

KOCAELİ ÜNİVERSİTESİ



Faculty of Engineering
Mechatronic Engineering Department

Veri Bilimi İcin Programlama

SINAV

Student name:

Adebayo Azeez Folayemi

Student number:

226113001

Exploratory Data Analysis (EDA) involves using visual tools such as: histograms, box plots, density plots, scatter plots, and correlation matrices to understand the distribution and variability of the data. We gain insights from the “Superstore” dataset by analyzing various plots and as such conclude with detailed report of our findings.

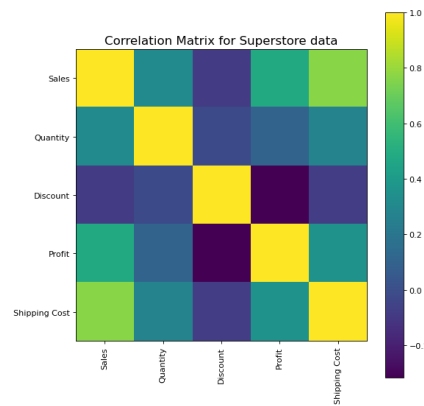


Figure 1: Sales performance by product category.

1. Statistical insights from the dataset by performing exploratory data analysis (EDA)?

- From the colors and their corresponding value codes, we see that "Sales" and "Discount" have a direct impact on our "Profit," with Sales showing a positive correlation and Discount showing a negative correlation, respectively. This is expected because as sales rise, profits rise as well, and as discounts are offered on sales, profit margins fall.
- Additionally, we can see that there is a positive correlation between "Sales" and "Quantity," which also makes sense because the more quantity, the greater the amount of sales, but this correlation appears to be weak.
- We can also observe that there is little to no difference in sales and quantity due to discounts, proving that the discount strategy is ineffective for certain items. As a result, Superstore needs to rethink its strategies for luring customers.

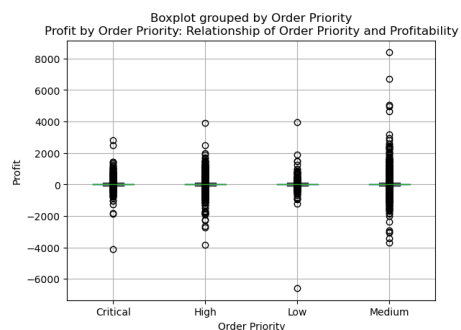


Figure 2: (a) Discount vs Profit plot. (b) Box-plot of Profit against Order Priority.

2. Is there a relationship between the discount offered and the profit margin?

- The negative sign indicates that there is an inverse relationship between the two variables. In other words, as the discount increases, the profit tends to decrease. Also, the magnitude of the correlation coefficient (-0.316) indicates the strength of the relationship. Since it is closer to zero, it suggests a moderate correlation.

3. What about a relationship between the Order Priority and the profit margin?

- Profits slightly higher when Order priority is Medium.

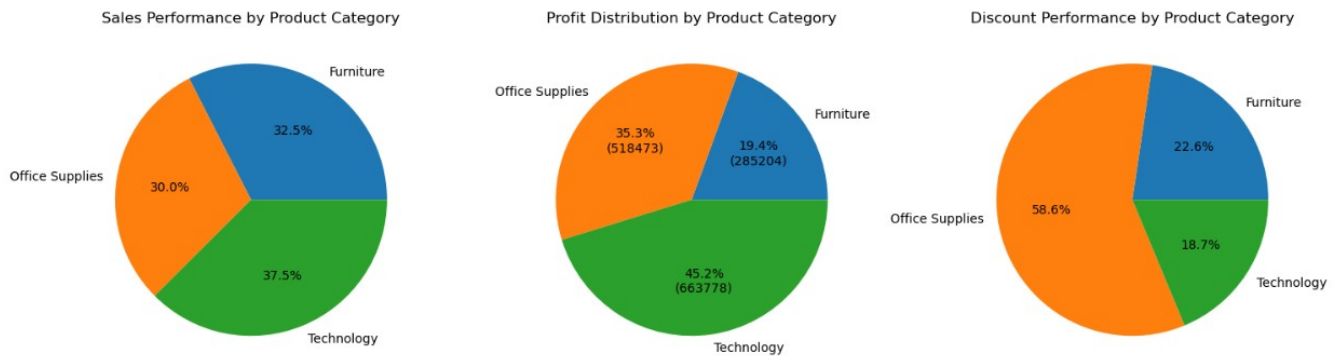
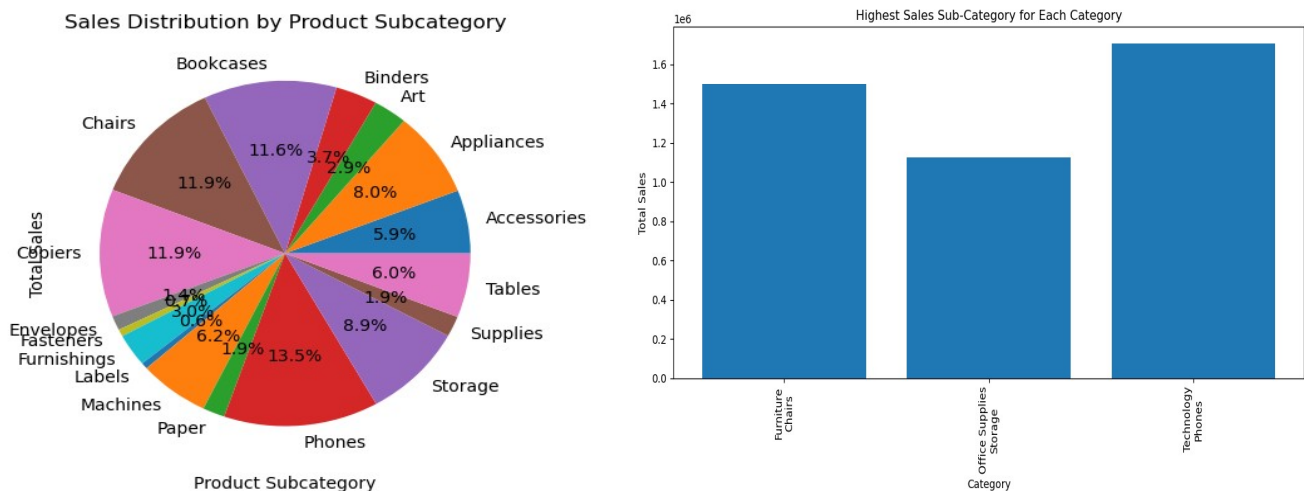


Figure 3: (a) Sales performance by product category. (b) Profit distribution by product category. (c) Discount performance by product category.



(d) Sales distribution by sub-category pie and (e) Highest Sales sub-category for each category respectively.

4. From the above Figures, we make the following observations:

I. Technology :-

- This category creates the most overall profit of 663.7kUSD, or 45.2% of the total profit earned by superstores, as well as the most sales, which make up around 37.5% of superstore sales. Due to shifting consumer preferences toward cutting-edge technologies, this industry is predicted to expand quickly in the next years.
- Additionally, the "**Technology**" category offers the lowest overall discount of 18.7% of all the categories. This could be due to the fact that it is the most expensive Category and that more expensive materials and specialized labour are needed throughout the manufacturing process.
- This Category has the highest overall sales and profits.

Its Sub-Category

- **Phones** as sub-category maintains the highest with 13.5% of the entire Superstore sales.

Its highest sale Product Name

- **Cisco TelePresence System EX90 Video conferenci** maintains the highest sales with 22.6K

II. Office Supplies :-

- This Category has the least overall sales as it is considered the least expensive category. The majority of the things in this category are compact and widely accessible at numerous other places.
- The most popular category is undoubtedly office supplies as it is one with the most discounts at 58.6%. This also leads it to have a somewhat good profit generated at 35.3% of Superstores' overall profit.

Its Sub-Category

Storage as sub-category maintains the highest with 8.9% of the entire Superstore sales.

Its highest sale Product Name

- **KitchenAid Stove, Silver** with 4624.290USD sales value are the most popular in this category, perhaps because each unit is expensive.

III. Furniture :-

- Among all categories this category has Lowest average profit generation per unit which is only 285.2k holding 19.4% profit.
- it is because the amount of profit margin is really low and at the same time.
- Even after having the minimum profit margin, we see that Furniture category has a fairly low discount of 22.6%

Its Sub-Category

Chairs as sub-category maintains the highest with 11.9% of the entire Superstore sales.

Its highest sale Product Name

- **Chromcraft Conference Table, Fully Assembled** with 5244.840USD made the most sales in the Furniture Category.

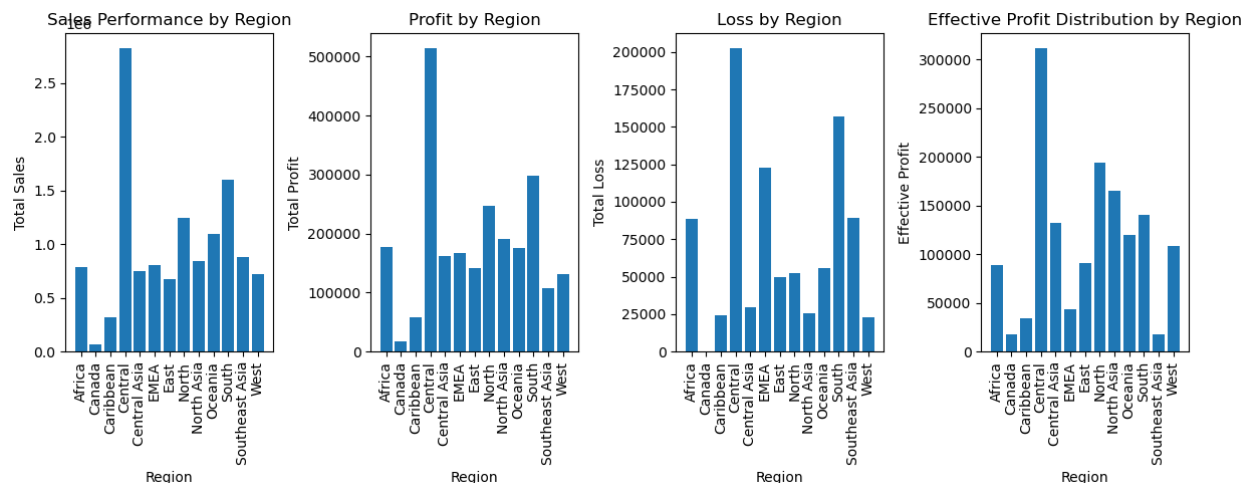


Figure 4: (a) Sales performance by product category. (b) Profit by Region. (c) Loss by Region. (d) Effective profit distribution by Region.

5. Which region's superstore makes the most money?

- Since sales has to do with quantity and price. When we look at the Profit that each region's superstores created, it is clear that the Central region's superstores generated the most overall earnings, with over 300kUSD effective profits and over 2.5x1e6 in sales.
- It is also clear that while Canada's sales performs least and as little profits returned, it is also interesting that it in fact made no losses and as such, a good business overall.
- Some regions also makes it clear that there is no direct proportionality between sales and profits as South made considerable high sales but turned in really low profits when compared to its loss.

6. Which Region Has Seen the Most Superstore Losses?

- It is obvious that "**Central**" region's Superstores have contributed the majority of losses, amounting to over 200kUSD incurred by Superstore.

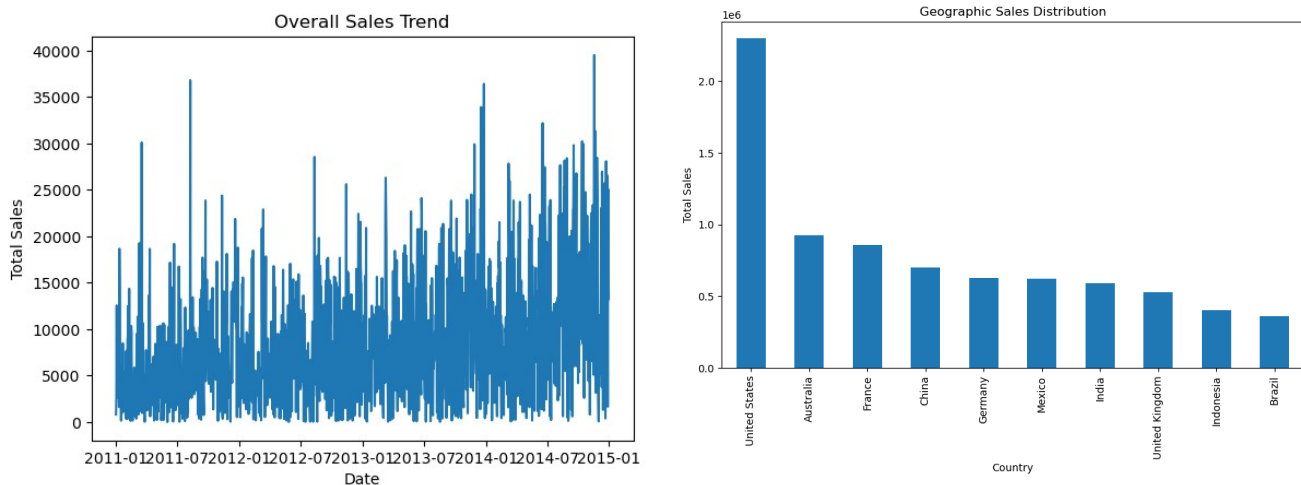
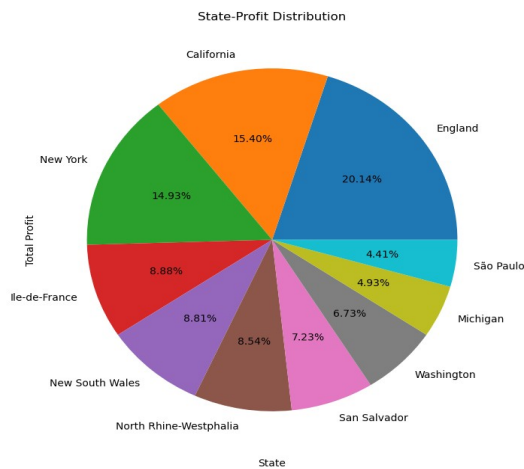


Figure 5: (a) Overall sales trend. (b) Geographic “country” sales distribution.



(c) State-Profit distribution pie chart.

7. Sales :-

- What is the overall sales trend for the supermarket?** The overall sales trend for the supermarket is positive or upward. This means that the sales are consistently growing or improving over the analyzed period.
- There is a huge disparity in the number of sales between the highest and lowest performing Countries As well as between States, Superstore has the highest profits in **England** and then the state of **California** (as shown in fig. 5c) with 20.14% and 15.40% respectively, but vice versa country-wise with **United States** outselling **United Kingdom** with over 1.5x1e6USD in the total sales generated by Superstore.

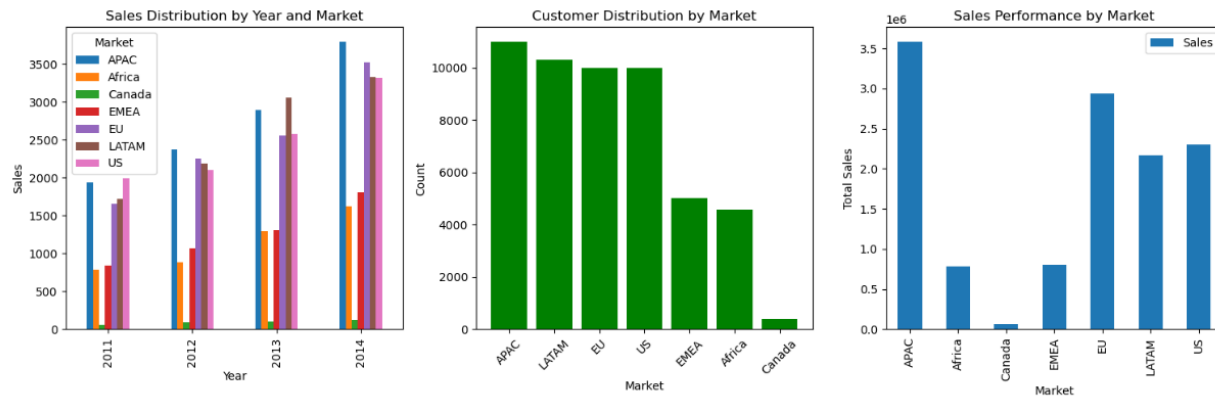


Figure 6: (a) Sales distribution by Year and Market. (b) Customer distribution by Market. (c) Sales performance by Market.

8. Market Analysis :-

- In every year, there is a somewhat steady increase in the general market cap for Superstore. And this also translated to profits as it was earlier shown that the trend was upward or positive.
- The population of people in the **APAC** market outweighs or outnumbers everywhere else. With the **Canada's** market yielding the least customer count, this data also helps it become clear why **Canada** has the least profits among all sales made by Superstore.
- Another interesting find is that, **LATAM** market being populous in terms of customer count than the **EU** market but not outperforming the **EU** when sales was considered.

Some other random information discovered during the course of the analysis are given below:

- The Number of one-time customers was discovered to be 17.
- The average shipping cost distribution for top 3 different countries with the highest average shipping cost are given as **Taiwan** 155.6, **Chad** 148.9, and **Lesotho** with 135.6.
- Some of the customers with the highest profit values are Tamara Chand, Raymond Buch, Hunter Lopez, Adrian Barton, and Sanjit Chand.

Conclusion

Exploratory Data Analysis (EDA) on the "Superstore" dataset was performed. A Jupyter Notebook file is accompanied with this submission containing all answers to the 9 questions asked. Some of which have been further described above with additional information on sales, market, profit, loss, discount and other valuables.