

Exploratory Data Analysis of The Global Superstore Dataset

NAPQUEENS - ANARIX.AI - ASSESSMENT

SAMRIDH VIKHAS S

20MIY0001



INTRODUCTION TO THE GLOBAL SUPERSTORE DATASET

data.head()																							
	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	...	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit				
0	32298	CA-2012-124891	7/31/2012	7/31/2012	Same Day	RH-19495	Rick Hansen	Consumer	New York City	New York	...	TEC-AC-10003033	Technology	Accessories	Plantronics CS510 - Over-the-Head monaural Wir...	2309.650	7	0.0	762.1845				
1	26341	IN-2013-77878	2/5/2013	2/7/2013	Second Class	JR-16210	Justin Ritter	Corporate	Wollongong	New South Wales	...	FUR-CH-10003950	Furniture	Chairs	Novimex Executive Leather Armchair, Black	3709.395	9	0.1	-288.7650				
2	25330	IN-2013-71249	10/17/2013	10/18/2013	First Class	CR-12730	Craig Reiter	Consumer	Brisbane	Queensland	...	TEC-PH-10004664	Technology	Phones	Nokia Smart Phone, with Caller ID	5175.171	9	0.1	919.9710				
3	13524	ES-2013-1579342	1/28/2013	1/30/2013	First Class	KM-16375	Katherine Murray	Home Office	Berlin	Berlin	...	TEC-PH-10004583	Technology	Phones	Motorola Smart Phone, Cordless	2892.510	5	0.1	-96.5400				
4	47221	SG-2013-4320	11/5/2013	11/6/2013	Same Day	RH-9495	Rick Hansen	Consumer	Dakar	Dakar	...	TEC-SHA-10000501	Technology	Copiers	Sharp Wireless Fax, High-Speed	2832.960	8	0.0	311.5200				

```
data.shape
```

```
(51290, 24)
```

FIRST STEPS WITH THE DATA

```
data.columns
```

```
Index(['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode',  
      'Customer ID', 'Customer Name', 'Segment', 'City', 'State', 'Country',  
      'Postal Code', 'Market', 'Region', 'Product ID', 'Category',  
      'Sub-Category', 'Product Name', 'Sales', 'Quantity', 'Discount',  
      'Profit', 'Shipping Cost', 'Order Priority'],  
      dtype='object')
```

- Missing Data: Postal Code column has almost 80% missing values; ignored for analysis.
- Duplicates: No duplicated rows.
- Data Types: Ship Date and Order Date converted from string to datetime format.
- Date Range:
- First order: January 1, 2011.
- Last order: December 31, 2014.

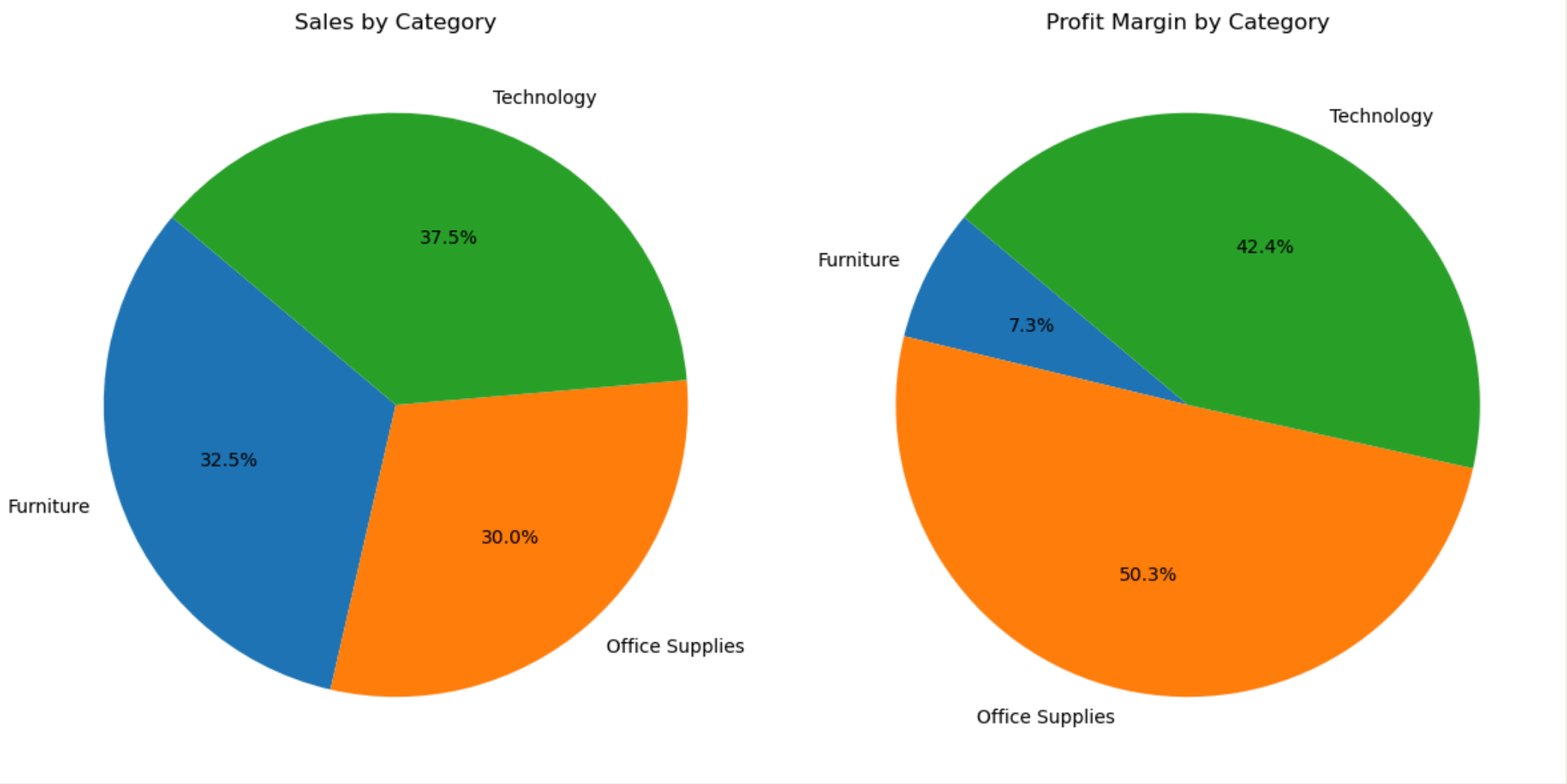
DATA PREPARATION

profit_margin	shipping_time
0.330000	0
-0.077847	2
0.177766	1
-0.033376	2
0.109963	1

New Columns Created:

- Profit Margin: Profit/Sales.
- Shipping Time: Ship Date – Order Date.

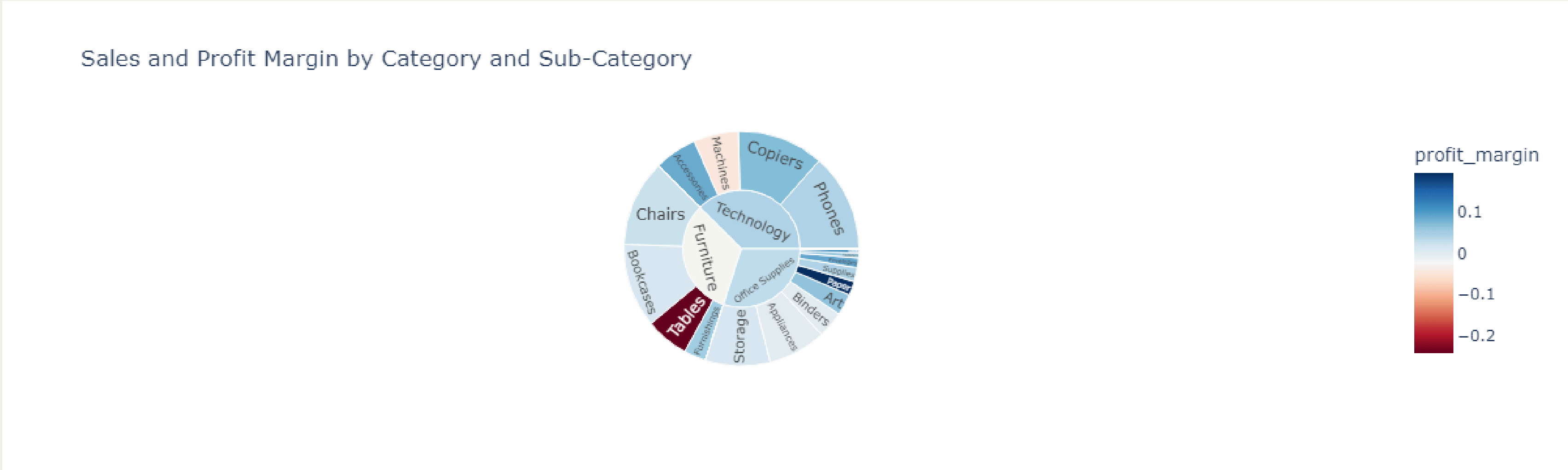
Sales and Profit Margin by Category



We can observe that **Technology(37.5%)** has the highest Sales followed by Furniture(32.5%), but the Highest Profit Margin is gained from **Office Supplies(50.3%)** followed by Technology(42.4%) making Furniture a very low-profitable Category

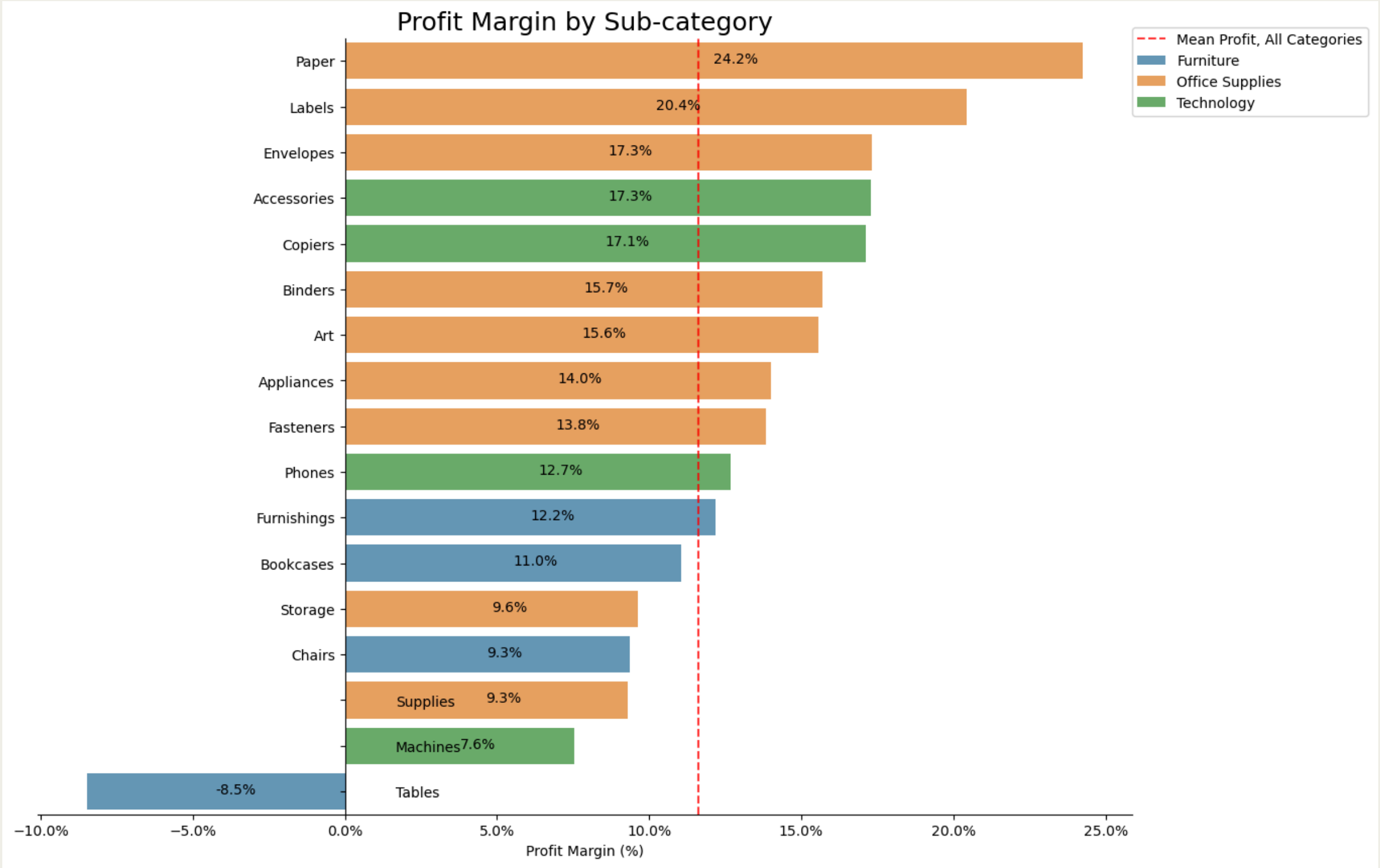
Sales and Profit Margin by Category and Subcategory

Visual: Sunburst chart.



Profit Margin by Subcategory

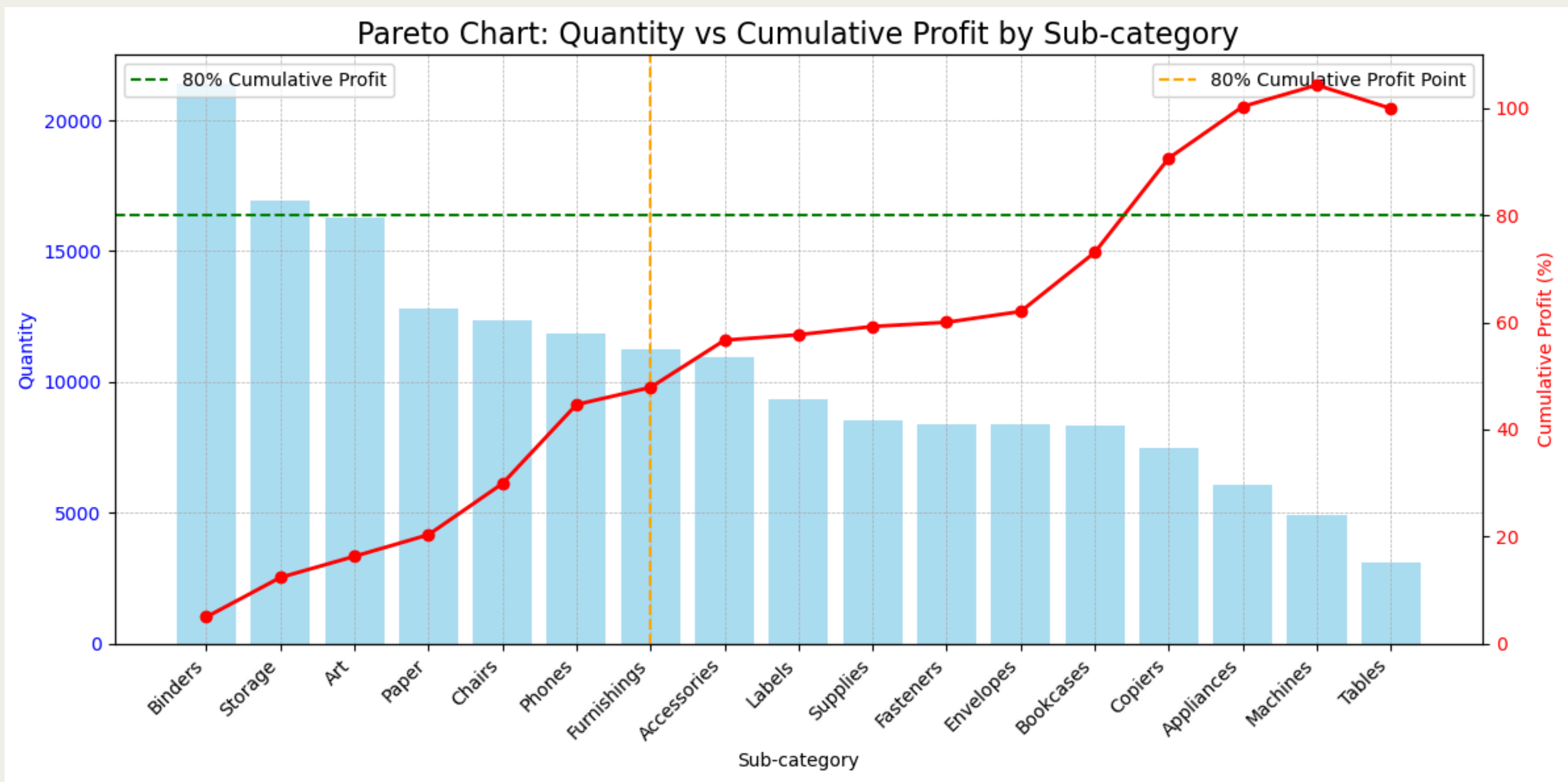
Visual: Bar chart with average profit margin reference line.



- Tables subcategory (Furniture) shows a loss.
- Mean profit margin: 11.6%.
- 11 out of 17 subcategories have profit margins above the mean.

Pareto Chart - Quantity vs. Profit by Subcategory

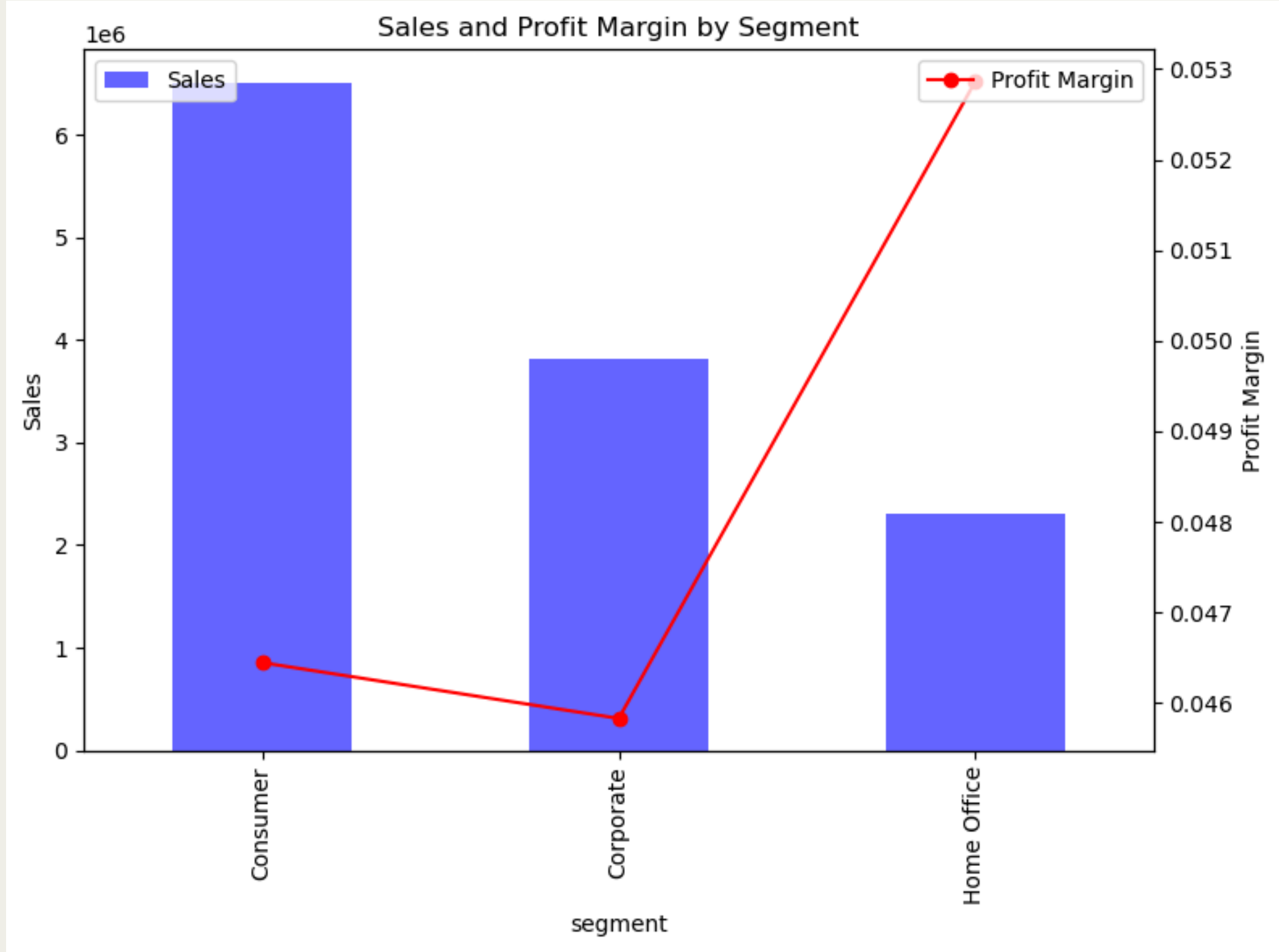
Visual: Pareto Chart



- The Pareto Rule (80% of profit from 20% of goods) is not followed.
- Recommendation: Focus on products with higher quantities and profits.

Sales and Profit Margin by Segment

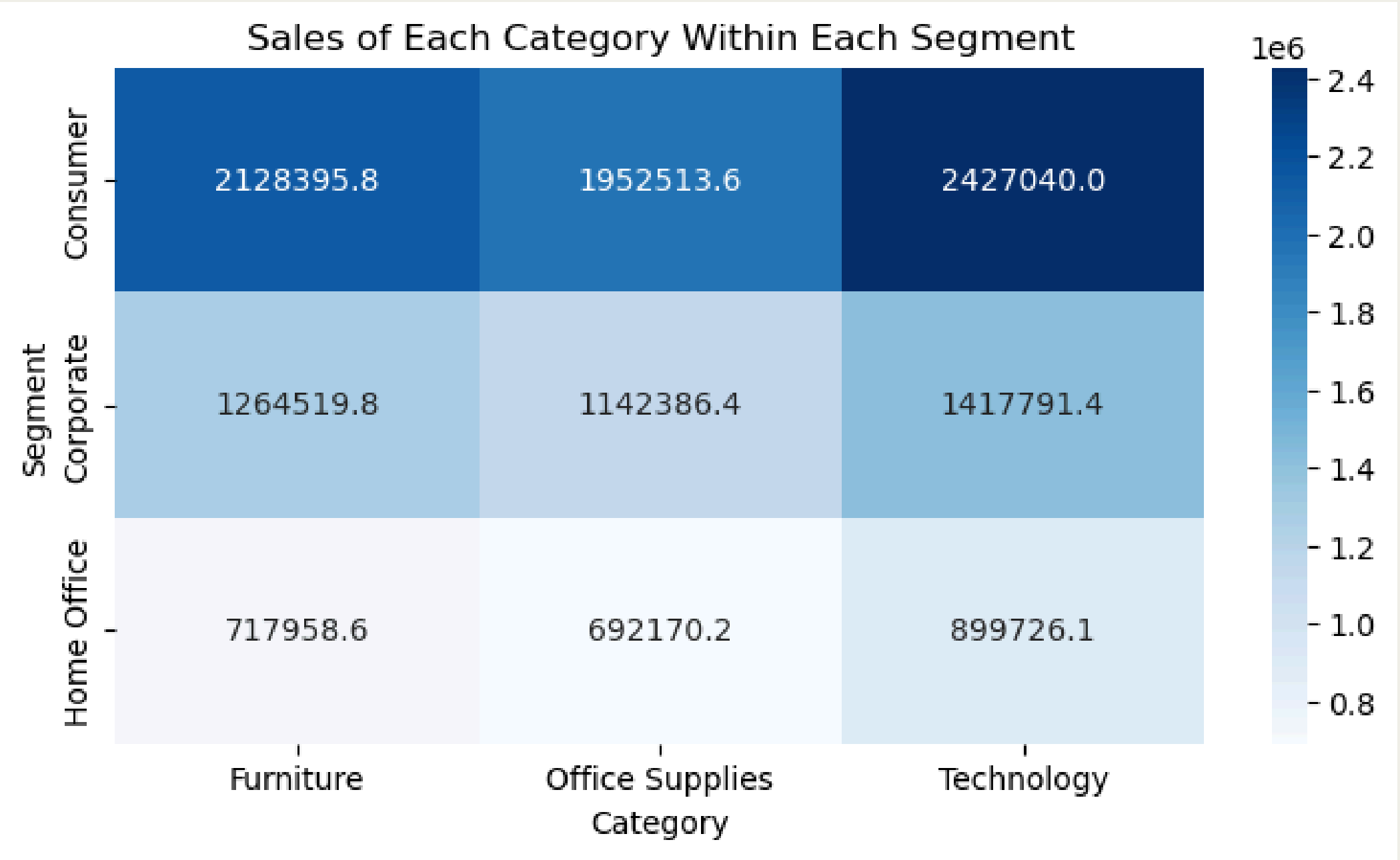
Visual: Bar and Line Chart



- Home Office segment boasts the highest profit margin at 5.3%.
- Despite this, the Home Office segment has the lowest sales volume.

Sales of Each Category Within Each Segment

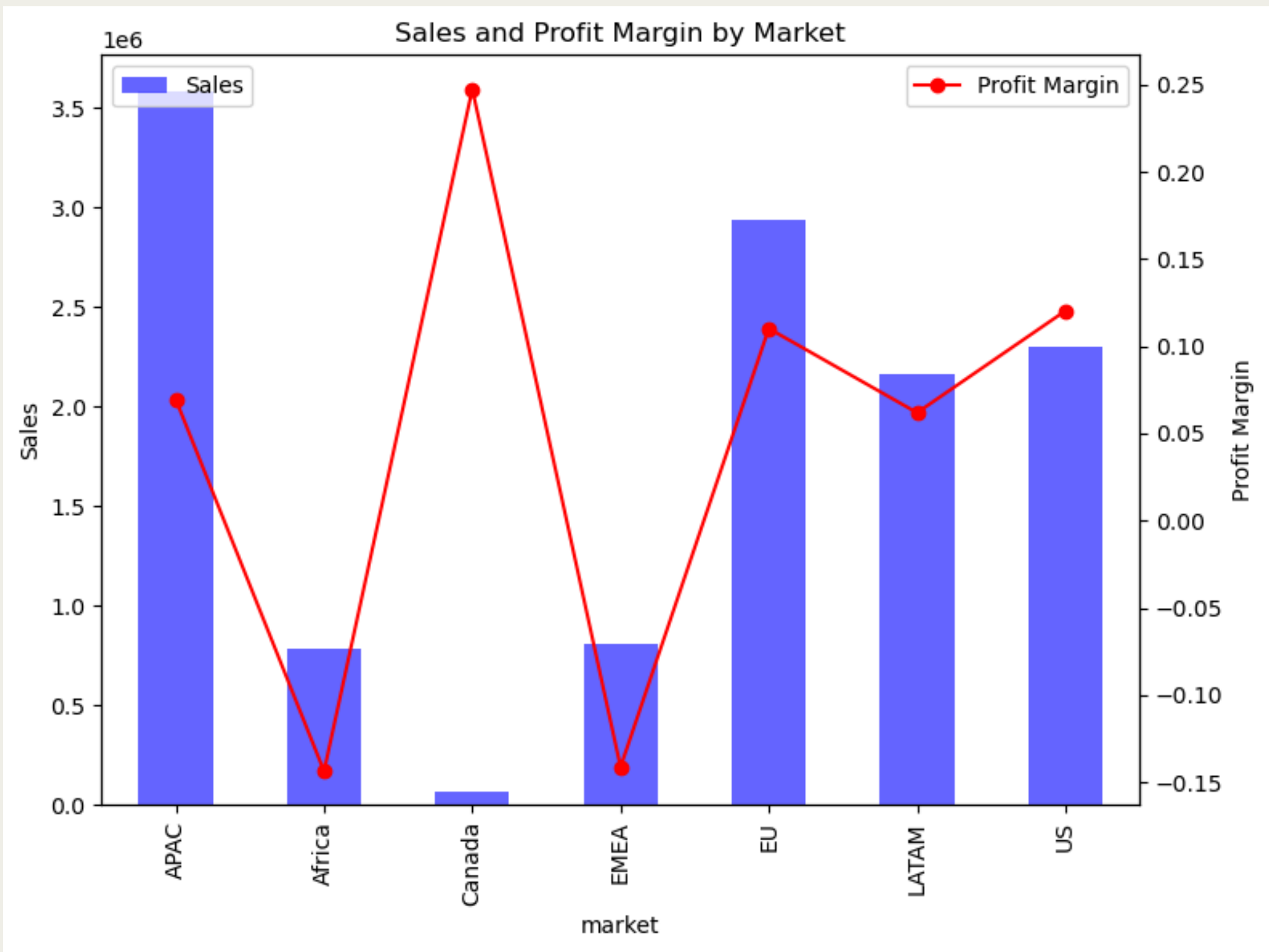
Visual: Pivot Heatmap



- The Consumer Segment purchases more Technology products than any other segment buying other products.
- Across all 3 segments Technology Sales are the highest

Sales and Profit Margin by Market

Visual: Bar Chart



- Canada has the highest profit margin at 25%, despite lower sales.
- Africa and EMEA markets exhibit negative profit margins.

Sales and Profit Margin by Market and Region

Visual: Sunburst Chart

Sales and Profit Margin by Market and Region



Average Shipping Cost by Market

market	shipping_cost
APAC	35.190430
Africa	19.215058
Canada	19.285495
EMEA	17.573221
EU	30.942235
LATAM	22.745153
US	23.831678

- APAC region incurs the highest shipping costs.
- EMEA region enjoys the lowest shipping costs.

Average Shipping Time by Market

market	shipping_time
APAC	3.969097
Africa	3.910399
Canada	3.677083
EMEA	3.933386
EU	4.008300
LATAM	3.996794
US	3.958875

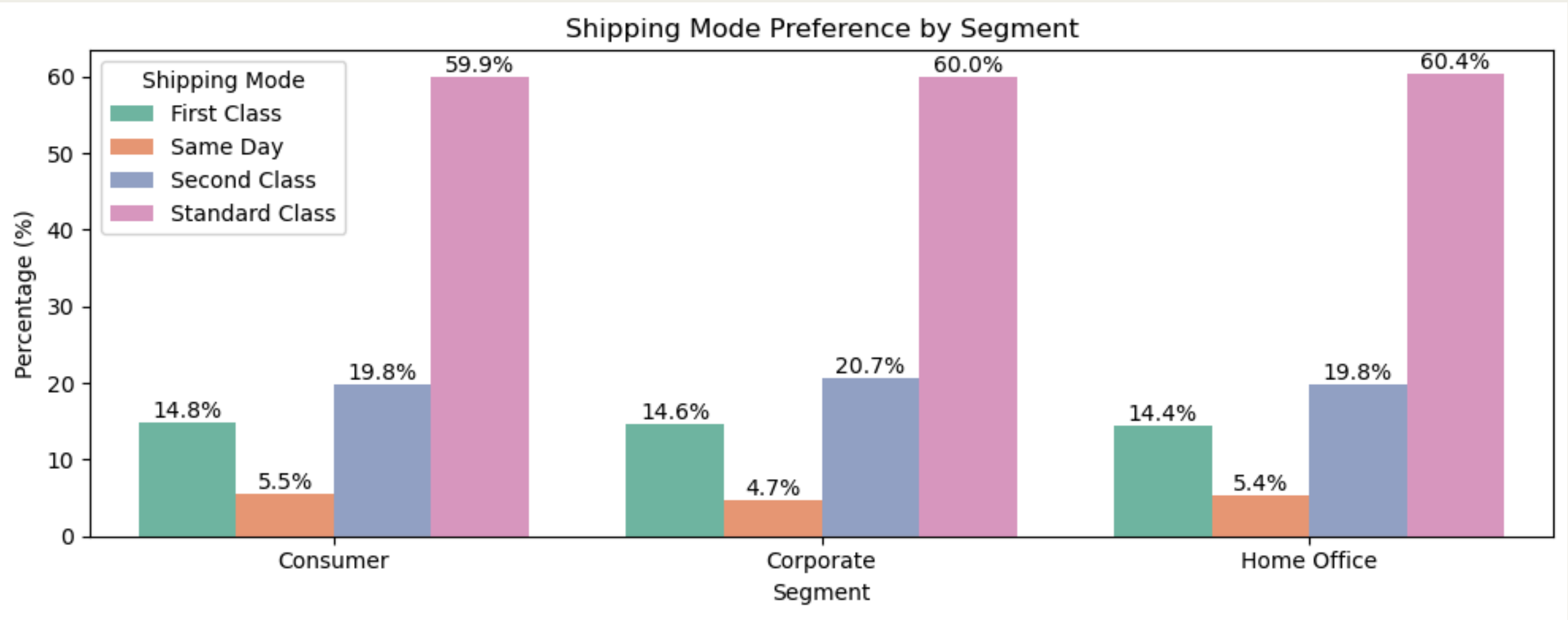
- Canada leads with the shortest shipping times.
- EU experiences the longest shipping times.

Most Frequent Customers by Order Count

customer_name	order_count
Muhammed Yedwab	108
Steven Ward	106
Gary Hwang	102
Patrick O'Brill	102
Bill Eplett	102
Harry Greene	101
Eric Murdock	100
Art Ferguson	98
Brosina Hoffman	97
Bart Watters	96

Shipping Mode Preference by Segment

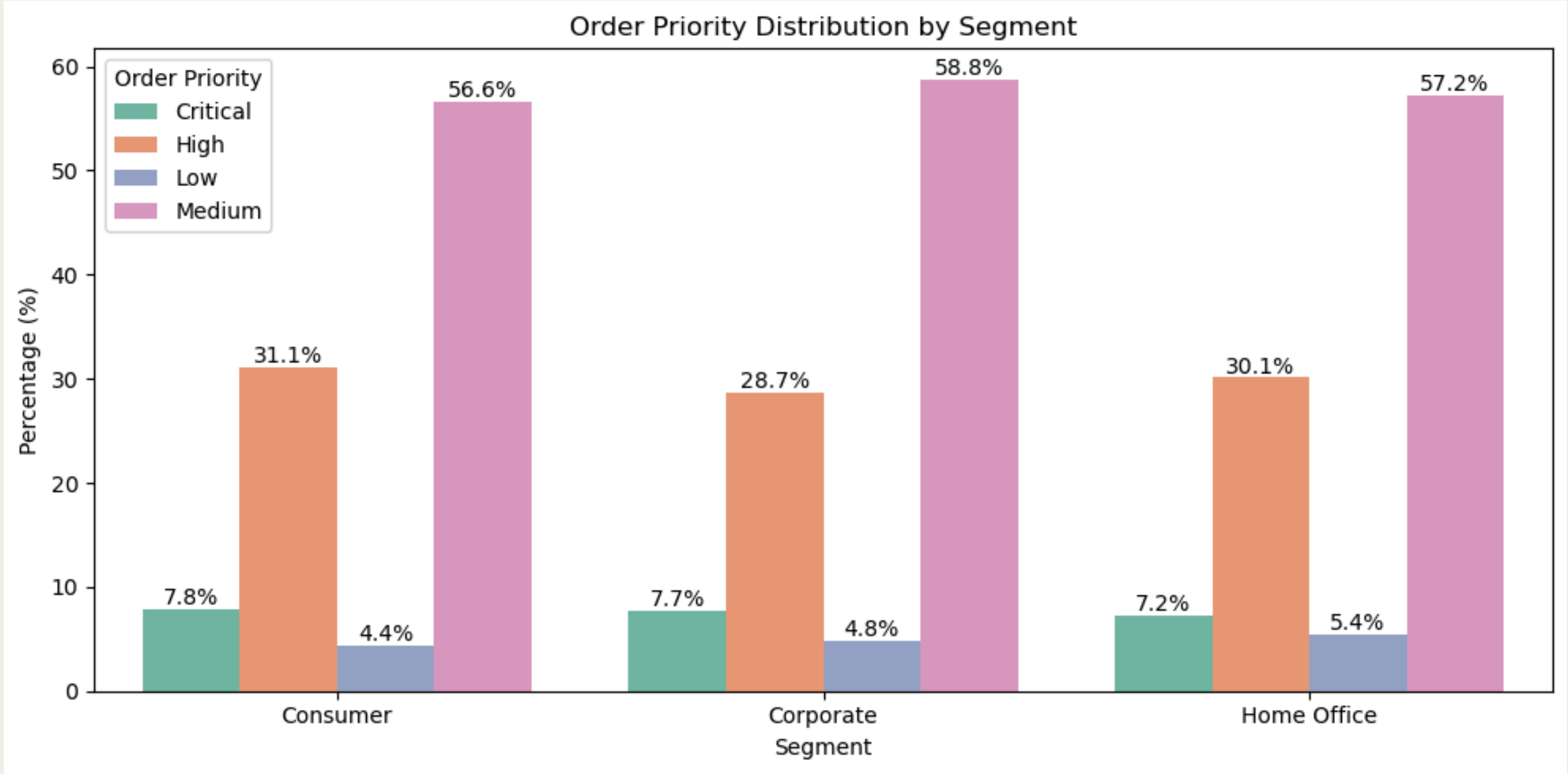
Visual: Bar Chart with Percent Values



- Standard Delivery is the preferred shipping mode across all segments.
- No significant variation in delivery class preference across segments.

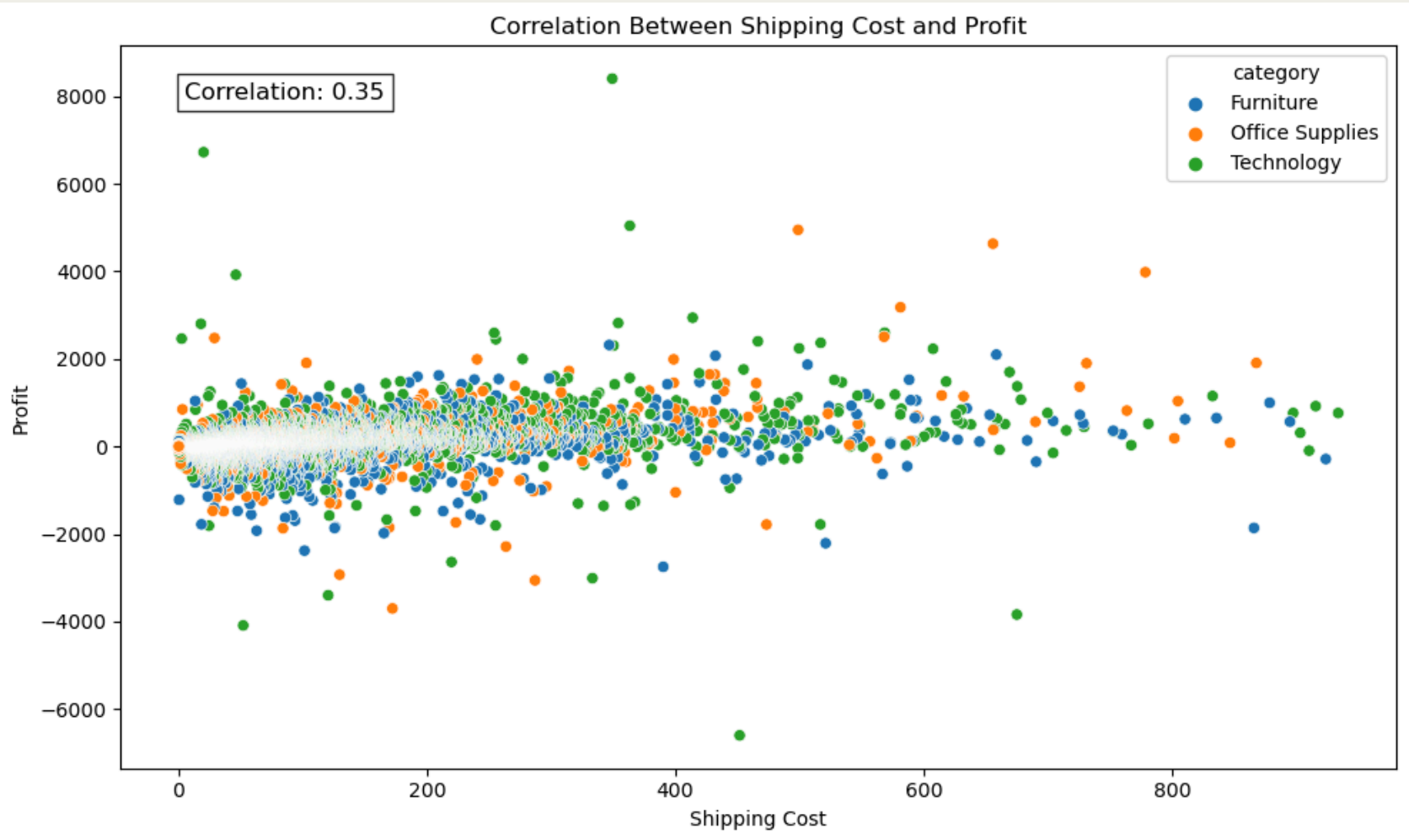
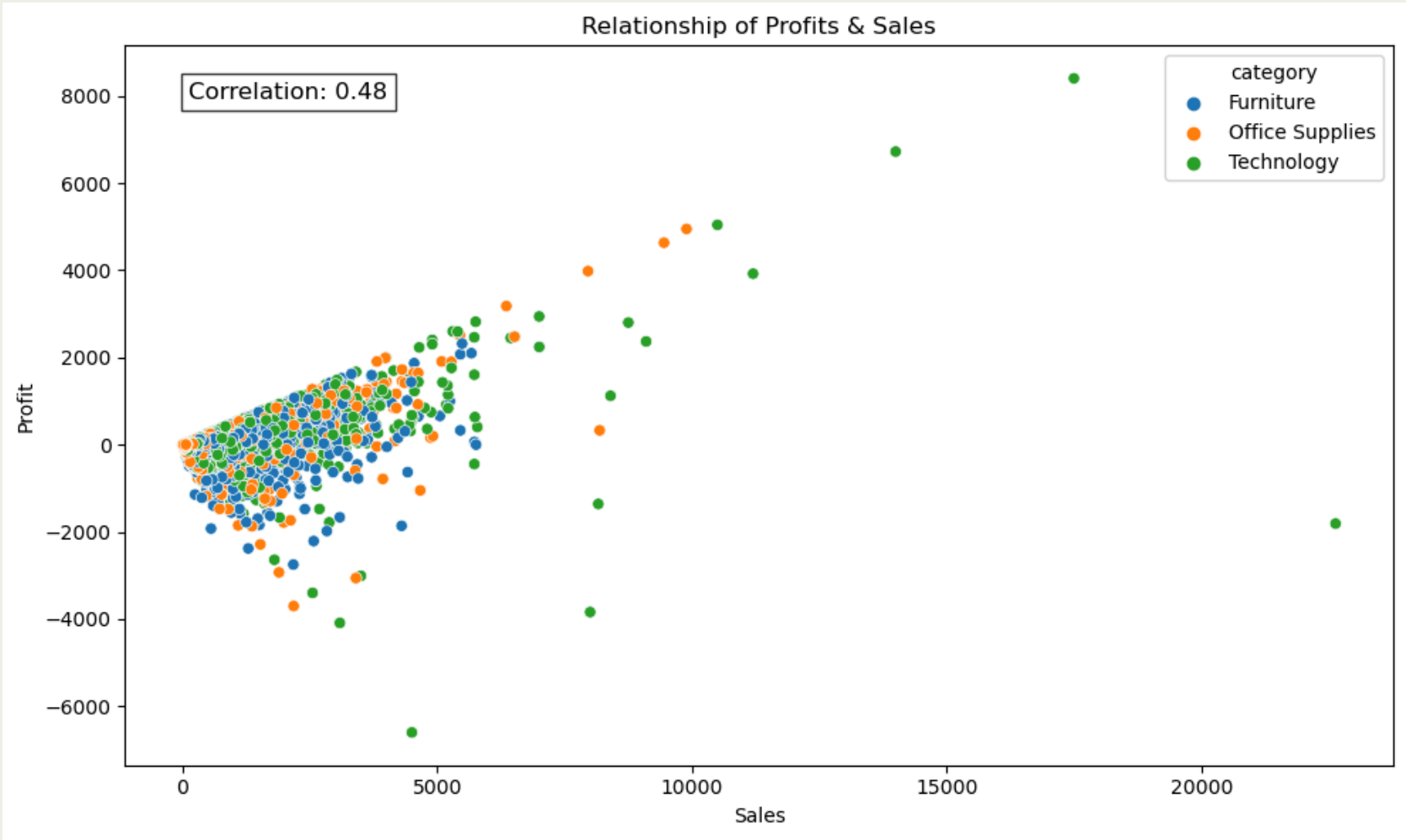
Order Priority Distribution by Segment

Visual: Bar Chart with Percent Values

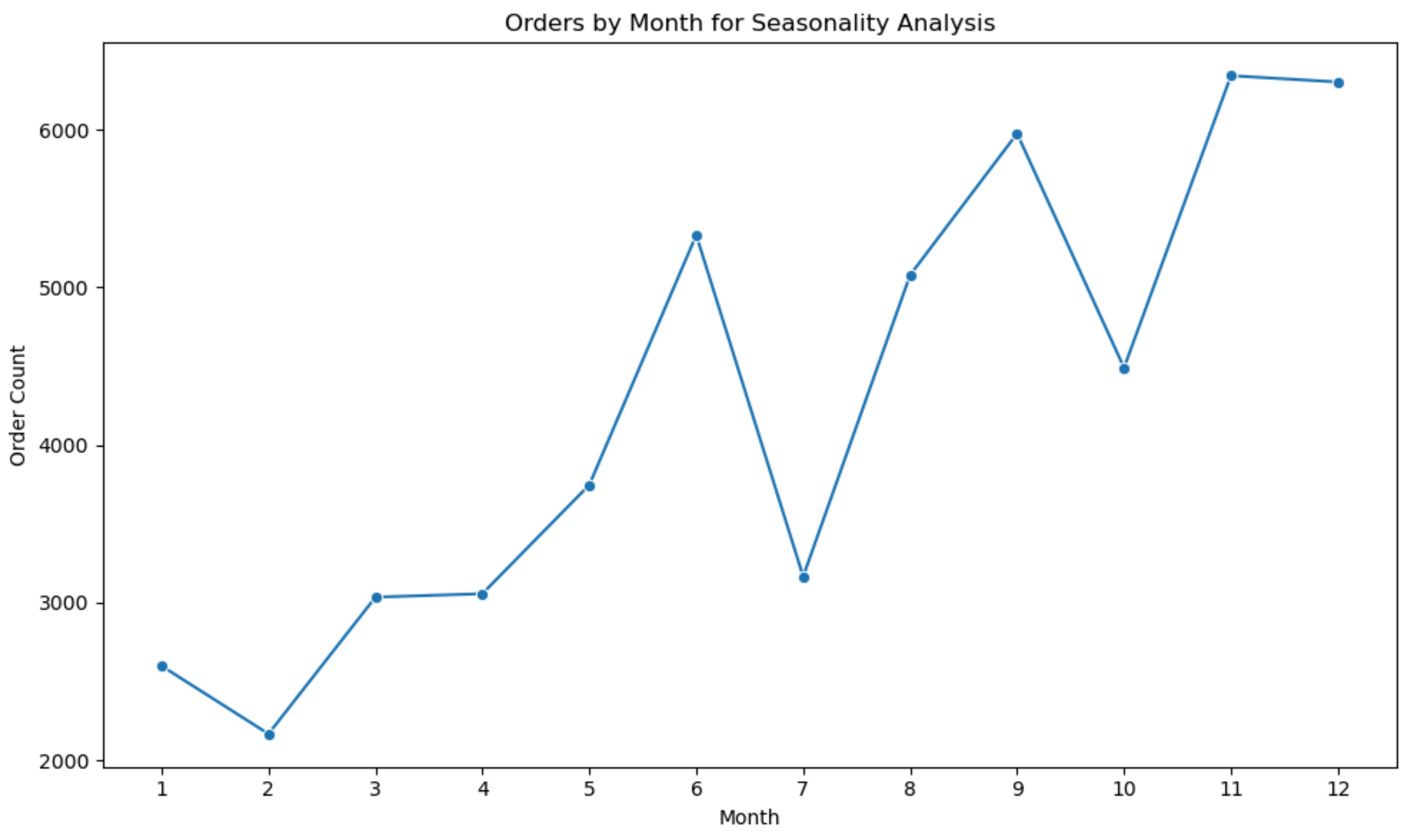


- Medium priority orders dominate across all segments.
- Similar distribution of order priority among segments.

Relationships in the Data

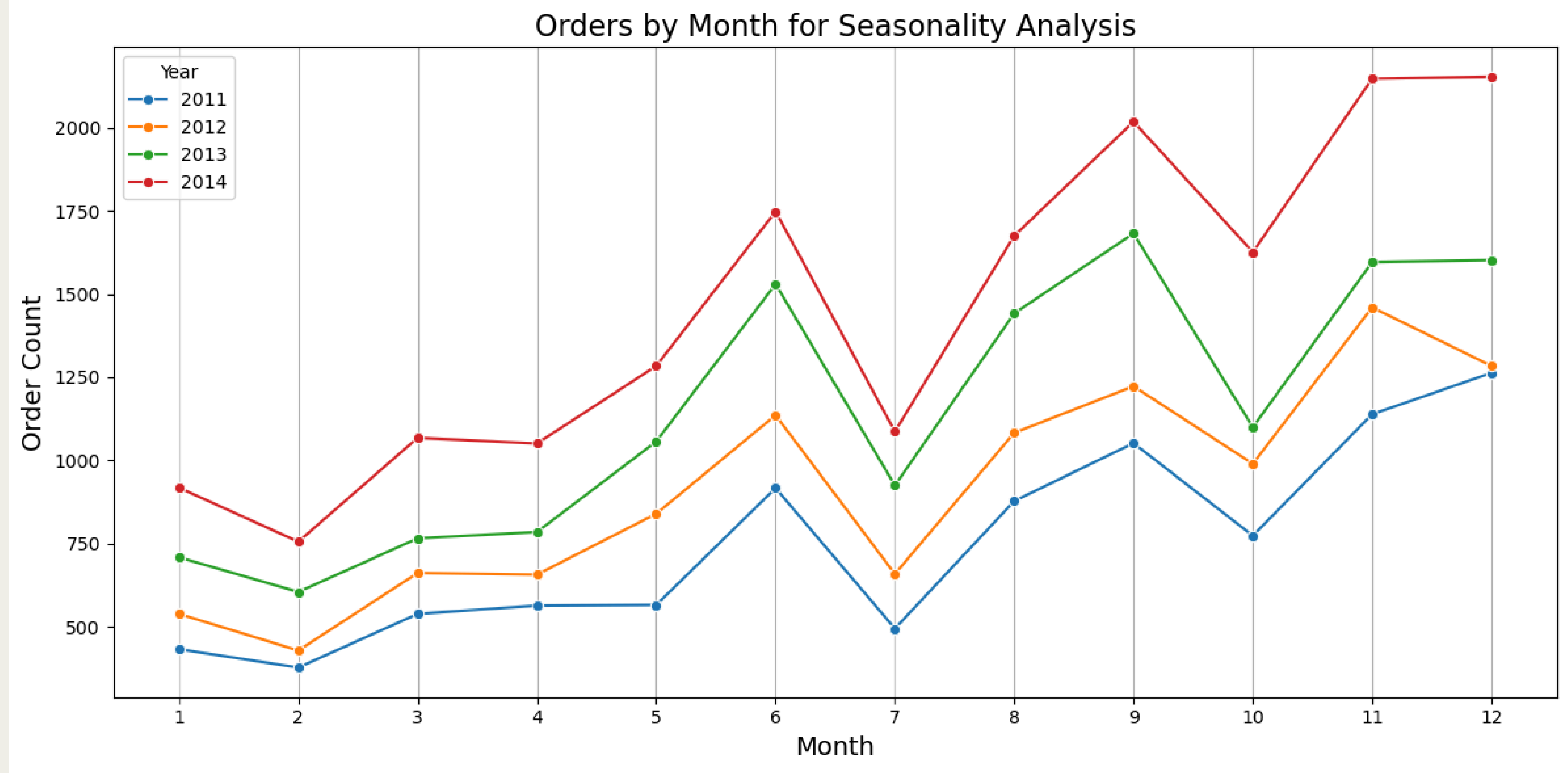


Orders by Month for Seasonality Analysis



- Consistent seasonal patterns: spikes in June, August, September, November, and December, with a dip in October.
- Patterns are stable across all years (2011-2014).

Orders by Month for Seasonality Analysis



LIMITATIONS AND FUTURE SCOPE OF EXPANSION

- **Absence of Customer Demographics**

- The dataset lacks detailed customer demographic information such as gender and age.
- Consequently, in-depth customer segmentation and profiling were not possible.

- **Time Constraints**

- Limited time prevented the execution of additional analyses, including association analysis and Recency, Frequency, and Monetary (RFM) analysis.
- Future work could benefit from these advanced analytical techniques for a more comprehensive understanding.

Thank you!

SAMRIDH VIKHAS S

20MIY0001

