



Superstore Analysis in the US

AUTHOR: EBUBE NWAIGBO

December 1, 2022

Problem Statement

- ❖ An early stage startup is venturing into the e-commerce industry but needs to understand the market first.
- ❖ The goal is to collect, clean and analyze data from a variety of sources so as to provide recommendations on how to succeed in the e-commerce Industry.

Data Description

The data was gotten from a github repository, you can find the link [here](#). The dataset contains sales transactions for 2015, from the analysis, the store operates in 4 regions and 49 states. Some of the insights gotten from the analysis may be applicable to other e-commerce platforms.

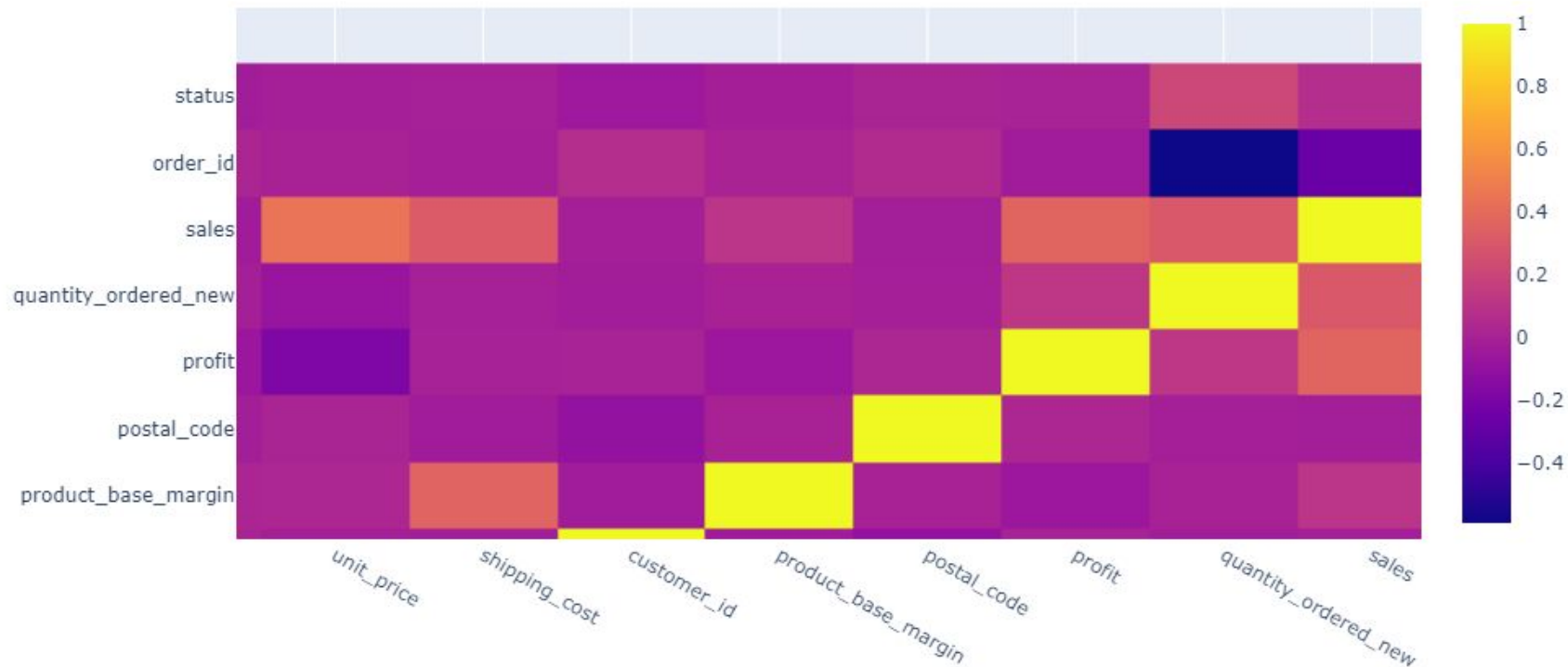
The Data consists of 1952 observations & 27 features:

- ❖ row_id
- ❖ order_priority
- ❖ discount
- ❖ unit_price
- ❖ shipping_cost
- ❖ customer_id
- ❖ customer_name
- ❖ ship_mode
- ❖ customer_segment
- ❖ product_category
- ❖ product_sub-category
- ❖ product_container
- ❖ product_name

- ❖ product_base_margin
- ❖ country
- ❖ manager
- ❖ region
- ❖ state_or_province
- ❖ city
- ❖ postal_code
- ❖ order_date
- ❖ ship_date
- ❖ profit
- ❖ quantity_ordered_new
- ❖ sales
- ❖ order_id
- ❖ status

A Correlation Heatmap

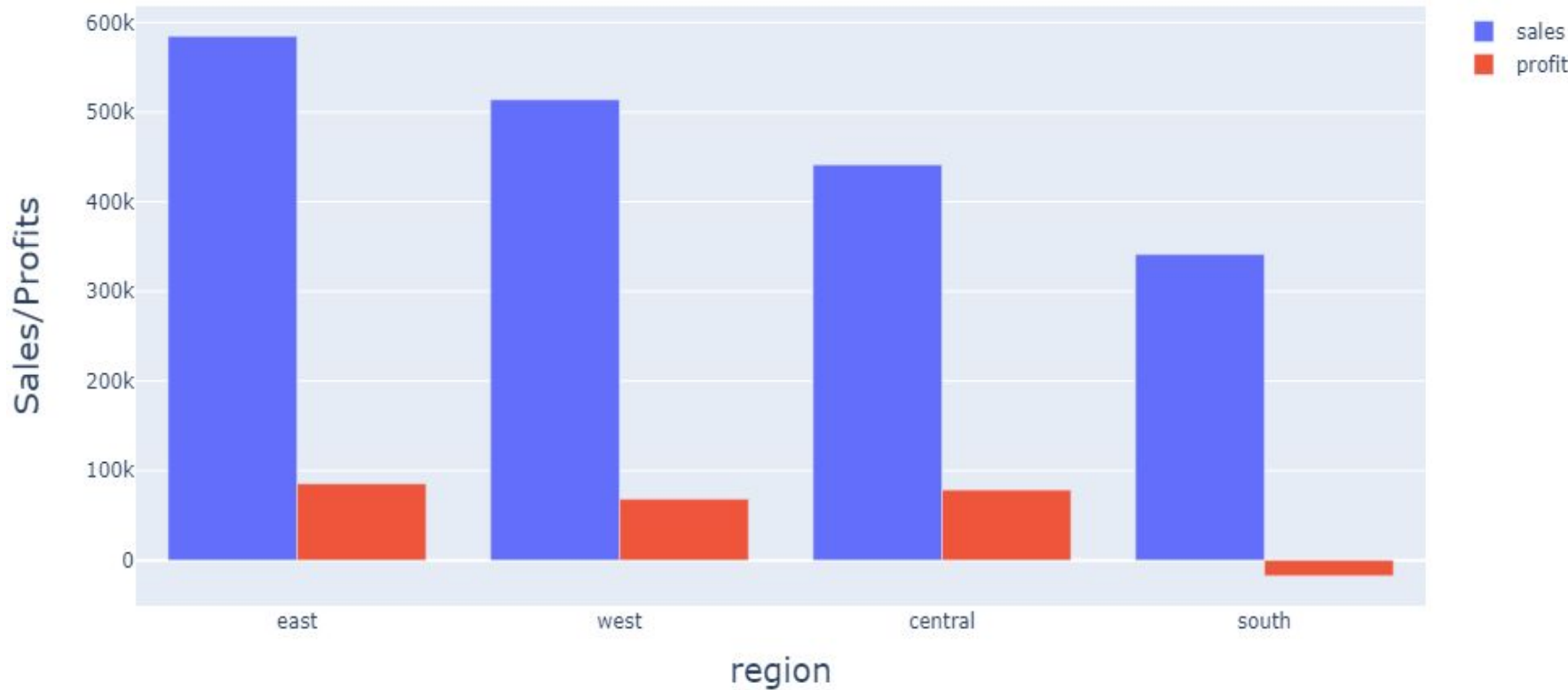
Correlation Heatmap



FINDINGS:

Variables like profit and sales; profit and unit_price; shipping_cost and sales; have a very strong correlation.

Sales/Profit per Region

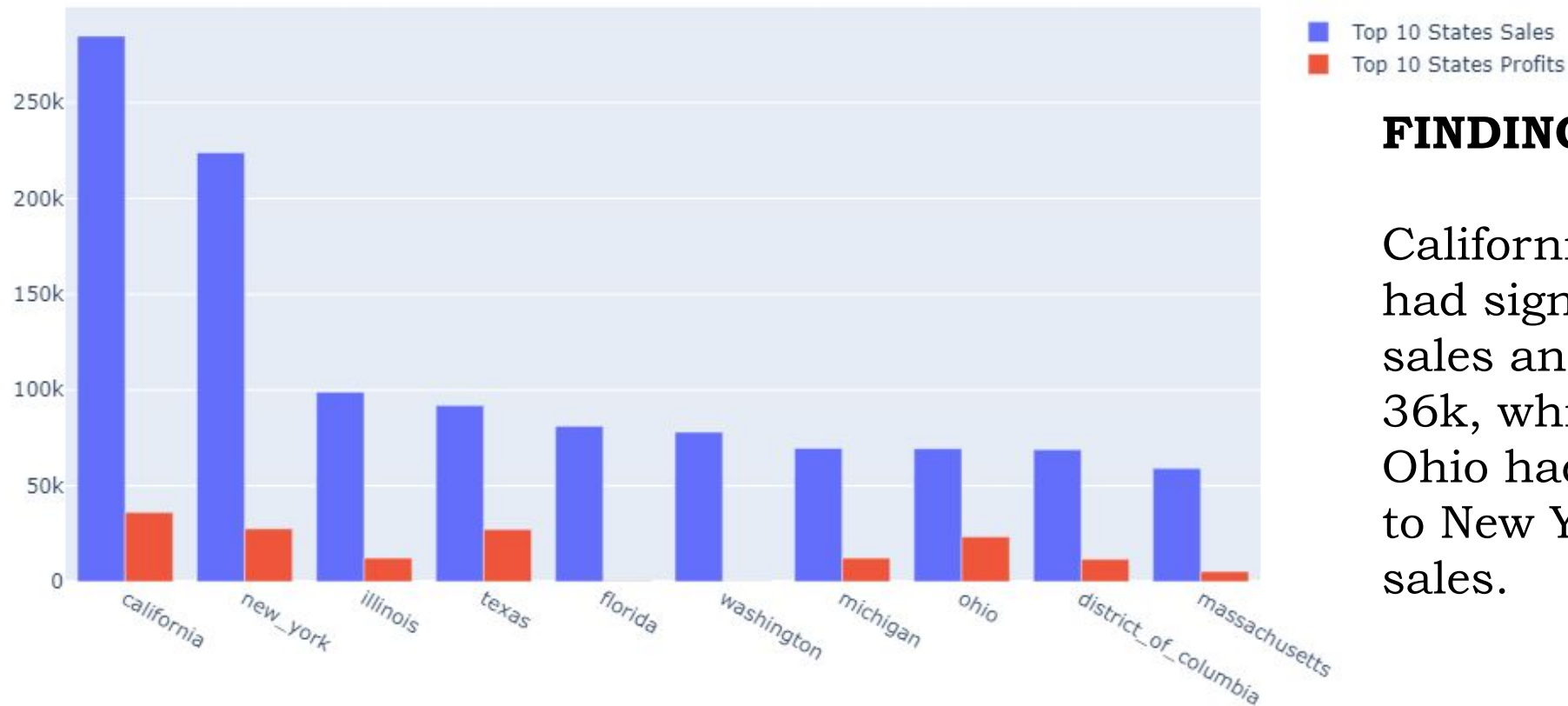


FINDINGS:

The east region had more sales and profit than any other region while the south, despite making sales, had a loss.

Sales/Profit for Top 10 States

Grouped Bar Plot For Sales and Profits
(For The Top Ten State Sales)

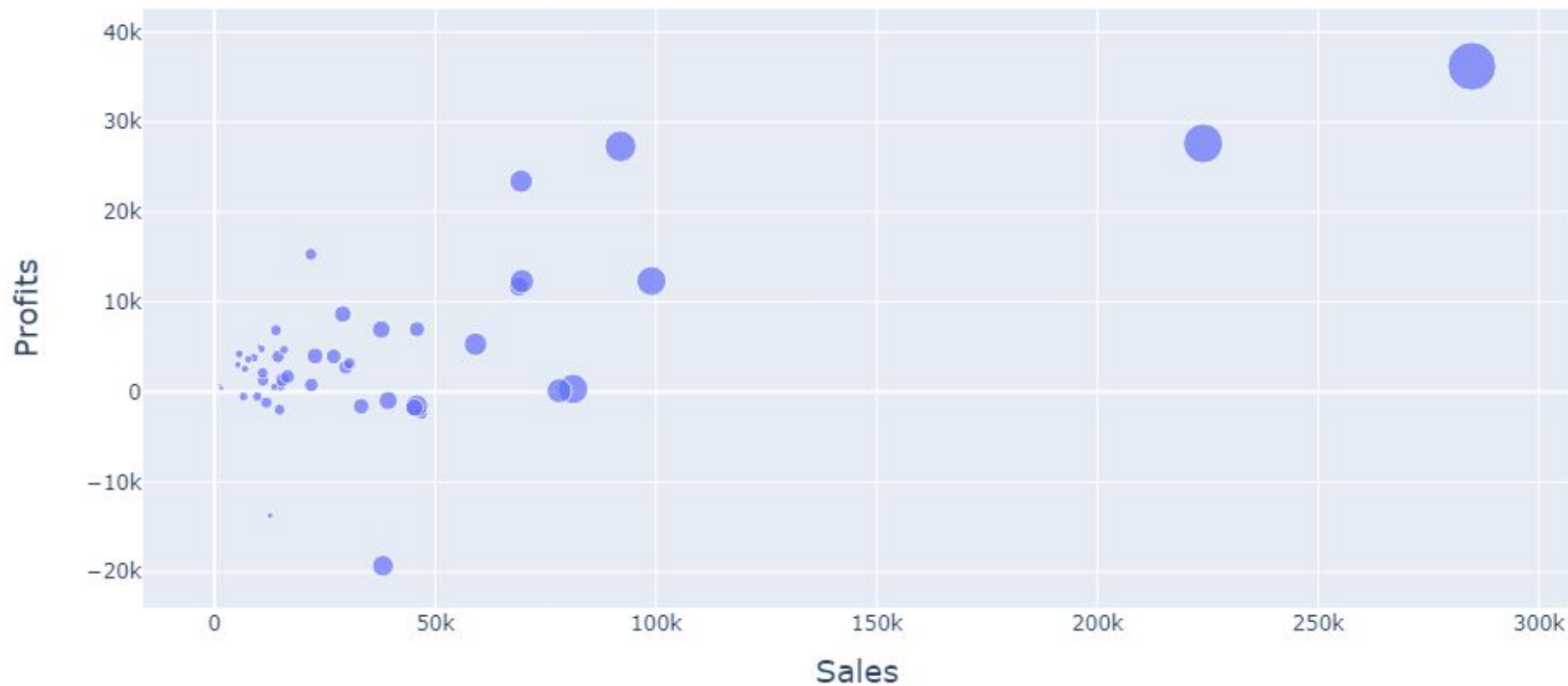


FINDINGS:

California and New York had significantly high sales and profits around 36k, while Texas and Ohio had similar profits to New York but low sales.

Sales/Profit for all States

Sales & Profits by States

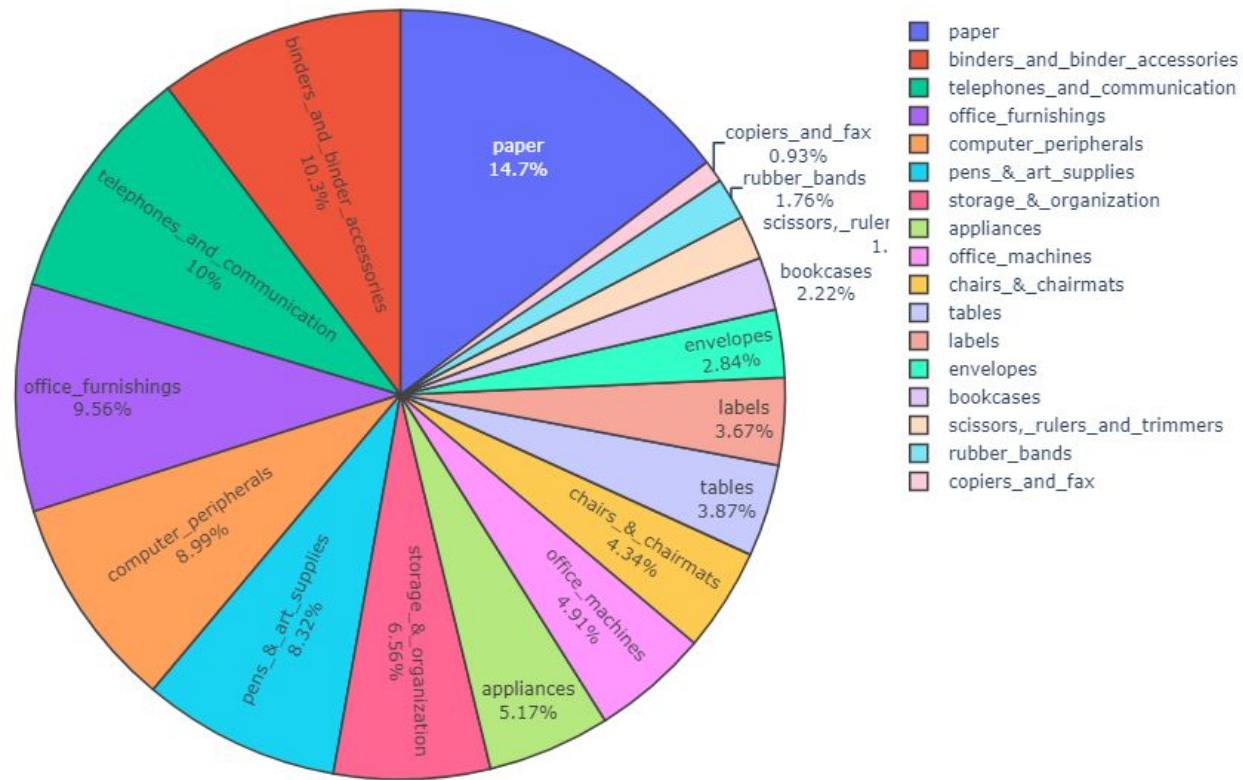


FINDINGS:

Only New York and California had sales of over 223k and 284k respectively, Majority of the states were clustered around 0k-100k.

Percentage Sales per Product Sub-Category

Pie Plot for Total Sales for each Product-Sub-Category

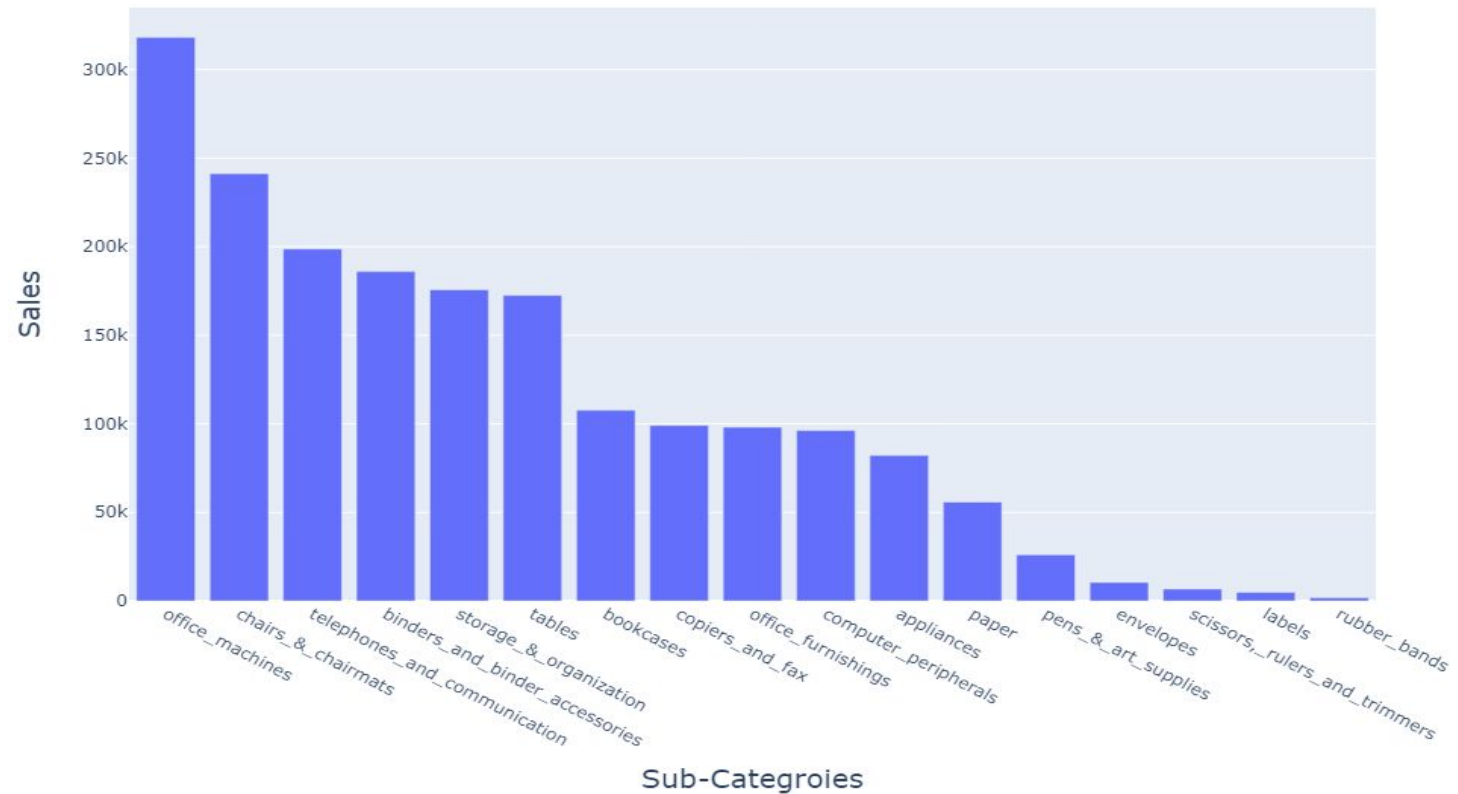


FINDINGS:

Tangible products seem to gain more purchases compared to smaller items like envelopes, bookcases, rubber bands, etc.

Product Sub-Category Sales

Bar Plot For Product-Sub-Categories Sales

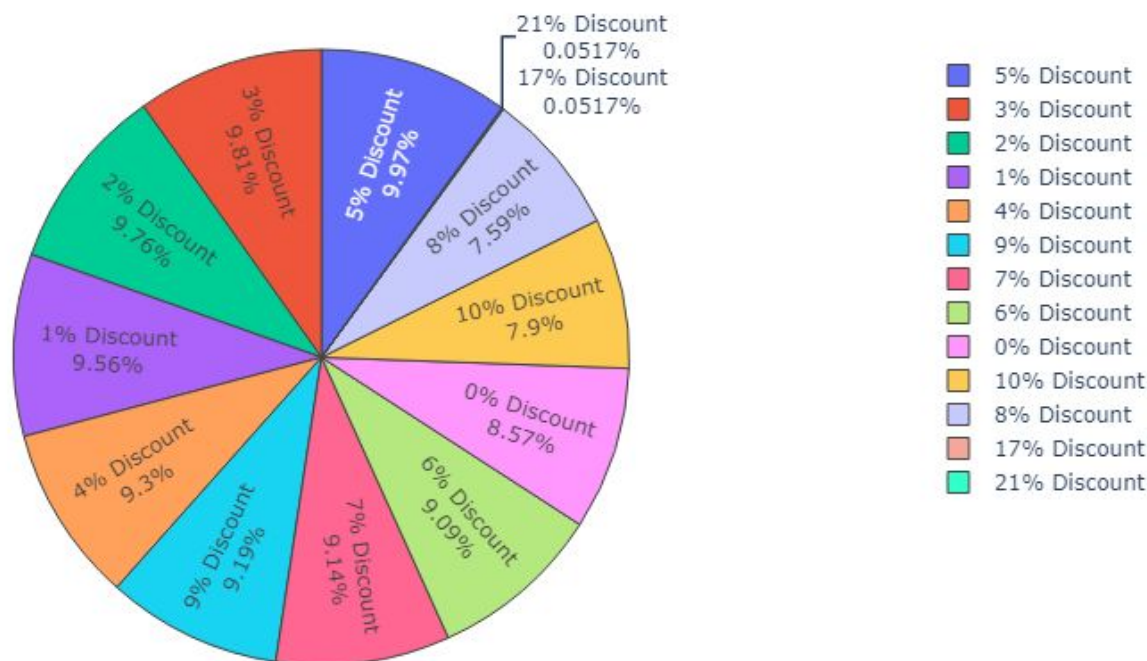


FINDINGS:

The top 3 highest sales came from office machines, chairs, chair mats, telephone and communication.

Percentage of Discount Given by the Stores

Discount Pie Plot



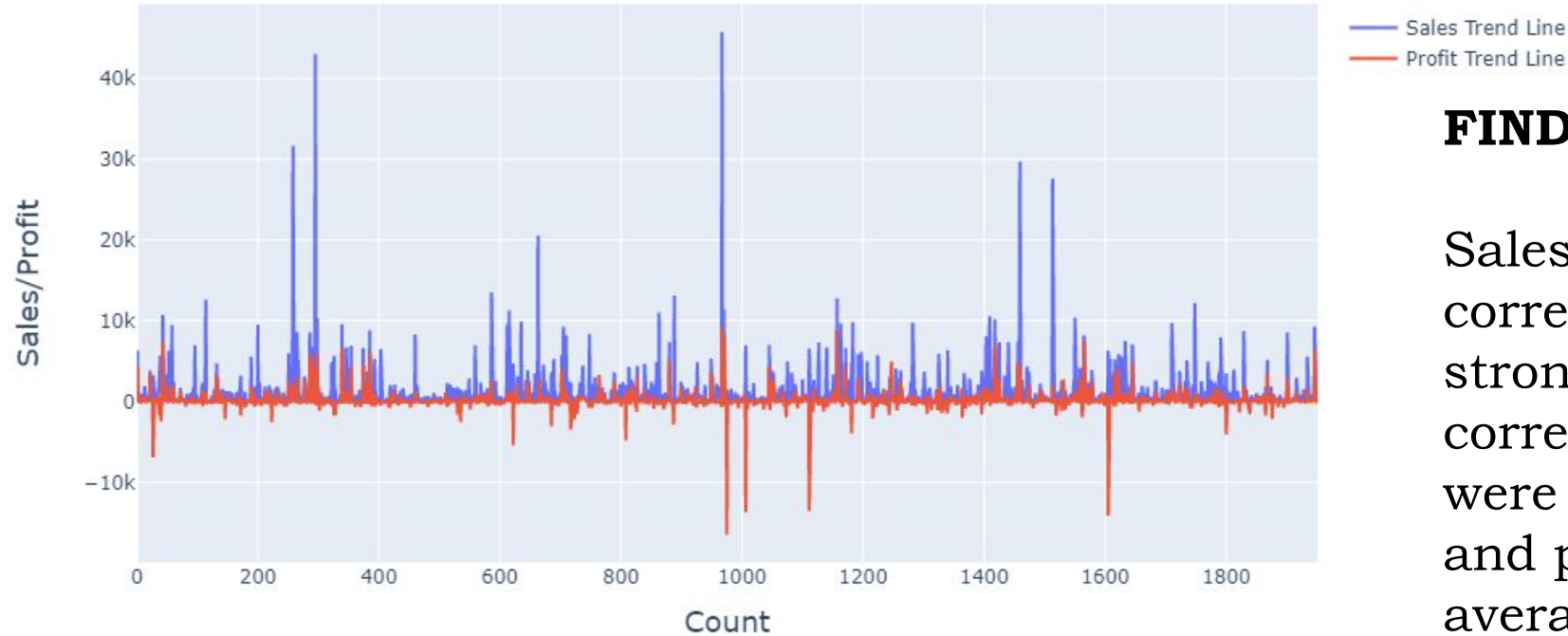
FINDINGS:

9.97% of the products' sub-categories got 5% discount while the others of relatively same proportion got higher discounts. The very least had the most discounts of 17% and 21% respectively.

These similar proportions questions the criteria for determining a discount.

Sales & Profit Trend

Sales/Profit Trend plot



FINDINGS:

Sales and profit have a correlation but not a very strong positive correlations. Most sales were between 0 - 48k and profit trends were on average during the sales period, with some negatives.

Key findings: Summary

- ❖ The `Eastern region` had more sales and profit amongst other other region. The `Southern region` had losses. Also, the type of pricing policy might differ from both regions and this can be used to increase profit.
- ❖ This analysis shows that `California` and `New York` had more sales and profit than other states, while `Florida` and `Washington` made losses.
- ❖ The highest sales made was about 45.7k but the profit was at its lowest during this period at about -16.7k. This could be a result of many things such as extreme discount sales, massive returns, late delivery timeframe.

Key findings: Summary (Cont.)

- ❖ The analysis also shows that more sales were made from office furniture, chairs, and chair mats, telephone and communications, etc while very low sales were made on rubber bands, labels, scissors, rubber, and trimmers.
- ❖ There was a massive quantity/count of products sold on paper, binders and binders accessories, telephone and communication, office furniture.
- ❖ The discount given to the highest product sub-category is 5%. While the least proportions had 17% and 21% respectively.

Top 3 Metrics to Determine Success of the Product

- ❖ **Sales Metrics:** This will help determine if the company is making sales or not. Here, we calculate sales cycle length, average selling price, average profit margin and average purchases value.
- ❖ **Financial metrics:** This should track sales turnover, profits, expenditures, assets, liabilities, and capital. They will help in improving operational efficiency.
- ❖ **Marketing metrics:** We target customer reach using conversion rate, SEO metrics and website traffic.

Problems the insight generated could solve for an e-commerce company.

- ❖ This analysis would help companies solve inventory problems based on what type of goods to sell first.
- ❖ Customers Behavior is a very important factor to consider in an e-commerce company, as this would help put recommendation systems in place which would lead to more sales.
- ❖ Storage location is another critical dilemma faced by most e-commerce companies, as they would love to be closer to their customers for faster delivery and reduced delivery services.

Business Recommendations

- ❖ The Eastern region of the country seems very favourable. Therefore, it could be the first location of the business and its stores, since they make more purchases.
- ❖ Implement recommender systems so as to understand the shopping patterns of the customers and recommend products tailored to their interests.
- ❖ Investing more in goods that have higher sales and profit. For example: Papers, Office furniture, Chairs, etc.
- ❖ Low selling goods like Trimmers, Staplers, etc. can gradually be adopted based on customer demands and cost price.

You can find the link to my github repository where this analysis was carried out [here](#).

Thank you!

