

Superstore study case -Retail Data Analytic-

Harish Muhammad

DQLab Live Class - Data Analyst Bootcamp
Python & SQL with Google Platform



Contents

- Background & Business understanding
- Analysis Process
- Data Understanding
- Data preprocessing (data cleansing)
- Data Analysis
- Insights and recommendations
- Data visualization (Dashboard)

Background

- Superstore is a retail chain that operates across various regions in the United States. It sells a wide range of products categorized under Furniture, Office Supplies, and Technology.
- The Superstore dataset provides a comprehensive view of sales and performance metrics for a retail superstore over a certain period. The dataset highlight information on orders, products, customers, and shipping details, etc.

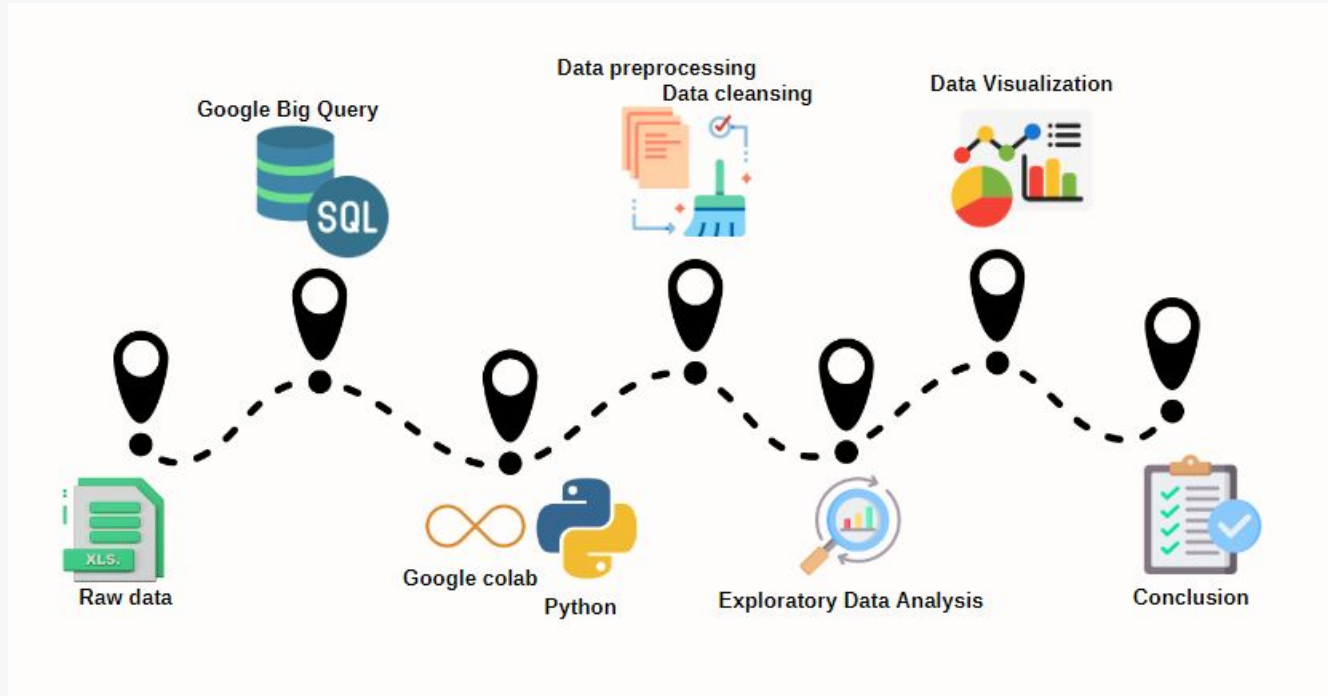


Objective

- The primary objective of our analysis is to uncover insights across key factors, offering actionable recommendations to enhance the Superstore's operations.
- We will delve into total order patterns, evaluate sales and profit trends, gain geographical insights, and analyze product performance to inform decision-makers in Superstore management about potential areas for improvement and optimization



Analysis process



Data Understanding

Data Understanding

The Superstore dataset contains

- 9994 rows
- 20 columns
- Mix of numerical and categorical data

```
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 20 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Order_ID            9994 non-null   object
1   Customer_ID         9994 non-null   object
2   Postal_Code         9994 non-null   int64
3   Product_ID          9994 non-null   object
4   Sales               9994 non-null   float64
5   Quantity            9994 non-null   int64
6   Discount            9994 non-null   float64
7   Profit              9994 non-null   float64
8   Category            9994 non-null   object
9   Sub-Category        9994 non-null   object
10  Product_Name        9994 non-null   object
11  Order_Date          9994 non-null   object
12  Ship_Date           9994 non-null   object
13  Ship_Mode           9994 non-null   object
14  Customer_Name       9994 non-null   object
15  Segment             9994 non-null   object
16  Country/Region      9994 non-null   object
17  City                9994 non-null   object
18  State               9994 non-null   object
19  Region              9994 non-null   object
dtypes: float64(3), int64(2), object(15)
```

Data Understanding

	Order_ID	Customer_ID	Postal_Code	Product_ID	Sales	Quantity	Discount	Profit	Category	Sub-Category	Product_Name	Order_Date	Ship_Date	Ship_Mode	Customer_Name	Segment
0	CA-2019-152156	CG-12520	42420	FUR-BO-10001798	261.9600	2	0.00	41.9136	Furniture	Bookcases	Bush Somerset Collection Bookcase	11/8/2019	11/11/2019	Second Class	Claire Gute	Consumer
1	CA-2019-152156	CG-12520	42420	FUR-CH-10000454	731.9400	3	0.00	219.5820	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs,....	11/8/2019	11/11/2019	Second Class	Claire Gute	Consumer
2	CA-2019-138688	DV-13045	90036	OFF-LA-10000240	14.6200	2	0.00	6.8714	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters b...	6/12/2019	6/16/2019	Second Class	Darrin Van Huff	Corporate
3	US-2018-108966	SO-20335	33311	FUR-TA-10000577	957.5775	5	0.45	-383.0310	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	10/11/2018	10/18/2018	Standard Class	Sean O'Donnell	Consumer
4	US-2018-108966	SO-20335	33311	OFF-ST-10000760	22.3680	2	0.20	2.5164	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	10/11/2018	10/18/2018	Standard Class	Sean O'Donnell	Consumer
5	CA-2017-115812	BH-11710	90032	FUR-FU-10001487	48.8600	7	0.00	14.1694	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Access...	6/9/2017	6/14/2017	Standard Class	Brosina Hoffman	Consumer

- **Order ID** : Unique Order ID for each Customer.
- **Customer ID** : Unique ID to identify each Customer.
- **Postal Code** : Postal Code of every Customer.
- **Product ID** : Unique ID of the Product.
- **Sales** : Sales of the Product.
- **Quantity** : Quantity of the Product.
- **Discount** : Discount provided.
- **Profit** : Profit/Loss incurred.
- **Category** : Category of the product ordered.
- **Sub-Category** : Sub-Category of the product ordered.
- **Product Name** : Name of the Product.
- **Order Date** : Order Date of the product.
- **Ship Date** : Shipping Date of the Product.
- **Ship Mod** : Shipping Mode specified by the Customer.
- **Customer Name** : Name of the Customer.
- **Segment** : The segment where the Customer belongs.
- **Country** : The Country of residence of the Customer.
- **City** : The City where Customer live.
- **State** : The state of residence of the Customer.
- **Region** : The Region where the Customer belong.

Data Cleansing

Descriptive statistics of numerical data

	Sales	Quantity	Discount	Profit
count	9994.000000	9994.000000	9994.000000	9994.000000
mean	229.858001	3.789574	0.156203	28.656896
std	623.245101	2.225110	0.206452	234.260108
min	0.444000	1.000000	0.000000	-6599.978000
25%	17.280000	2.000000	0.000000	1.728750
50%	54.490000	3.000000	0.200000	8.666500
75%	209.940000	5.000000	0.200000	29.364000
max	22638.480000	14.000000	0.800000	8399.976000

Descriptive statistics elaborate the statistical features from Superstore dataset

Descriptive statistics of non-numerical data

	Order_ID	Customer_ID	Product_ID	Category	Sub-Category	Product_Name	Ship_Mode	Customer_Name	Segment	Country/Region	City	State	Region
count	9994	9994	9994	9994	9994	9994	9994	9994	9994	9994	9994	9994	9994
unique	5009	793	1862	3	17	1817	4	793	3	1	531	49	4
top	CA-2020-100111	WB-21850	OFF-PA-10001970	Office Supplies	Binders	Staple envelope	Standard Class	William Brown	Consumer	United States	New York City	California	West
freq	14	37	19	6026	1523	48	5968	37	5191	9994	915	2001	3203

Handling missing value

```
# Calculating the percentage of the missing values to all data  
df.isnull().sum() / len(df) * 100
```

```
Order_ID      0.0  
Customer_ID   0.0  
Product_ID    0.0  
Sales         0.0  
Quantity      0.0  
Discount      0.0  
Profit        0.0  
Category      0.0  
Sub-Category  0.0  
Product_Name  0.0  
Order_Date    0.0  
Ship_Date     0.0  
Ship_Mode     0.0  
Customer_Name 0.0  
Segment       0.0  
Country/Region 0.0  
City          0.0  
State         0.0  
Region        0.0  
dtype: float64
```

Calculating the number or percentage of missing value "Nan"

Exploratory Data Analysis (EDA)

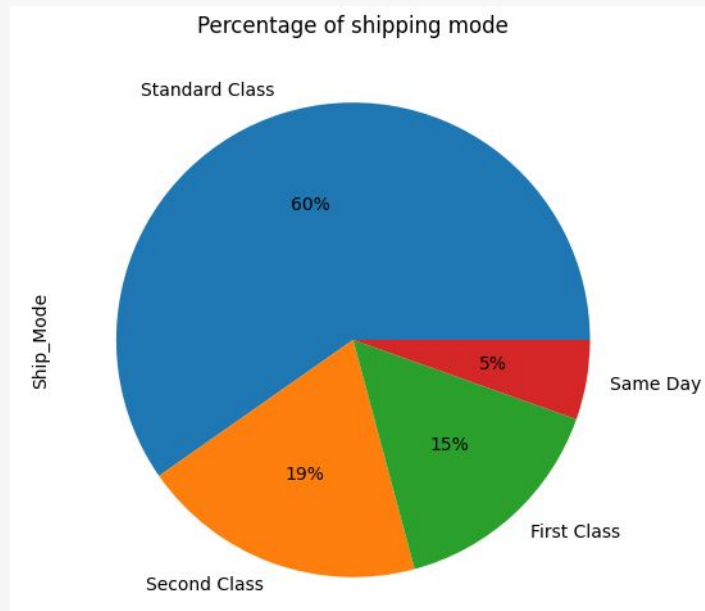
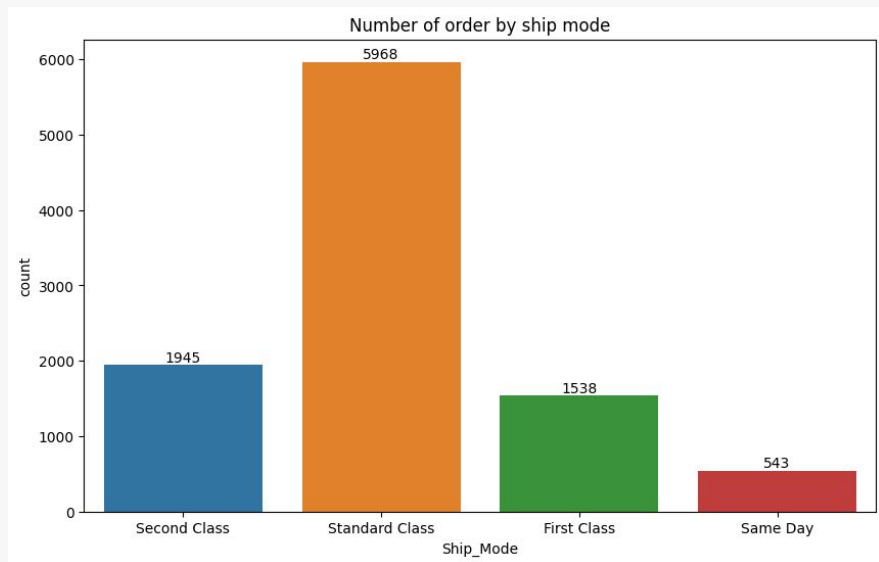
Order based analysis

Sub-themes:

- Shipping mode
- Market segment
- Category
- Sub-category
- Region performance
- State performance

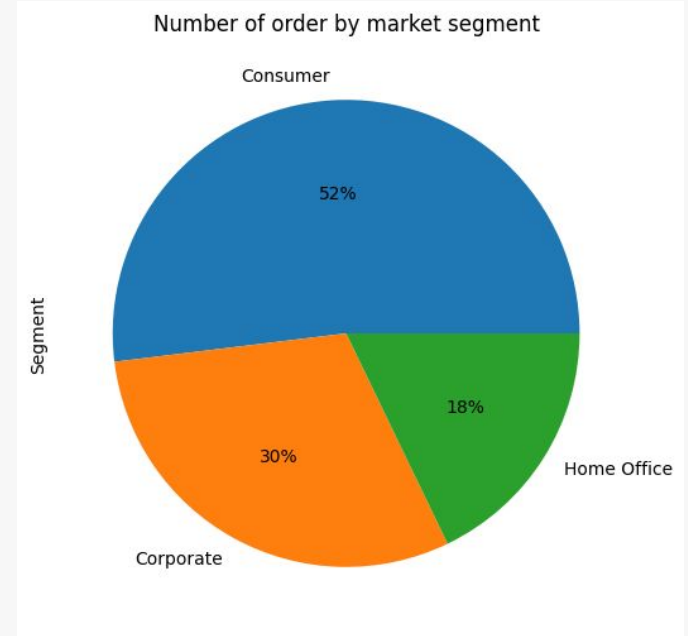
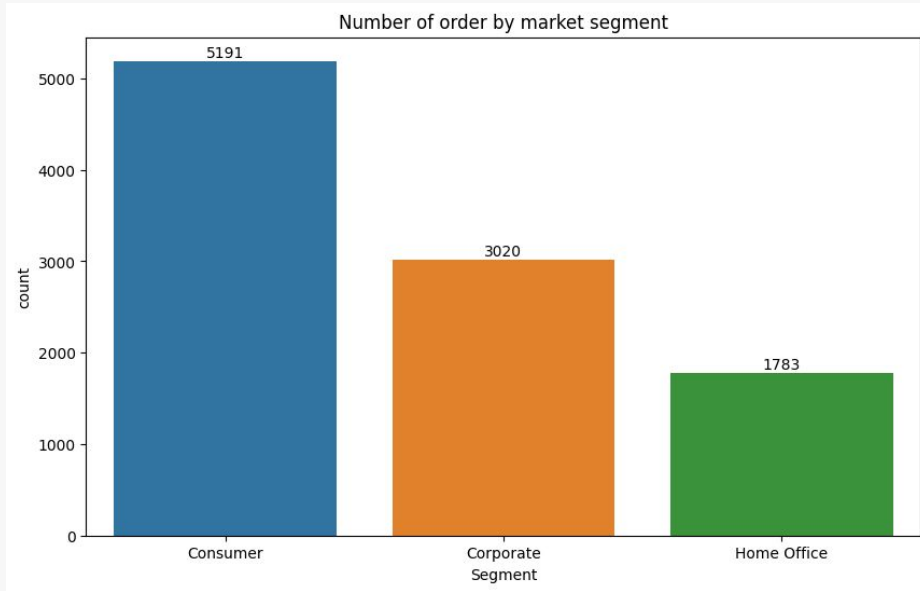
Sub-theme: Shipping mode

The most popular shipping mode is the standard class



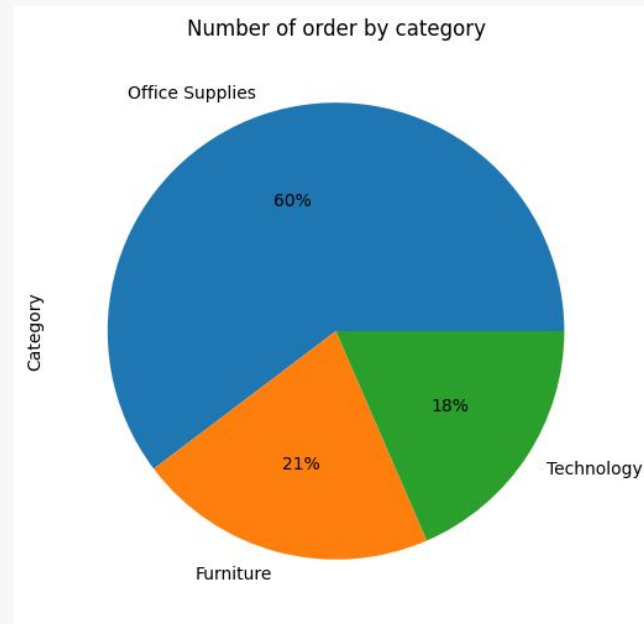
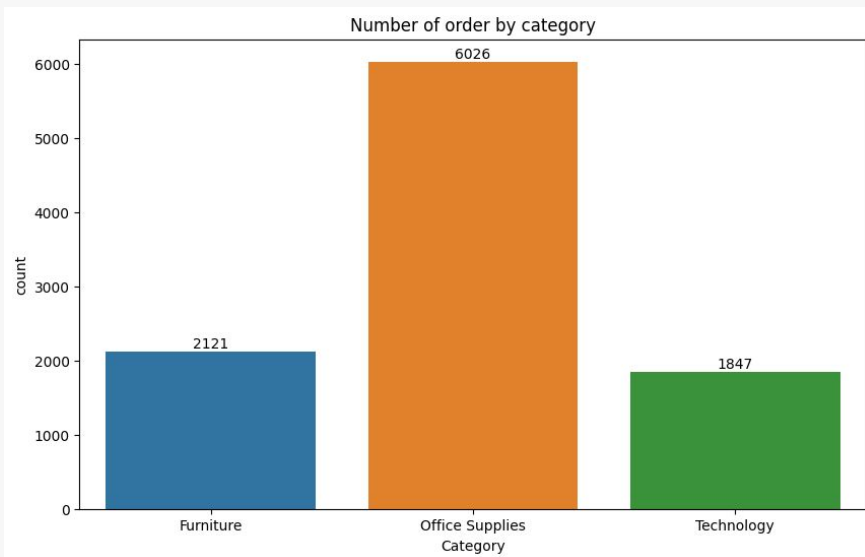
Sub-theme: Market segment

The highest number of order was dominated by private consumer segment



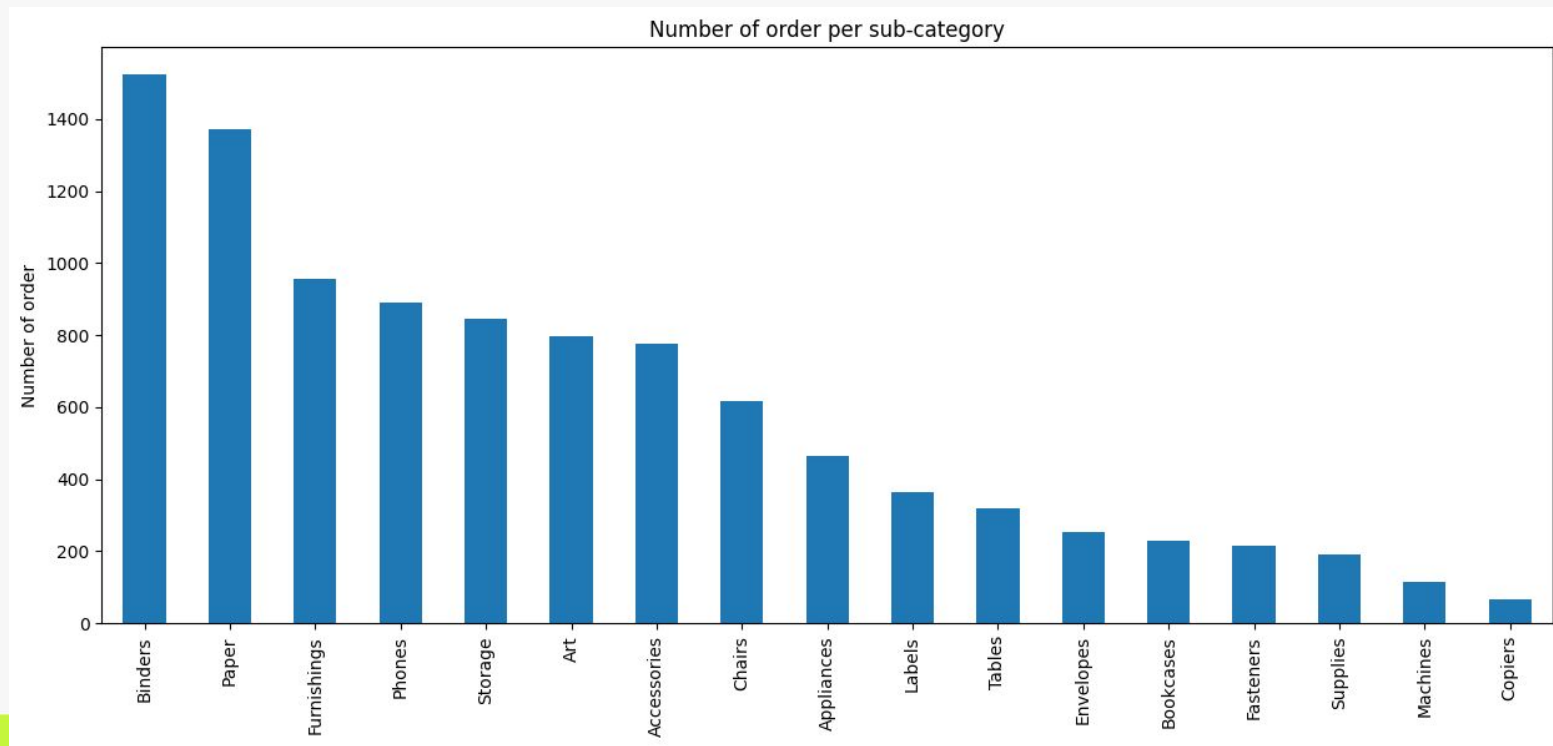
Sub-theme: Category

Office supplies was the most popular category



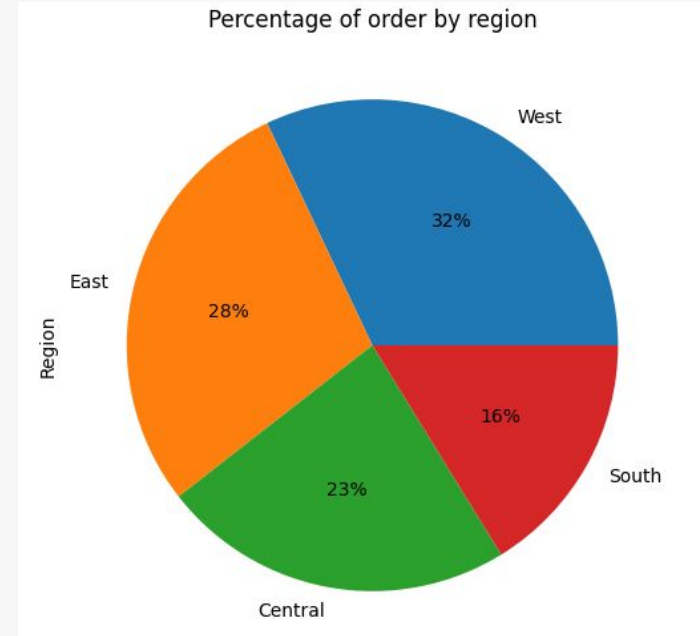
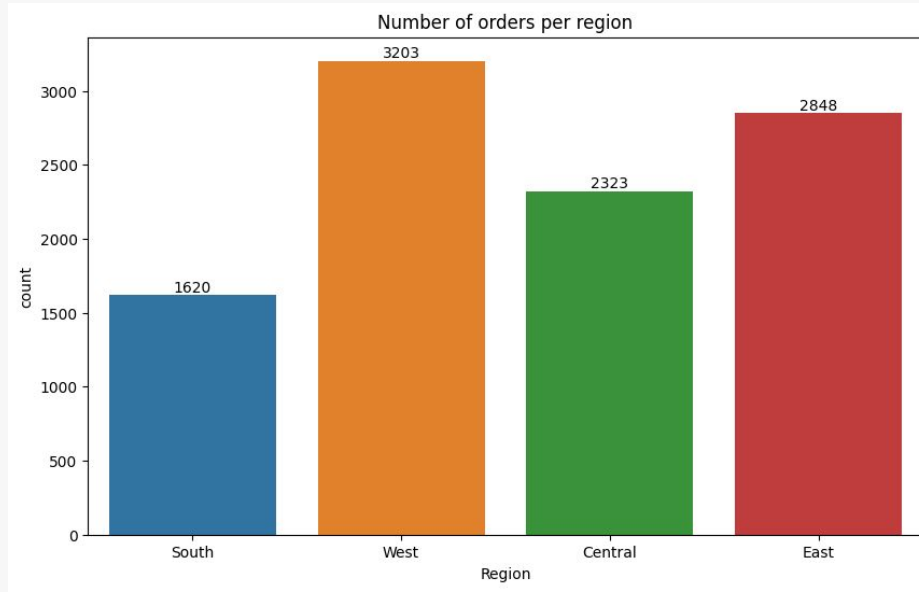
Sub-theme: Sub-category

Binders and paper were the most popular sub-category

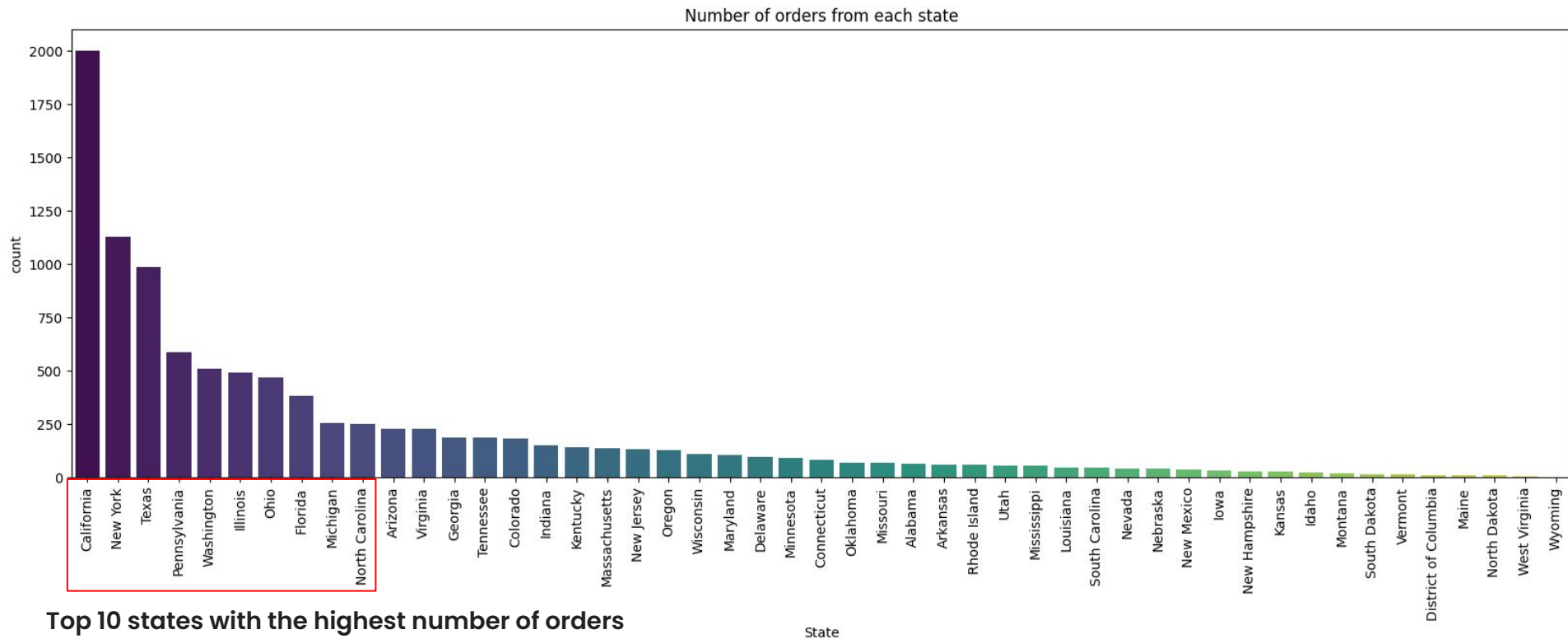


Sub-theme: Region performances

West region dominated the highest number of order in 2017-2020



Sub-theme: State performances



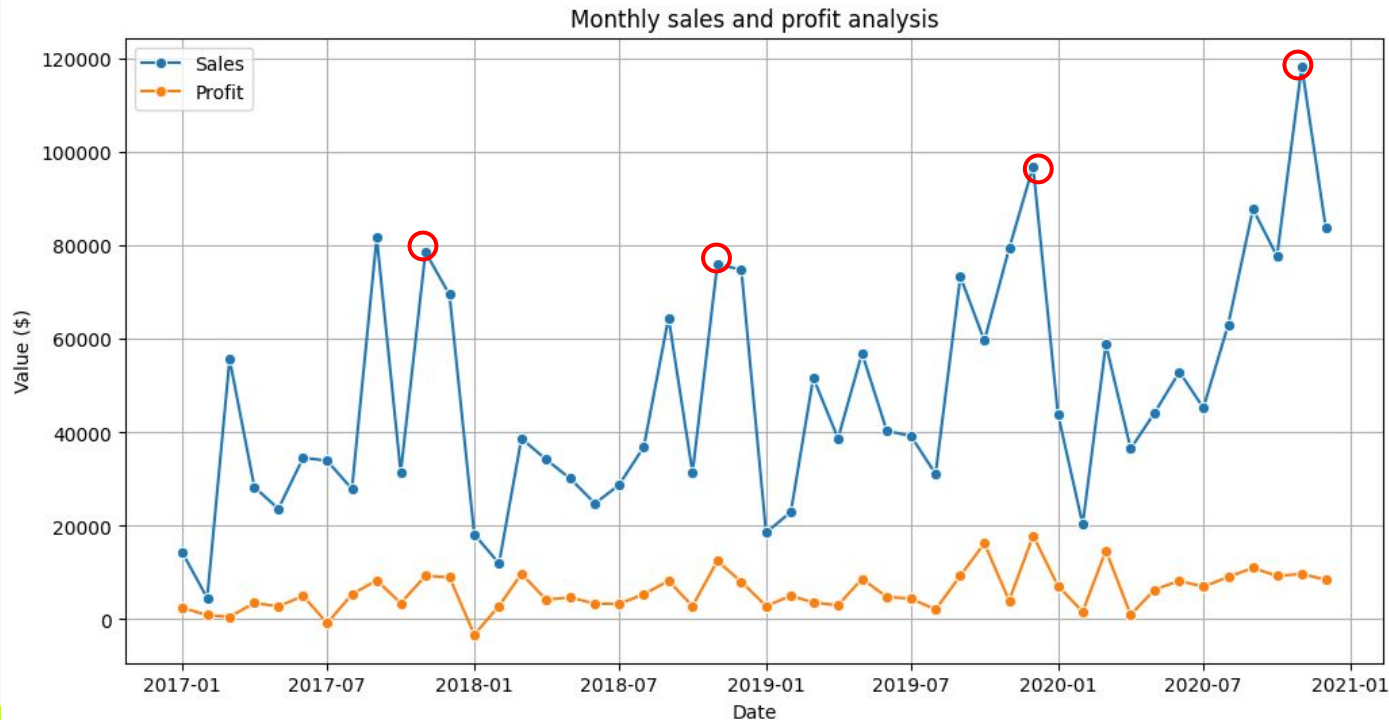
Top 10 states with the highest number of orders

Sales and profit analysis

Sub-themes:

- Time series trend
- Market segment
- Category
- Sub-category
- Region performance
- State performance

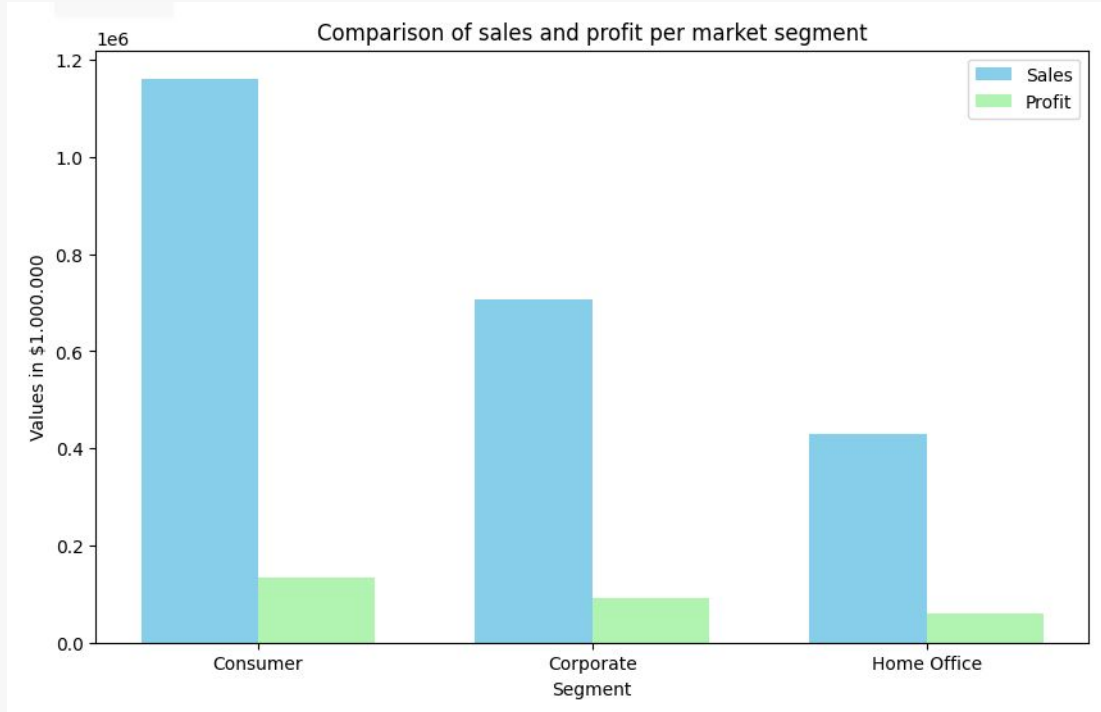
Sub-theme: Sales and profit trend



From 2017-01 to 2020-12, there were sales and profit increase up to 588,8% and 346.2%

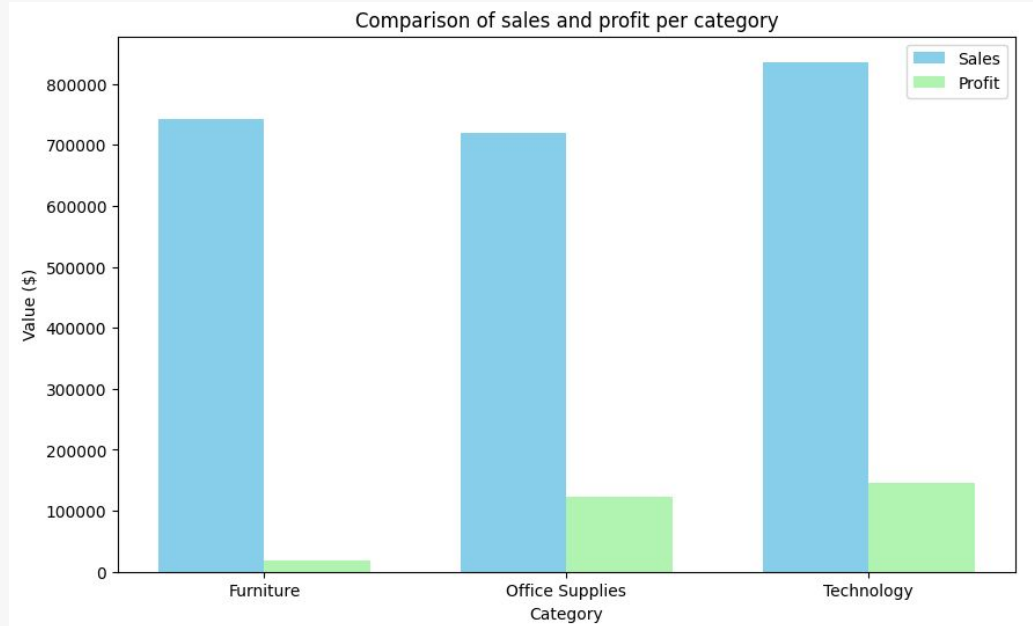
Noticeable yearly pattern of sales increase to the peak, 4th quarter, between October and December

Sub-theme: Market Segment



The market segment of private consumer generated the highest sales and profit of superstore

Sub-theme: Category

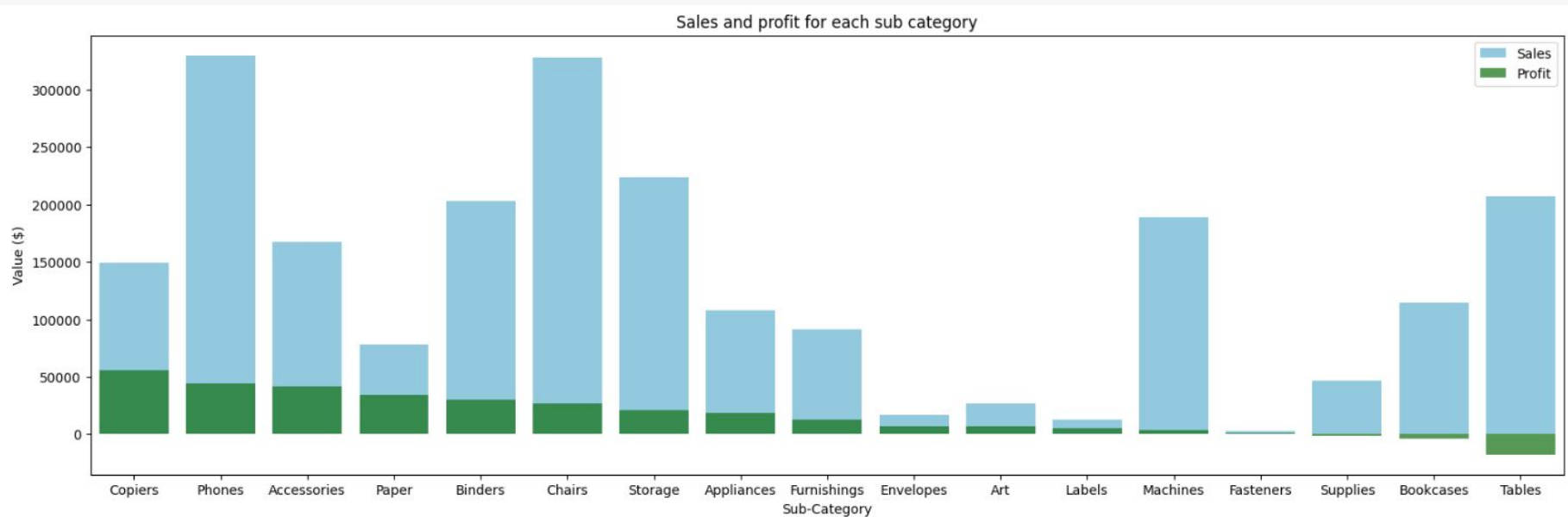


The technology category contributed to highest value of sales and the most profitable category

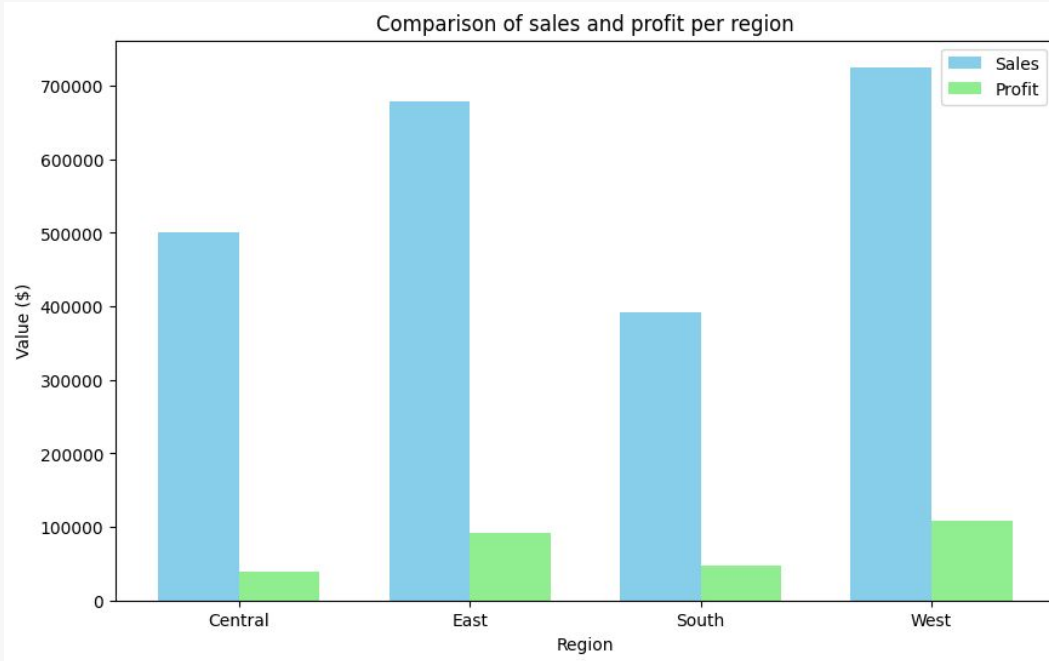
Sub-theme: Sub Category

Phones and chairs contributed to the highest sales value

However, copiers contributed to be the most profitable sub category



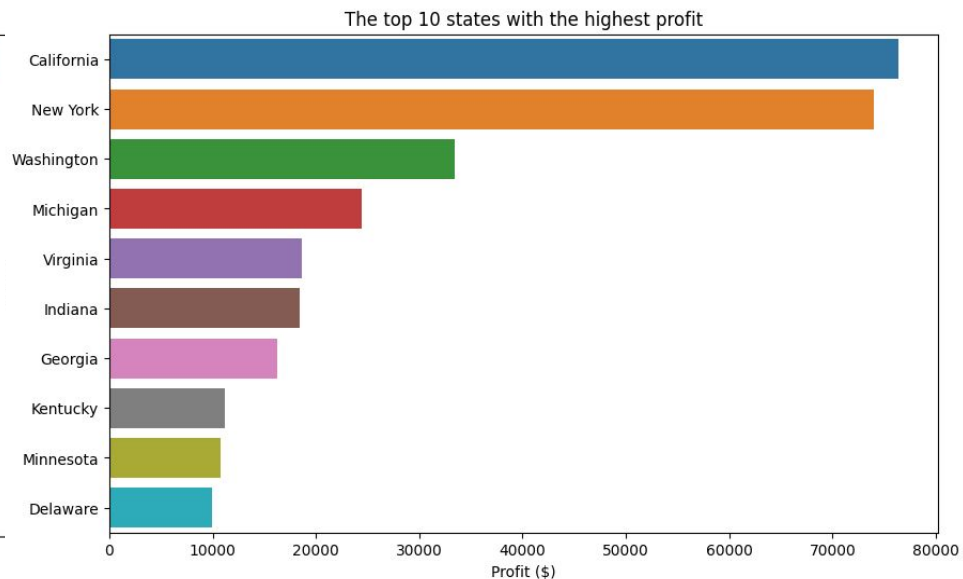
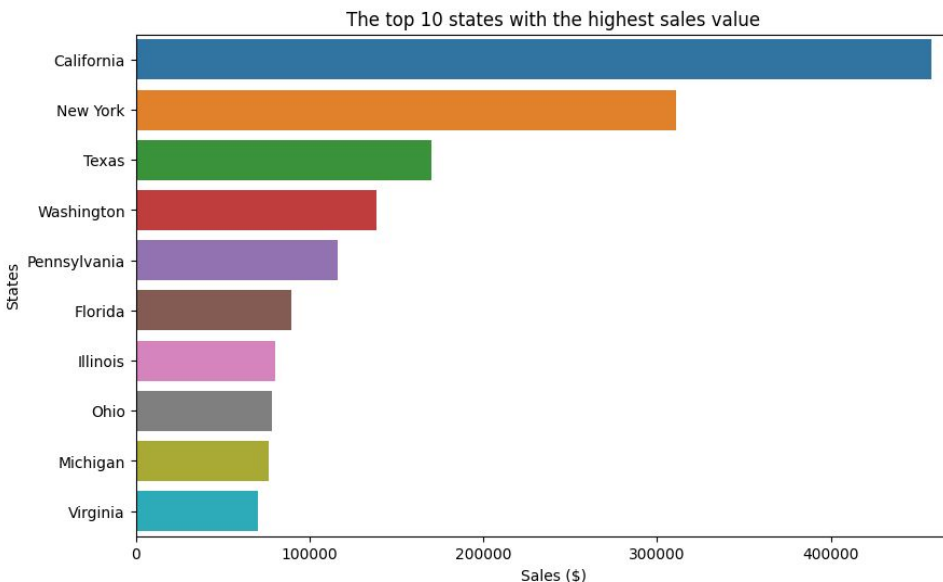
Sub-theme: Region



The west region generated the highest sales and profit

Sub-theme: State performance

California and New York have the highest sales and profit

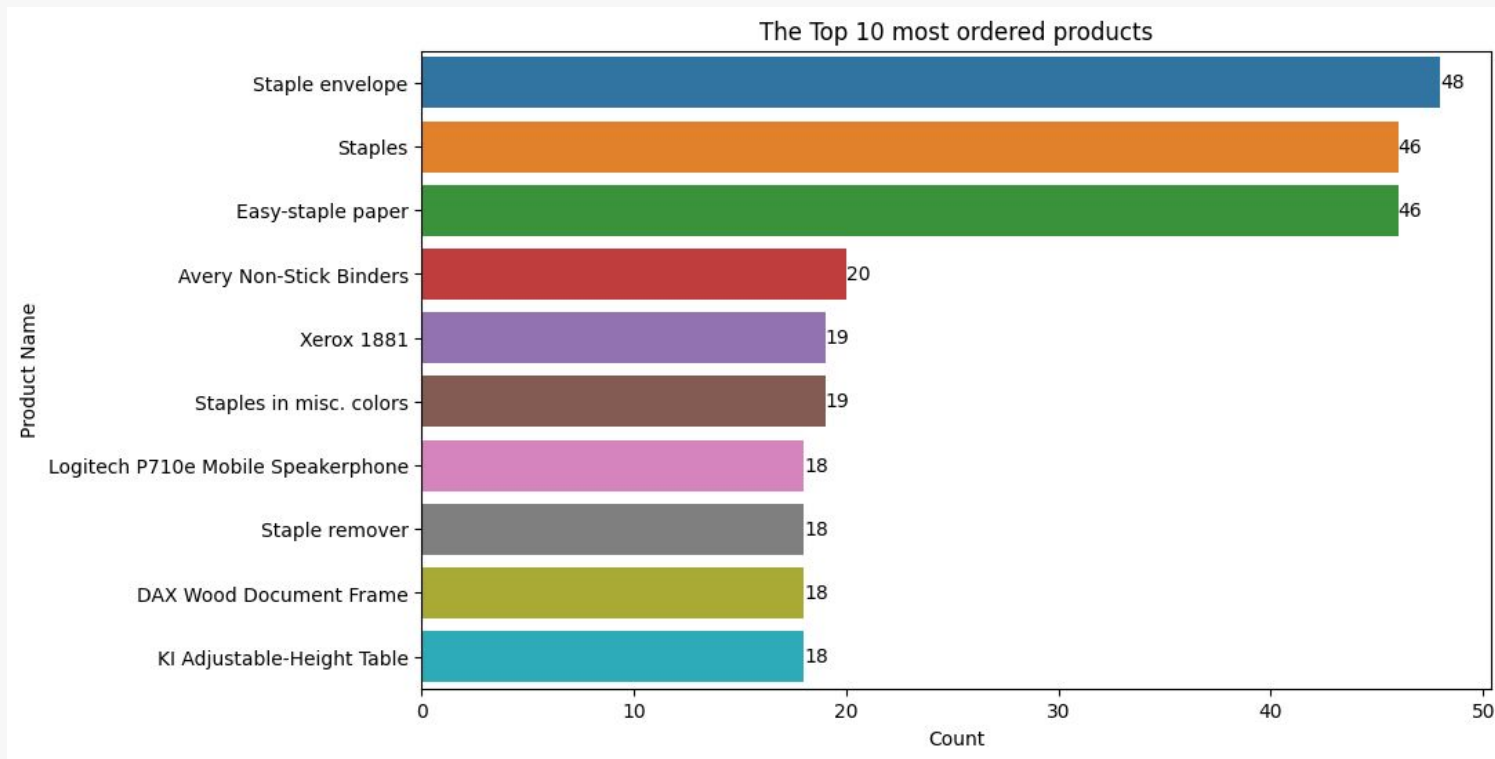


Product performance analysis

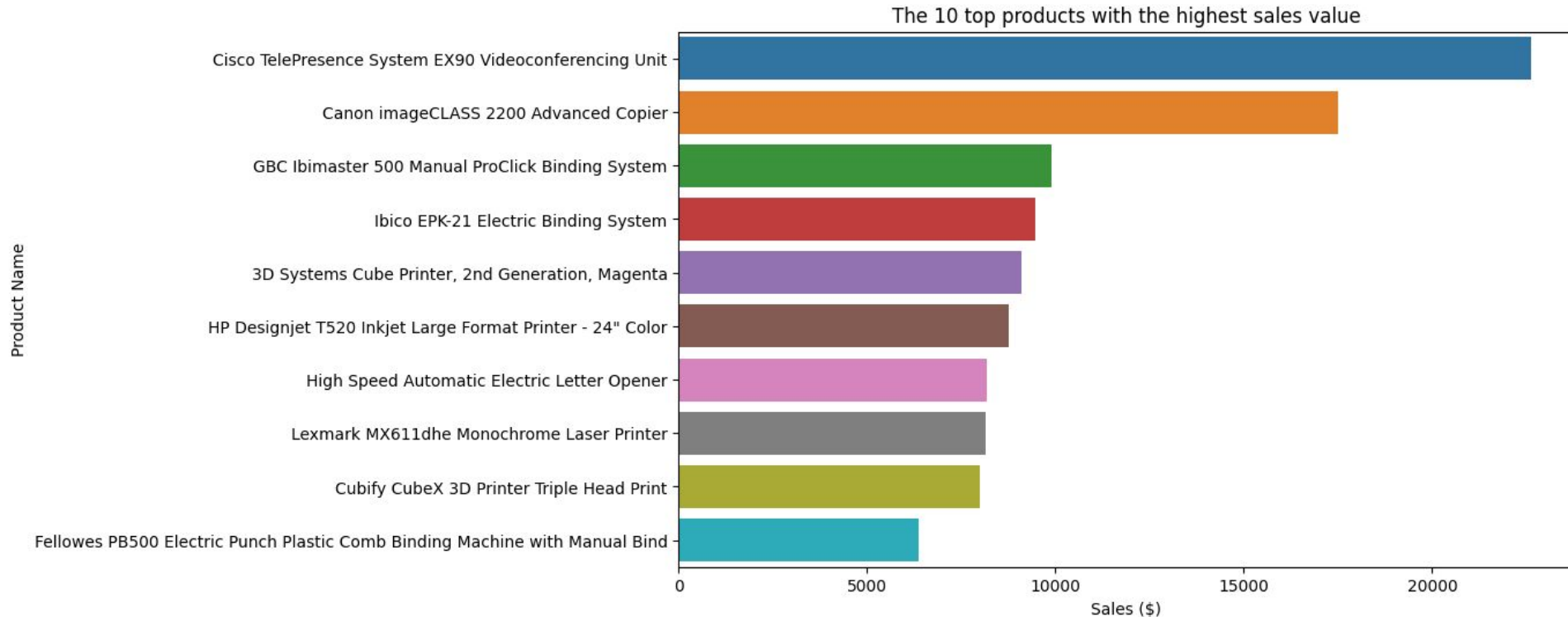
Sub-themes:

- Top 10 most popular products
- Top 10 products with highest sales value
- Top 10 most profitable products
- Top 10 least profitable products

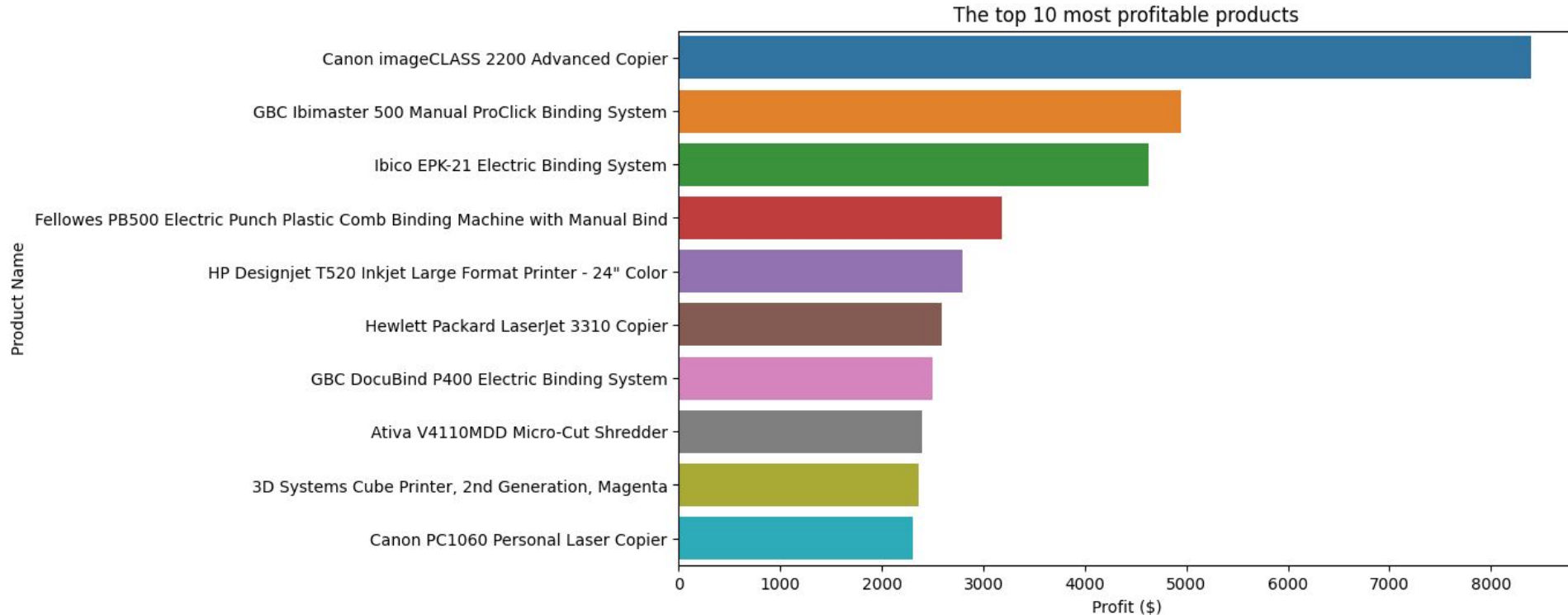
The top most ordered products of superstore



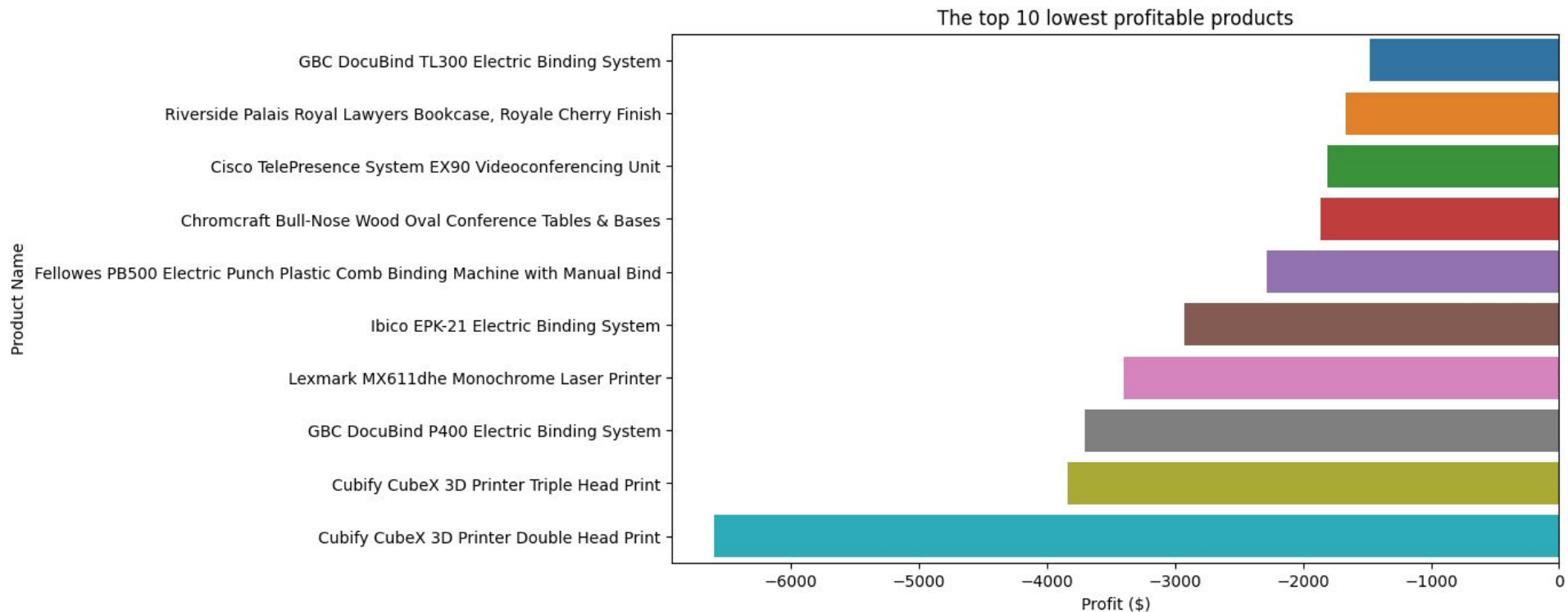
The top products with the highest sales value in Superstore



The top most profitable products of superstore



The least profitable products of superstore



Insights and recommendations

- **Shipping mode**

The most popular shipping mode is the standard class. It is recommended to improve the standard class quality of service to maintain good relationship with customers and improve customer loyalty

- **Market segment**

The highest number of order was dominated by private consumer segment. Furthermore, this segment generated the highest sales and profit of superstore.

- **Category and sub-category**

Office supplies were the most popular or the most ordered in category criteria. Moreover, Binders and paper were the most popular sub-category. However, the technology category contributed to highest value of sales and the most profitable. It is recommended to improve the sales strategy of items from technology to generate more profit for the company.

- **Geographical insights**

West region contributed to the highest number of order and generated the highest sales and profit contribution to the Superstore. From the whole states, California and New York contributed to the highest number of order and have the highest sales and profit. It is recommended to explore more the top performing states strategy as lesson learnt and to be adapted in the less performing states

- **Sales and profit trends**

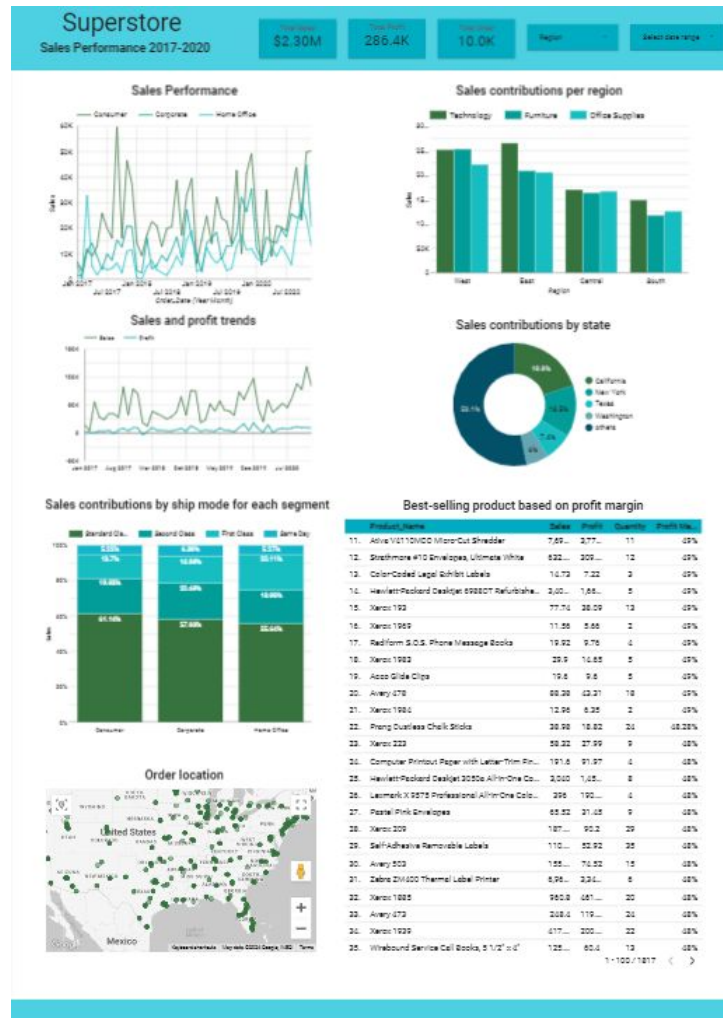
From 2017-01 to 2021-12, there were sales and profit increase up to 588,8% and 346.2%. Noticeable yearly pattern of sales increase to the peak, 4th quarter, between October and December. It is recommended to stock more products near the end of the year

- **Product insights**

Phones and chairs contributed to the highest sales value. However copiers contributed to be the most profitable sub category. It is recommended to stock more the top ten most ordered products, especially during the peak season and conduct more marketing for them in the less performance area. It is recommended to improve the marketing process/strategy to sell more the top 10 most profitable products, to get more profit. It is recommended to fix or to eliminate the least profitable products, especially that cause some losses to the company

Data Visualization

Dashboard of Superstore



Thanks!

My Contact



<https://github.com/harishmuh>



www.linkedin.com/in/harish-muhammad-7b600b102/



harishmuh@gmail.com



dqlab



www.dqlab.id