Superstore Analysis

by Isnaini Mufidhatul Mughni

Contents

- Business and Data Understanding

- Exploratory Data Analysis

- Data Visualization

- Product Analysis

- Customer Analysis

- Conclusion and Recommendation

Business Understanding

SuperStore, an online home & office supply retailer operating in 48 states, faces increased market competition and seeks to improve its profits and profit margin.

Problem

The executive team would like to understand which products, subcategories & categories they should target or avoid.

Analytical Approach

Using the analysis is to understand sales patterns, identify areas for improvement, and provide strategic recommendations.

Dataset Definition

- Order ID : A unique identifier for each order.
- Order Date : The date when the order was placed.
- Ship Date : The date when the order was shipped.
- Customer : The name or identifier of the customer who placed the order.
- Manufactory : The manufacturer of the product.
- Product Name: The name of the product ordered.
- Segment : The market segment to which the customer belongs (e.g., Consumer, Corporate, Home Office).
- Category : The category of the product (e.g., Furniture, Office Supplies, Technology).
- Region : The geographical region where the order was placed (e.g., East, West, Central, South).
- Zip : The zip code of the customer's location.
- City : The city where the customer is located.
- State : The state where the customer is located.
- Country : The country where the customer is located.
- Discount : The discount applied to the order.
- Profit : The profit earned from the order.
- Quantity : The number of units ordered.
- Sales : The total sales amount for the order.
- Profit Margin: The profit margin percentage for the order.

Data Understanding

There are 9,994 orders with 5,009 unique data points, which means that customers made transactions with different products or on different days. There are 793 unique customers, with William Brown being the customer who made the most purchases. The country is, of course, only the United States.

Product

There are 3 segments, 3 categories, 17 subcategories, and 1,849 types of products. The most purchased product is the Staple Envelope, with 48 transactions.

Geographical Data

The data consists of 4 regions, with the West region having the highest sales. There are 531 unique cities, with New York City having the most entries, and 49 states, with California having the highest sales.

Data Cleaning & Preparation

- Handling missing data
- Ensuring data format consistency

0 Missing Values

```
order_id        0
order_date      0
ship_date       0
customer        0
manufactory     0
product_name    0
segment         0
category        0
subcategory     0
region          0
zip             0
city            0
state           0
country         0
discount        0
profit          0
quantity        0
sales           0
profit_margin   0
order_year      0
order_month     0
ship_year       0
ship_month      0
dtype: int64
```

Data Format
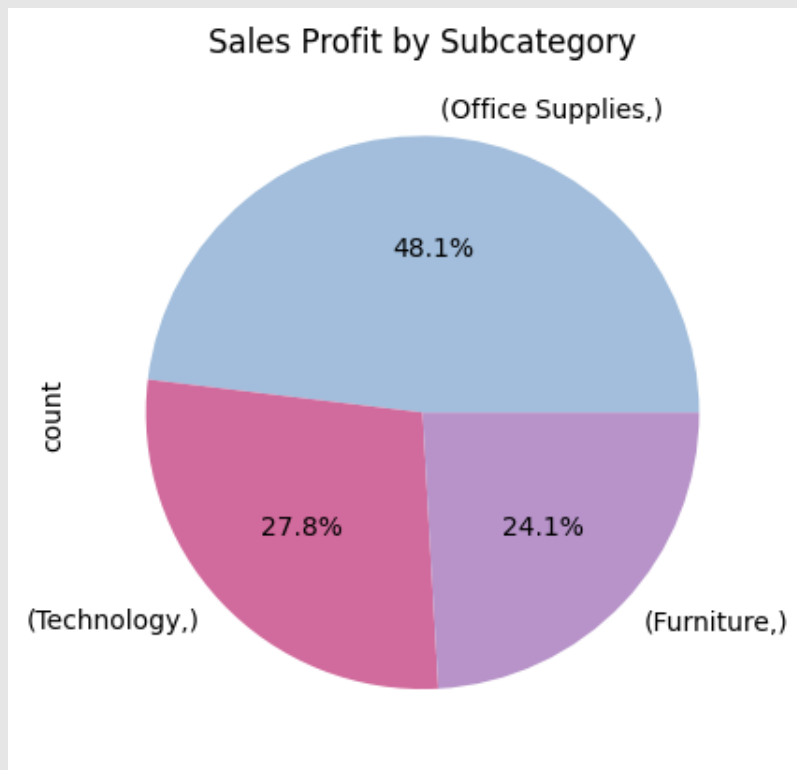
```
0   order_id       9994 non-null   object
1   order_date     9994 non-null   datetime64[ns]
2   ship_date      9994 non-null   datetime64[ns]
3   customer       9994 non-null   object
4   manufactory    9994 non-null   object
5   product_name   9994 non-null   object
6   segment        9994 non-null   object
7   category       9994 non-null   object
8   subcategory    9994 non-null   object
9   region         9994 non-null   object
10  zip            9994 non-null   int64
11  city           9994 non-null   object
12  state          9994 non-null   object
13  country        9994 non-null   object
14  discount       9994 non-null   float64
15  profit         9994 non-null   float64
16  quantity       9994 non-null   int64
17  sales          9994 non-null   float64
18  profit_margin  9994 non-null   float64
19  order_year     9994 non-null   int32
```

Exploratory Data Analysis

Top Categorical Values

	count	top	freq
order_id	5009	US-2023-100111	14
customer	793	William Brown	37
manufactory	182	Other	1893
product_name	1849	Staple envelope	48
segment	3	Consumer	5191
category	3	Office Supplies	6026
subcategory	17	Binders	1523
region	4	West	3203
city	531	New York City	915
state	49	California	2001
country	1	United States	9994

The order ID 'US-2023-100111' is the most recorded, but it belongs to Seth Vernon. However, the customer who made the most purchases is William Brown. This is because William Brown made many purchases with different order IDs, so he does not have the highest single order ID

Sales Profit by Subcategory

- (Office Supplies,) 48.1%
- (Technology,) 27.8%
- (Furniture,) 24.1%

count

Sum of Profit by Subcategory

None,None
- (profit, sum)
- (sales, sum)

Sum of Profit

Subcategory: Accessories, Appliances, Art, Binders, Bookcases, Chairs, Copiers, Envelopes, Fasteners, Furnishings, Labels, Machines, Paper, Phones, Storage, Supplies, Tables

Highest profit is earned in Copiers while Selling price for Chairs and Phones is extremely high compared to other products.

People don't prefer to buy Tables and Bookcases from Superstore. Hence these departments are in loss.
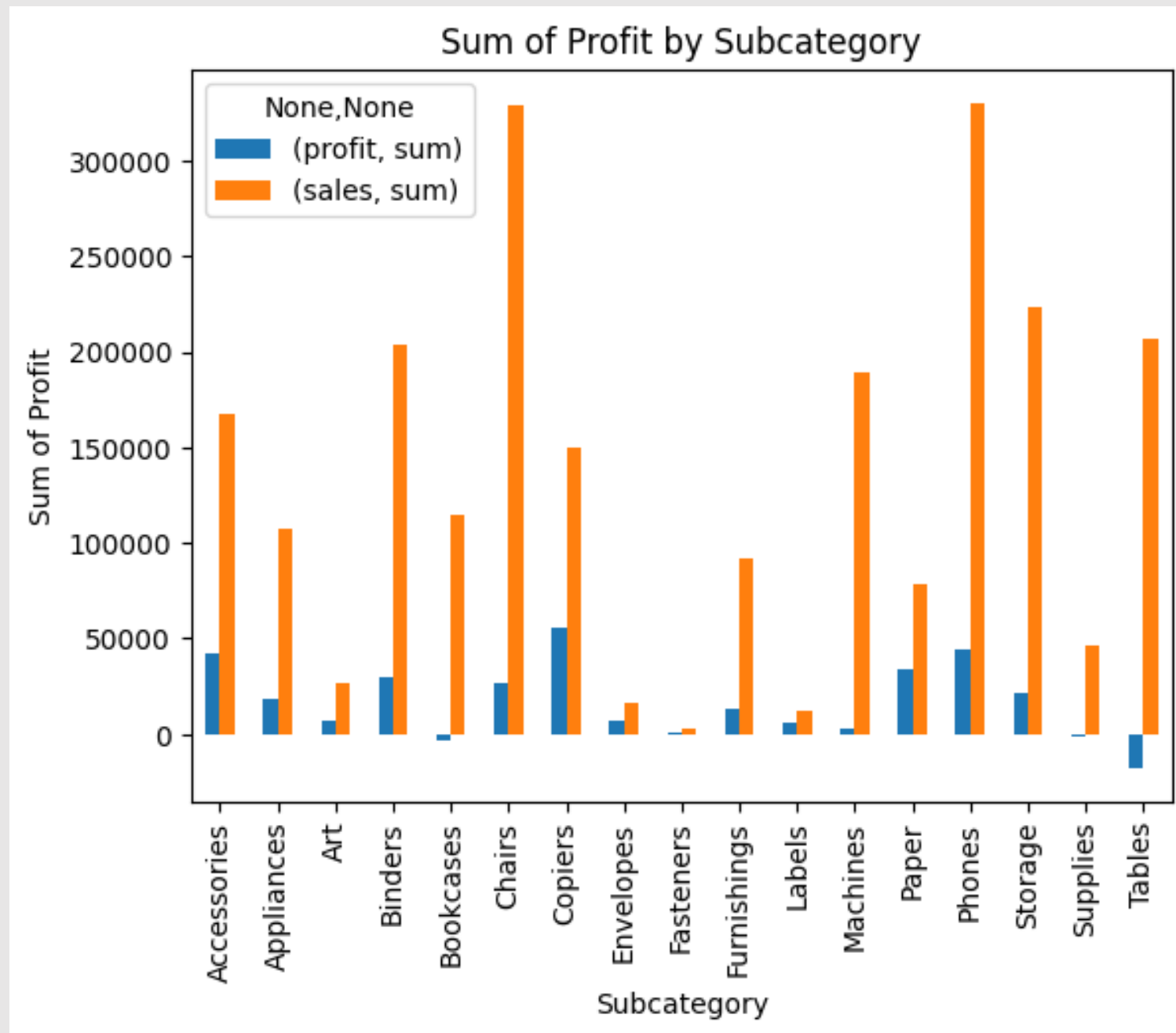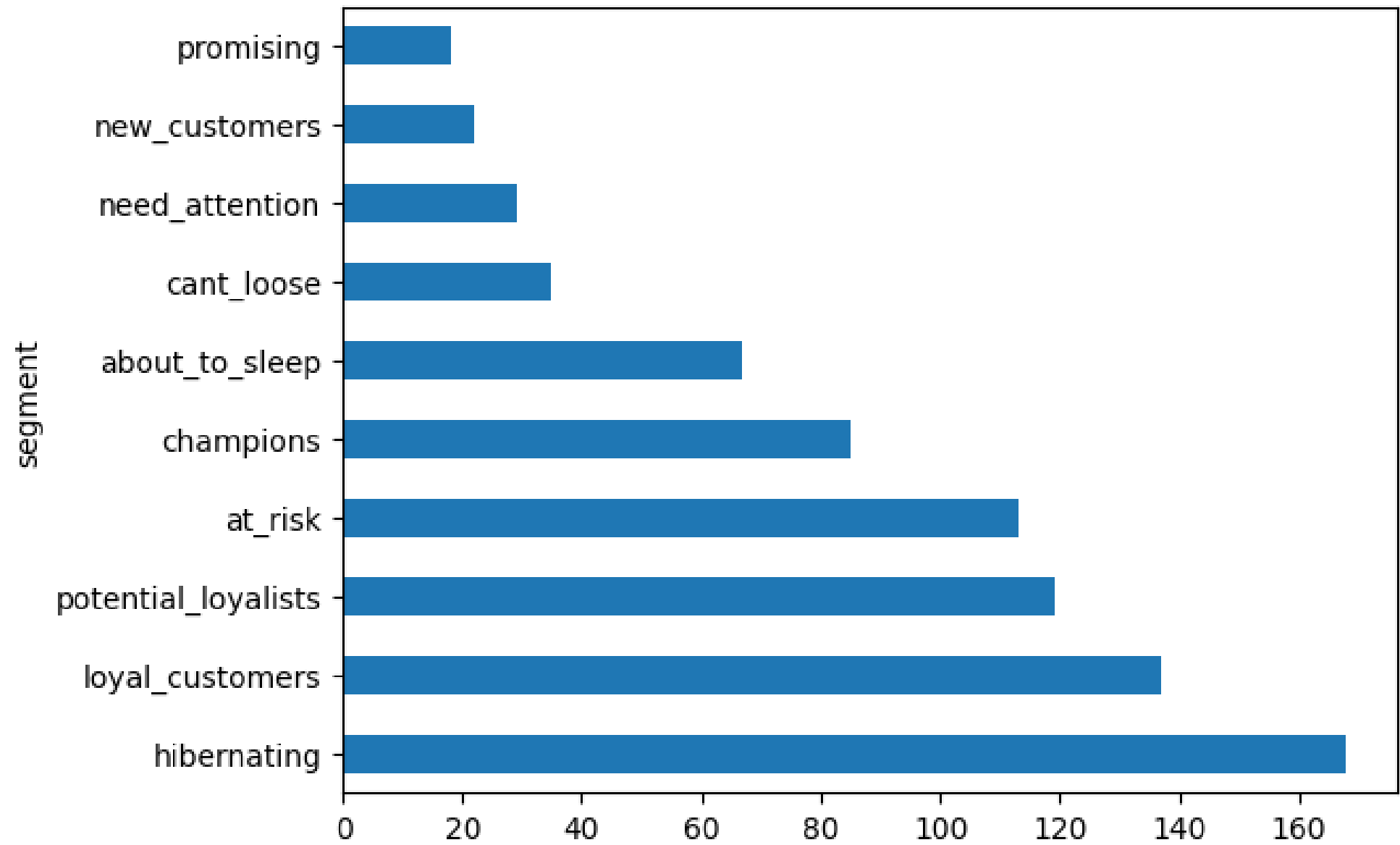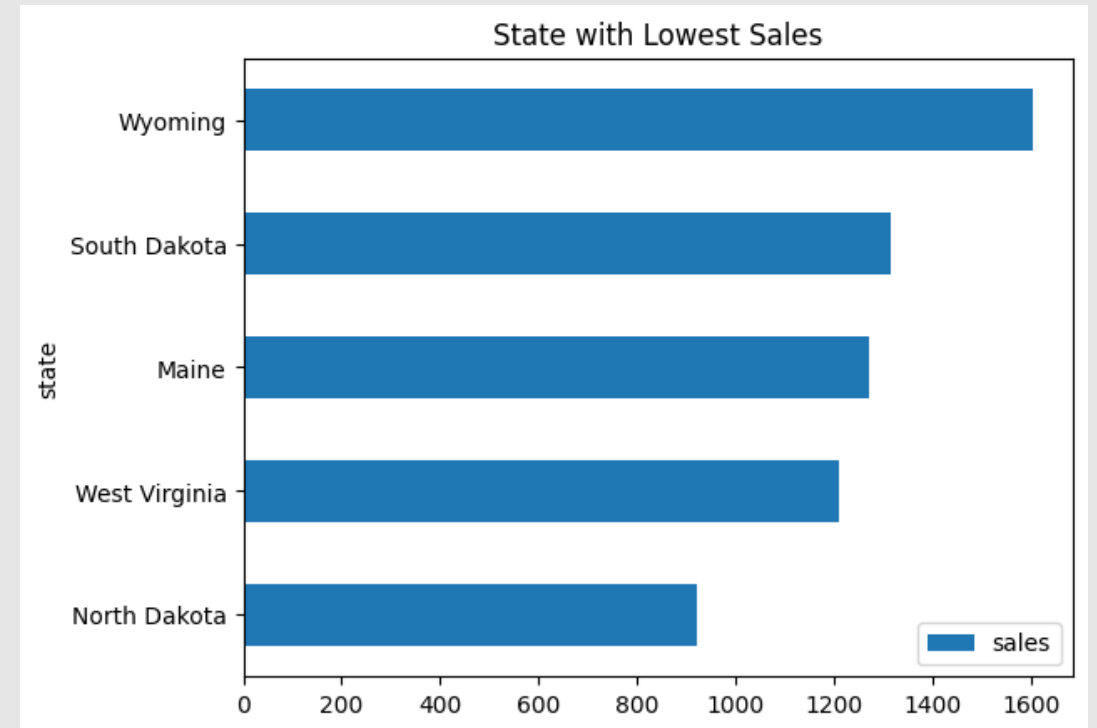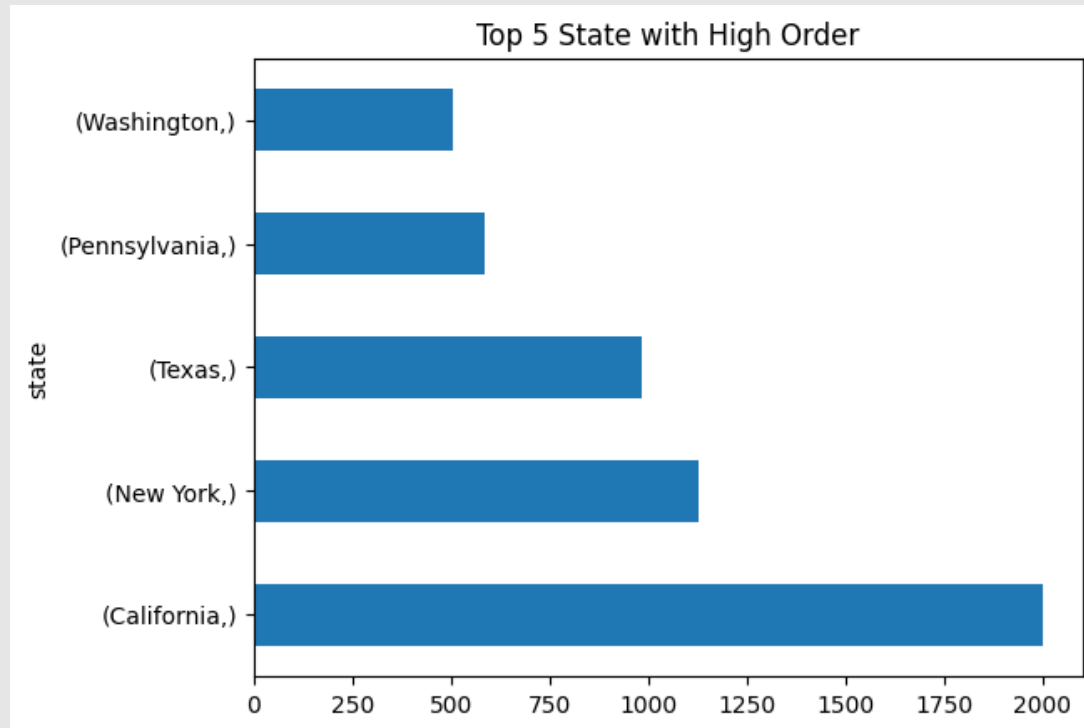
Top 10 Profitable Customers

Segment
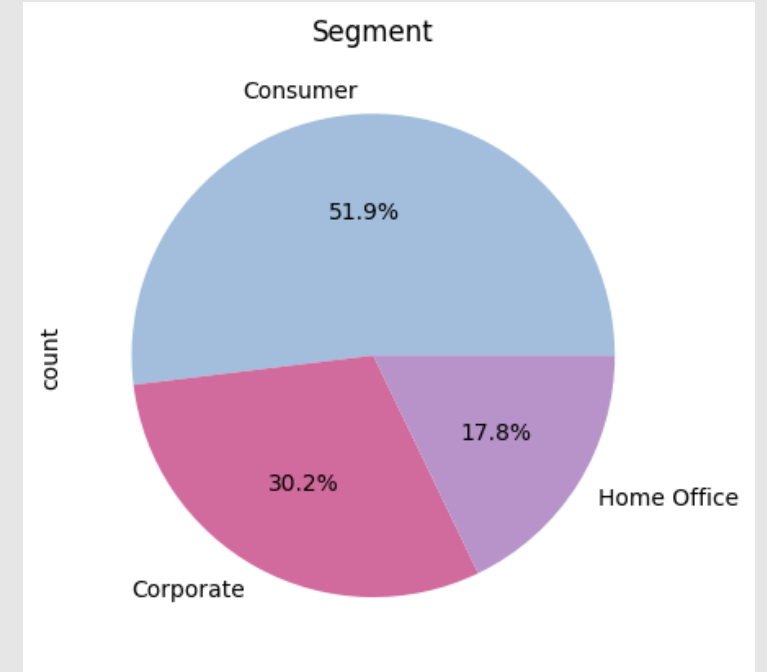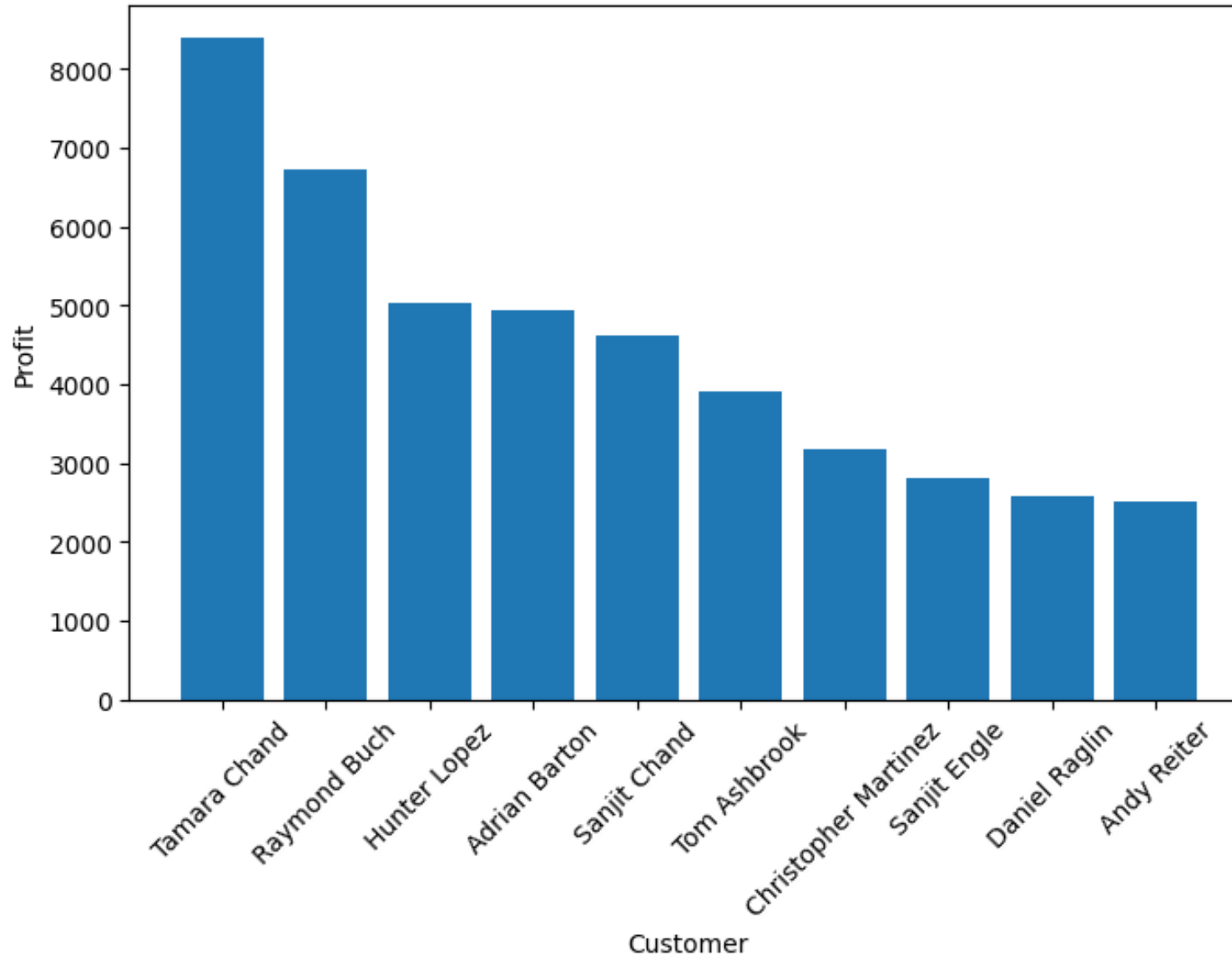
A profitable customer is identified by calculating the profit generated by each customer, indicating their potential for repeat orders. This data can be used to tailor marketing strategies, enhance customer retention efforts, and identify key customer segments that drive the most revenue.
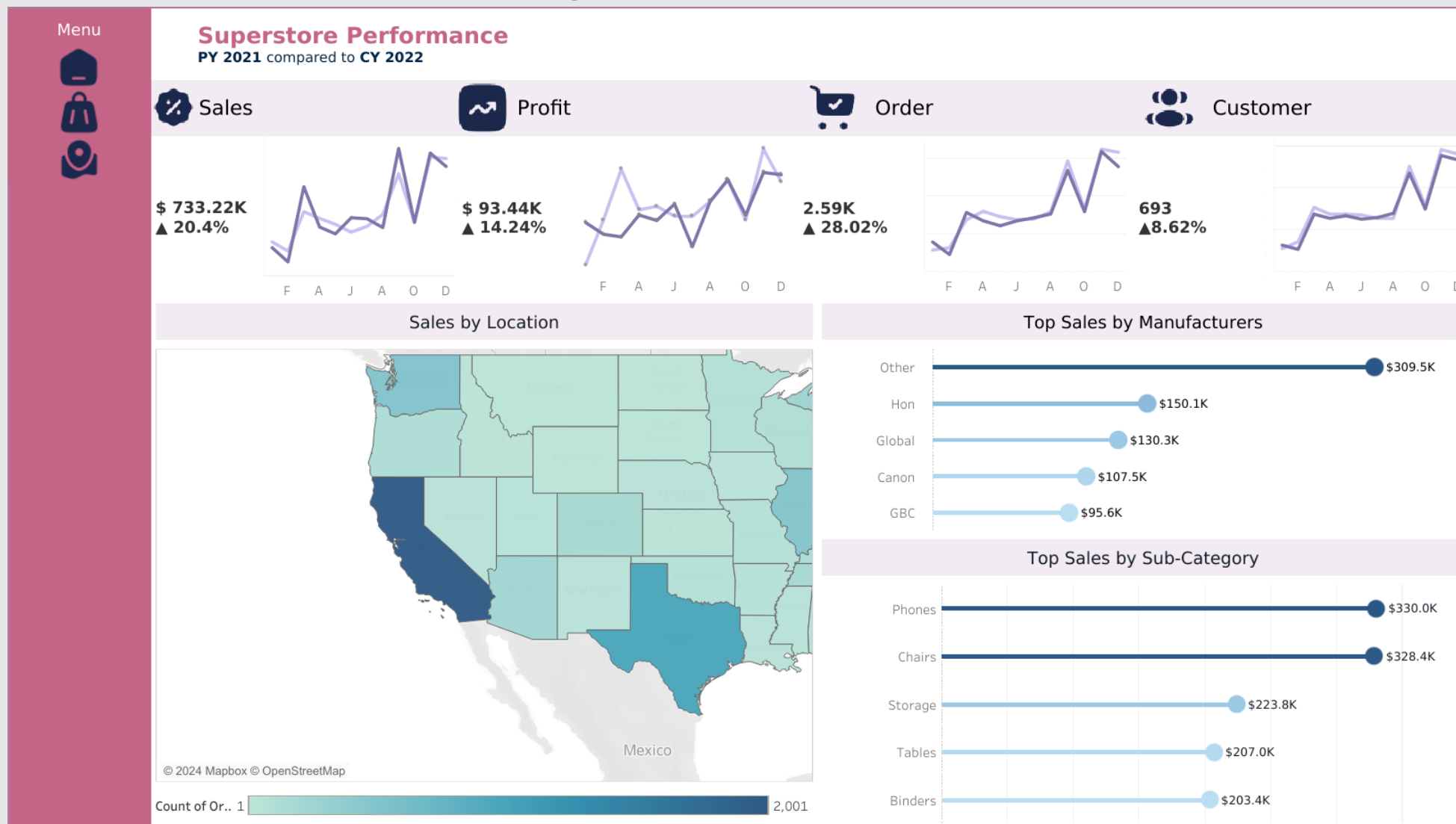
Count Customer Segment

we can focus our sales efforts on the states with high orders and further analyze the factors contributing to the low sales in the states with the lowest sales.

Data Visualization

Click this image to switch to the Tableau dashboard

Product Analysis

Products with High Sales & Profits

product_name	profit
GBC Ibimaster 500 Manual ProClick Binding System	4946.3700
Fellowes PB500 Electric Punch Plastic Comb Bin...	3177.4750
Hewlett Packard LaserJet 3310 Copier	2591.9568
Fellowes PB500 Electric Punch Plastic Comb Bin...	1906.4850
Fellowes PB500 Electric Punch Plastic Comb Bin...	1906.4850
GBC DocuBind TL300 Electric Binding System	1453.1238
Hewlett Packard LaserJet 3310 Copier	1439.9760
Fellowes PB500 Electric Punch Plastic Comb Bin...	1270.9900
GBC DocuBind TL300 Electric Binding System	1264.7559
Fellowes PB500 Electric Punch Plastic Comb Bin...	1143.8910
Hewlett Packard LaserJet 3310 Copier	1007.9832
Hewlett Packard 610 Color Digital Copier / Pri...	899.9820
GBC DocuBind TL300 Electric Binding System	843.1706
GBC DocuBind TL300 Electric Binding System	843.1706
GBC DocuBind TL300 Electric Binding System	843.1706

Subcategory with High Sales & Profits

subcategory	profit
Chairs	746.4078
Binders	274.4910
Phones	204.1092
Phones	113.6742
Chairs	87.3504
Storage	238.6530
Storage	34.3548
Accessories	65.9780
Envelopes	36.6930
Appliances	93.5816
Phones	52.6344
Phones	42.0186
Chairs	51.4764
Chairs	54.1764
Chairs	284.9800

Stars: Products with high revenue and high sales volume.

The calculation is done by finding the third quartile (Q_3) of the 'profit' data and the count of sales (value.count()) of the 'product' data. Then, the output result is sorted based on the profit.

Product Analysis

Products with Low Sales & Profits

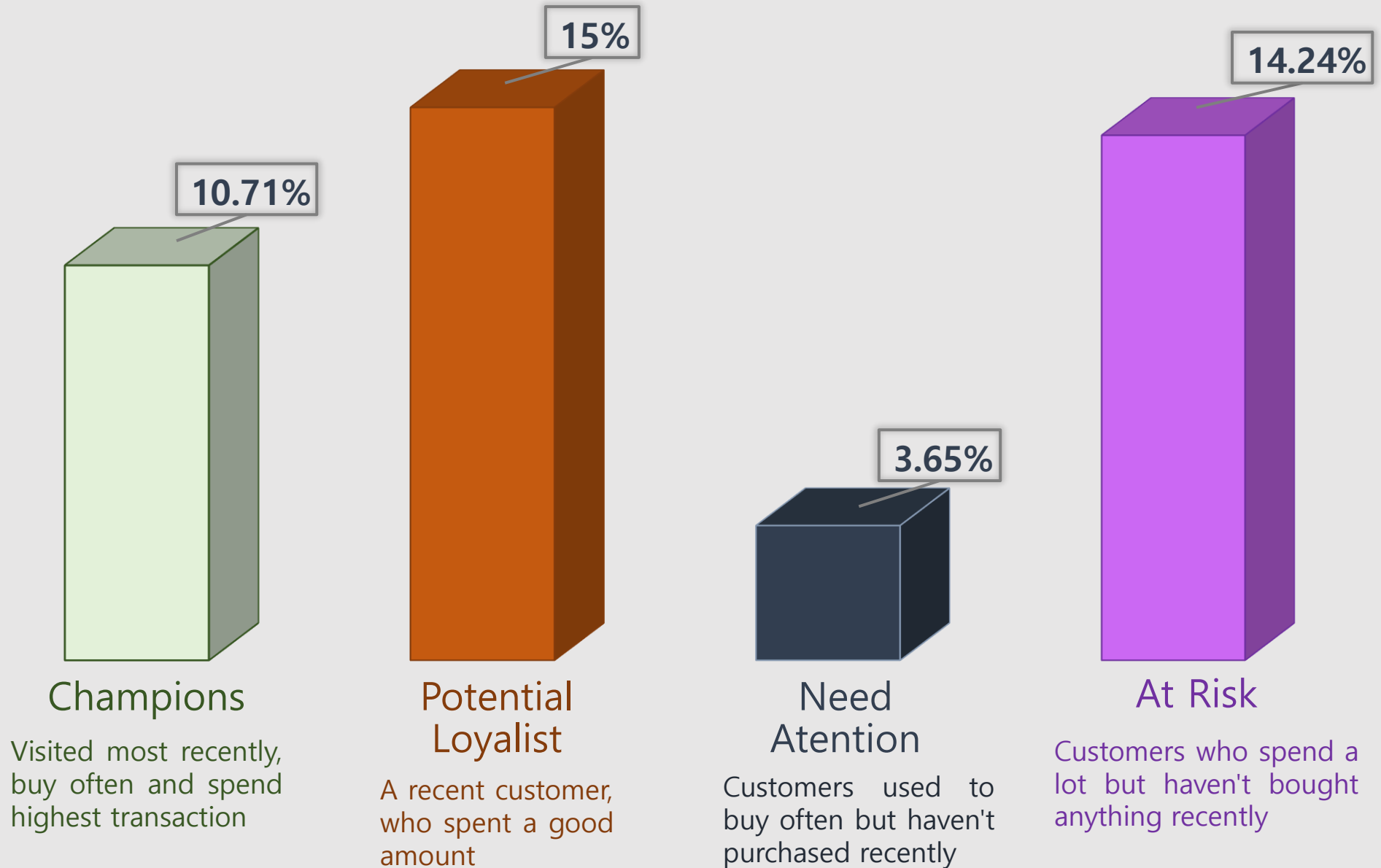| product_name	profit
Cubify CubeX 3D Printer Double Head Print	-6599.9780
Cubify CubeX 3D Printer Triple Head Print	-3839.9904
Ibico EPK-21 Electric Binding System	-2929.4845
Cubify CubeX 3D Printer Double Head Print	-2639.9912
Cisco TelePresence System EX90 Videoconferenci...	-1811.0784
Cisco 9971 IP Video Phone Charcoal	-950.4000
HP Designjet T520 Inkjet Large Format Printer ...	-944.9946
Zebra GK420t Direct Thermal/Thermal Transfer P...	-938.2800
Epson TM-T88V Direct Thermal Printer - Monochr...	-935.9595
High Speed Automatic Electric Letter Opener	-786.0144
Okidata MB491 Multifunction Printer	-643.7100
3D Systems Cube Printer, 2nd Generation, White	-571.9956
BoxOffice By Design Rectangular and Half-Moon ...	-566.5625
Swingline SM12-08 MicroCut Jam Free Shredder	-479.9880
BoxOffice By Design Rectangular and Half-Moon ...	-420.0000

Subcategory with Low Sales & Profits

category	subcategory	profit
Technology	Machines	-6599.9780
Technology	Machines	-3839.9904
Office Supplies	Binders	-2929.4845
Technology	Machines	-2639.9912
Technology	Machines	-1811.0784
Technology	Machines	-950.4000
Office Supplies	Supplies	-786.0144
Furniture	Tables	-566.5625
Furniture	Tables	-420.0000
Furniture	Bookcases	-376.6932
Office Supplies	Appliances	-251.1864
Furniture	Bookcases	-247.7988
Office Supplies	Appliances	-225.5568
Office Supplies	Appliances	-197.5752
Furniture	Bookcases	-190.8522

Dogs: Product with low revenue and low sales

The calculation is done by finding the third quartile (Q_1) of the 'profit' data and the count of sales (value.count()) of the 'product' data. Then, the output result is sorted based on the profit.
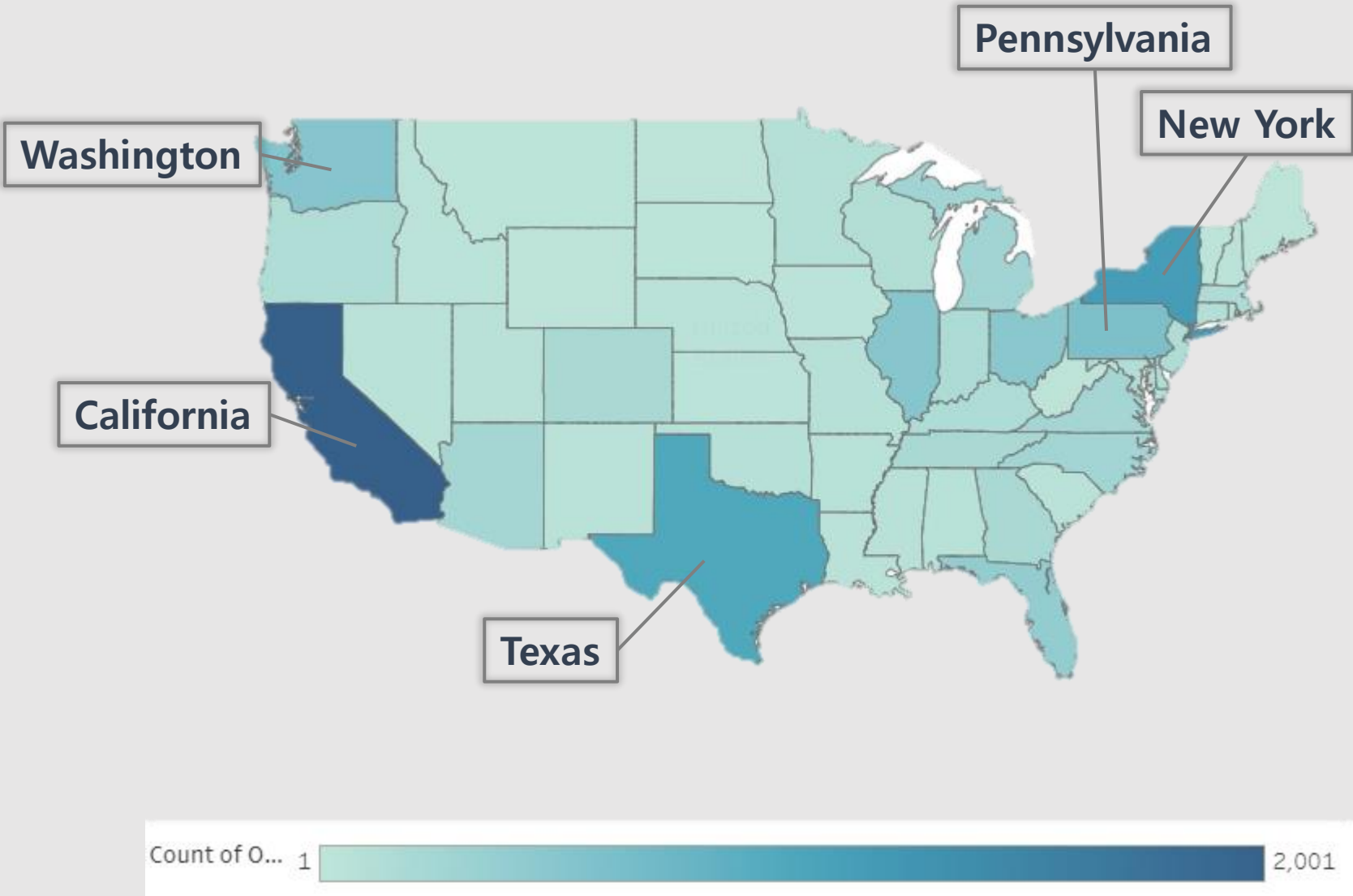
Customer Analysis

10.71%

15%

3.65%

14.24%

Champions
Visited most recently, buy often and spend highest transaction

Potential Loyalist
A recent customer, who spent a good amount

Need Atention
Customers used to buy often but haven't purchased recently

At Risk
Customers who spend a lot but haven't bought anything recently

Conclusion and Recommendation

The main goal of our exploratory data analysis (EDA) is to identify areas, product, and customer where we can enhance our sales and overall profitability.

To boost our total profit, we should focus on the Southern states as they generated less revenue. Additionally, increasing sales of underperforming product types, such as Supplies, Bookcases, and Tables, can help increase our overall earnings. Furthermore, the Furniture category has not been very profitable, so efforts to improve this category could also be beneficial. Addressing these aspects can lead to better overall sales and profit.