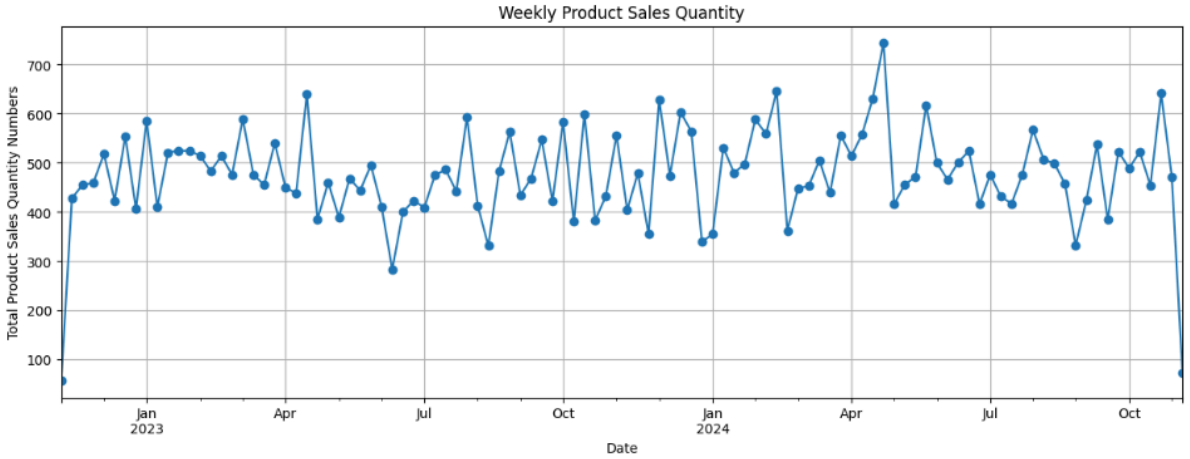
In the assignment, firstly, both data sets were examined in detail.

Only some graphs and their results will be explained in this report. Because all the codes in the assignment were explained in detail in the comment lines. Since Jupyter Notebook is used, all graphics and comments are included in the codes.

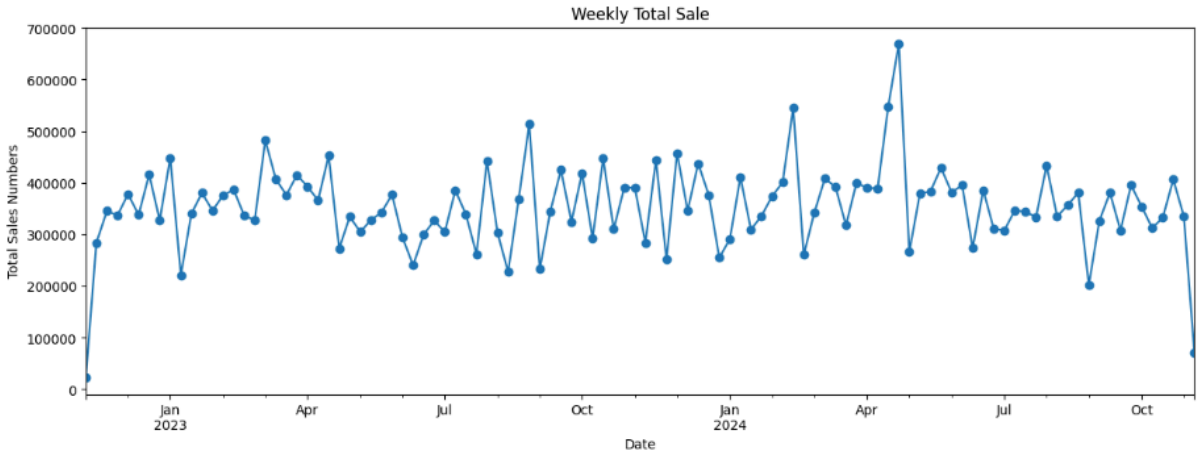
TIME SERIES ANALYSIS

- Weekly Product Sales Quantity



According to these graph, Highest sales week is 2024-04-28 and highest product quantity sales amount is 744. Lowest sales week is 2022-11-06 and lowest sales amount is 57
It was observed that other values were generally between 650 and 300.

- Weekly Total Sales



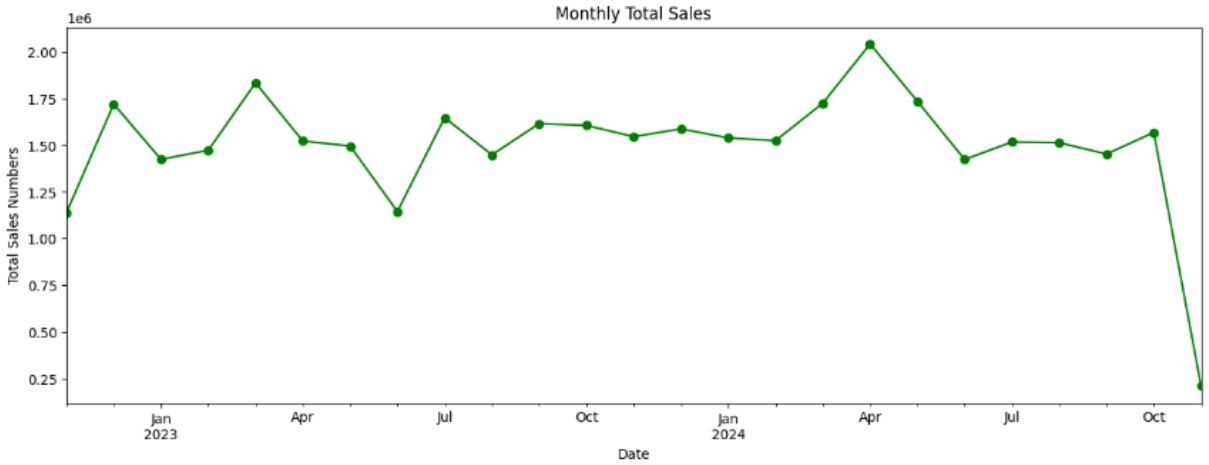
According to these graph, highest total sales week is 2024-04-28 and highest total sales amount is 668963.85. Lowest sales week is 2022-11-06 and lowest sales amount is 22763.97

According to these results, it was seen that the total weekly sales amount and product number results were directly proportional to each other.

- Monthly Product Sales Quantity

According to these graph, highest total product sales month is 2024-04-30 and amount is 2539. Lowest total product sales month is 2024-11-30 and amount is 251. Unlike these values, monthly product sales are between 1800 and 2500, except for June.

- Monthly Total Sales



According to these graph, highest total product sales month is 2024-04-30 and amount is 2040868.98. Lowest total product sales month is 2024-11-30 and amount is 211246.16.

There is also a direct proportion between monthly product sales and total sales.

The numbers make this situation certain, but when you look at the graphs, the weekly product-total sales / monthly product-total sales graphs are the same. This shows that there is a direct proportion not only in the maximum and minimum values, but in all other week / month values.

Sales have always decreased in the last month and weeks. Therefore, the number of days in the last month was checked. The last month is 29 days. So the reason for the decrease is not the lack of data in the month.

Categorical and Numerical Analysis

According to the results in this section, the electronics category constitutes the majority of sales. It is the highest sales category in every age group. With a percentage of 48.51, this category constitutes almost half of the sales.

The highest product sales were in the 50+ group. In general, sales rates decreased as the age group decreased.

When we look at the total sales and spending amounts, it is seen that women buy more products. Except for the 26-35 age group, women always spent more.

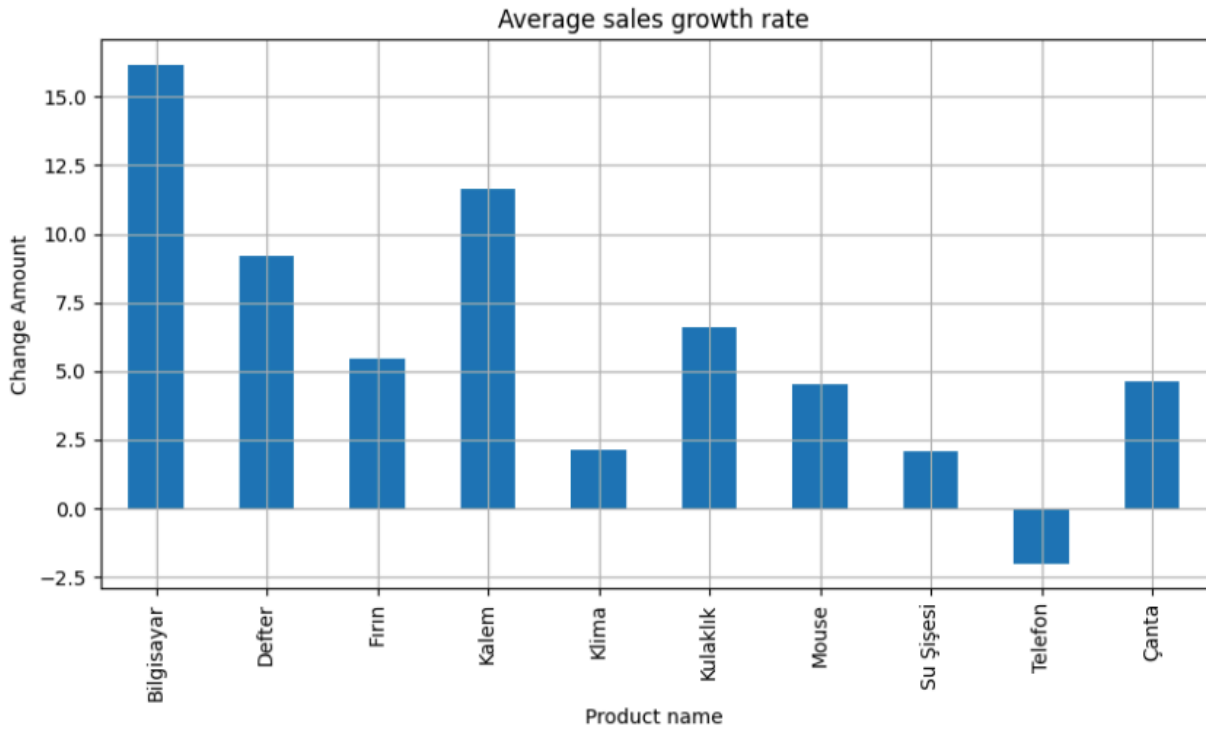
The product women buy the most is phones, while men buy bags.

General comment;
Considering that there are customers interested in electronics, increasing product advertisements on social media can be a method to increase sales.

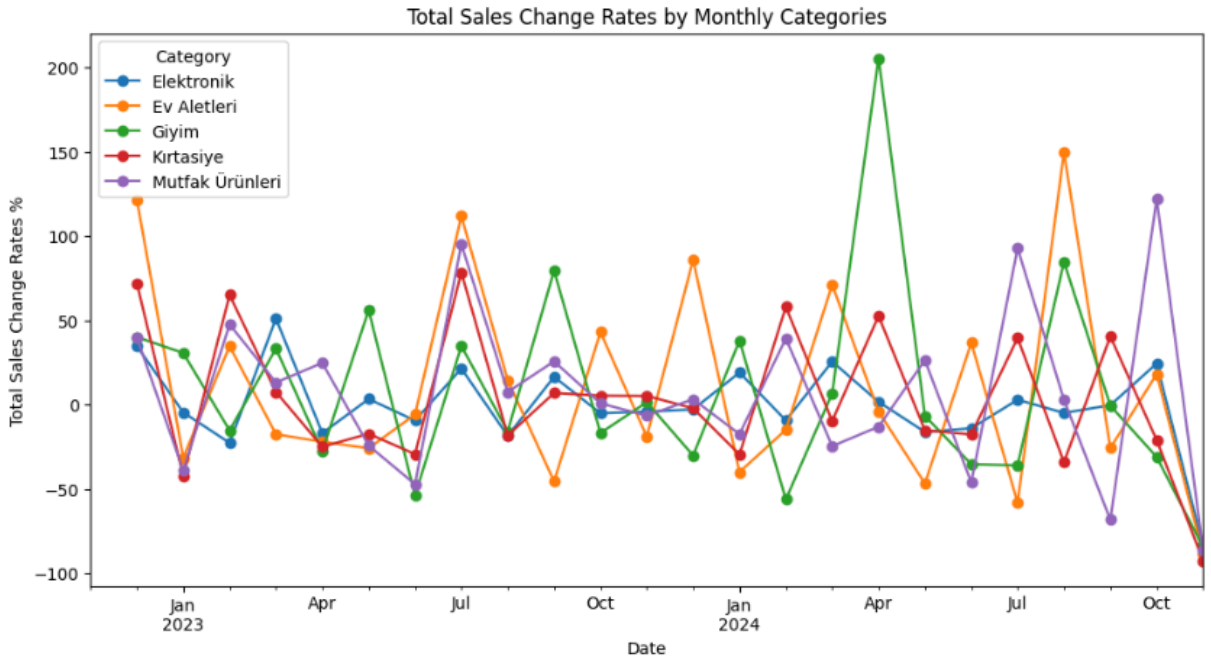
Women generally like to shop more. Therefore, it would be more effective to increase marketing to the 26-35 age group.

Advertisements targeting both women and men should be increased in the 18-25 age group. Advertising activities can be carried out to increase product sales targeting the needs of young people.

Advanced Data Manipulation

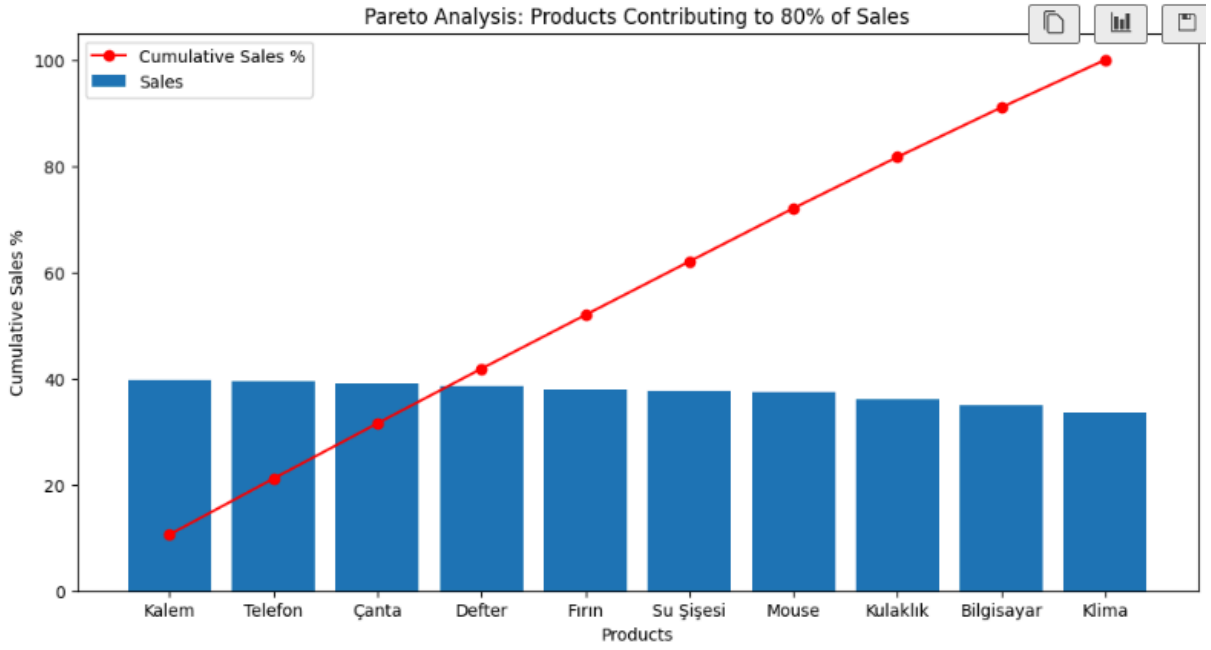


Although the highest product sales in every category are electronics, there is a decrease in the monthly percentage of phone sales. The reason for the decrease in phone sales may be; decrease in demand, seasonal reasons, price changes and increase in competition. Since there are rapid developments in technological products such as phones, a negative change may occur.
Advertisements should be made to increase sales for other products that are below 5%, primarily phones



This chart shows monthly changes by category. Separate charts were created for each in the code and made clearly visible.

PARETO ANALYSIS

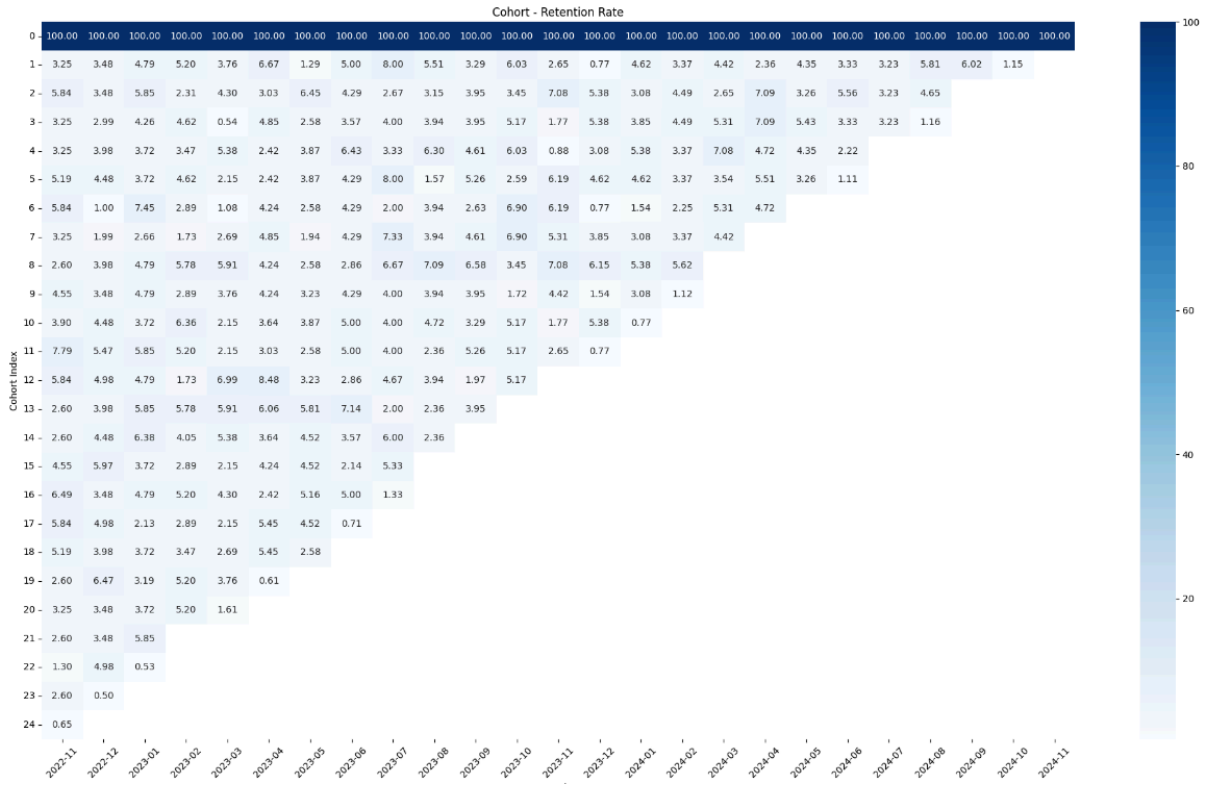


The sales amount of each product is shown in the blue columns. The number of sales is listed in increasing order from left to right.

The cumulative sales line is represented by the red line. The slope formed by this line indicates which product has the most total sales. The steeper parts of the slope mean a greater contribution to sales.

A strategy change can be made according to the highest and lowest selling products. New marketing strategies can be developed to increase lower selling products.

COHORT ANALYSIS



Each row represents a group of users who started using the service for the first time. For example, cohort index 0 (2022-11) represents users who started using the service in November 2022. Cohort index 1 (2022-12) represents users who arrived in December 2022. This is how users are grouped.

In November 2022, the first buyers returned 100%, 1 month later 3.25%. There is a cross-relationship between the data

In the 2023-04 month, there is 8.48% in the 12th cohort, which is a very high value. If there was a campaign this month, other retentions corresponding to this month would also be high. So, there was no campaign this month. It was thought that people returned after 1 year. February was examined to measure the accuracy of this thesis. Those who bought the product in the 2023-02 month did not return after 1 year. So, there is no such situation.

Those who bought the first product in the 2023-07 month seem to be more satisfied than others.

Healthy interpretations cannot be made by saying that retention is low in recent months.

General comment:

Even if campaigns are made, they are generally not successful. Because even if retention are high, their continuity is low. There is no stability, if the customer likes it, they return.

PREDICTION MODEL - LINEAR REGRESSION

The necessary explanations about the structure of the codes are made in the comment lines.

```
Mean Squared Error (MSE): 9725383991.79
R-squared (R²): -0.01
```

Since the regression model was used, MSE and R-squared evaluations were made instead of accuracy, precision and recall values.

As a result of model training, MSE being very high and R-squared being negative indicates that the model is not making correct predictions. Normally, various steps could be taken to increase model accuracy. Feature engineering steps can be applied and model parameters can be optimized.

Since the values in the data set are randomly generated, these accuracy values can be called normal. Since the data set is generated with random values, the model cannot be expected to perform well.