



Superstore Sales Analysis

Fajrur Rahman Suprpto
fajrur.rs@gmail.com

Outline

- Sales Analysis
 - About Dataset
 - Business Question
 - Data preprocessing
 - Exploratory Data Analysis and Visualization
 - Recommendation
-

Sales Analysis

Many people say that sales analysis is an important part of marketing management.

Sales analysis is an activity of analyzing or arranging information about market condition.

About Dataset

This dataset contain behavior data on January 2015 - December 2018 from a global superstore. There are 18 columns and 9,800 rows in this dataset. The dataset can be found on [Kaggle](#).

```
RangeIndex: 9800 entries, 0 to 9799
Data columns (total 18 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Row ID              9800 non-null   int64
1   Order ID            9800 non-null   object
2   Order Date          9800 non-null   object
3   Ship Date           9800 non-null   object
4   Ship Mode            9800 non-null   object
5   Customer ID         9800 non-null   object
6   Customer Name       9800 non-null   object
7   Segment             9800 non-null   object
8   Country             9800 non-null   object
9   City                9800 non-null   object
10  State               9800 non-null   object
11  Postal Code         9789 non-null   float64
12  Region              9800 non-null   object
13  Product ID          9800 non-null   object
14  Category            9800 non-null   object
15  Sub-Category        9800 non-null   object
16  Product Name        9800 non-null   object
17  Sales               9800 non-null   float64
dtypes: float64(2), int64(1), object(15)
```

Business Question

This project focus on 5 questions below:

- Total sales by date
 - Sales trends/monthly sales
 - Total sales by product
 - Total sales by location
 - Customers with highest number of purchase
-

Data Preprocessing

Remove Duplicates

There are no duplicates row in this dataset.

Drop Columns

Drop Row ID column, because we won't use it.

Check missing value

There are 11 records with NaN values in Postal Code. Because it won't affect our sales analysis, we will ignore them.

Data Preprocessing

Convert columns to datetime format

Convert Order Date and Ship Date column to datetime format.

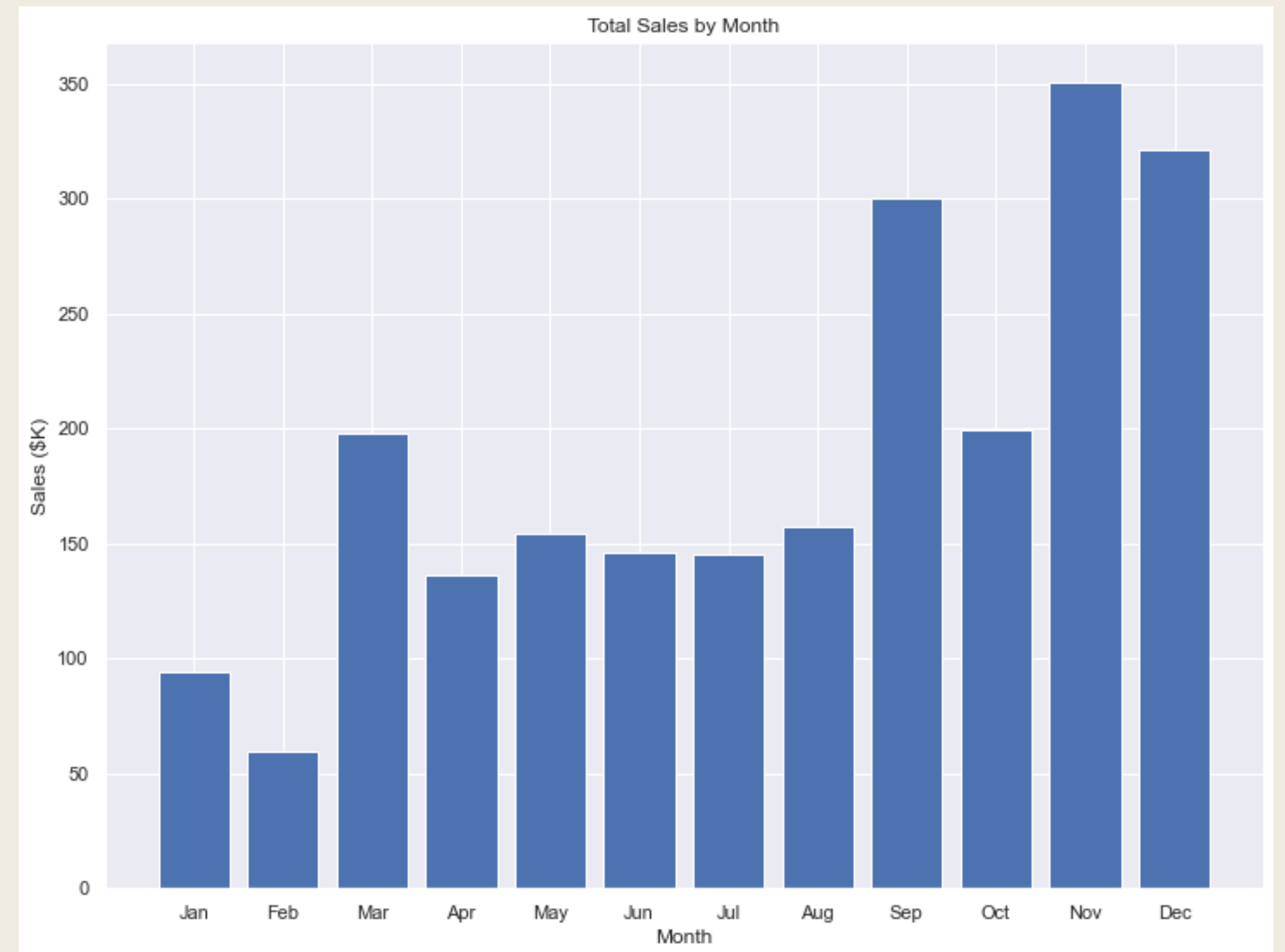
Add columns for analysis

Add Day, Month, Year, and Sales (K) to new column.

EDA and Visualization

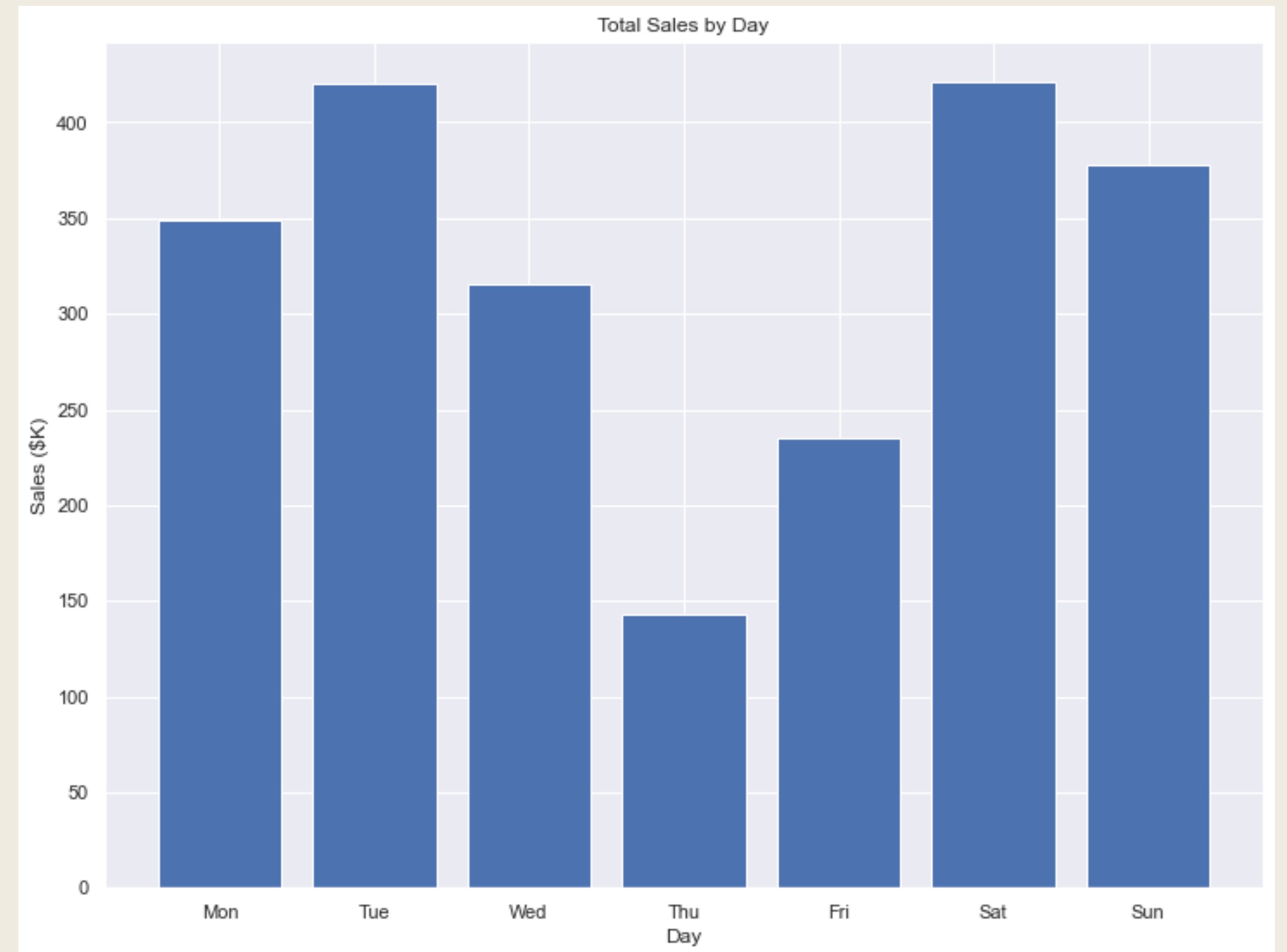
September, November, and December bring the highest sales. November and December may be related to Christmas.

Otherwise, February brings the lowest sales.



EDA and Visualization

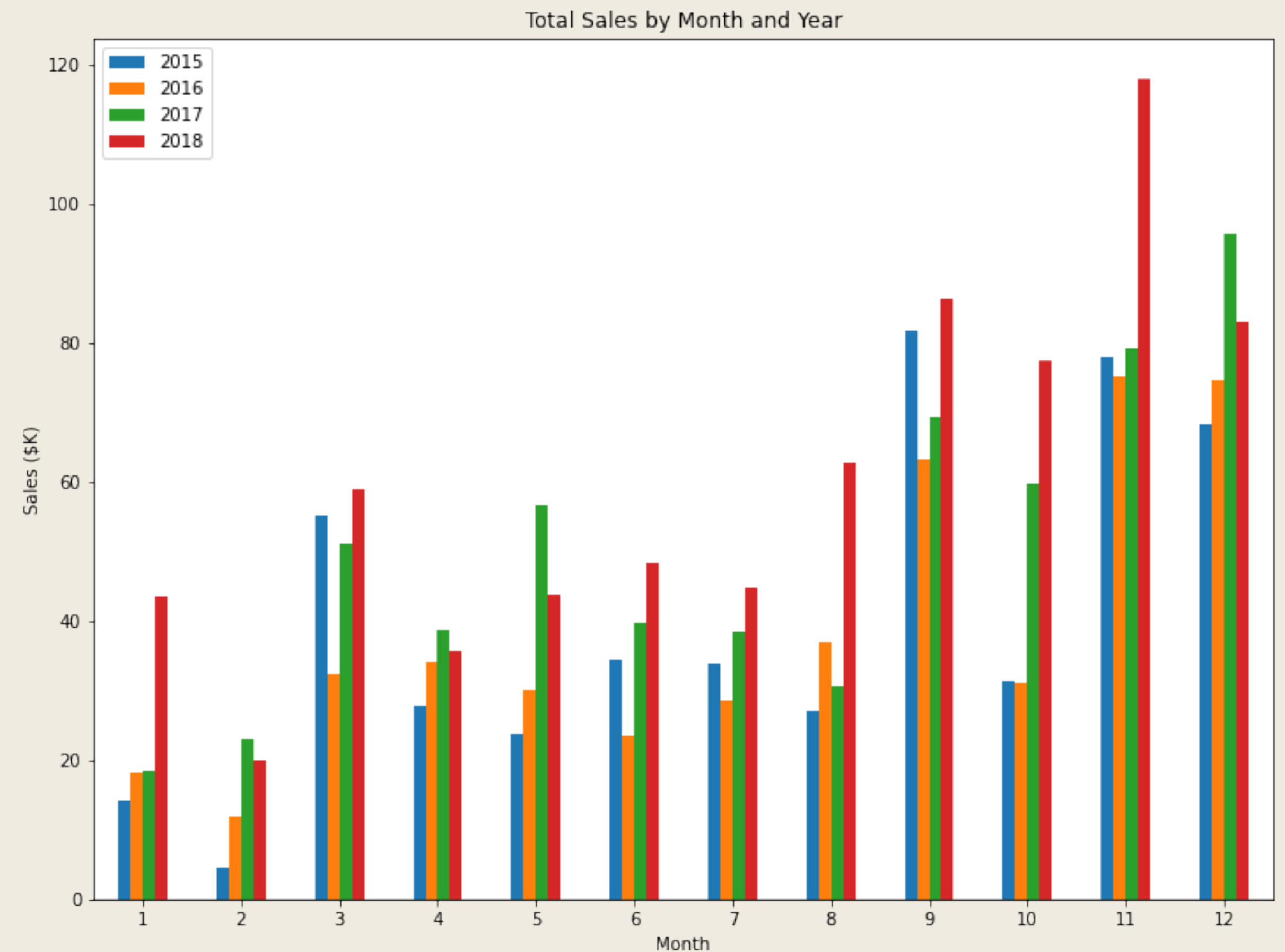
Saturday generates the highest sales, followed by Tuesday.



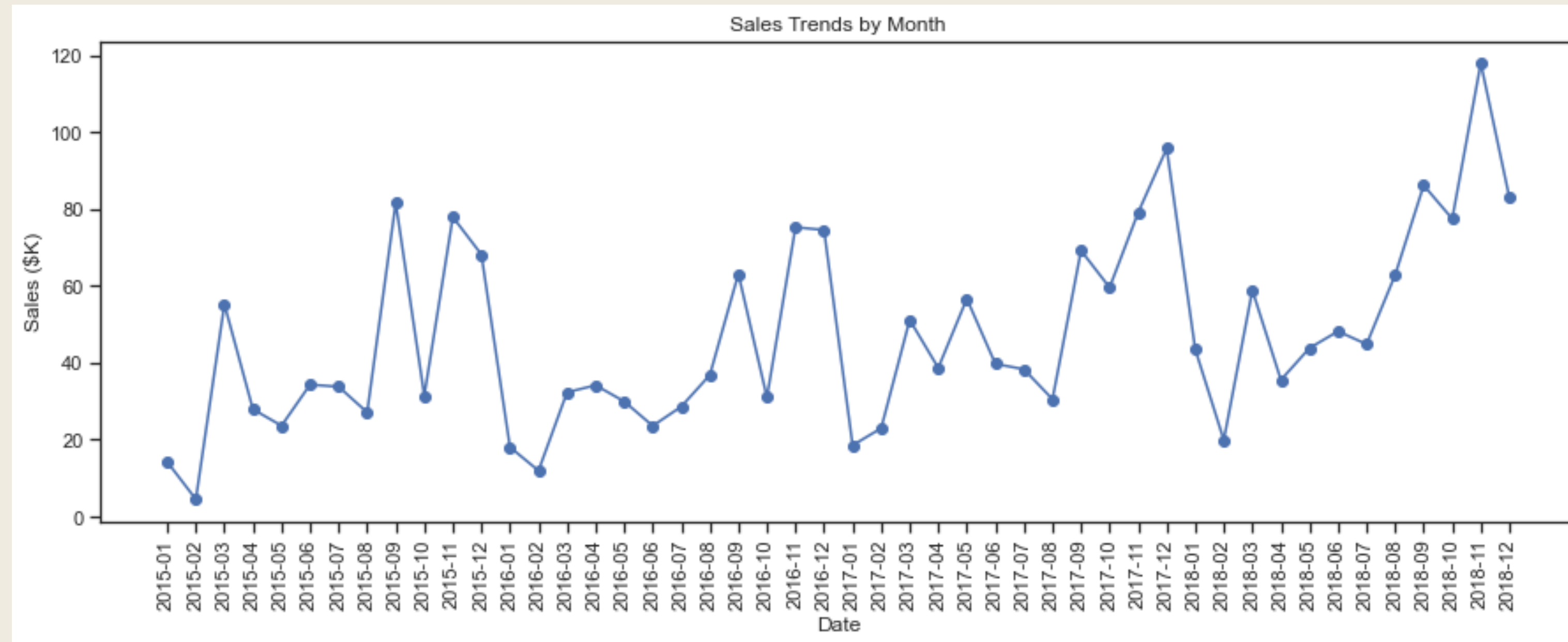
EDA and Visualization

Monthly sales are highest in September, November, and December.

Otherwise, January and February are the lowest monthly sales.



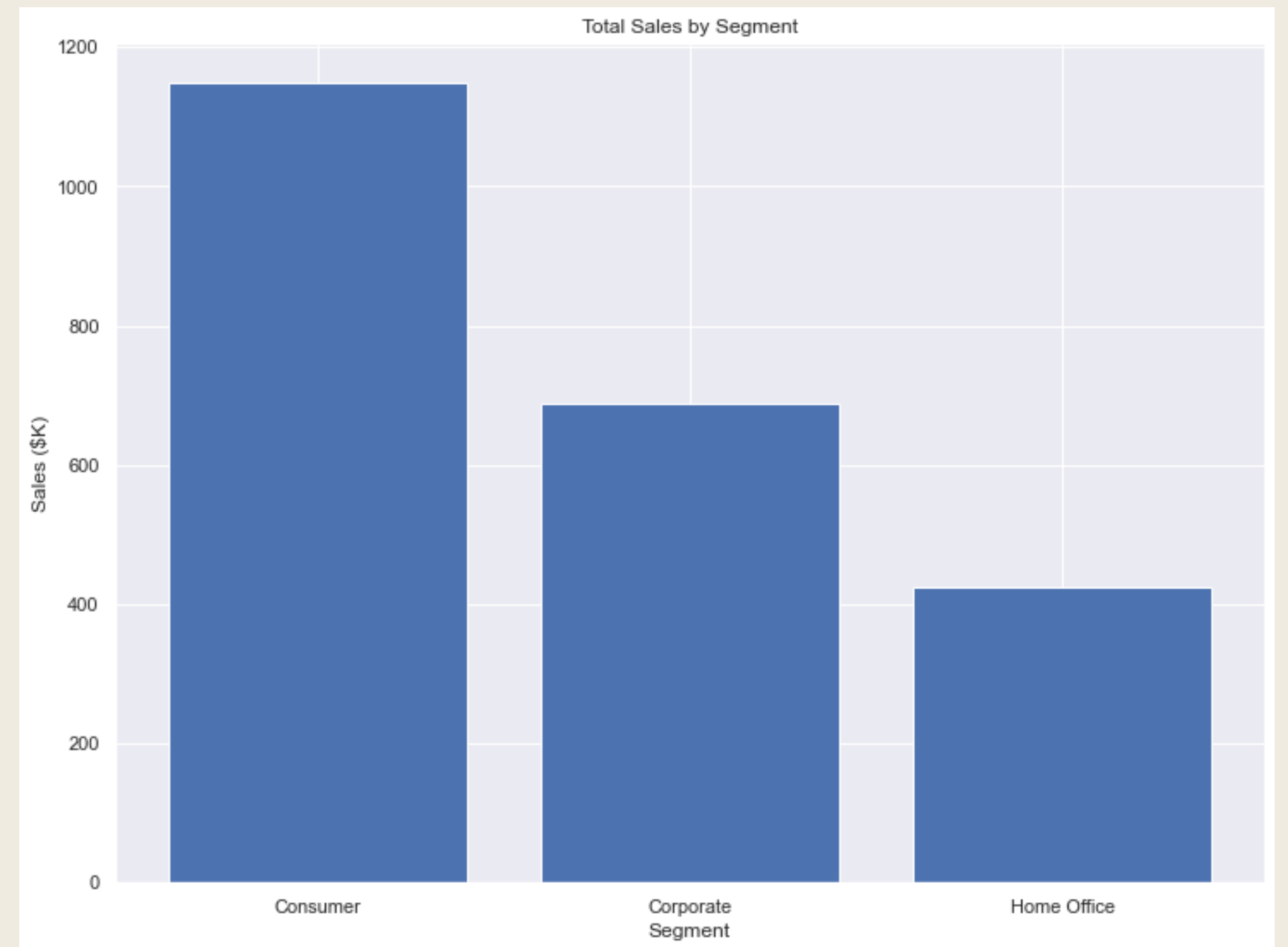
EDA and Visualization



There was an increase and decrease in sales for 4 years. The lowest sales is in February 2015, and the highest sales is in November 2018. There was always a decrease in sales in December to January.

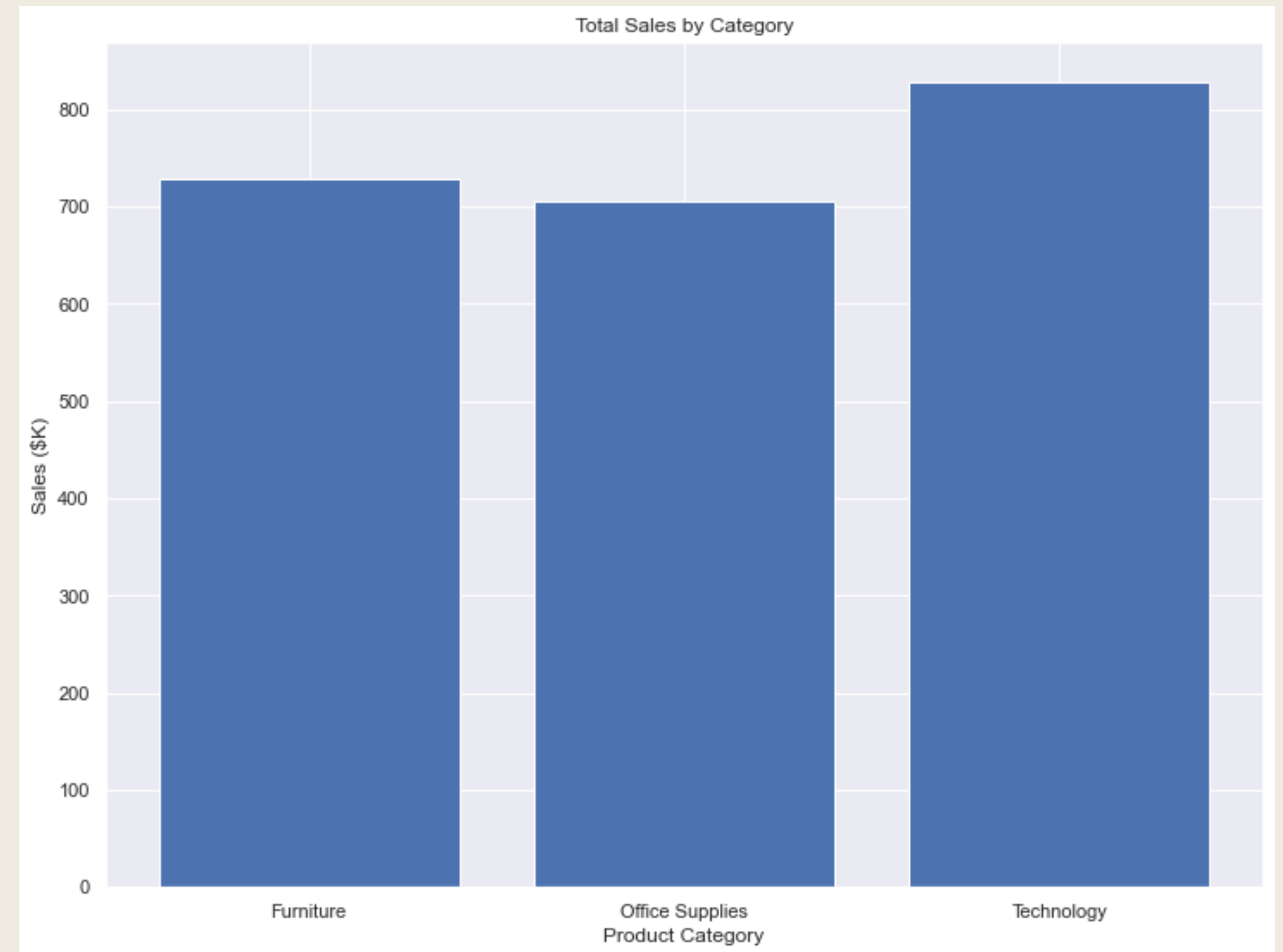
EDA and Visualization

Consumer segment bring the highest sales. Home Office bring the lowest sales, it may due to the lack of promotion in Home Office segment.



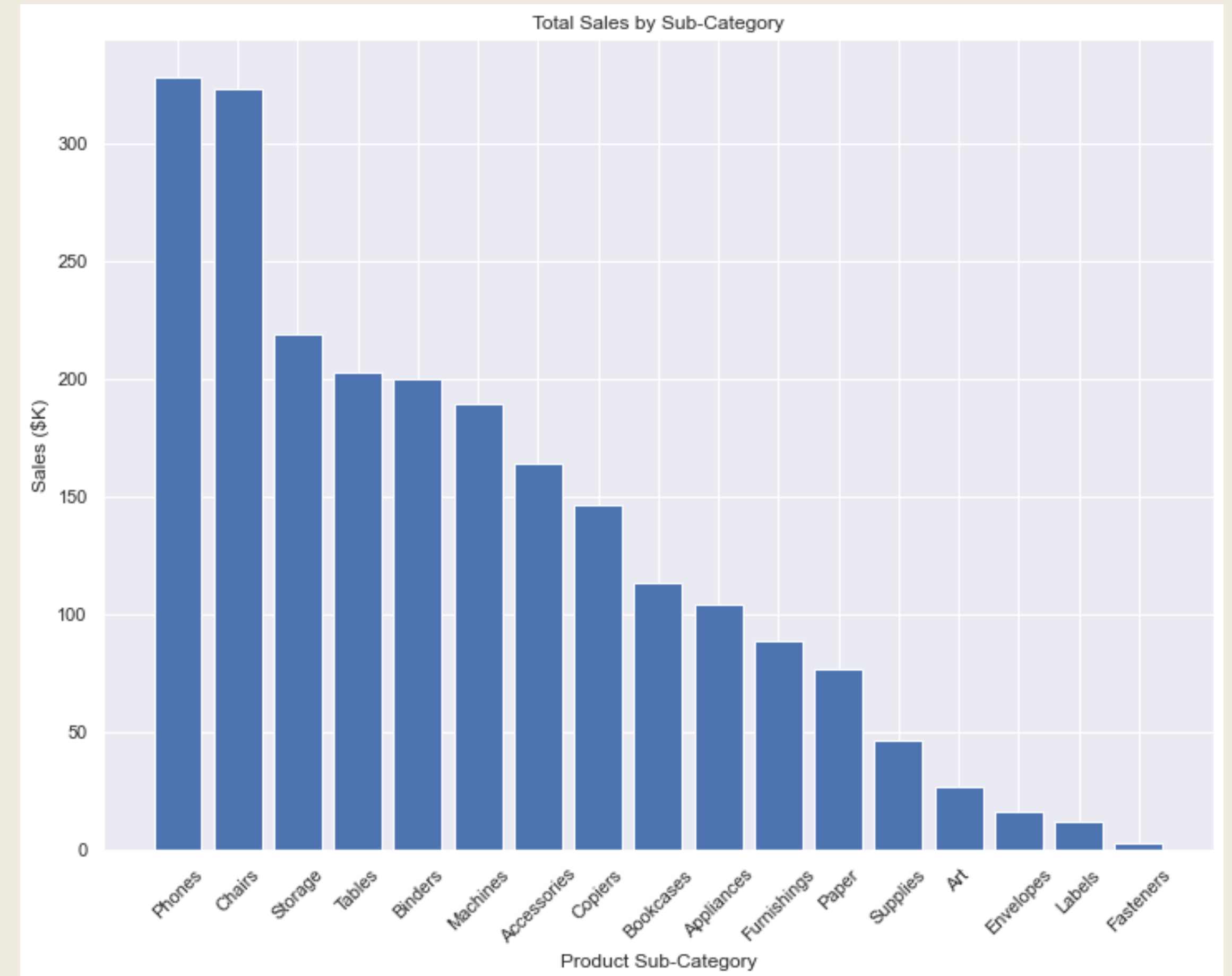
EDA and Visualization

The highest sales is products on Technology category, for more than \$ 800,000 sales.

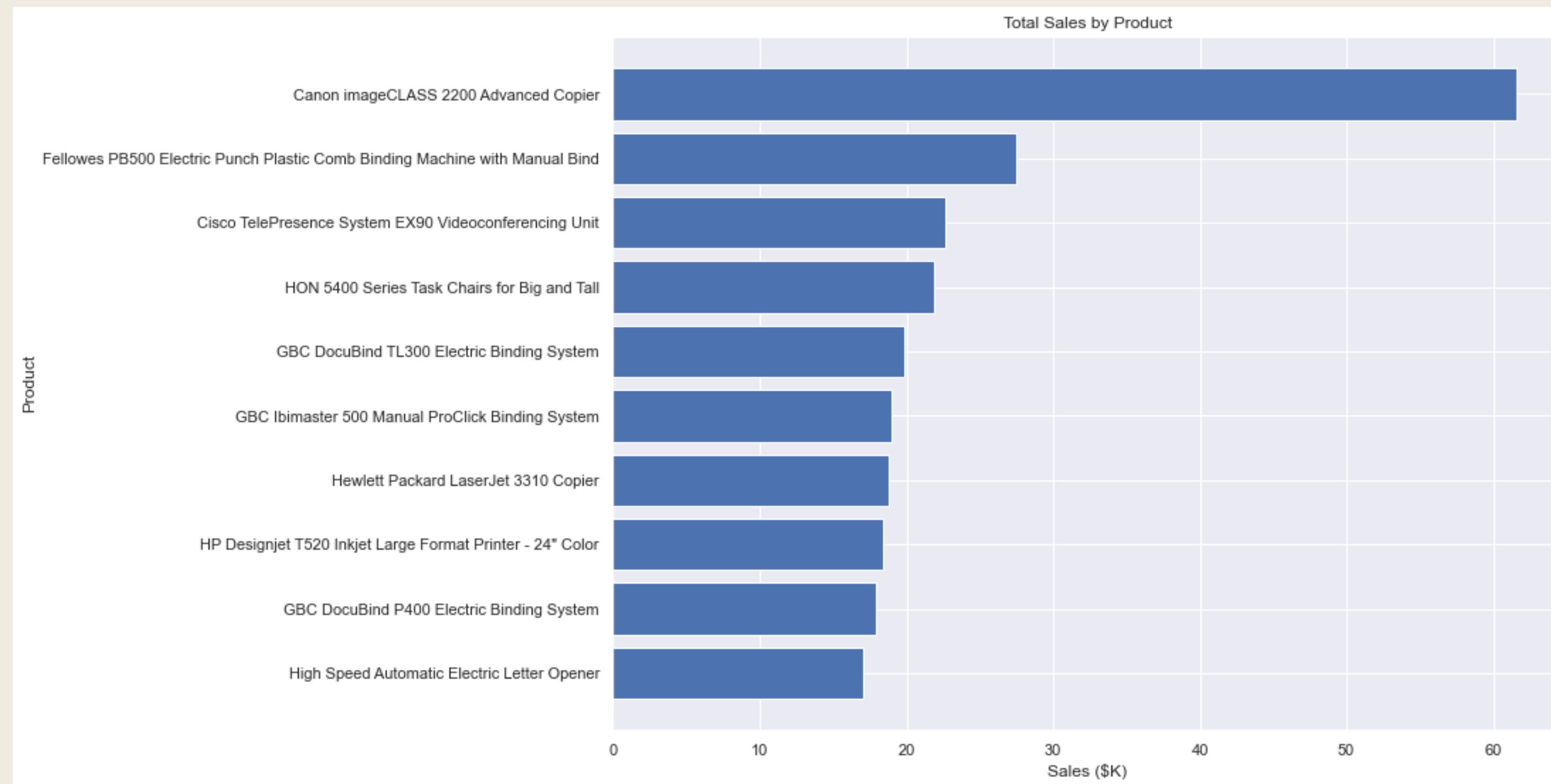


EDA and Visualization

There are 17 sub-categories in this store. Phones become the highest sales, followed by chairs.



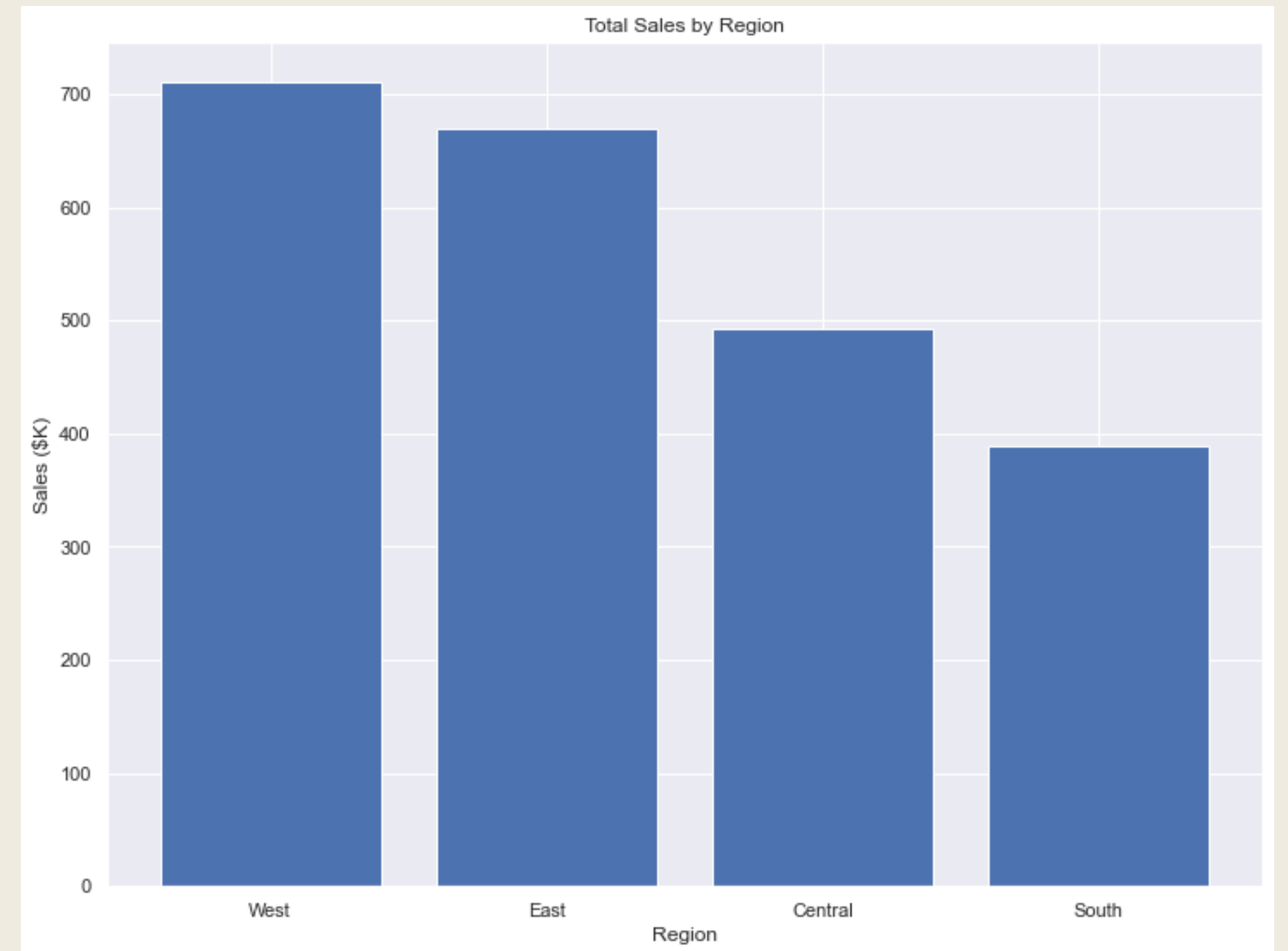
EDA and Visualization



There are 10 products with highest sales. Canon imageCLASS 2200 Advanced Copier generated the highest sales for about \$ 61,000.

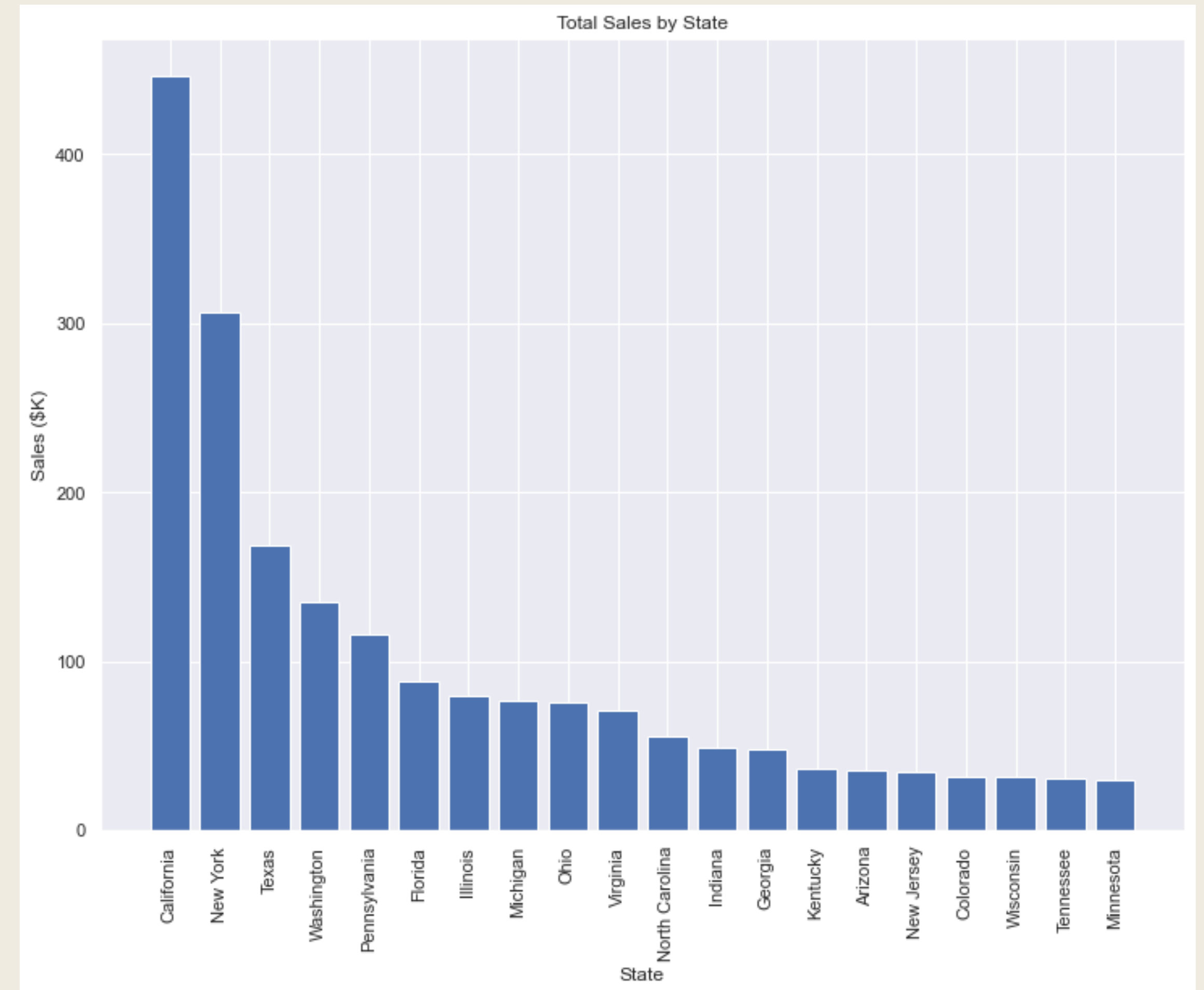
EDA and Visualization

West Region is the highest sales,
and South Region is the lowest
sales.



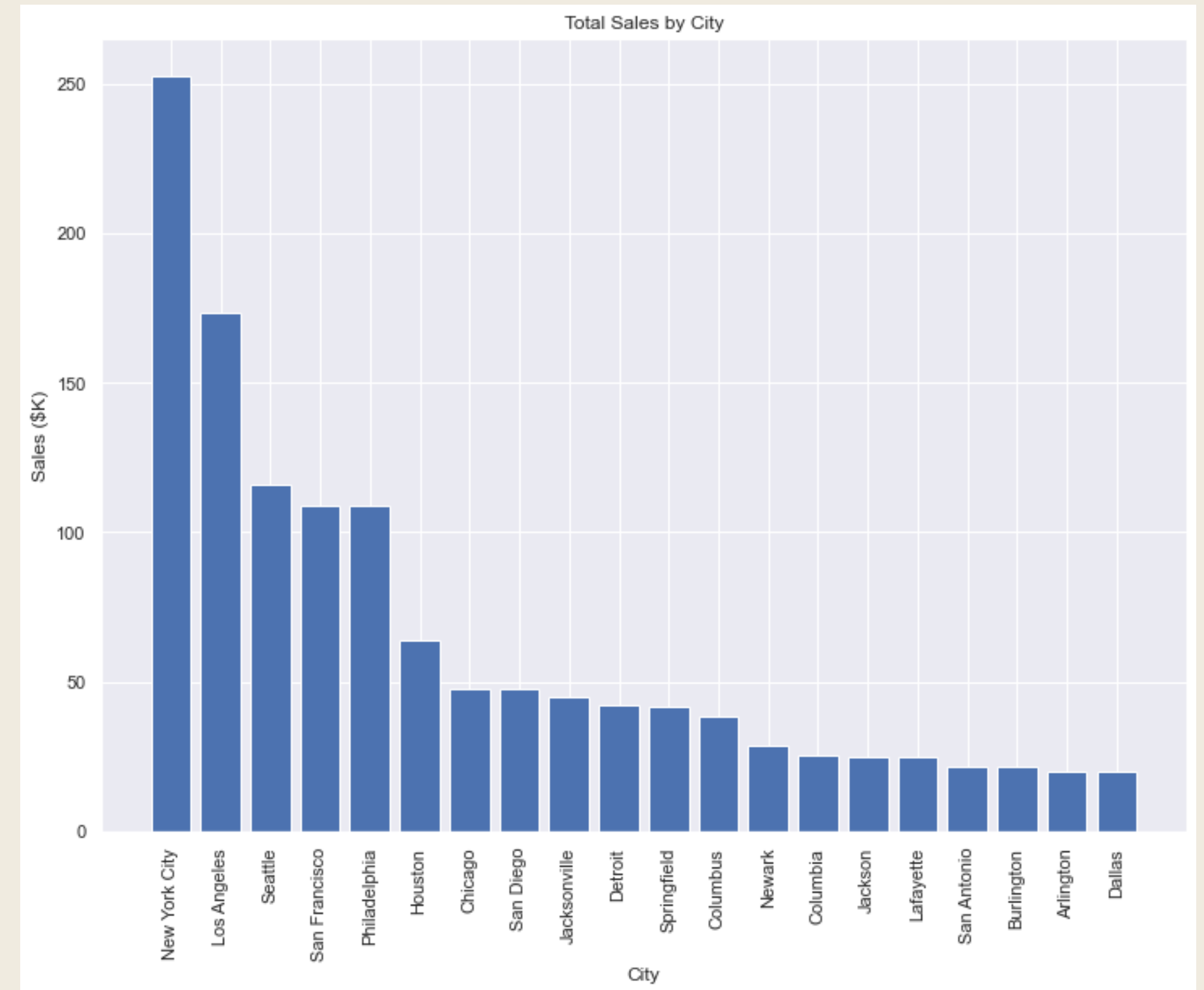
EDA and Visualization

There are 20 states with highest sales. Most of the total sales come from California and New York.



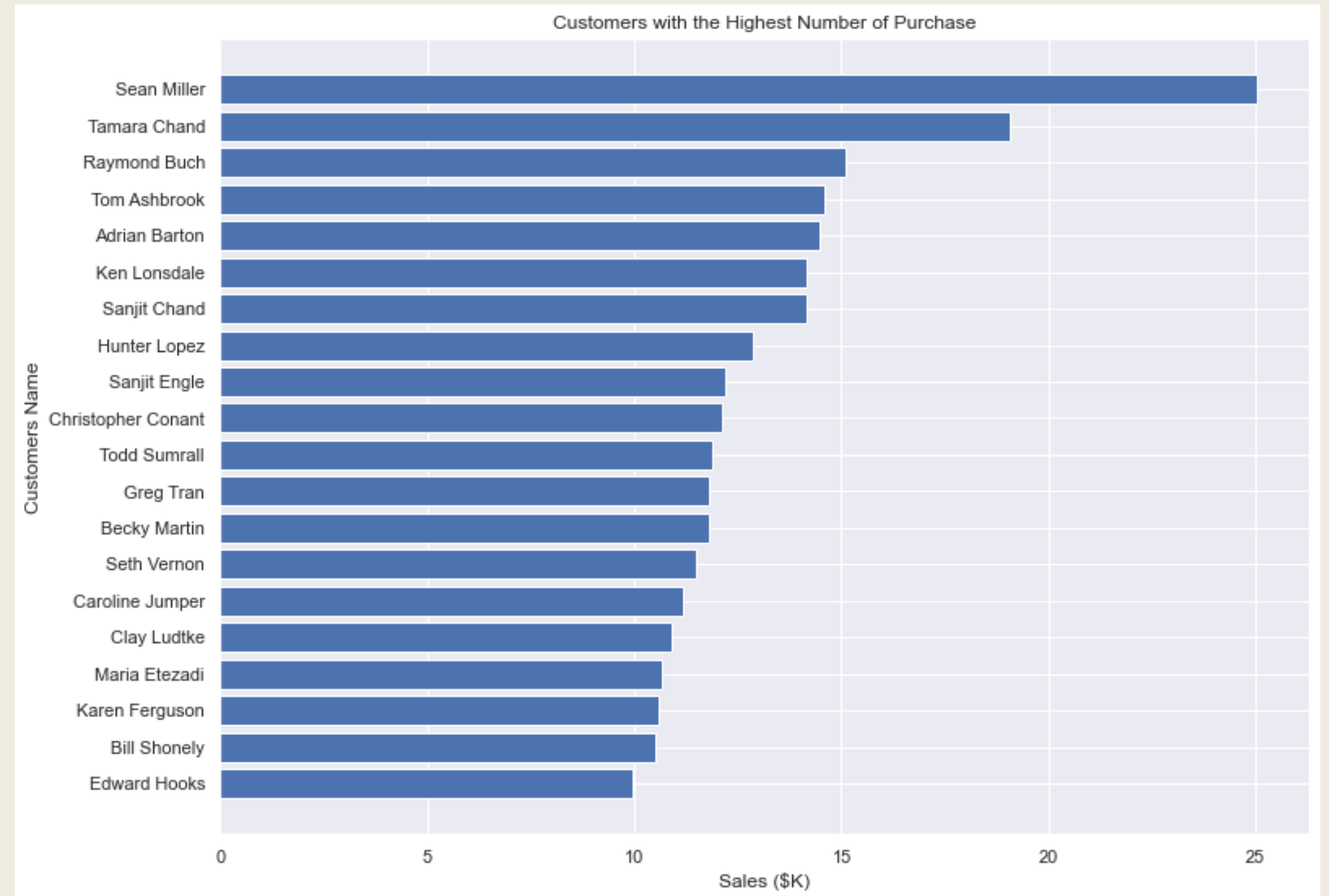
EDA and Visualization

There are 20 cities with highest sales. New York City become the most highest sales, followed by Los Angeles, and Seattle.



EDA and Visualization

There are 20 customers with the most purchases products. These customers buy more or higher-value products than the other customers.



Recomendation

After doing analysis and visualization, we can observe that:

- Make interesting offers at the beginning of the year
 - Increase the promotion on Thursday
 - More focusing on Home Office segment to increase the revenue
 - Maintain stock of Technology category, which make up the majority of sales
 - Make a bundling for category that have the low sales
 - Make promotion or discount for shipping cost to South region
 - Give loyalty program to consumer with the most highest purchase
-