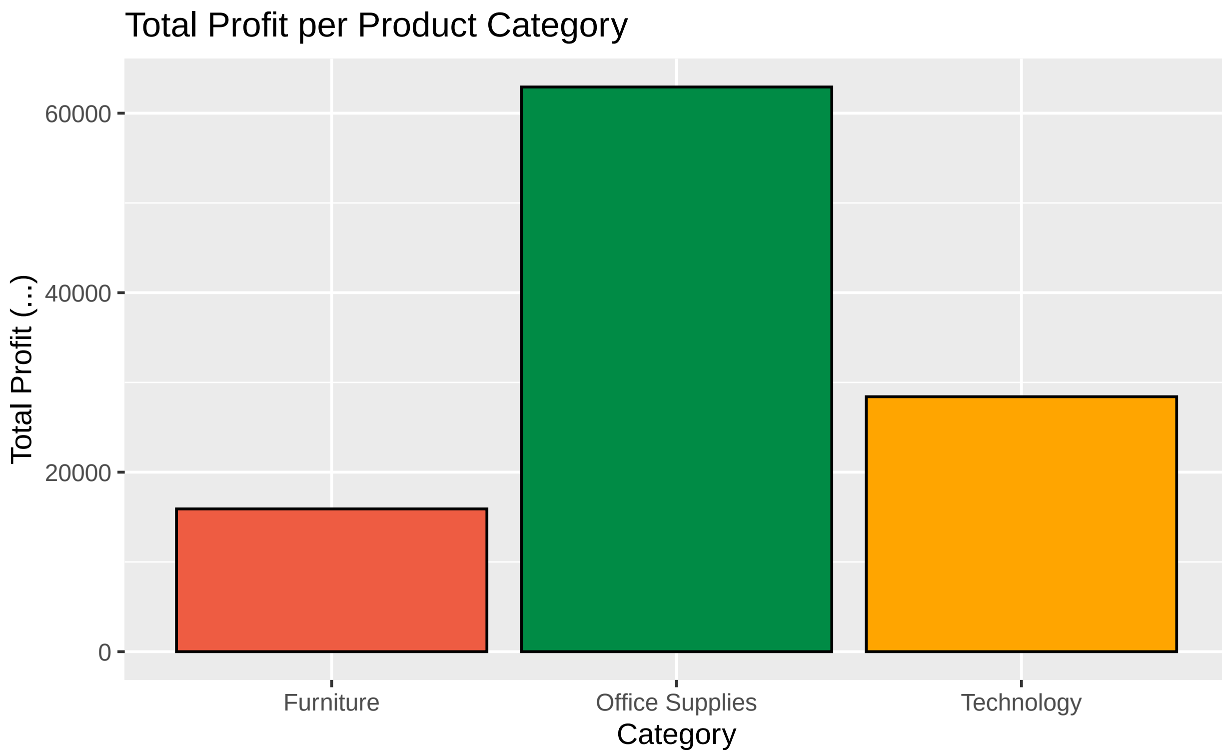
Profit Visualizations and Analyzis



This Profit distribution histogram, is a bell shaped histogram that show that most common values for the profit earned goes from (-5 to 20).

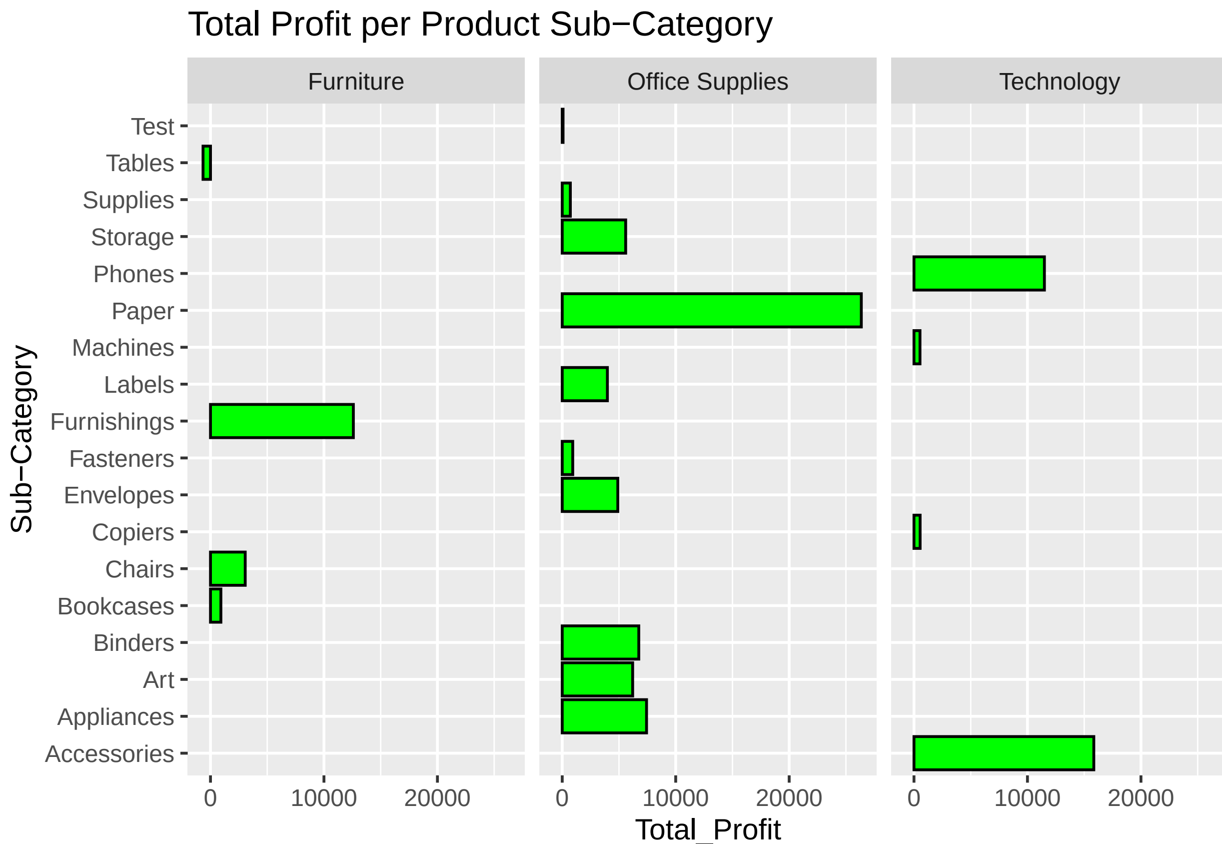
Due to the amount of values a median was used and this way we can see that the most frequent amount of profit earned gather together around the median which lies in the value of 7.25 EUR profit.

Visual analysis of profit by category



We can see that the most profitable product category is Office Supplies too, with 60,174.87€ of Profit over Sales. On the other hand, the Furniture Category passes to be the least profitable of the 3, with only 27,590.33€ of Profit.

Visual analysis of profit by subcategory

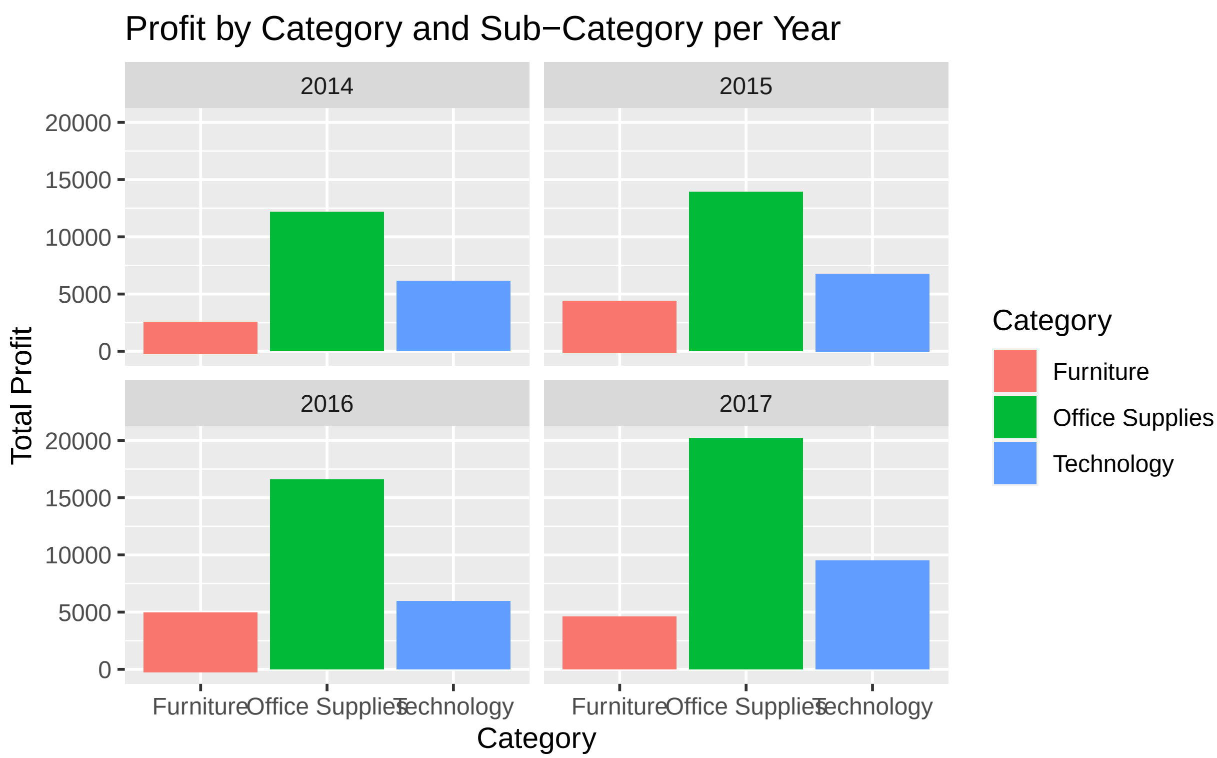


These graphs show the distribution of profits by subcategory. The product that generates the most profit in the Furniture category is Furnishings, in Office Supplies it's the paper and in Technology, the most profitable product is Accessories. We can suggest to AEKI to produce more of these projects to generate more profit.

The product that generates the least profit in the Furniture category is Tables, in Office Supplies it's the Tests and in Technology it's the Machines and Copiers. This confirms why we saw that these are the least selling products in the last analysis. We confirm that AEKI should seriously consider stop producing these products.

Profit analysis for AEKI Orders - BREAKDOWN BY YEAR

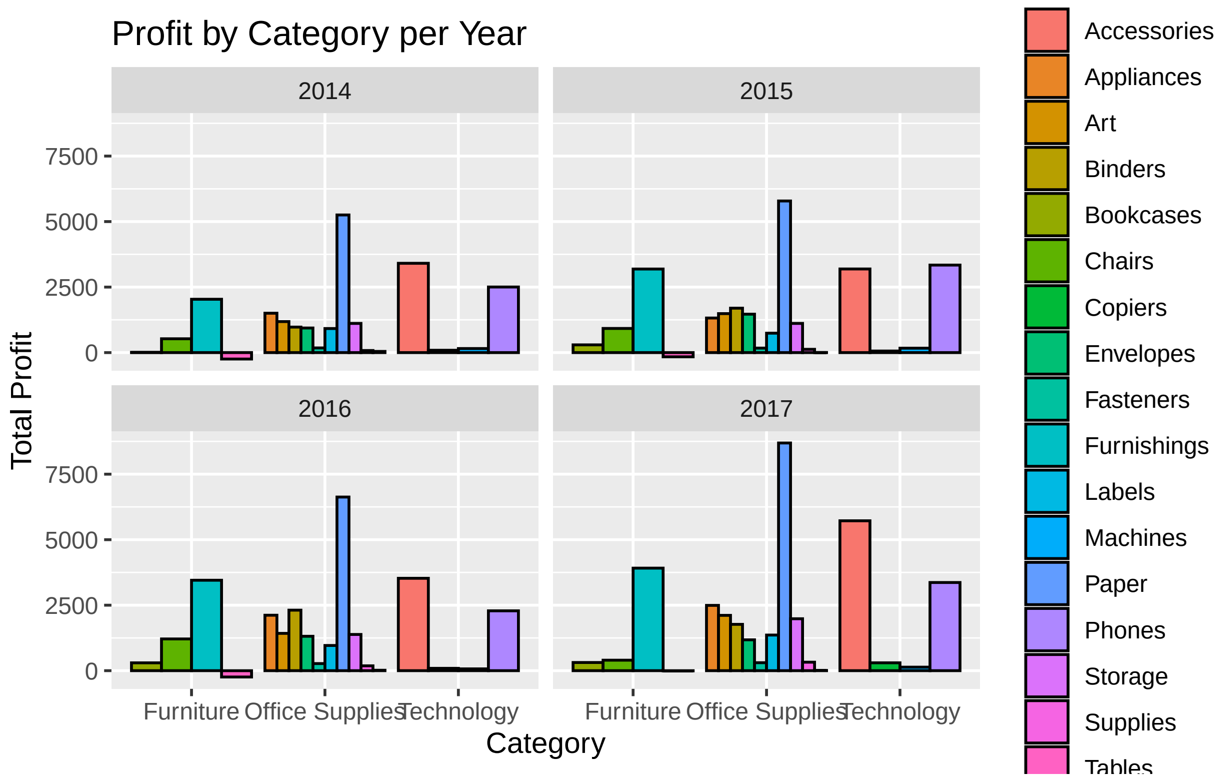
Profit by Category per Year



Profit also presents a general increasing trend over time. For 2014 and 2015, profit per category remains almost the same, except for Furniture which significantly increased, percentually, compared to the other 2 categories.

Then, in 2016, profitability of Office Supplies rose, but the Furniture and Technology categories got almost stuck.

Lastly, in 2017, Office Supplies and Technology had a really big growth in profitability, but, surprisingly, the Furniture category had a slight decrease. Therefore, as the sales for the Technology category also had a big increase in sales turnover, we can assume that, since 2017, consumer wants started to change and technological products started to have an increase in demand. Furthermore, the increase in profitability of technological products could come from the high prices charged on these kind of products, as people will still be willing to pay for them.

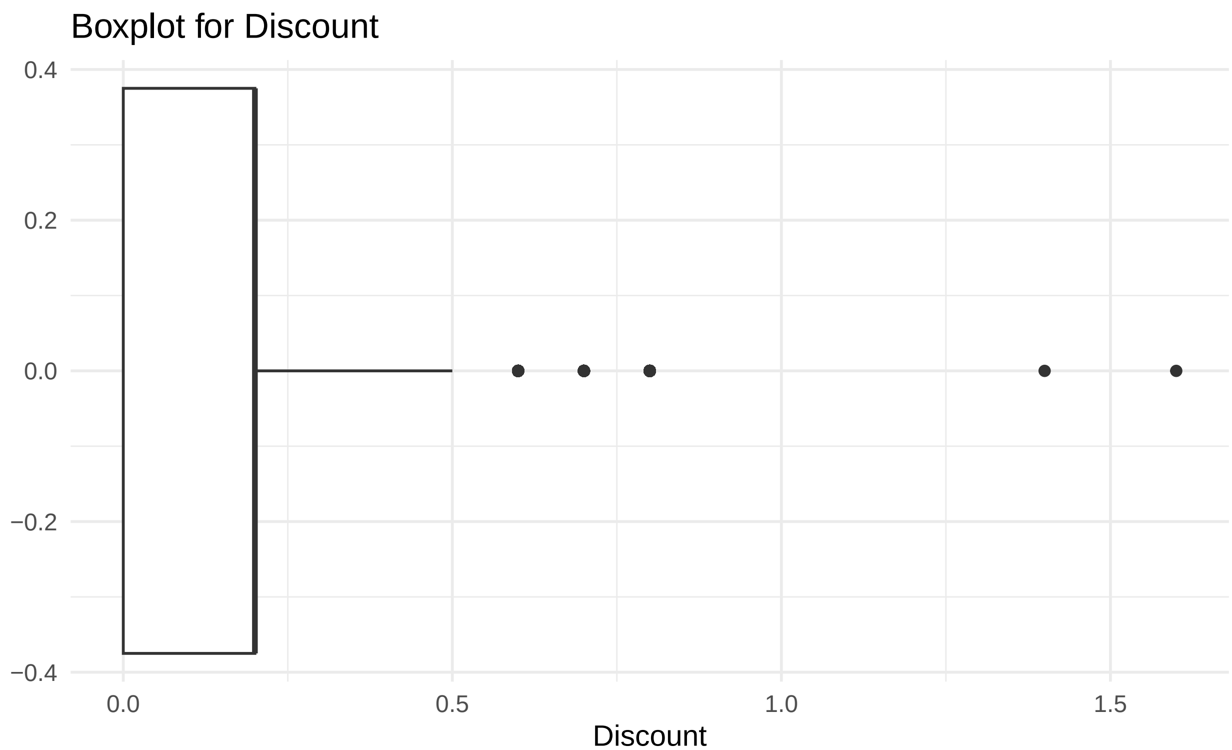


For profits, subcategories in general haven't changed a lot over years, although they show an slight increase. However, Paper and Accessories have had a huge increase in profitability since 2014.

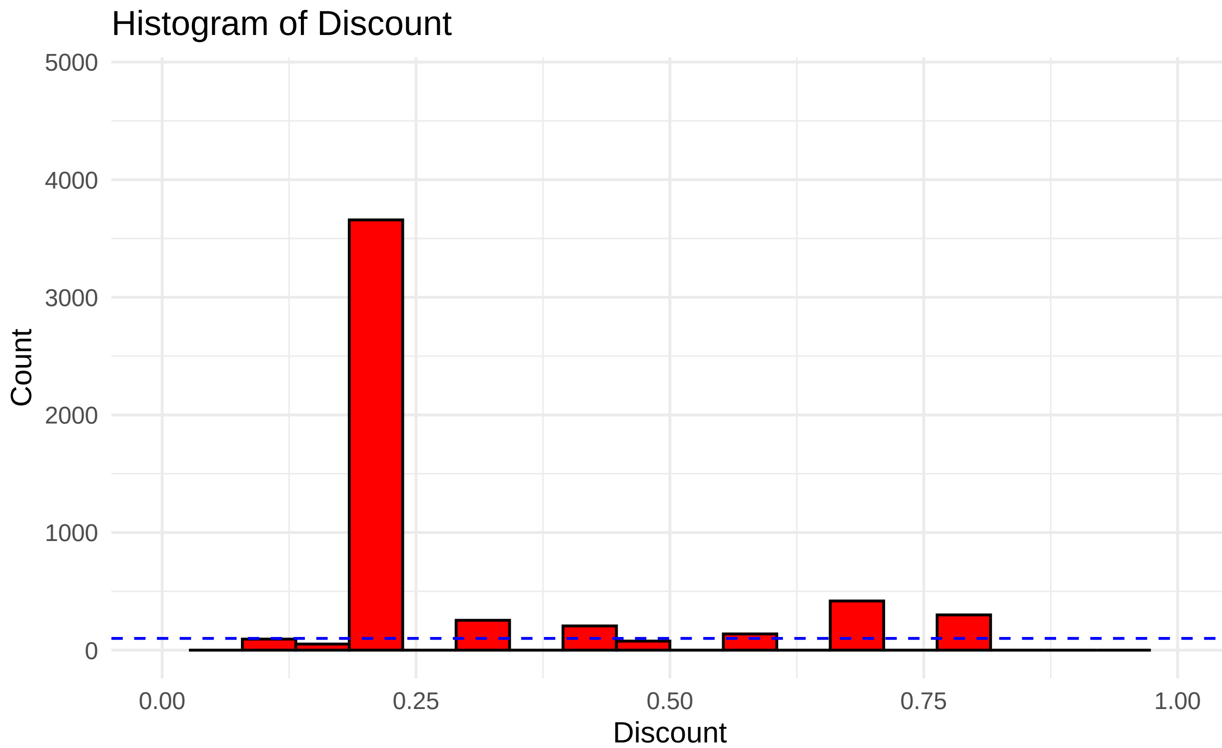
This big increase could be explained by the low prices set on these products, what lead to higher sales as more people can afford; the high profit margins, as the costs of production are minimum; and, probably, the high frequency of purchase, as they run out and need to be bought again.

Discount

VISUALISING AND ANALYSING OUTLIERS FOR DISCOUNT

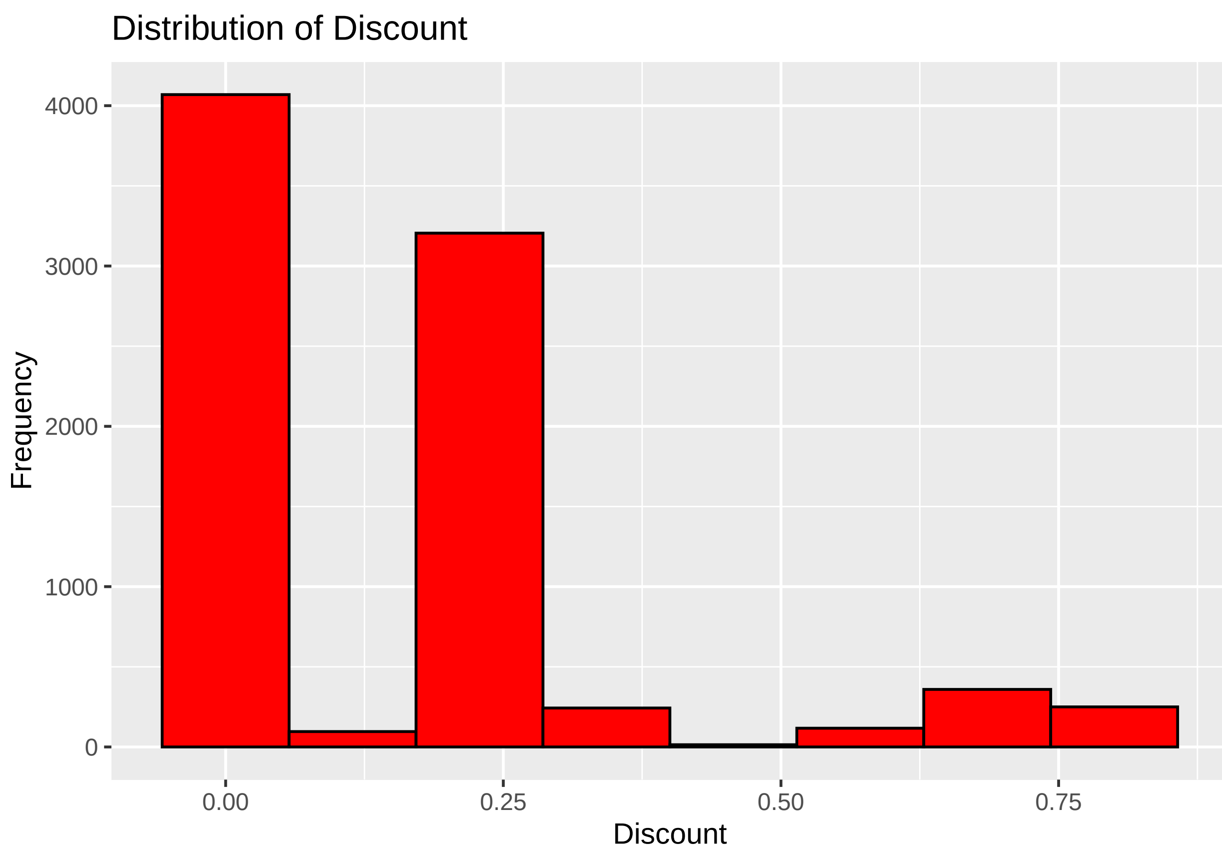


According to the boxplot, Discount > 0.5 are outliers



By plotting, we deduce that the range of Discount 0 to 1 is the correct one, as the Count for those Discounts that are mathematically considered outliers (>0.5), follows our condition of Count>100. Although, according to the outlier formula, every Discount greater than 0.5 is an outlier, we think that every discount equal or lower than 1 should be accepted as they are plausible. Therefore, the only observations that we consider to be outliers are those with Discount>1 (1.6 & 1.4)

Univariate Analysis on Discount



This discount distribution histogram with an irregular shape is showing that the most used discount is either NO discount or a 20% discount.

Which are the products and product categories / subcategories in which we applied more discount in relative and in absolute values?

For PRODUCTS

```
Group_by_Product_ID <- Orders_Products %>%
  na.omit() %>%
  group_by(`ID Product`) %>%
  summarise(count = n(), Mean_Price_Per_Unit = mean(Price_Per_Unit), Relative_Discount = mean(Discount)
  arrange(desc(Relative_Discount))
```

The Product IDs with the highest mean relative discount applied is OFF-AP-10002203 with 0.8 discount (80%). The Product ID with the highest mean absolute discount applied is FUR-TA-10003238 with 150.42€ discount applied.

For Poduct Categories

```
Group_by_category <- Orders_Products %>%
  na.omit() %>%
  group_by(Category) %>%
```

```
  summarise(count = n(), Mean_Price_Per_Unit = mean(Price_Per_Unit), Relative_Discount = mean(Discount)
  arrange(desc(Relative_Discount))
```

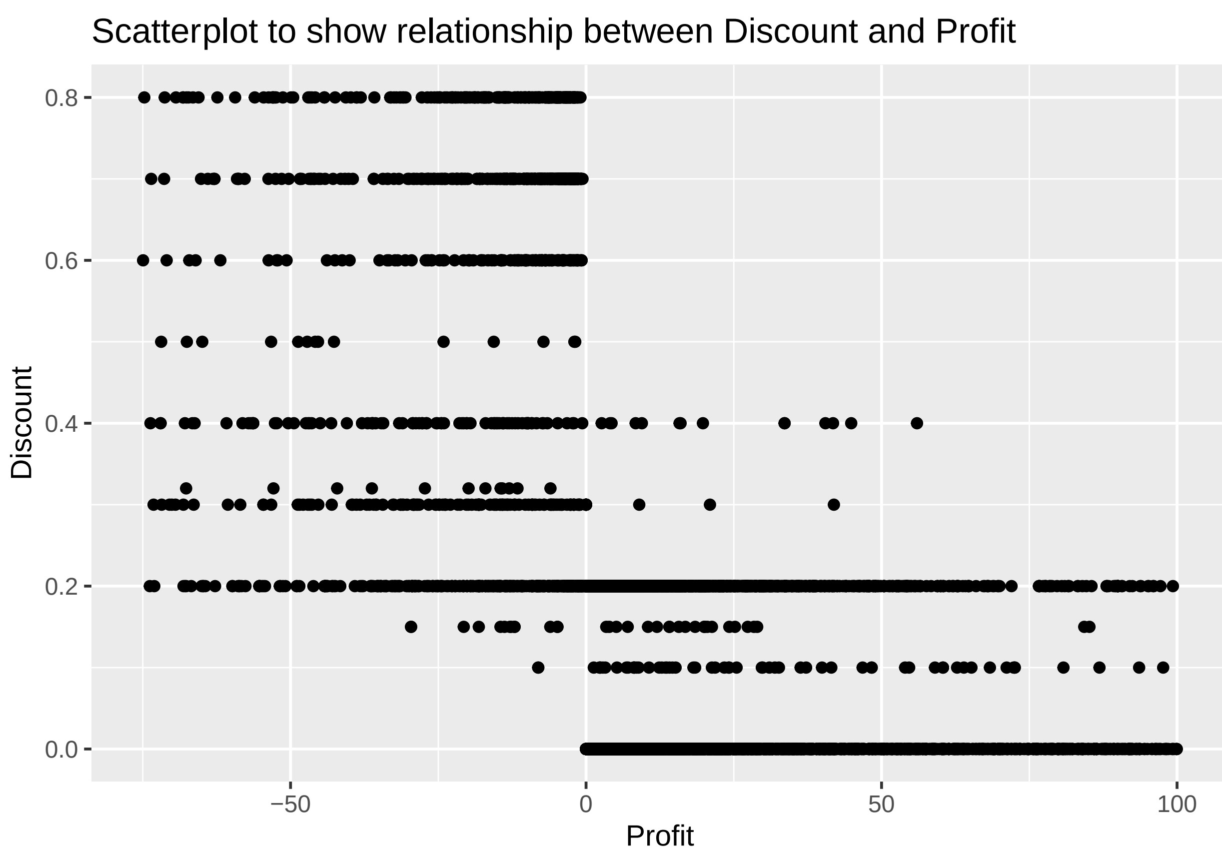
The product category with the highest mean relative discount applied is the Office Supplies with 0.1568 discount (15.68%). The product category with the highest mean absolute discount applied is Furniture with 9.3€ average discount on its products.

For Product Subcategories

```
Group_by_subcategory <- Orders_Products %>%
  na.omit() %>%
  group_by(`Sub-Category`) %>%
  summarise(count = n(), Mean_Price_Per_Unit = mean(Price_Per_Unit), Relative_Discount = mean(Discount)
  arrange(desc(Relative_Discount))
```

The product subcategory with the highest mean relative discount applied is the Binders with 0.3781 discount (37.81%) The product subcategory with the highest absolute discount applied is Copiers with 49.66€ average discount on its products.

Relationship between Discount and profit

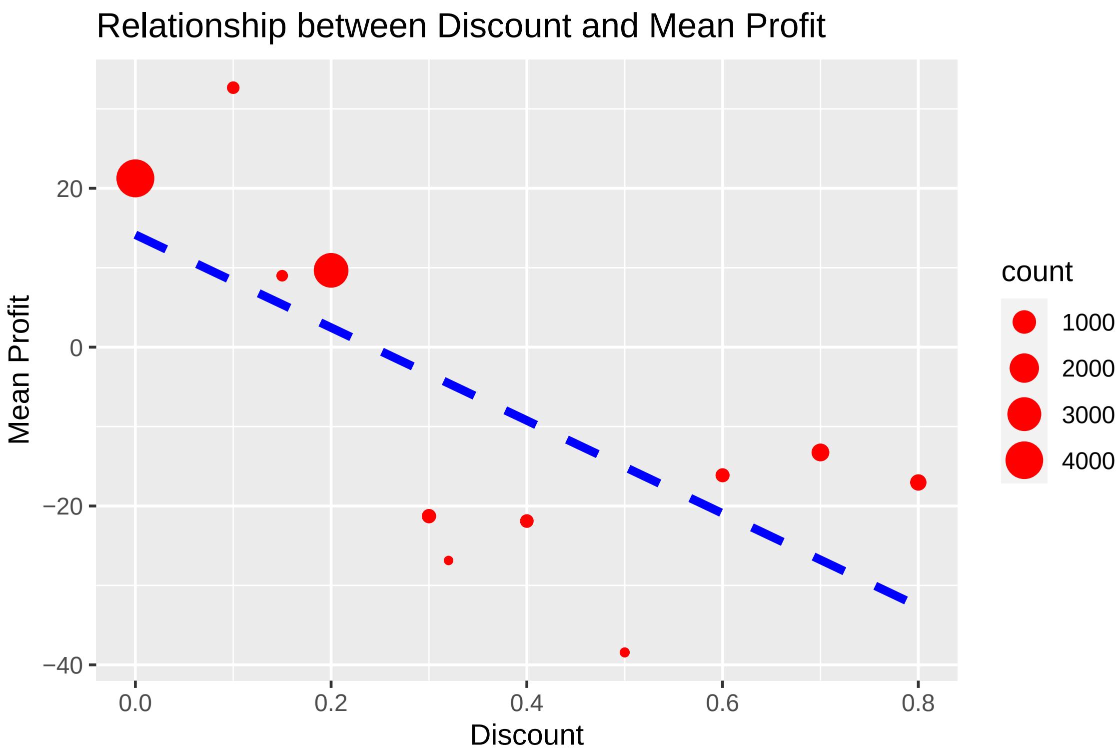


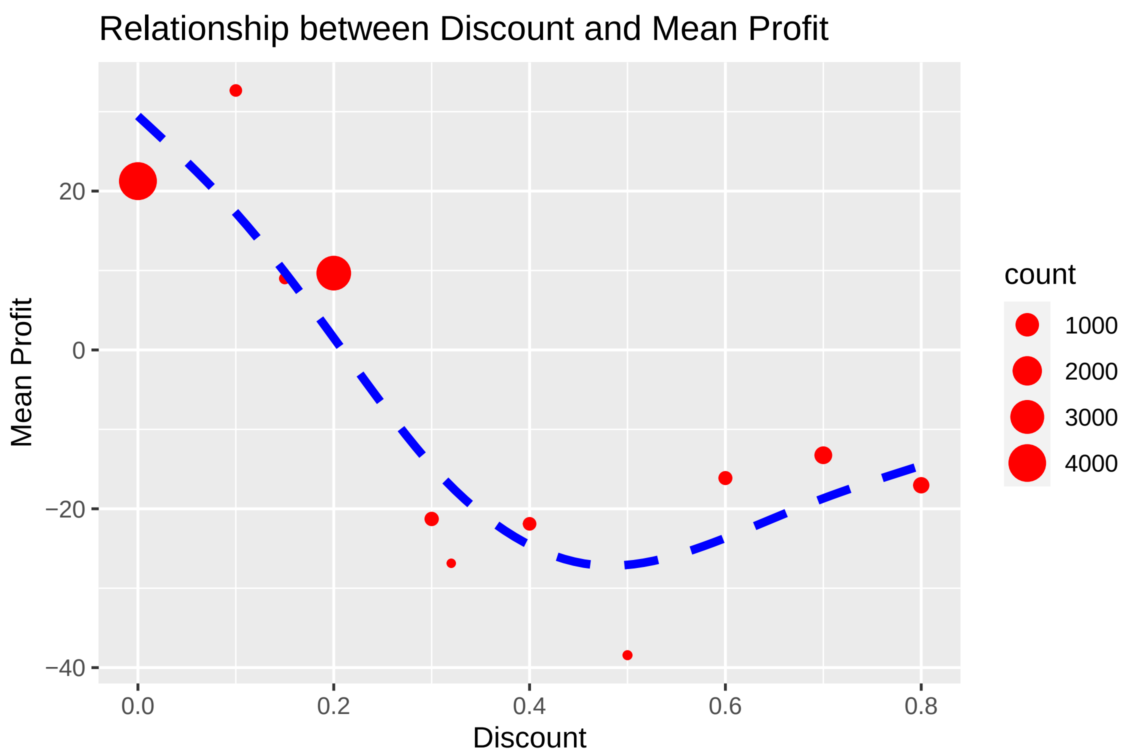
No clear pattern or relationship can be derived thus further analysis will be needed.

Code to find mean profit for each amount of discount applied

```
discount_profit_grouped <- Clean_AEKI_Orders %>%
  select(Discount, Profit) %>%
  group_by(Discount) %>%
  summarise(Mean_profit = mean(Profit), count=n())
```

Illustrating the data





For better analysis we have plotted both types of method ("gam" and "lm"). Both methods show a clear descending trend, what means that higher the Discount applied, lower the Mean Profit achieved. However, the linear model (lm) represents better this tendency as it is not influenced by the Count variable.

Analyzing correlation between variables (Discount and Profit)

```
cor(Clean_AEKI_Orders$Discount, Clean_AEKI_Orders$Profit)
```

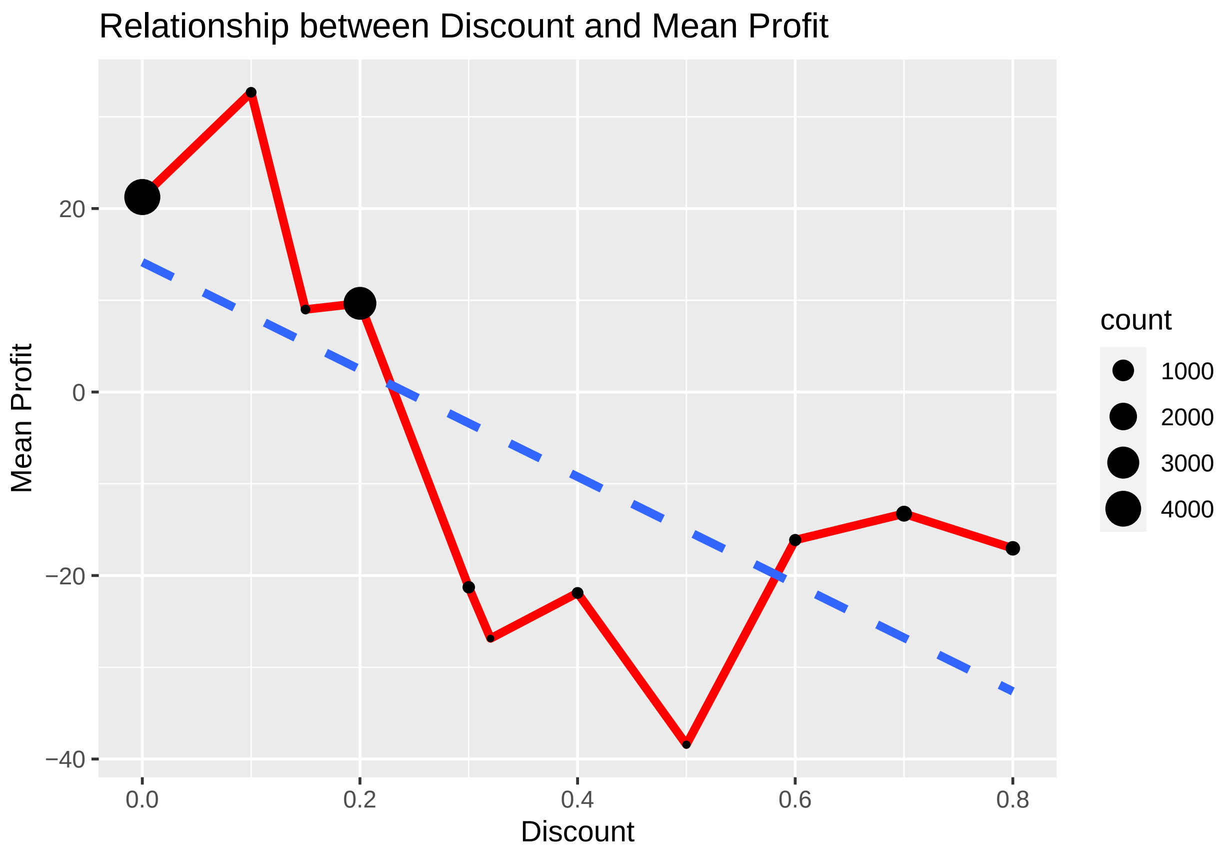
```
## [1] -0.4622241
```

-0.4665864 correlation.

The negative value of correlation between AEKI's discount and profit means that as one of the variables increases, the other decreases. However, the value is relatively low, closer to 0 than 1, so it will only cause a moderate decrease in comparison to the increase. However, we cannot assume that one causes the other from this correlation alone. But, as both variables are closely related to demand and revenue, we can interpret that Discount is the cause for the fluctuation in profit.

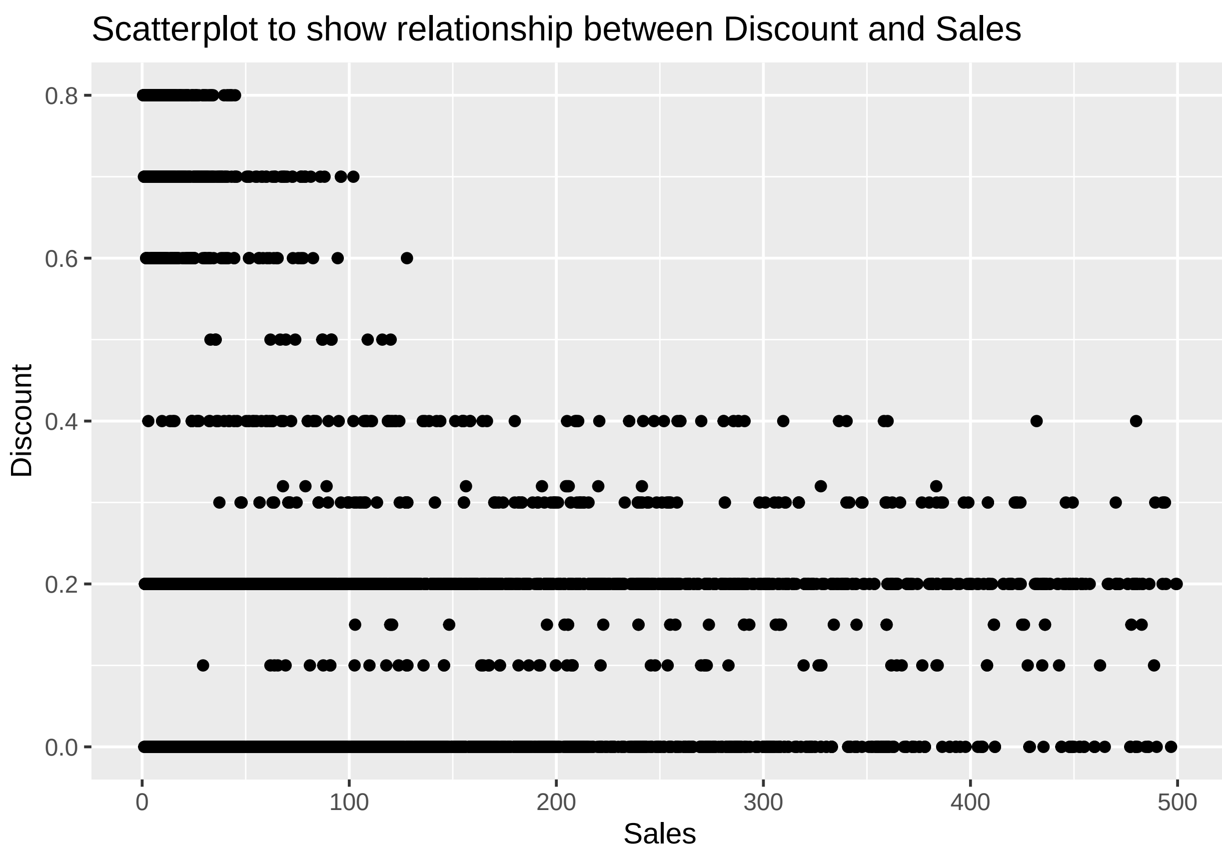
In conclusion, it is likely that discounts may lead to higher sales, but reduced profit margins.

Further illustration of relashionship



With this plot, we can better appreciate the fluctuation of the Mean Profit as the Discount applied grows.

Relationship between Discount and Sales

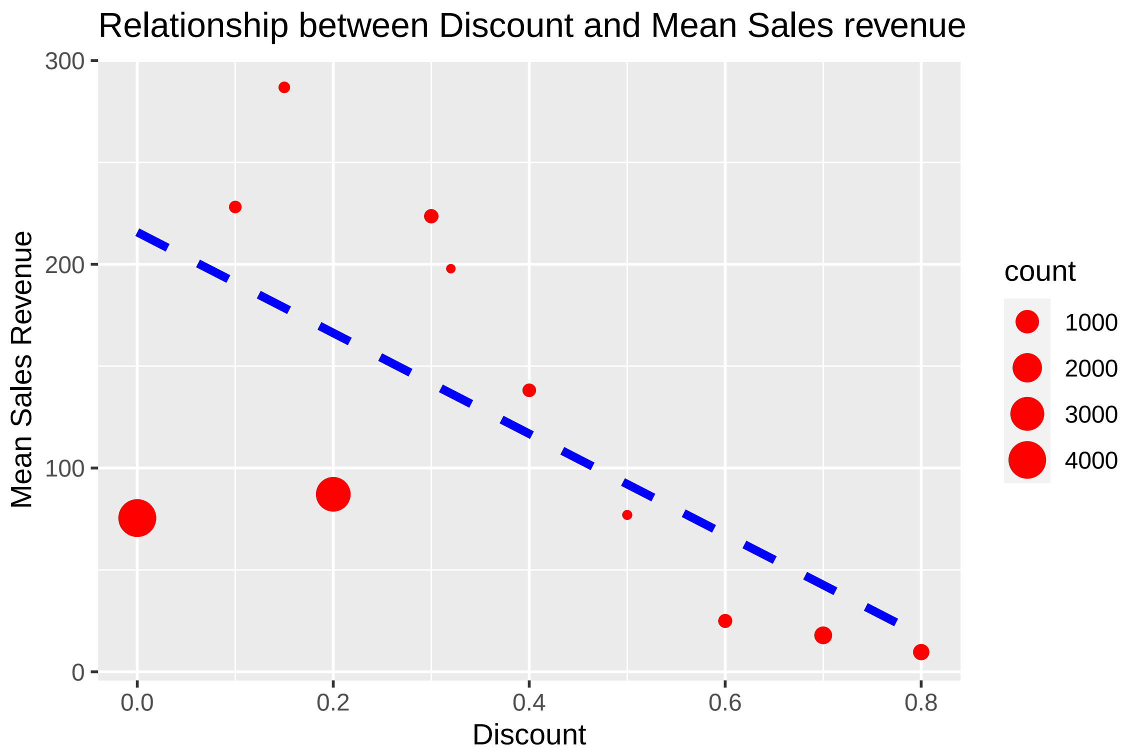


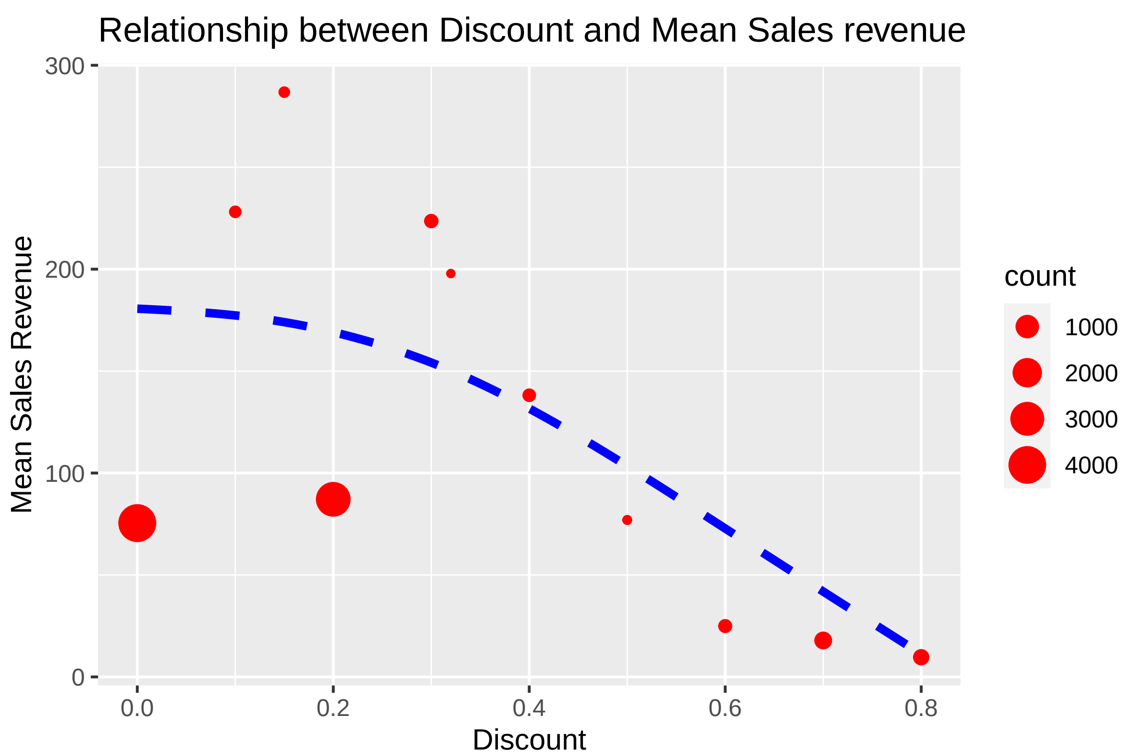
Not very useful as no clear pattern can be derived.

Code to find mean sales (revenue) for each amount of discount applied

```
discount_sales_grouped <- Clean_AEKI_Orders %>%
  select(Discount, Sales) %>%
  group_by(Discount) %>%
  summarise(Mean_Sales_Revenue = mean(Sales), count=n())
```

Illustrating the data





As for the relation Discount-Profit, for better analysis we have plotted both types of method ("gam" and "lm"). Both methods show a descending trend, what means that higher the Discount applied, lower the Mean Sales revenue. However, the linear model (lm) represents better this tendency as it is not influenced by the Count variable. Furthermore, the "gam" method remains steady for the first 20%, what suggest that, from that point on, the grow of the discount will impact heavier on the Mean Sales revenue, while little discounts do not affect dramatically on it.

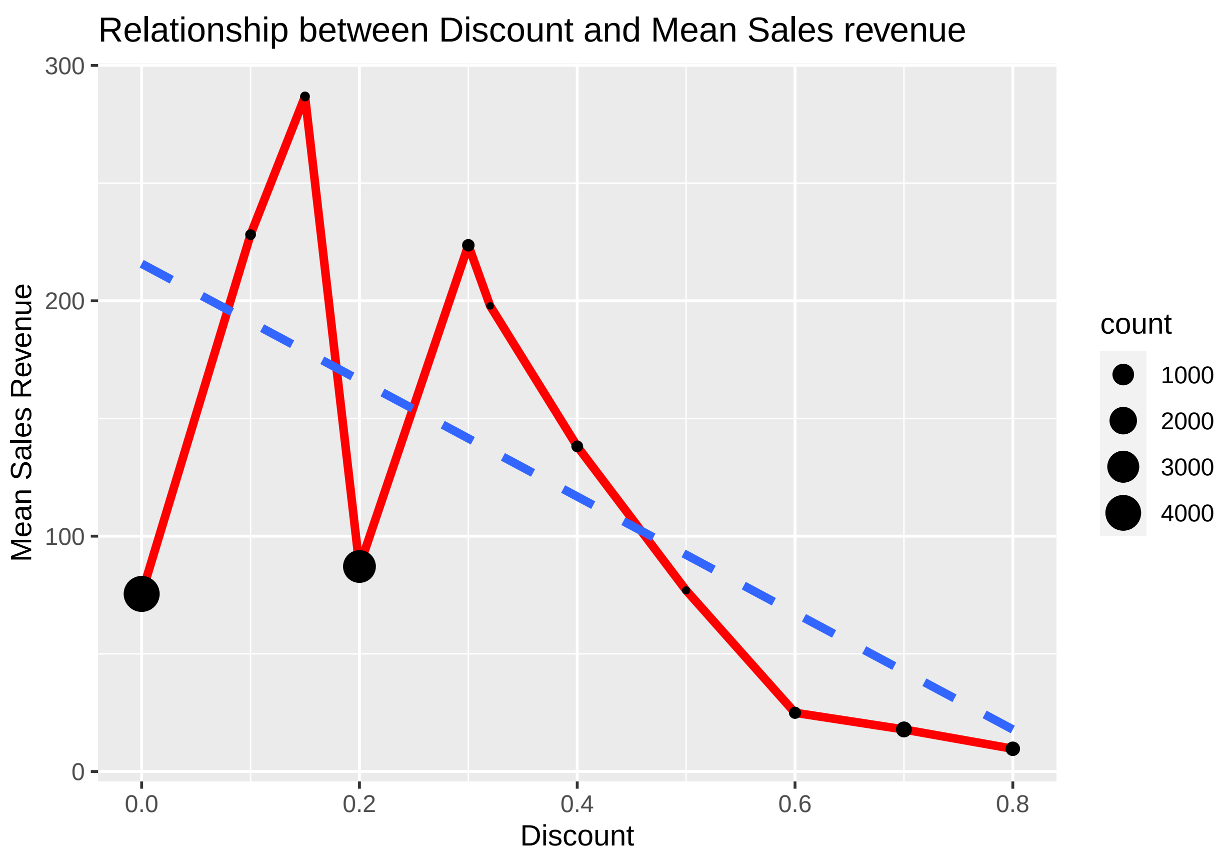
Analyzing correlation between variables (Discount and Sales)

```
## [1] -0.1101005
```

-0.1159012 correlation.

The negative value of correlation between AEKI's discount and sales means that, as one of the variables increases, the other decreases. However, the value is considerably low, really close to 0, so it will only cause a moderate decrease in comparison to the increase. However, we cannot assume that one thing causes the other from this correlation alone. But, as both variables are closely related to demand and revenue, we can interpret that Discount is the cause for the fluctuation in Sales, as equal to profits. In conclusion, it is likely that discounts may lead to higher sales, but lower sales revenue.

Further illustration of relashionship



With this plot, we can better appreciate the fluctuation of the Mean Sales Revenue as the Discount applied grows.

RETURNS ANALYSIS

Merging Clean Orders and Returns data sets:

Cleaned orders dataset with column stating whether returned or not.

```
Orders_Returns <- Clean_AEKI_Orders %>%
  left_join(AEKI_Returns, by = "Order ID") %>%
  mutate(Returned = ifelse(is.na(Returned), "No", Returned))
```

We have created a new dataset by merging AEKI_ORDERS and AEKI_RETURNS by the Order ID.

Furthermore, for the new variable (Returned) added to AEKI_ORDERS by the merge, we have change all NA value for "No" to make it easier for us to work and understand.

Dataset with only Orders which were returned:

```
Returned_orders <- Clean_AEKI_Orders %>%
  left_join(AEKI_Returns, by = "Order ID") %>%
  filter(Returned=="Yes")
```

In this new dataset, we only have the orders and products that have been returned (Returned = Yes).

For the breakdown of returns we will work with the dataset only including returned products.

We are going to analyse each aspect of product returns in order to look for potential improvement areas.

Returns breakdown by product.

```
returns_product <- Returned_orders %>%
  group_by(`ID Product`) %>%
  summarise(count = n()) %>%
  arrange(desc(count))
```

108 products had been returned multiple times, with 13 products being returned 3 times.

555 different products were returned.

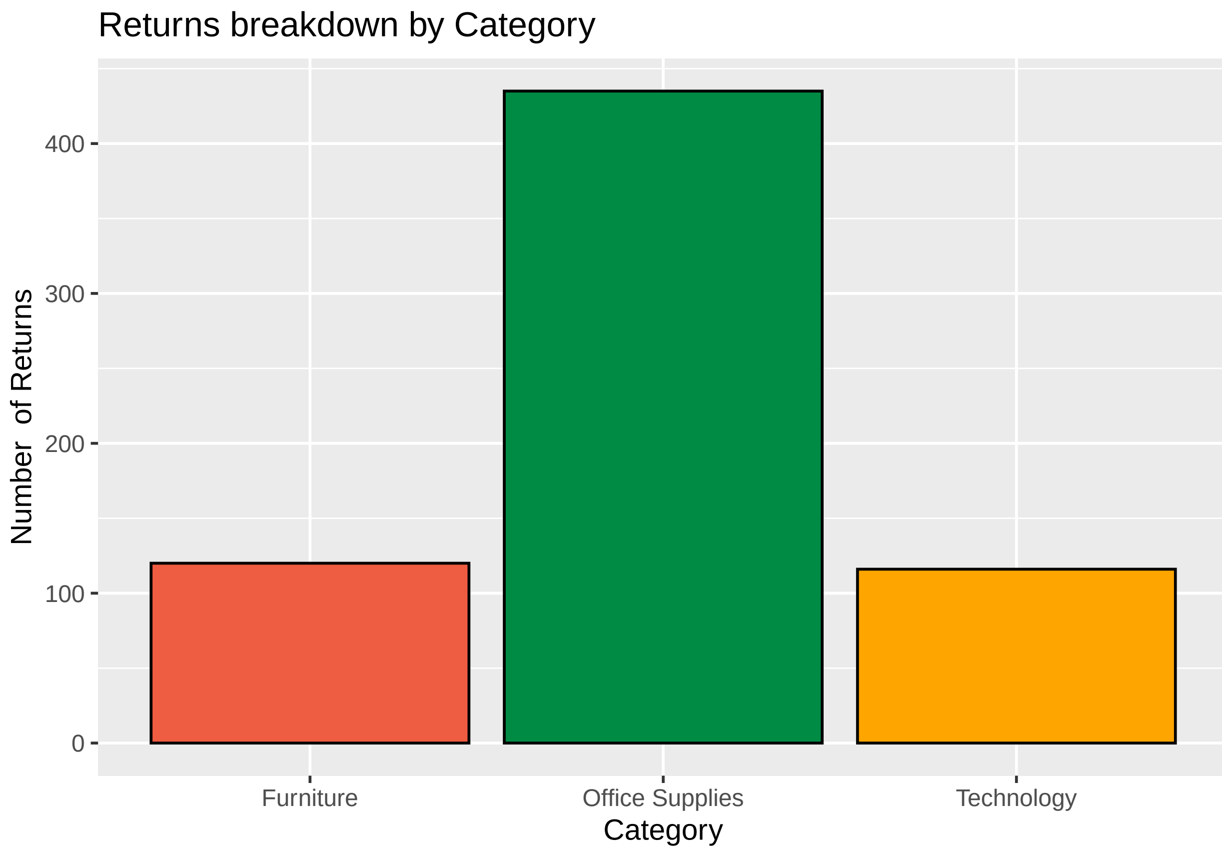
However, trying to graphically represent all these returns is not very useful and hard to visualize. Therefore, we are going to make a Category and subcategory breakdown, which we think that might be much more insightful.

```
Returned_type <- Returned_orders %>%
  left_join(unique_products, by = c("ID Product"))
```

Returns breakdown by category: We have created this new dataset to have the returned products and the category and subcategory to which they belong to, all together.

Count of returns per category:

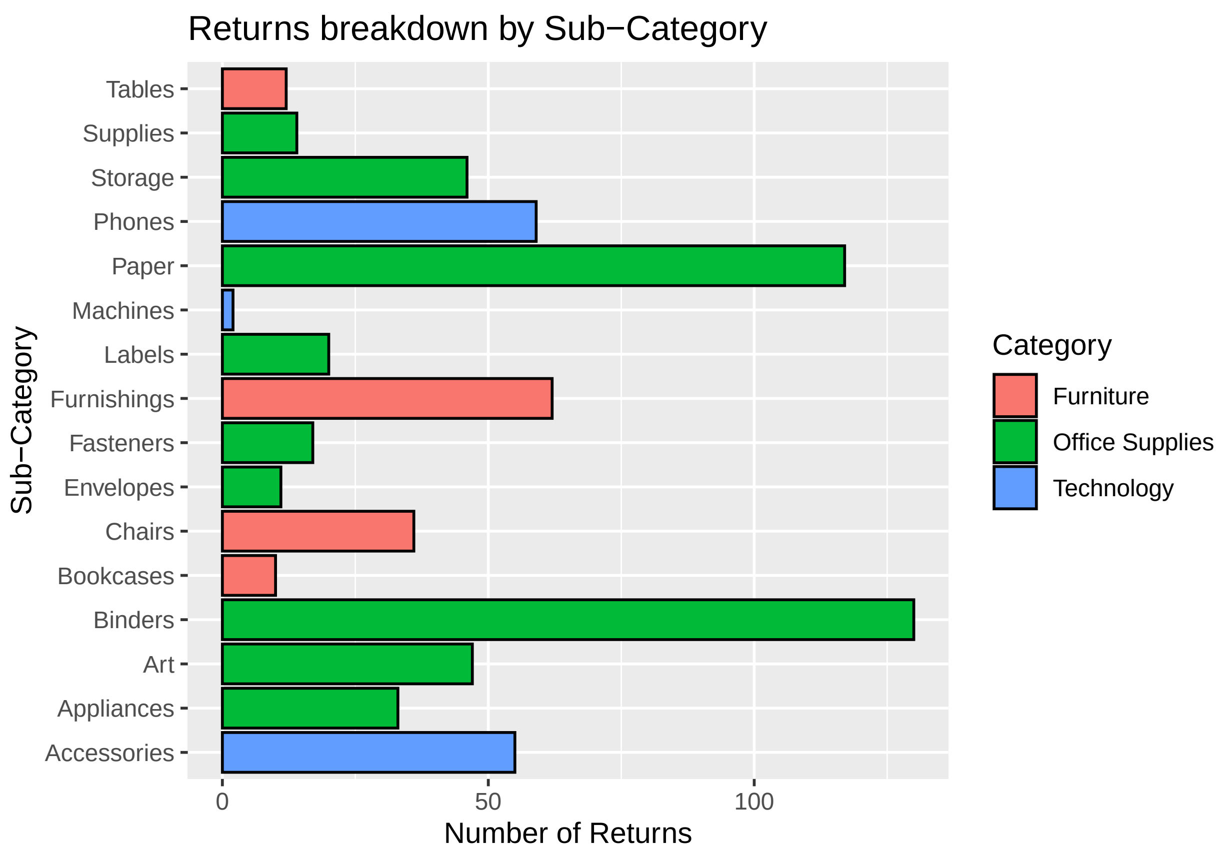
```r
returns_category <- Returned_type %>%
  na.omit() %>%
  group_by(Category) %>%
  summarise(count = n())
```



The most returned product category is Office Supplies with 435 returned products. This was quite to be expected, as it is also the most sold category. In conclusion, the number of returns follows the same rankings as the number of sales per category.

Count of returns per sub-category:

```r
returns_sub_category <- Returned_type %>%
  na.omit() %>%
  group_by(Category, `Sub-Category`) %>%
  summarise(count = n())
```

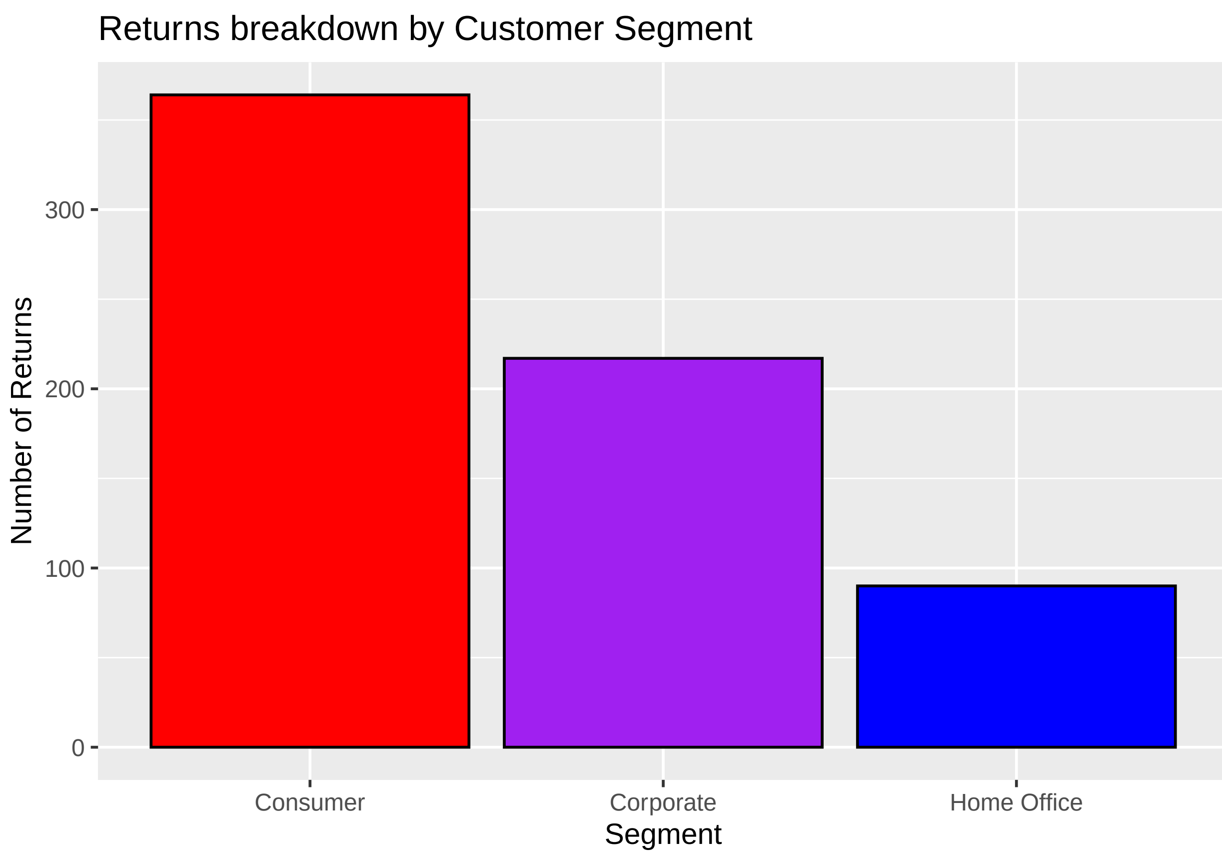


The subcategory with the highest number of returns is Binders with 130 products returned. We can appreciate that 2 subcategories (Test and Copiers) are not represented in this graph, what means that they haven't been returned a single time. In conclusion, we can see that the graph follows the same rankings as the number of sales per subcategory too.

Relationship between returns and customer segments:

```
segment_returns <- Returned_type %>%
  na.omit() %>%
  group_by(Segment) %>%
  summarise("Number of Returns" = n())
```

With this new dataset we get the number of returns per customer segment.

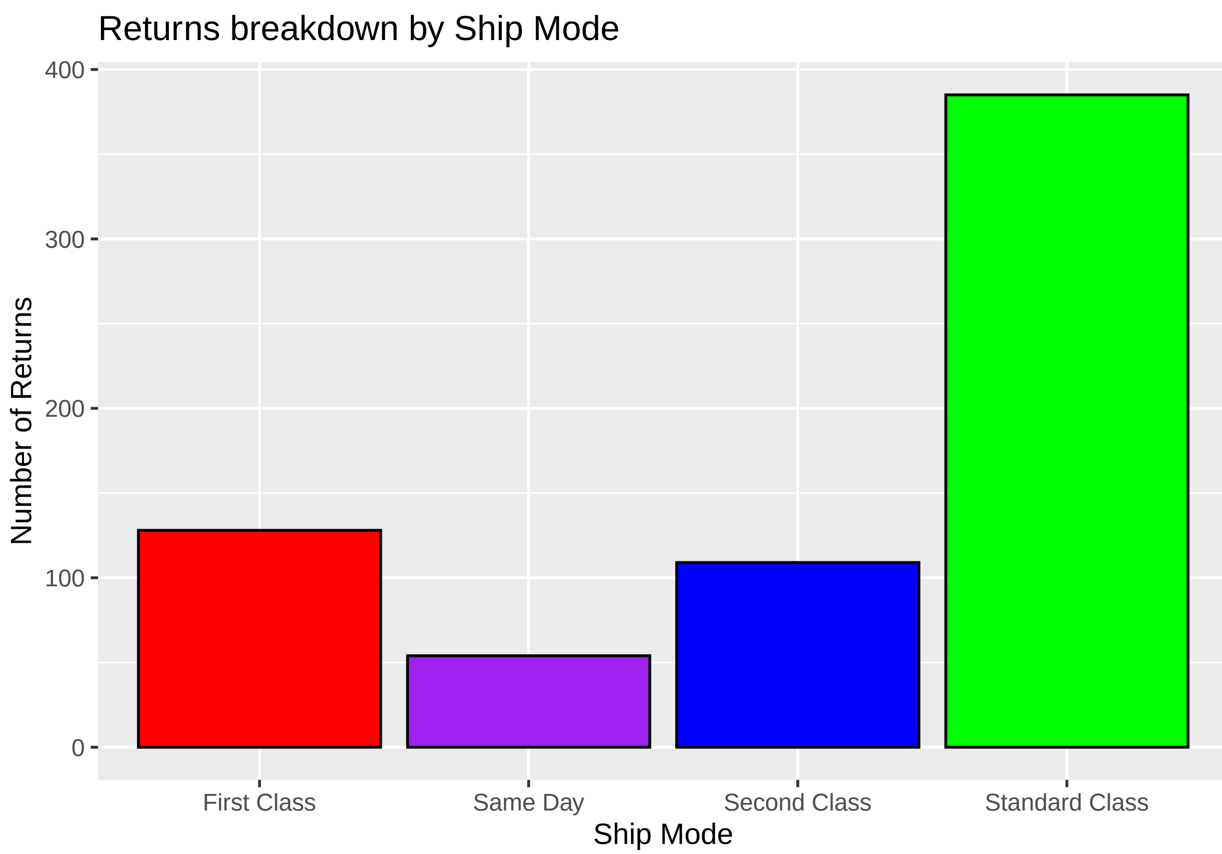


Consumers are the customer segment with the highest number of returns with 364 returned products. This could be due to the fact that they are also the ones with most purchased products.

Also, it could be caused by the types of products that this customer buys. Most likely, consumers will be the ones purchasing products with higher chances of return due to its quality of whatever, but products for corporates or home offices better products might be sold.

Relationship between returns and ship mode:

```
ship_mode_returns <- Returned_type %>%
  group_by(`Ship Mode` ) %>%
  summarise("Number of Returns" = n())
```

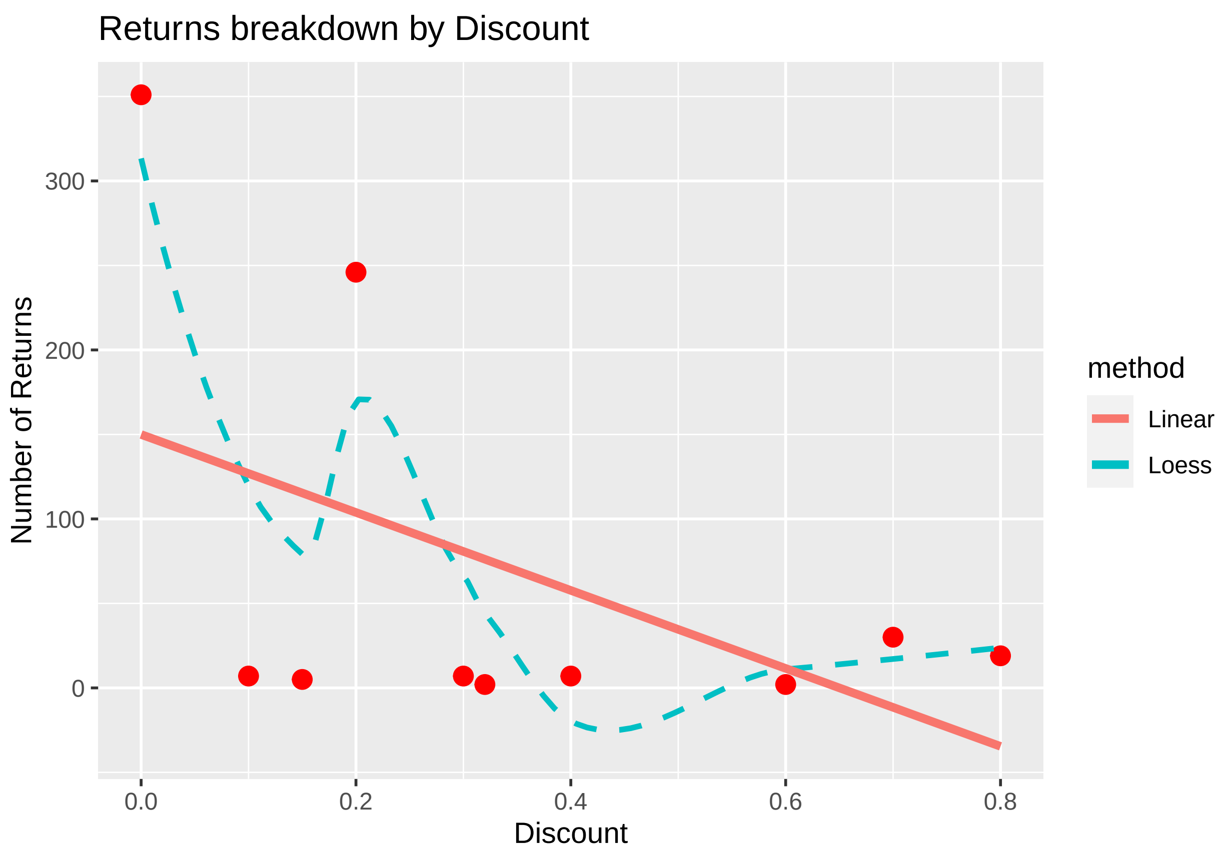


Standard class ship mode is the one with highest returns. Most likely, products shipped with Standard Class mode are not as high quality as those shipped with First or Second Class, what could be the reason for this big difference in the number of returns. In addition, the returns for products shipped with First and Second Class, the number of returns are practically the same. Therefore, (if the number of products sold of each mode were similar) AEKI has an opportunity to reduce costs by shipping more products with Second Class instead of First Class.

Relationship between returns and Discount:

```
discount_returns <- Returned_type %>%
  group_by(Discount ) %>%
  summarise("Number of Returns" = n())
```

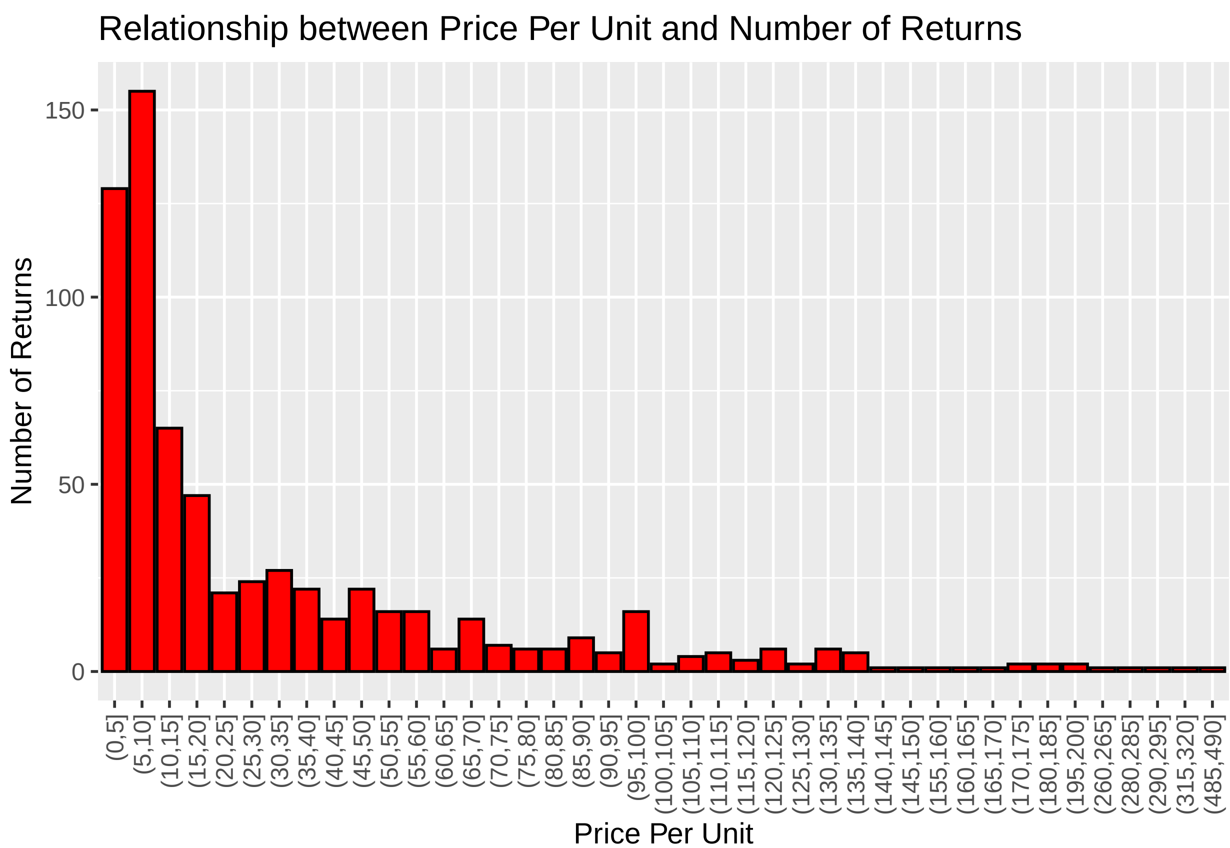
```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```



There is a decreasing trend, as the discount applied on the products increases. Most returns are concentrated in only 2 types of discount, 0% and 20% discount. This is probably influenced by the bigger amount of products sold with those amounts of discount. However, this relation doesn't give us any valuable information.

Relationship between returns and Price (grouped every 5€):

```
price_returns <- Returned_type %>%
  mutate(Price_Per_Unit = cut(Price_Per_Unit, breaks = seq(0, max(Price_Per_Unit) + 5, by = 5))) %>%
  group_by(Price_Per_Unit) %>%
  summarise("Number of Returns" = n())
```



This graphs shows the most complete representations from all of the above. We can clearly see that, as the unitary price increases, the number of return decreases. Most likely, this is influenced by the quality of the products sold. Cheaper products tend to have lower quality than more expensive ones what lead to higher customer dissatisfaction and more returns. Also, returns could be influences by their ease of return. For example, a stapler (cheap product) is easier and more comfortable to return than a Bookcase (expensive product) due to its size, leading to higher returns. Thus, AEKI's objective is to minimize the returns for these cheap a tiny products.

Demographic Analysis

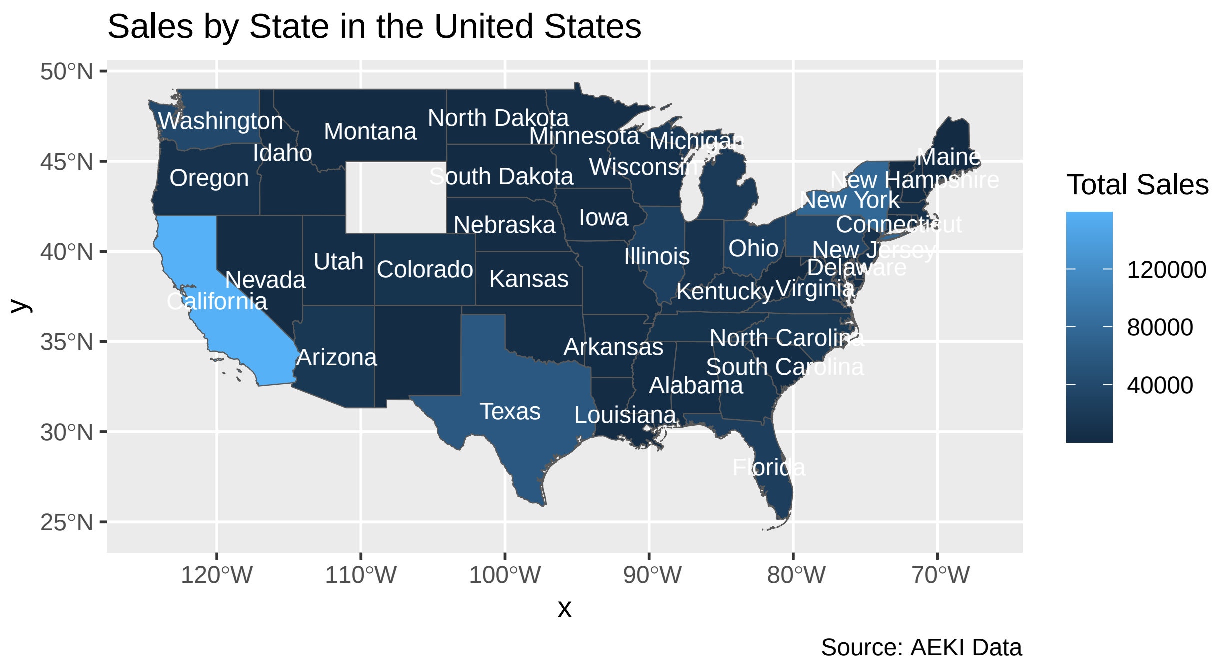
Demographic Sales analysis

```
us_states <- ne_states(country = "united states of america", returnclass = "sf")

state_sales <- Orders_Products%>%
  group_by(State) %>%
  summarise(Total_Sales = sum(Sales))

us_states_sales <- merge(us_states, state_sales, by.x = "name", by.y = "State")
```

Sales per State Graphics

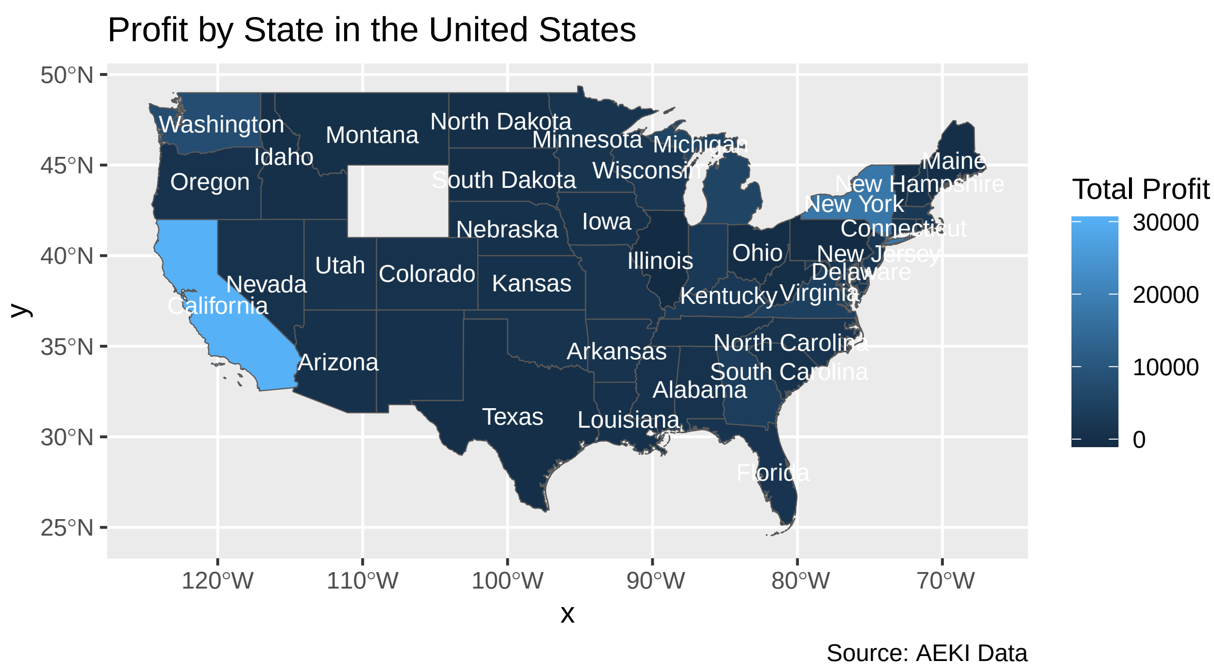


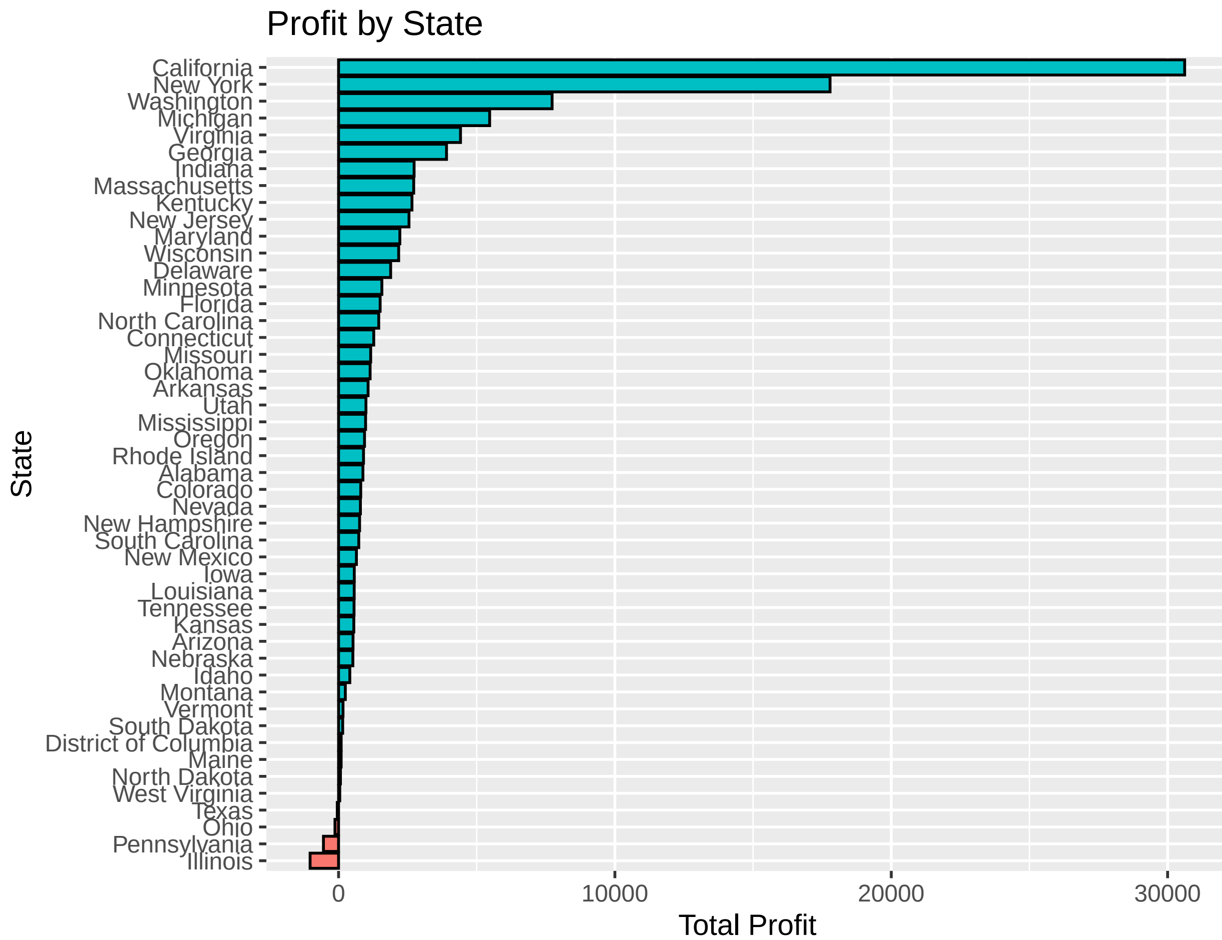
Both of this graphs represent the amount for total sales by state. The top states by sales are: California, New York, Texas, Washington, Pennsylvania, Illinois, Florida, Ohio, Arizona, Michigan, North Carolina, and Virginia.

Profits per State Graphics

```
state_profit <- Orders_Products %>%
  group_by(State) %>%
  summarise(Total_Profit = sum(Profit))

us_states_profit <- merge(us_states, state_profit, by.x = "name", by.y = "State")
```


Profit by State in the United States
Washington
Montana
North Dakota
Minnesota
Michigan
Idaho
Oregon
South Dakota
Wisconsin
Maine
New Hampshire
New York
Connecticut
Nebraska
Iowa
Utah
Colorado
Kansas
Illinois
Ohio
New Jersey
Delaware
Nevada
California
Kentucky
Virginia
Arizona
Arkansas
North Carolina
South Carolina
Alabama
Texas
Louisiana
Florida
50°N
45°N
40°N
35°N
30°N
25°N
y
120°W
110°W
100°W
90°W
80°W
70°W
x
Total Profit
30000
20000
10000
0
Source: AEKI Data



Now both of this graphs represent the total profit per state. In which Profit wise the states that generate more profit for the company are “California, New York, Washington, Michigan, Virginia, Georgia, Indiana, New Jersey, Kentucky, and Massachusetts.

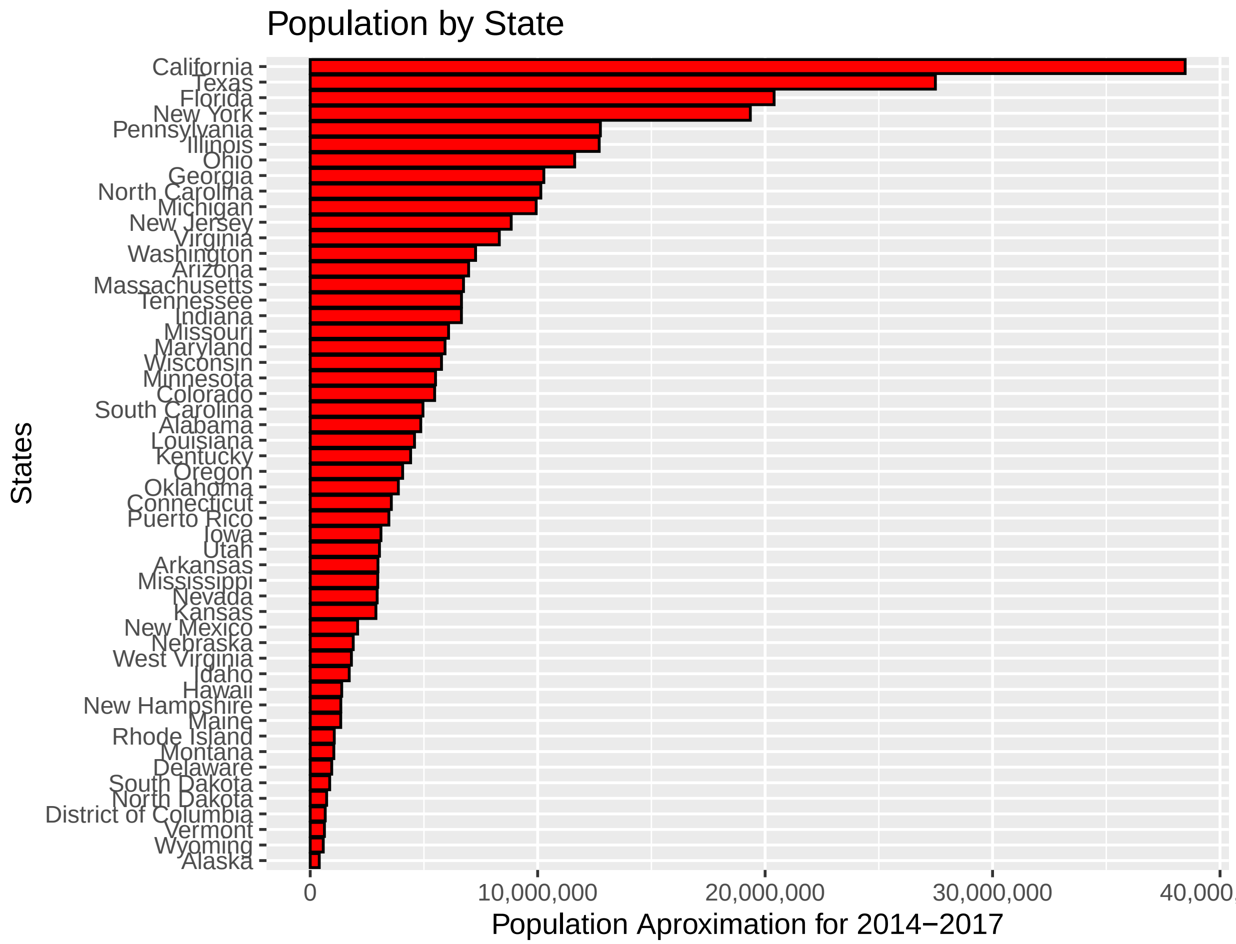
Observations for Sales and Profit Relationship per State:

From this diagram, we found out that the top states in terms of profit and sales are **California, New York, Washington, Michigan, and Virginia**. These are all tested markets that all but have great growth potential. Especially **Michigan, Washington, and Virginia**.

On the other hand, we found some markets that are not very profitable for Aeki: Pennsylvania, North Carolina, and Florida. Furthermore, Texas, Illinois, and Ohio are yielding negative profitability. These are markets we will study more in-depth and would recommend Aeki either to opt-out from these states or implement some changes we will discuss further in this report

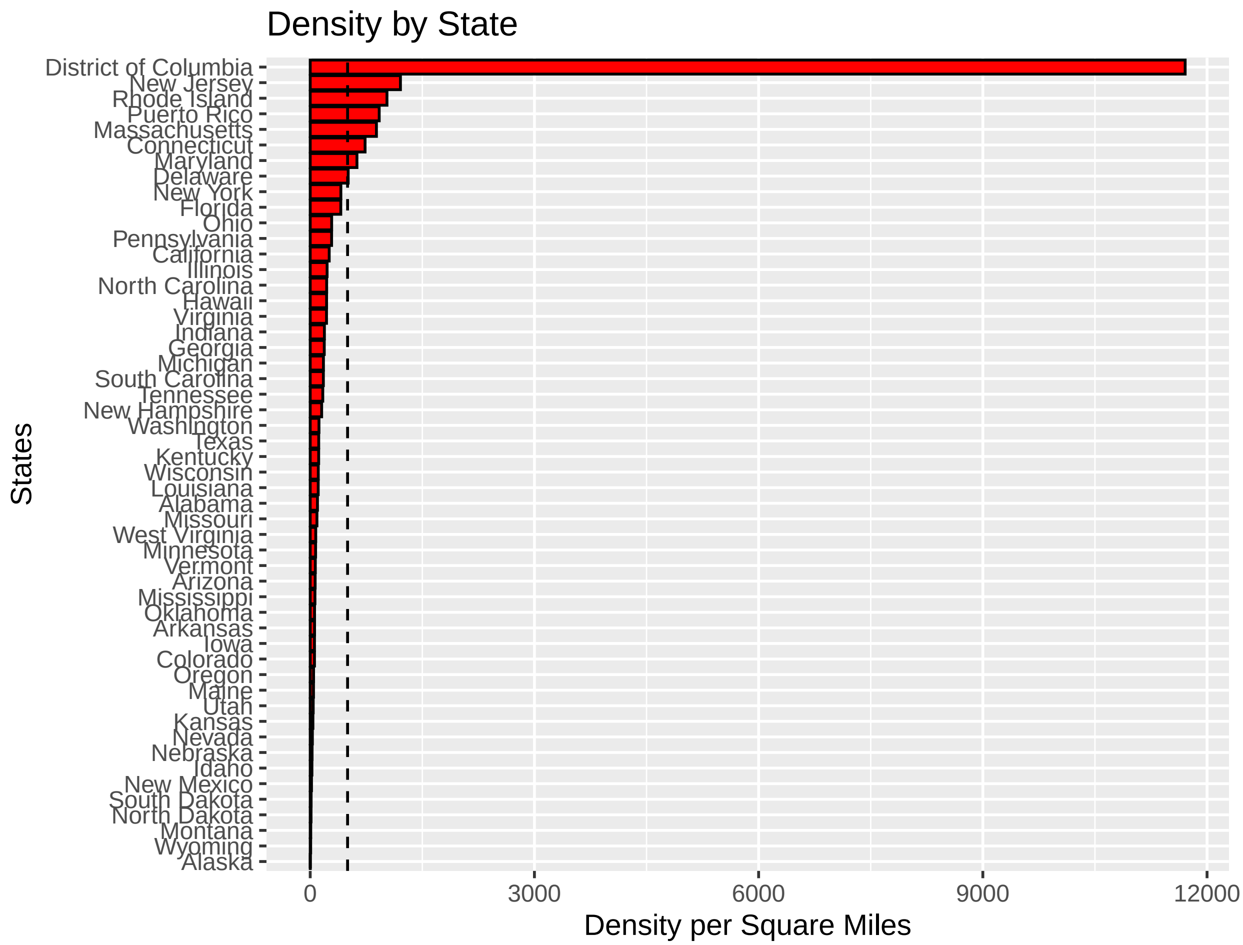
Lastly, we found several markets that have great potential due to their profitability but have very few sales to become representative in terms of total sales. These states with great potential are Georgia, Kentucky, Indiana, New Jersey, and Massachusetts. We recommend that Aeki invest more resources from states like Texas and Illinois into the previously mentioned states to increase their revenue and thus generate more profit.

Total Population per State



The states with the most population are California, Texas, Florida, New York, Pennsylvania, Illinois, Ohio, Georgia, North Carolina and Michigan which roam over the 10 million.

Density per State



In terms of density per state we have the district of Columbia with a massive density over any other state reaching over ten thousand habitants per square mile. Furthermore, We have New Jersey, Rhode Island, Puerto Rico, Massachusetts, Connecticut, Maryland and Delaware with over 500 habitants per square mile.

Observations Between Financial and Demografic Characteristics per State

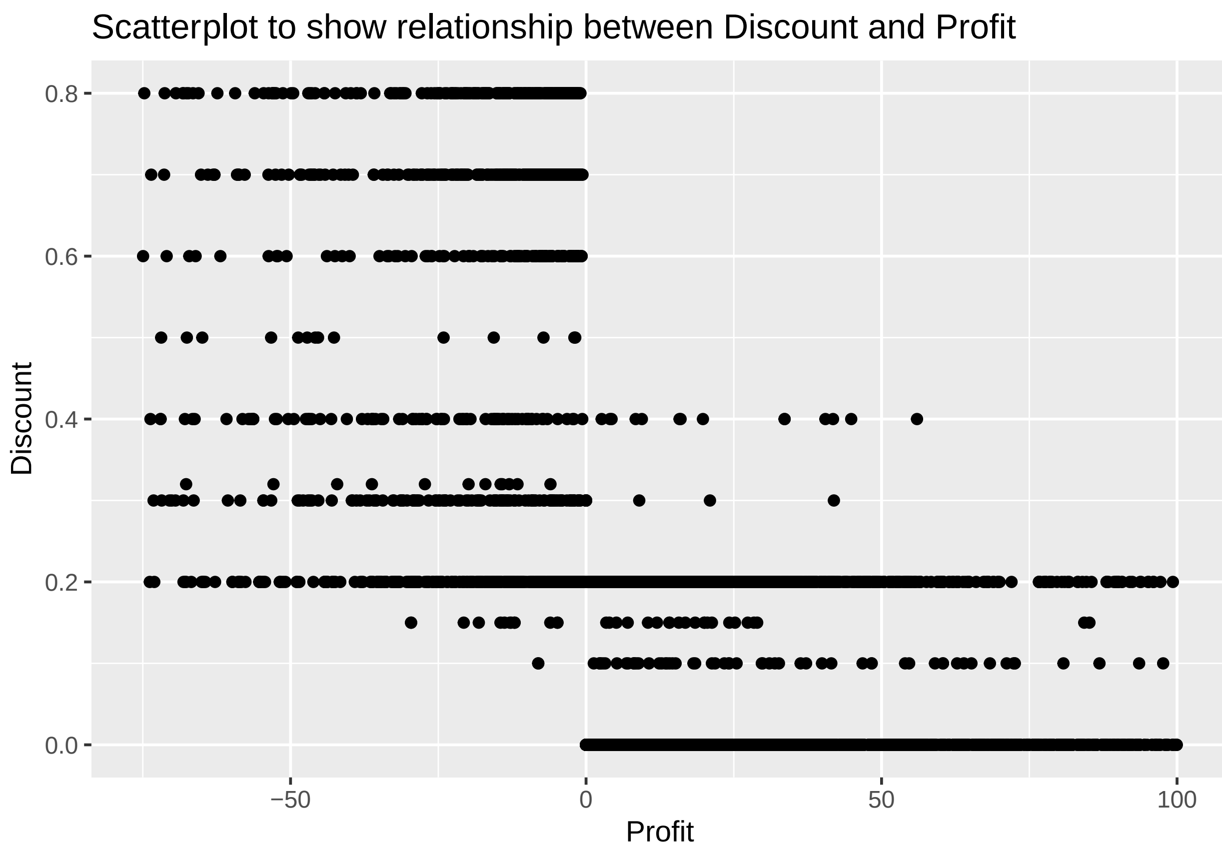
```
Sales_Demographics<- inner_join(state_sales, AEKI_Demographics)

correlation_coefficient <- cor(Sales_Demographics$Population_Aproximation_2015, Sales_Demographics$Total
```

In conclusion, we find that the density has a relationship with the total sales and profit per state. But we do find a strong relationship between the population of each state and the total sales as they have a **correlation factor of 0.911**.

Correlation of financial data

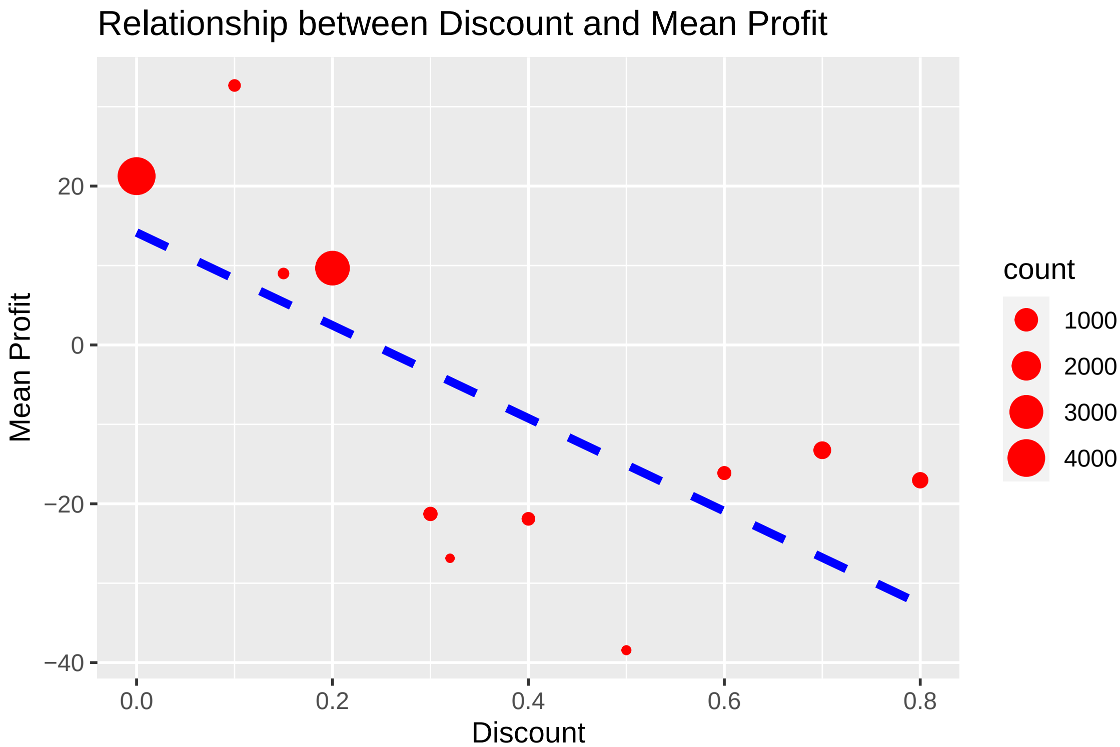
Relationship between Discount and profit



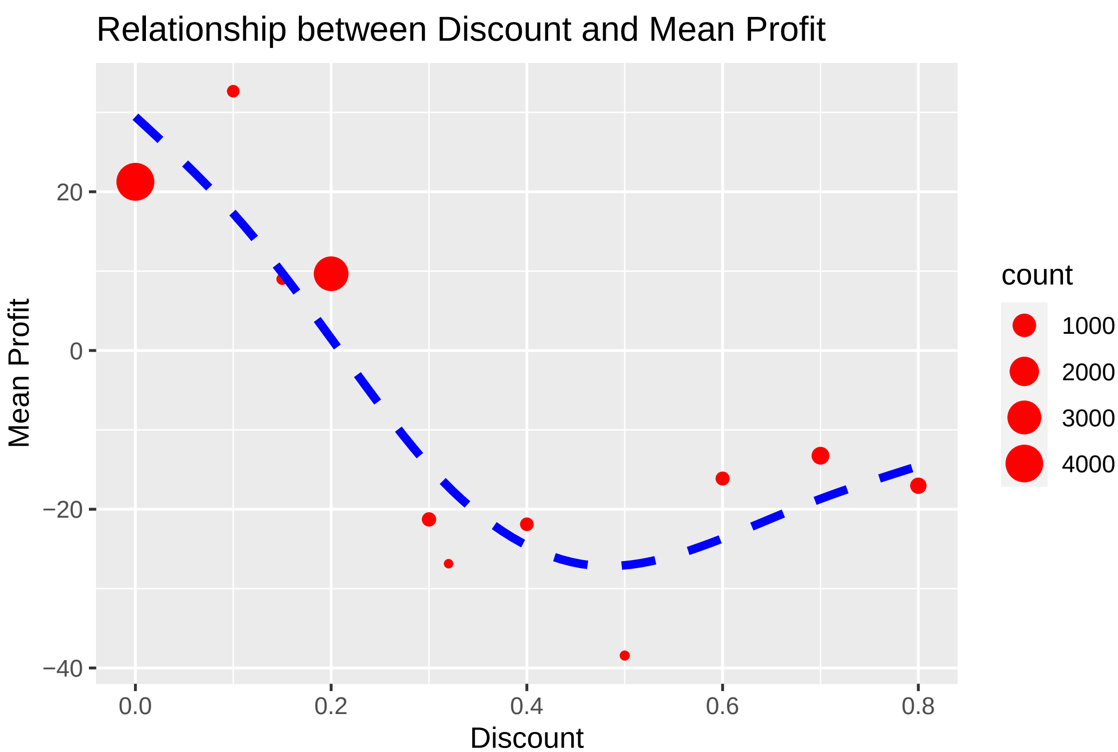
No clear pattern or relationship can be derived. Thus further analysis will be needed.

Code to find mean profit for each amount of discount applied

```
discount_profit_grouped <- Clean_AEKI_Orders %>%
  select(Discount, Profit) %>%
  group_by(Discount) %>%
  summarise(Mean_profit = mean(Profit), count=n())
```



Illustrating the data



For better analysis we have plotted both types of method ("gam" and "lm"). Both methods show a clear descending trend, what means that higher the Discount applied, lower the Mean Profit achieved. However, the linear model (lm) represents better this tendency as it is not influenced by the Count variable.

Analyzing correlation between variables (Discount and Profit)

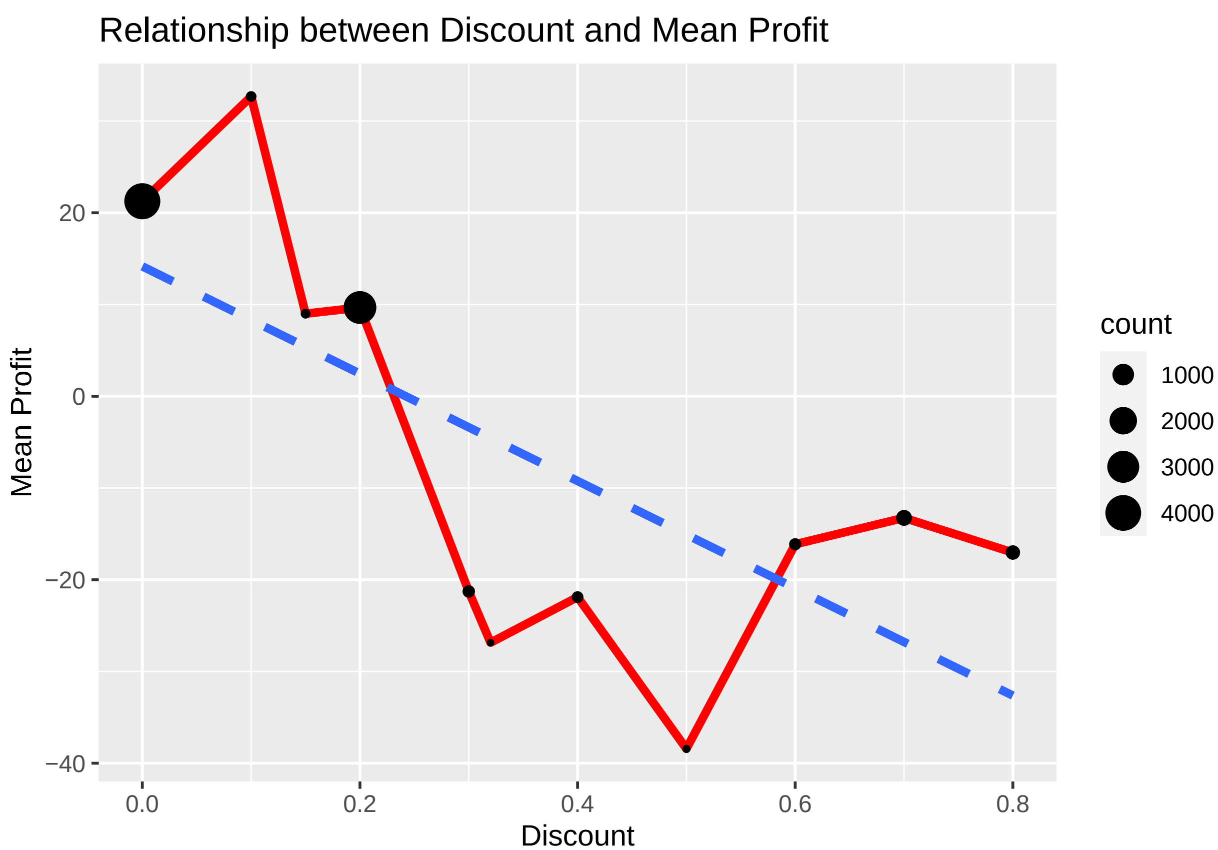
```
## [1] -0.4622241
```

-0.4665864 correlation.

The negative value of correlation between AEKI's discount and profit means that as one of the variables increases, the other decreases. However, the value is relatively low, closer to 0 than 1, so it will only cause a moderate decrease in comparison to the increase. However, we cannot assume that one causes the other from this correlation alone. But, as both variables are closely related to demand and revenue, we can interpret that Discount is the cause for the fluctuation in profit.

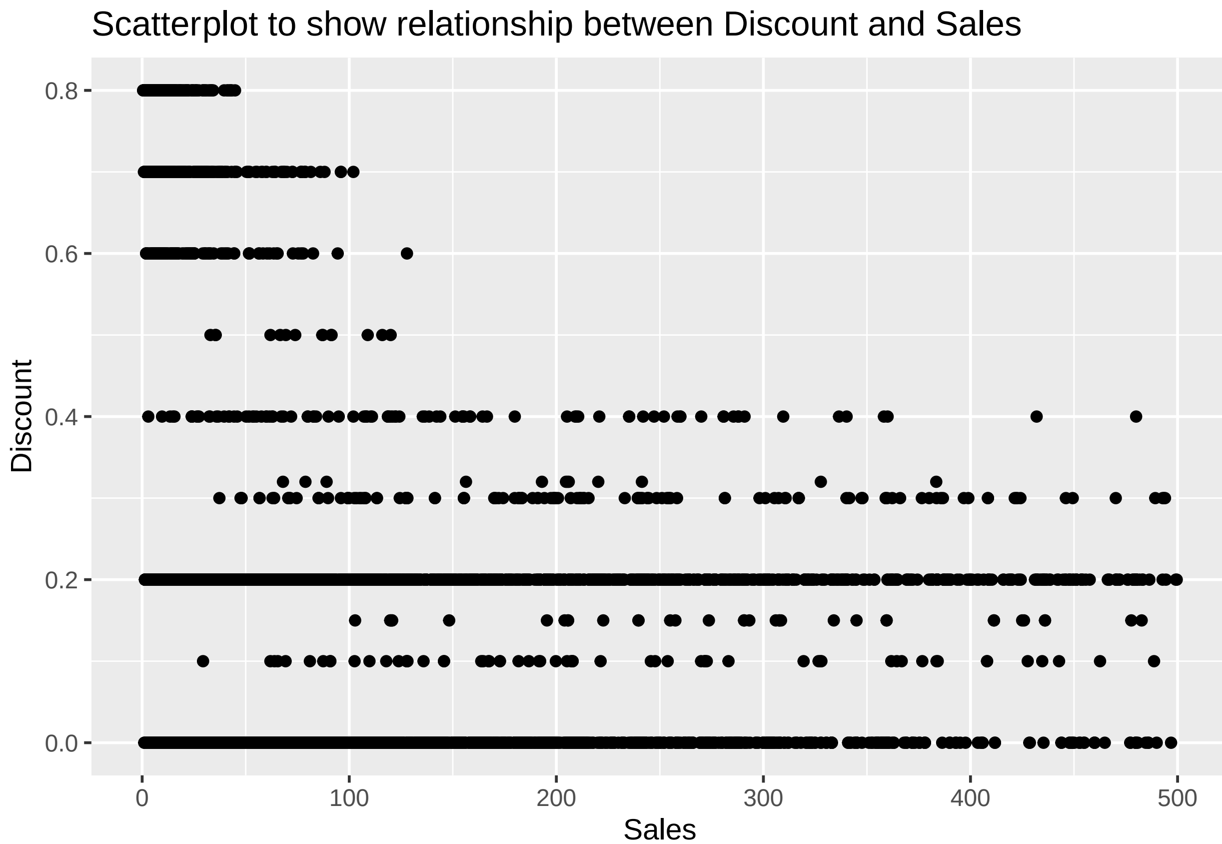
In conclusion, it is likely that discounts may lead to higher sales, but reduced profit margins.

Further illustration of relationship



With this plot, we can better appreciate the fluctuation of the Mean Profit as the Discount applied grows.

Relationship between Discount and Sales

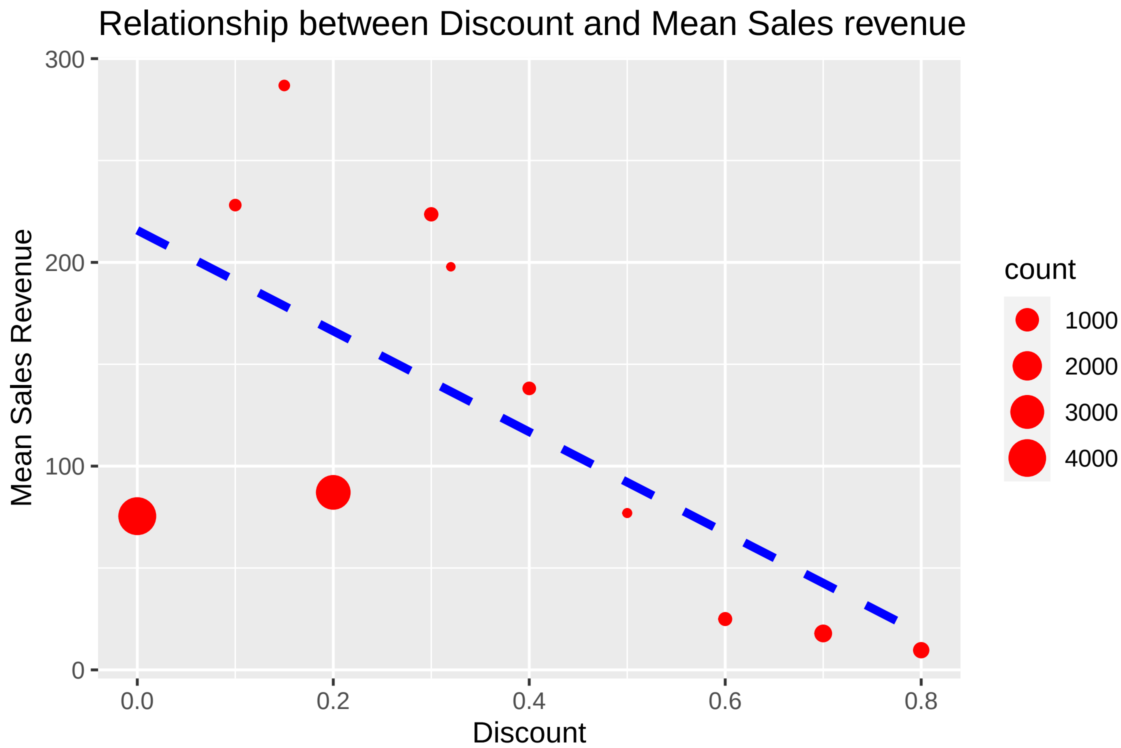


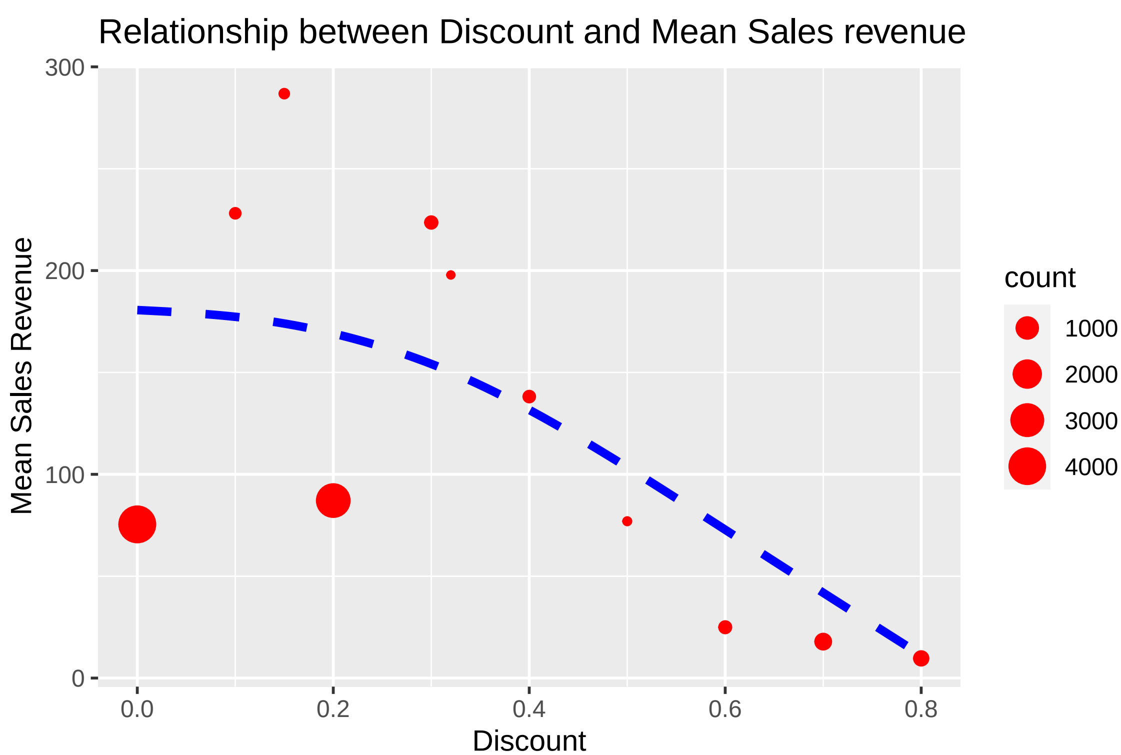
As per profit, in sales there is no clear pattern or relationship that can be derived. Thus further analysis will be needed.

Code to find mean sales (revenue) for each amount of discount applied

```
discount_sales_grouped <- Clean_AEKI_Orders %>%
  select(Discount, Sales) %>%
  group_by(Discount) %>%
  summarise(Mean_Sales_Revenue = mean(Sales), count=n())
```

Illustrating the data





As for the relation Discount-Profit, for better analysis we have plotted both types of method ("gam" and "lm"). Both methods show a descending trend, what means that higher the Discount applied, lower the Mean Sales revenue.

However, the linear model (lm) represents better this tendency as it is not influenced by the Count variable. Furthermore, the "gam" method remains steady for the first 20%, what suggest that, from that point on, the grow of the discount will impact heavier on the Mean Sales revenue, while little discounts do not affect dramatically on it.

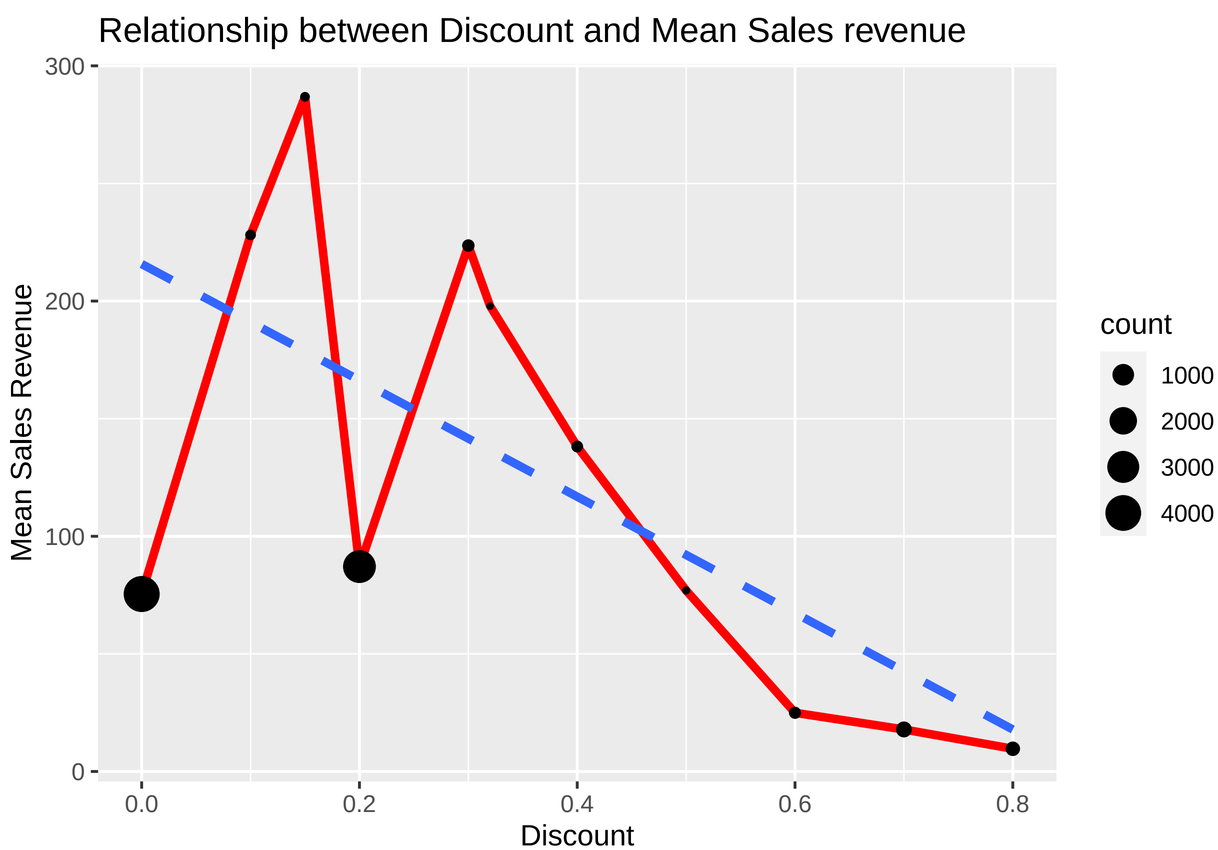
Analyzing correlation between variables (Discount and Sales)

```
## [1] -0.1101005
```

-0.1159012 correlation.

The negative value of correlation between AEKI's discount and sales means that, as one of the variables increases, the other decreases. However, the value is considerably low, really close to 0, so it will only cause a moderate decrease in comparison to the increase. However, we cannot assume that one thing causes the other from this correlation alone. But, as both variables are closely related to demand and revenue, we can interpret that Discount is the cause for the fluctuation in Sales, as equal to profits. In conclusion, it is likely that discounts may lead to higher sales, but lower sales revenue.

Further illustration of relationship



With this plot, we can better appreciate the fluctuation of the Mean Sales Revenue as the Discount applied grows.

Relationship between Sales and Profit

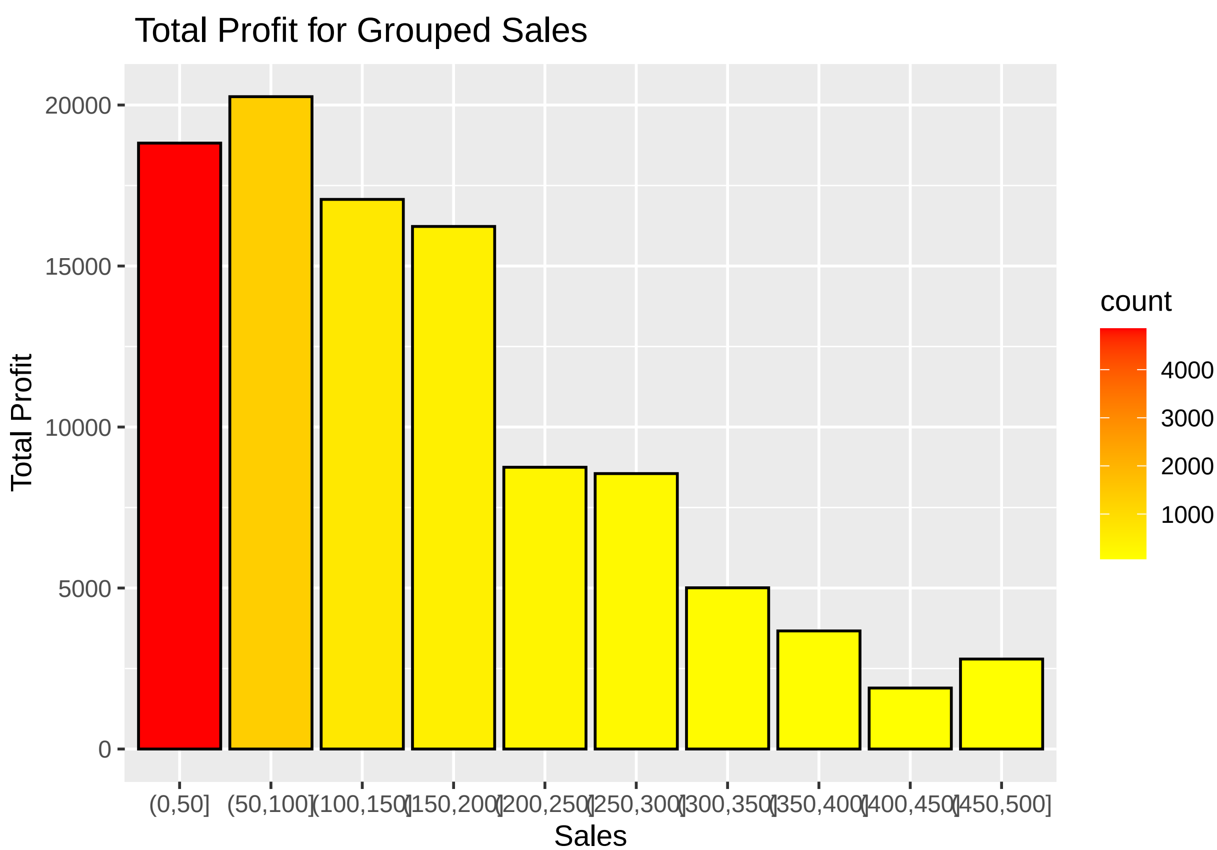
```r
grouped_sales <- Clean_AEKI_Orders %>%
  mutate(Sales = cut(Sales, breaks = seq(0, max(Sales) + 50, by = 50))) %>%
  group_by(Sales) %>%
  summarise(count = n(), total_Profit = sum(Profit, na.rm = TRUE), mean_profit = total_Profit/count)
```

In order to make it easier to work with and make conclusions, we have grouped sales values in breaks of 50. This helps us to have less observations and build better relationships than having each one of the sales values of the Clean_AEKI_Orders dataframe.

In addition, we have determined that 50 is a pretty acceptable length to make the intervals. 50 is big enough to make our analysis easier, but it isn't that big to harm our studio.

```r
ggplot(grouped_sales, mapping = aes (x = Sales, y = total_Profit))+
  geom_col(aes(fill = count), color="black")+
  scale_fill_gradient(low = "yellow", high = "red")+
  labs(title = " Total Profit for Grouped Sales", y= "Total Profit")
```

Total profit per group of sales



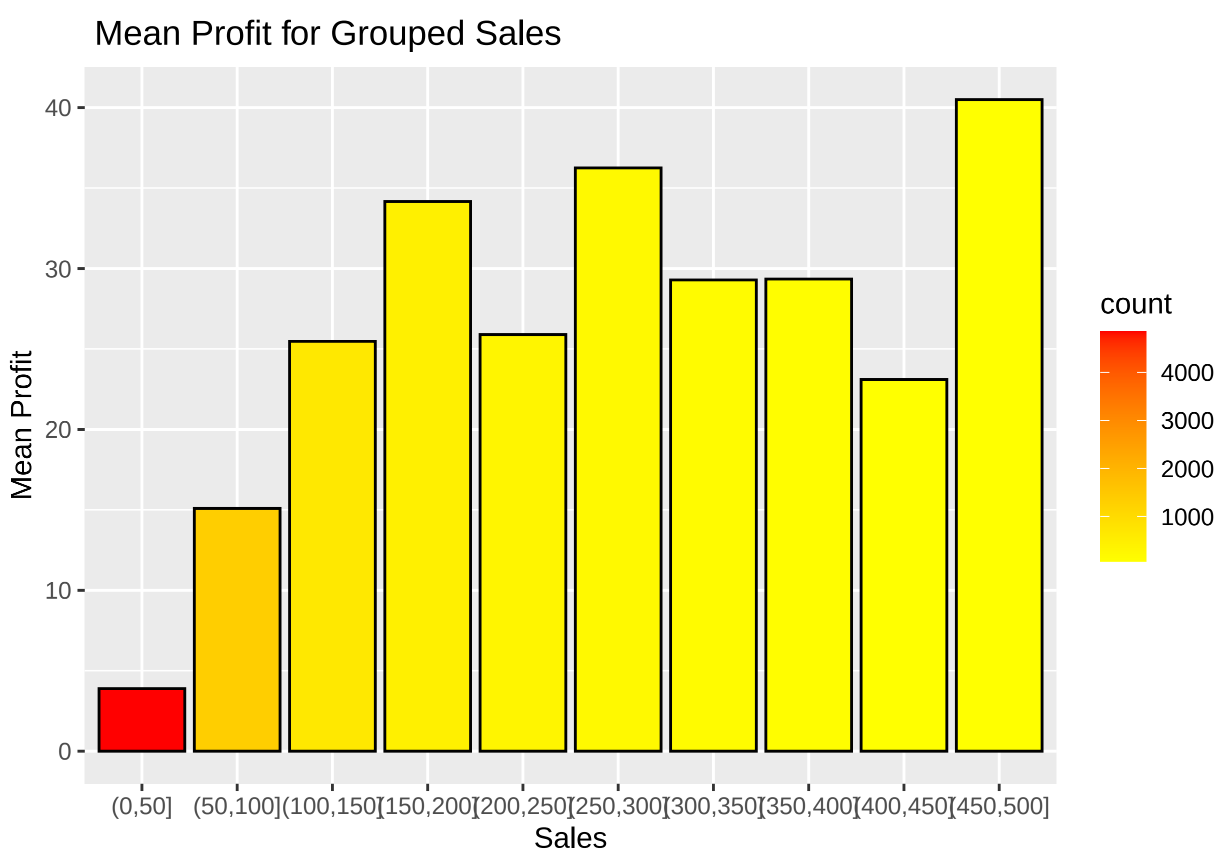
This graph represents the total profit achieved by AEKI in relation with sales values. The range of sales values that bring most profits is 50-100€.

We can see that, for lower sales revenue values there are higher profits, creating a right-skewed shape, as most profits are situated within the first 200€ in sales revenue. However, this doesn't necessarily mean that those ranges of sales are the most profitable ones, in terms of net profits per sale.

Most probably, this is due to the amount of sales made in that ranges of revenue per sale, because the lower the value of the sales, the higher the number of transactions. Hence, we are going to plot the relationship between the mean profits per transaction and the grouped sales revenue values, to find which are the most profitable ranges in terms of net profit.

```
ggplot(grouped_sales, mapping = aes (x = Sales, y = mean_profit))+
  geom_col(aes(fill = count), color="black")+
  scale_fill_gradient(low = "yellow", high = "red")+
  labs(title = " Mean Profit for Grouped Sales", y= "Mean Profit")
```

Mean profit per group of sales



Now, we can appreciate that what the range that generated higher profits is the one with he lowest profit per transaction. This doesn't mean that it is a bad or unprofitable range of sales.

As expected, the range of sales value with the highest mean profit is 450-500€, which is the maximum achieved by AEKI. The maximum mean profit is 40.49€ (Sales Value = 450-500€), followed by the range 250-300€, with 36.25€ of mean profit, and the range 150-200€ with 34.16€.

To sum up, although the range of sales values 450-500€ is the one bringing the highest profit per transaction, it is the second lowest profit earner for AEKI. Furthermore, low sales values going from 0 to 100€, although they don't have that high mean profits, they are the greatest revenue earners for AEKI.

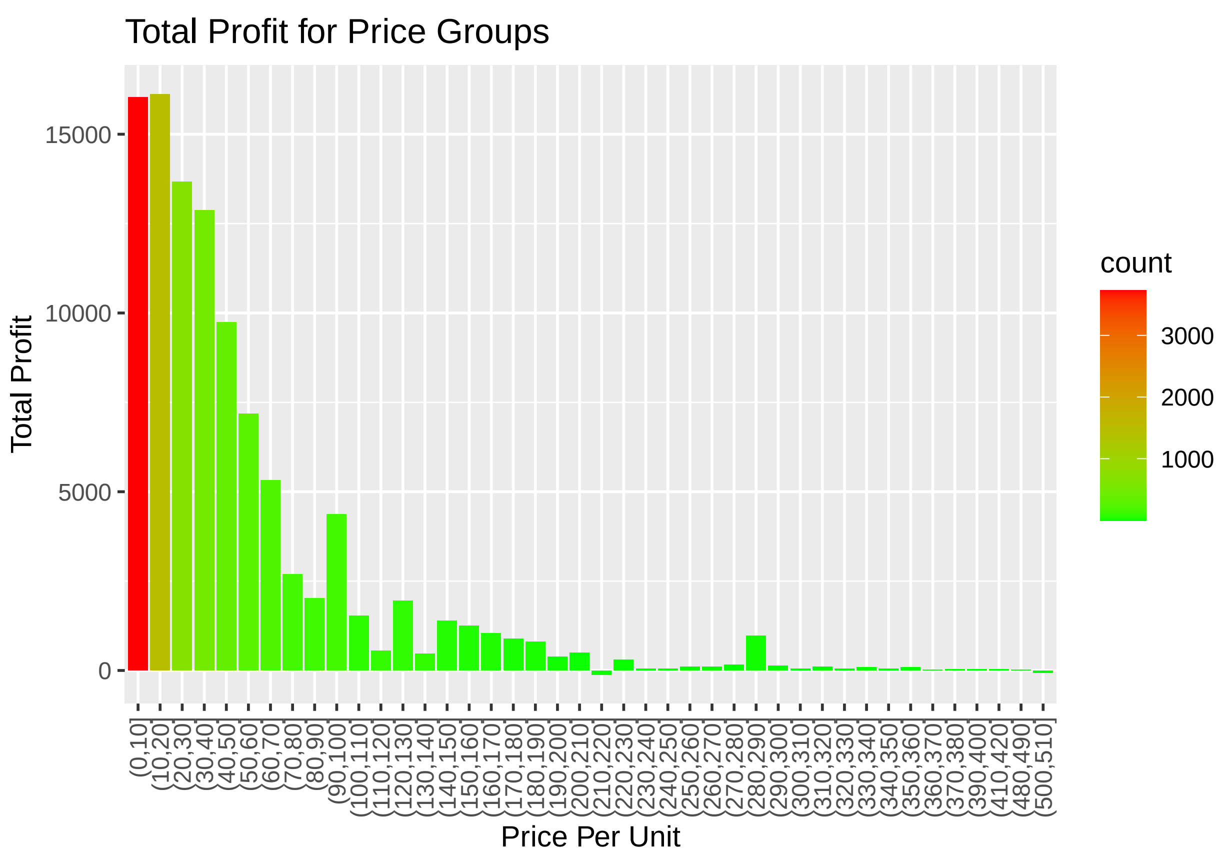
Lastly, the most valuable observation we get from this representation is that the range of sales going from 150€ to 200€ has the third highest mean profit (really close to the second one). Thus, considering that it is a relatively low price/sales value range, AEKI could focus on promoting products that could fit that interval to increase sales and maximize profits.

Relationship between Price and Profit

```r
grouped_prices <- Clean_AEKI_Orders %>%
  mutate(Price_Per_Unit = cut(Price_Per_Unit, breaks = seq(0, max(Price_Per_Unit) + 10, by = 10))) %>%
  group_by(Price_Per_Unit) %>%
  summarise(count = n(), total_Profit = sum(Profit, na.rm = TRUE), mean_profit = total_Profit/count)
```

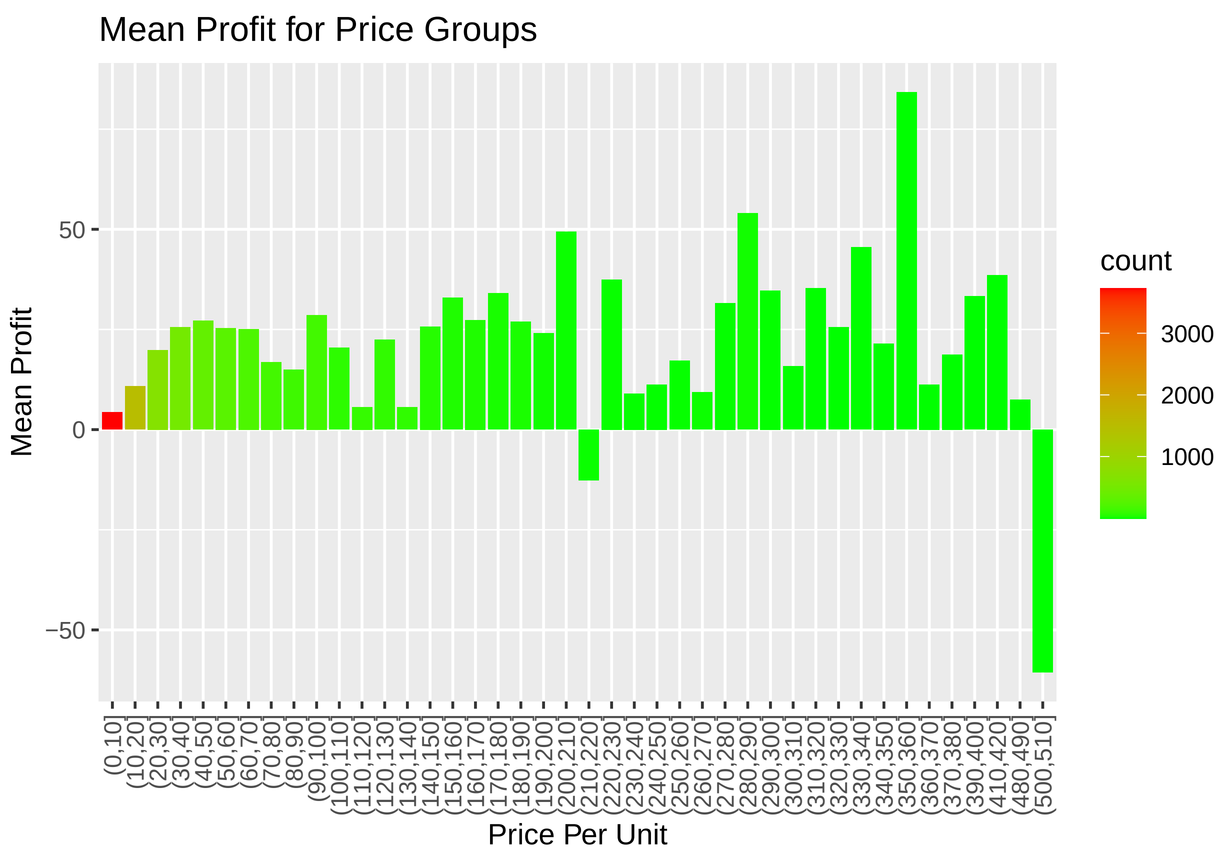
Group by price per unit As well as in the previous analysis, we have divided all the prices of products in groups of 10€ to facilitate the use of data and build better relationships. In this case, we have decide that ranges of length 10 would be suitable.

Total profit for price groups



As well as in the graph for total profit against sales value, price groups with lower value have the highest profits in total. With this graph we can conclude that the products that are most sold by AEKI are the ones going from 0 to 50€ per unit.

Mean profit for price groups



Also, as expected, as the prices increase, the mean profits increase too. However, a surprising observation is that for the the most expensive products sold by AEKI (500-510€), they have significant losses with a mean profit of -60.61€ per product sold. Products with unitary prices between 210€ and 220€ also have losses, but lower -12.63€ per product sold.

In conclusion, although products with very low prices don't have that big mean profits, they are the ones bringing the most money to AEKI. Furthermore, agreeing with the previous analysis made on profits per sales values, products ranging between 150€ and 200€ have really good mean profits, what creates a good opportunity for AEKI to promote these products in order to maximize profitability.

Margins

Based on the sales and profit analysis we were able to identify the best selling and most profitable sub-categories off products for AEKI. This products that we will call "winners" are: Furnishings, Paper, Storage, Appliances, Binders, Art, Accessories and Phones. This are the product categories who had the most sales and profits above 50 thousand Euros.

```
## Adding missing grouping variables: `Category`
```

Category	Sub-Category	AVG_Price	AVG_Profit	Margins	Total_Profit	Total_Sales
Furniture	Furnishings	24.192168	13.253246	0.1877748	12590.584	67051.51
Office Supplies	Appliances	38.070938	20.414913	0.1977604	7431.028	37575.91
Office Supplies	Art	9.107062	7.788106	0.2441855	6207.120	25419.69
Office Supplies	Binders	11.594504	4.788253	0.1479043	6746.648	45614.96
Office Supplies	Paper	14.463149	19.139858	0.4276225	26355.584	61632.83
Office Supplies	Storage	45.389359	7.999033	0.0649662	5591.324	86065.15
Technology	Accessories	39.257135	23.406915	0.2028885	15846.481	78104.37
Technology	Phones	69.533997	17.763767	0.1121884	11493.157	102445.16

This table contain the financial data for the "winners" subcategories

From these 8 "winners" products the company has received a total of € 92,262 in profit out of € 107,242 which comes from a total of 19 product categories. This means that these 8 products account for 86.% of the company's profits.

Calculations of Winners Financial Data

x
0.1086009

Now that we have collected the mean, standard deviations and IQR we will proceed to analyze the margins profit and avarage price for the "winners". sub-categories

Margin: By taking one standard deviation from the mean of the winner's margin we are looking to have 68% of the the observations. Creating a range that will cover the most successful margins per product for AEKI. 0.09 < Margin < 0.31

In this table with the established profit margins, we see that it covers 6 of the winners. With this pattern, we see that Paper was left out due to a very high margin, and then Storage also was cut out due to having very low margins. For new inclusions, we have Machines, Copiers, and Tests. These 3 sub-categories have the margins and potential to be winners but need more total sales to generate significant profit and become winners.

Observations: The most effective products in terms of generating profit for the company are those which have a profit margin between 9% to 31%. We will encourage AEKI to promote the growth and sell for items in this range and take into consideration that for this product to become greatly profitable they need volume. Most products that have been effective have a volume per sub-category of over a thousand items with a mean of 2357 items.

x
6.86006

Profit: Lastly, we repeat the procedure for the AVG_Profit of products to understand the behavior of these determined "Winners". 7.4 < AVG Profit < 21.1

Observations: In terms of profit "winner" products tend to relay 7.4 to 21.1 dollars in profit per item.

Average Price: Now to understand the best price we need to separate it by categories due to the big price difference seen between them

Technology -> mean = 122.8 -> applied standard deviation = 9.36 to 236.1 The average price in which products perform the best for the category Technology ranges from 9.36 to 236.1

Furniture -> mean = 17.6 -> applied standard deviation = 4 to 31.2 The average price in which products perform the best for the category of furniture ranges from 4 to 31.2

Office Supplies -> mean = 96 -> applied standard deviation = 46.6 to 145.4 The average price in which products perform the best for the category office supplies ranges from 46.6 to 145.4

Observations and Recommendations

Sales

Observations

Revenue Range and Sales Concentration: The analysis shows that AEKI's most common revenue range is between 12 and 100 euros, with an average sales concentration at 80.28 euros. This highlights the main price point that attracts customers.

Category Performance: Office Supplies are the leading category, accounting for 42.12% of AEKI's sales. In contrast, the Phones subcategory, though less sold, contributes a substantial 14.84% to total sales. This indicates varying consumer preferences within different product categories.

Sales Trajectory (2014-2017): All categories maintained stability until 2016, followed by a significant upswing, especially in Office Supplies. The year 2017 was particularly successful, marking a 30.47% increase in sales and a 22.53% boost in profit. This period reflects AEKI's best financial performance.

Impact of Technological Innovation: The rapid increase in Phone sales revenue over a short period underscores the influence of technological advancements on consumer choices.

Monthly Sales Pattern: There is an overall upward trend in sales, but a noticeable dip occurs every October. This pattern suggests specific challenges that need addressing.

Recomendations

Strategic Planning for Sales Dip: To address the October sales, we recommend AEKI to implement strong marketing strategies and special promotions during this month to increase sales volume.

Optimizing Product Offerings: AEKI should focus on optimizing its product range, particularly capitalizing on the success of the category Office Supplies and on Technology due to the growing demand for Phones.

Extending Peak Sales Periods: We recommend developing strategies to extend the duration of peak sales periods, possibly by introducing new products or promotions during traditionally slower months.

Profit

Observations

Consistent Profitability: The profit distribution histogram shows a stable median around 7.25 EUR, indicating consistent profitability.

Category Performance: Office Supplies emerge as the most profitable category, contributing significantly to profits. On the other hand, the Furniture category shows lower profitability in relationship to other categories.

Subcategory Insights: Within the Furniture category, Furnishings, Paper, and Accessories are key profit contributors, whereas Tables, Tests, Machines, and Copiers are less profitable.

Profit Trends: There is a general upward trend in profit, with 2017 marking significant growth, especially in Office Supplies and Technology. This suggests a shift in consumer preferences towards these categories.

Challenges in Furniture Category: Despite overall growth, the Furniture category experienced a decrease in profitability in 2017.

Monthly Profitability and Cost Management: Fluctuations in monthly profitability, influenced by varying costs in low-season months, point out areas for improvement in cost management.

Recommendations

Capitalize on High-Profit Categories: AEKI should focus on expanding and enhancing high-profit categories like Office Supplies and its sub-categories Furnishings, Paper, and Accessories.

Reconsider Low-Profit Products: We recommend a strategic review or exit from less profitable products such as Tables, Tests, Machines, and Copiers, to streamline the product line and allocate resources more effectively.

Adapt to Consumer Trends in Technology: Staying attuned to the increasing consumer interest in technology and their willingness to pay higher prices in this category will be crucial.

Address Furniture Category Challenges: The company should re-consider changing or scaling down their product line in Furniture. As it represents a very high cost for the company

Cost Management: To stabilize monthly profitability fluctuations, AEKI should review their cost management strategies and look to optimize and adjust costs during low-season months.

Discount

Observations

Negative Correlation Between Discounts and Profit/Sales Revenue: A descending trend indicates that higher discounts correlate with lower mean profit and sales revenue. This suggests that while discounts might increase sales volume, they negatively impact profitability.

Impact of Discounts: There is a clear descending trend showing that higher discounts lead to lower mean profit. This negative correlation, indicated by a correlation coefficient of -0.4665864, suggests that increasing discounts moderately decreases profits and sales revenue.

Fluctuation in Mean Profit and Sales Revenue with Discounts: As discounts increase, there's a noticeable fluctuation in both profit and sales revenue, highlighting the need for a balanced approach to discounting.

Recommendations

Moderate Discount Strategy: Given the negative impact of higher discounts on profits, AEKI should adopt a more moderate discount strategy that doesn't significantly decrease profit margins. It was found that the optimal range in which fluctuations in sales and profit increase while not having great impact in margins is between 0% and 20%.

Seasonal Discounts: Instead of broad discounting through the year, we encourage AEKI to implement the discounts in in months with lower transactions such as October.

Margin Based Discounts: We recommend AEKI to use discounts on those products with higher margins. As it will still incentive purchasing and volume while not suffering losses on certain products.

Correlation of Financial Data

Observations

Distribution of Profits and Sales: The right-skewed distribution indicates that higher profits are concentrated at lower sales revenue values. However, the most profitable ranges per transaction are identified in the 450-500€, 250-300€, and 150-200€ ranges.

Profitability vs. Transaction Volume: While the 450-500€ range has the highest profit per transaction, it's not the most significant profit earner due to low transaction volume. Furthermore,this price range show significant losses associated with its products. On the other hand, products priced between 0 and 50€ are the most sold and bring in the highest total profits.While the 150-200€ range, wich has the third-highest mean profit, offers a balance of profitability and volume.

Recommendations

Continue focus on Lower-Priced Products: Given that products in the 0 to 50€ range are the most sold and return the highest total profits, AEKI should continue their current strategy and prioritize this price segment.

Strategy for the most Profitable Price Ranges: AEKI should promote products within the 150-200€ range more aggressively, as this range offers a good balance between profitability and consumer demand.

High-Priced Products Strategy: The company should review its strategy for high-priced items in the ranges 210-220€ and 500-510€, to avoid substantial losses. Furthermore, we recommend AEKI to either focus resources and their marketing strategy towards this products to increase their sales volume. Or lower their product line and re-direct their focus towards the products in lower price ranges.

Returns

observations

Shipping-Related Returns: There is a strong trend of products shipped with standard mode being returned more frequently. This raises concerns about this specific shipping method and the condition in which these products are reaching customers. There also exists the possibility of them orders being inferior quality products.

High Return Rate for Binders: Binders have been returned 130 times, indicating a significant issue with this product category which negatively impacts customer satisfaction.

Negative Relationship between Price and Returns: A strong negative correlation exists between the unit price of products and the number of returns, suggesting that lower-priced products are returned more frequently.

To take into consideration: However, these conclusions should not be taken without consideration of the nature of products in each category, sub-category, shipping mode or price range. For example, products shipped with First class are likely to be higher quality, more expensive products. Therefore, customer satisfaction is expected to be higher.

Recomendations

Investigate Shipping Issues: We recommend AEKI to conduct an in-depth investigation into the reasons behind the high return rate of products shipped via standard mode and products under the sub-category of Binders. We encourage to look into shipping processes, product handling, and customer feedback to determine if the issues are related to shipping conditions or product quality.

Focus on Improving Lower-Priced Products: To address the negative trend of returns with price, we advise enhancing the quality of lower-priced products or evaluation removing certain products from the lower-end line.

Demografics

Observations

Top Performing States: In terms of both sales and profit the top states are: California, New York, Washington, Michigan, and Virginia.

Under performing Markets: The less profitable states are Pennsylvania, North Carolina, and Florida . Additionally, although these states have shown great sales volume they have also yield a poor or negative profit: Texas, Illinois, and Ohio

Population Size and Sales: There is a strong correlation with a factor of 0.911 between the population of a state and total sales, suggesting that more populous states tend to generate higher sales.

Recommendations

Invest in High-Potential States: AEKI should consider reallocating resources from under performing states like Texas and Illinois to states with high profitability but low sales, such as Georgia, Kentucky, Indiana, New Jersey, and Massachusetts, to capitalize on untapped potential.

Re-evaluate Strategy in Under performing States: We recommend AEKI to analyse their cost structure on states like:Pennsylvania, North Carolina, Florida, Texas, Illinois, and Ohio to understand their under performance. Furthermore, AEKI should consider either exiting these markets or implementing changes bring performance up.

Enhance Presence in Top States: Due to the growth through 2014 to 2017 we recommend AEKI to continue investing in and expanding operations in the top-performing states, especially focusing on growth opportunities in Michigan, Washington, and Virginia.

Targeted Marketing Based on Population: Given the strong correlation between population and sales, AEKI should focuse its marketing strategies in high population states to maximize sales potential.

Conclusion

In concluding AEKI's comprehensive analysis, focusing on key strategic recommendations is essential for driving the company's success. Here are the most relevant recommendations for AEKI's overall success:

Optimizing High-Performing Sub-Categories From the company's 16 sub-categories, **8 of them account for 86%** of the companies total profits: Furnishings, Paper, Storage, Appliances, Binders, Art, Accessories, and Phones. Thus we strongly encourage AEKI to focus on and invest in the high-performing product sub-categories.

Prioritize Lowe Priced Products with High Volumes: Based on the analysis AEKIS products that bring the most total profit and sales are in the lower price ranges **0-200€**. AEKI should focus and adhere to products priced between 0 and 50€, ensuring a steady revenue stream. While intensify its promotional efforts for products in the 150-200€ range. This price range, while not the highest in terms of sales volume, offers a significant balance between profitability and consumer demand. Allowing AEKI to capitalize not only in volume but also greater margins. Making the company more profitable.

Reallocate Resources to High-Potential States: We recommend the company to reallocate resources and marketing efforts from under performing states (such as Texas and Illinois) to those with untapped potential like Georgia, Kentucky, Indiana, New Jersey, and Massachusetts. These regions show high profitability but currently contribute less to total sales, indicating room for significant growth.

Adjust Discount Strategy: The company should adjust their discount strategy to fit the range between *0%-20%*. Given the negative impact of higher discounts on profitability. AEKI should aim for moderate discount levels, applying the discount strategy in lower transaction months and apply discount on products with higher profit margins.

Evaluate Changes for Underperforming Underperfoming Categories and Sub-Categories: For the furniture category and the 8 categories that account for 14% of the companies profit. AEKI should consider reducing their product line in set sub-categories to reduce costs and allocate resources on the more beneficial ones. Modifying their product lines to fit modern trends and consumers demands. Lastly, opting out of some of those sub-categories and expand their product line into a more profitable market segment.