IBM SKILLSBUILD PROJECT

NAME: LALNUNDIKA DARLONG
COLLEGE NAME: SHRI RAM COLLEGE OF COMMERCE
COLLEGE STATE: DELHI
INTERNSHIP DOMAIN: DATA ANALYTICS
INTERNSHIP DATE: 12 JUNE - 24 JULY 2023

PROBLEM STATEMENT

We were given a data set containing information pertaining to the transactions and other details of an American super store which we had to analyze.

AGENDA

1. *Project Overview*
2. *End Users*
3. *Solution & Value Proposition*
4. *Customization*
5. *Modelling*
6. *Results*
7. *Links*

PROJECT OVERVIEW

Purpose

To analyze the data in the Super Store database to draw meaningful conclusions

Scope

Using Python and Excel to conduct the data analysis

Objectives

To Find:

1. **Top Selling Product Types**
2. **Top Most Profitable Sales**
3. **Top Discounted Products**
4. **Top Sales & Profits**
5. **Sale & Discount Correlation**

END USERS

Super Store Owner(s)

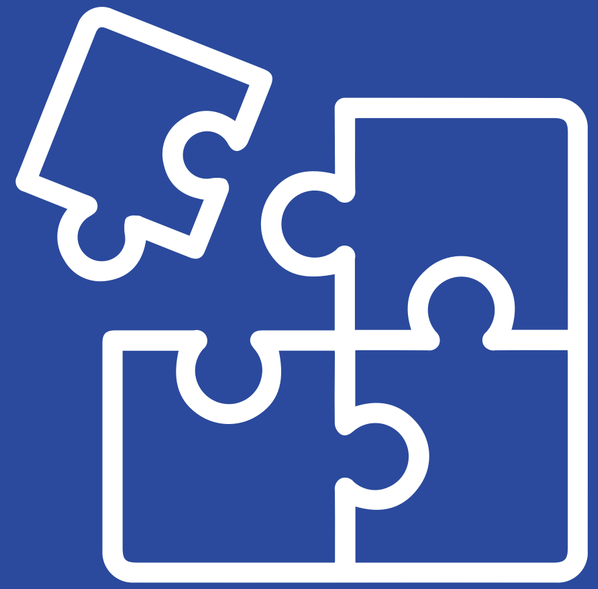
Potential Investors

Marketing Team

Sales & Outreach Team

SOLUTION

By analyzing the data to draw conclusions out of sales and profits by region and the correlation between sales and the discounts given, the end users can gauge the long-term feasibility of the superstore and make their decisions accordingly.



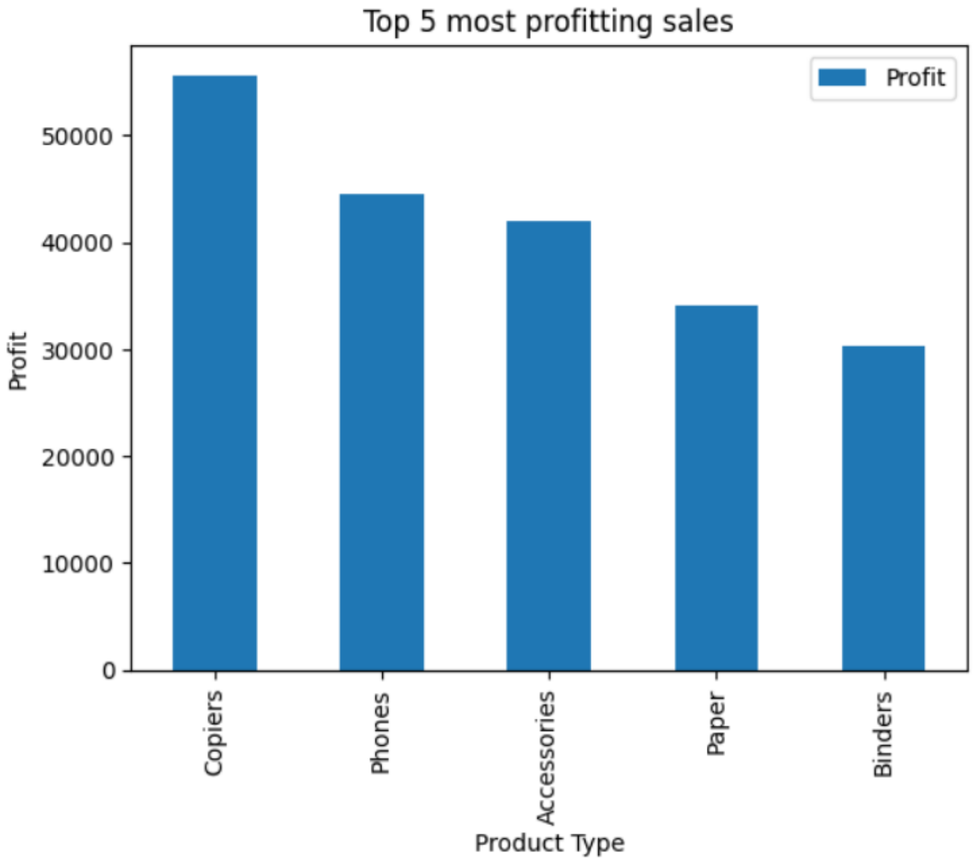
CUSTOMIZATION

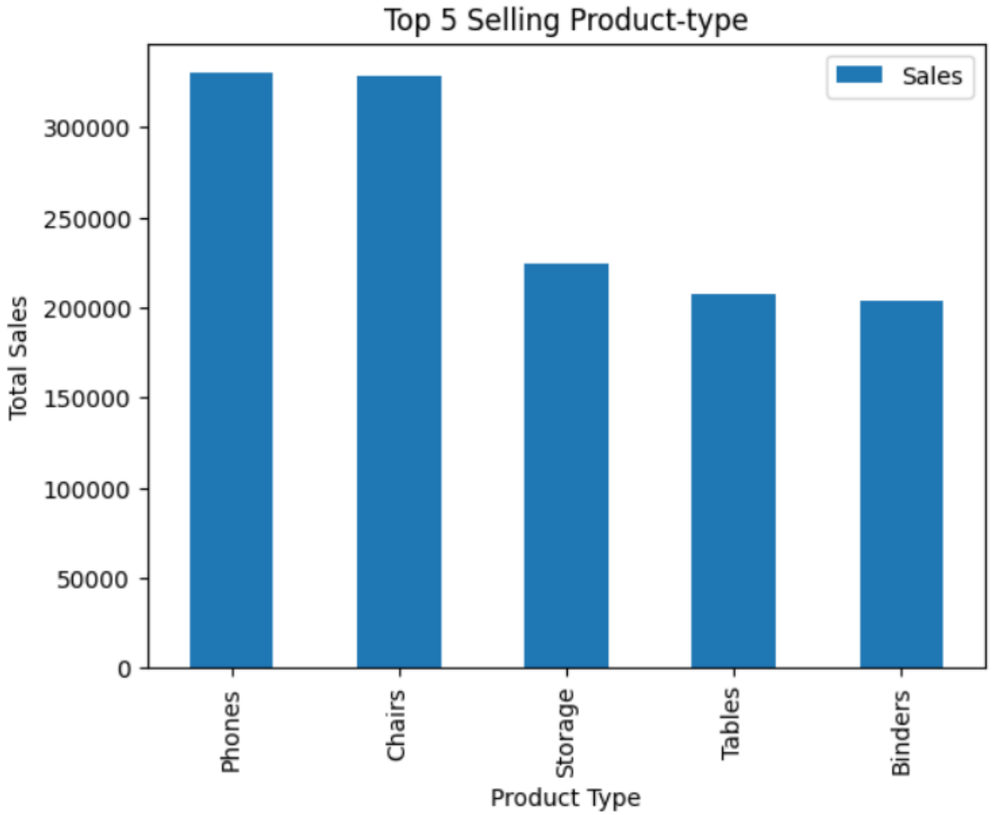
Having worked with a fellow intern on this project, we chose to present the data visualization in the standard bar charts as well as construct a scattergraph to show the correlation between sales and discounts which is something we believe to be somewhat unique.

MODELING

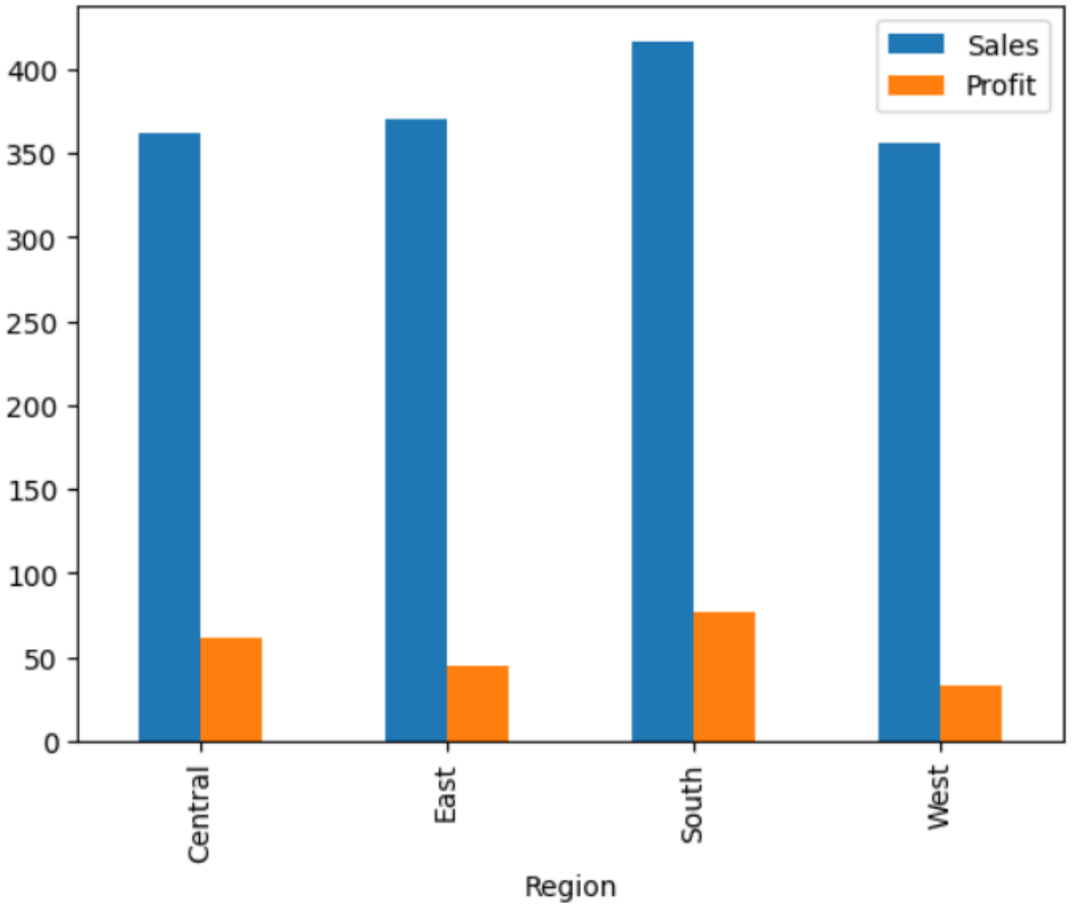
Pattern recognition, Custom filtering, and data visualization through graphs and charts to showcase trends and highlight differences in datapoints depending on the area.

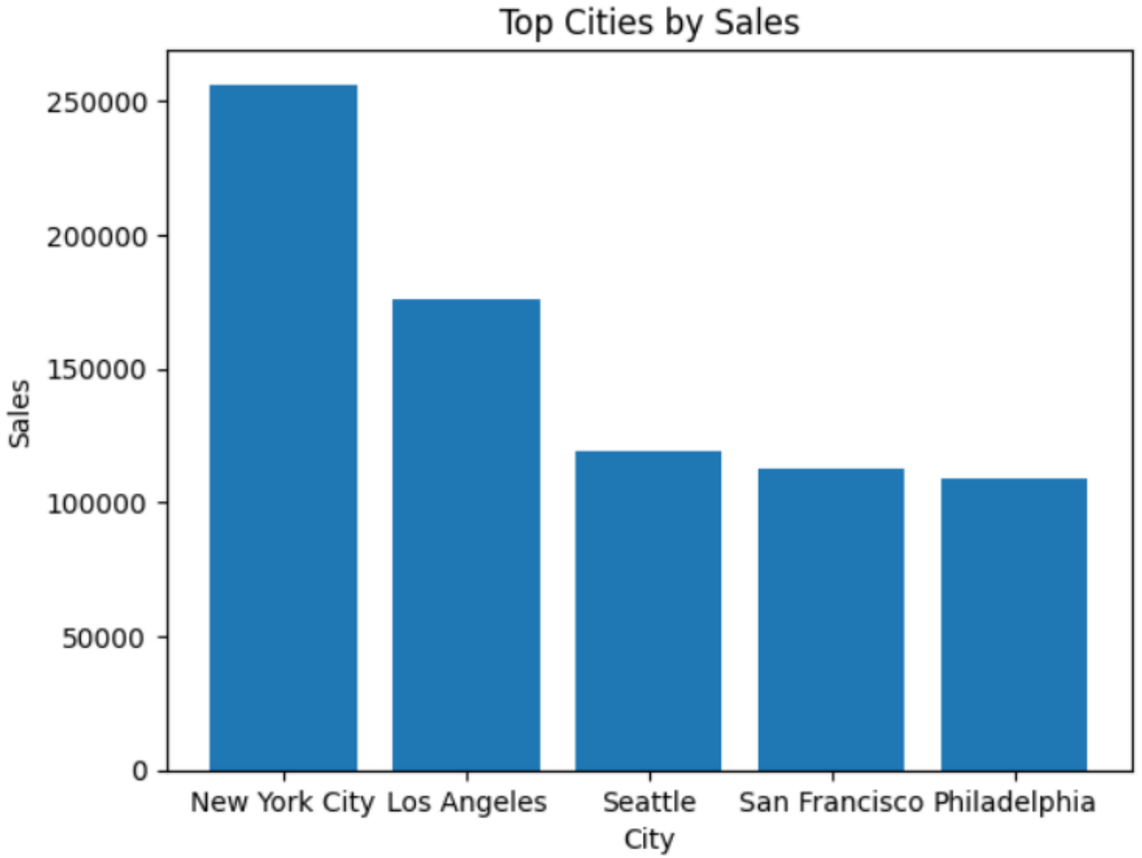
RESULTS – 1


Top 5 most profitting sales
Profit
50000
40000
30000
20000
10000
0
Profit
Copiers
Phones
Accessories
Paper
Binders
Product Type

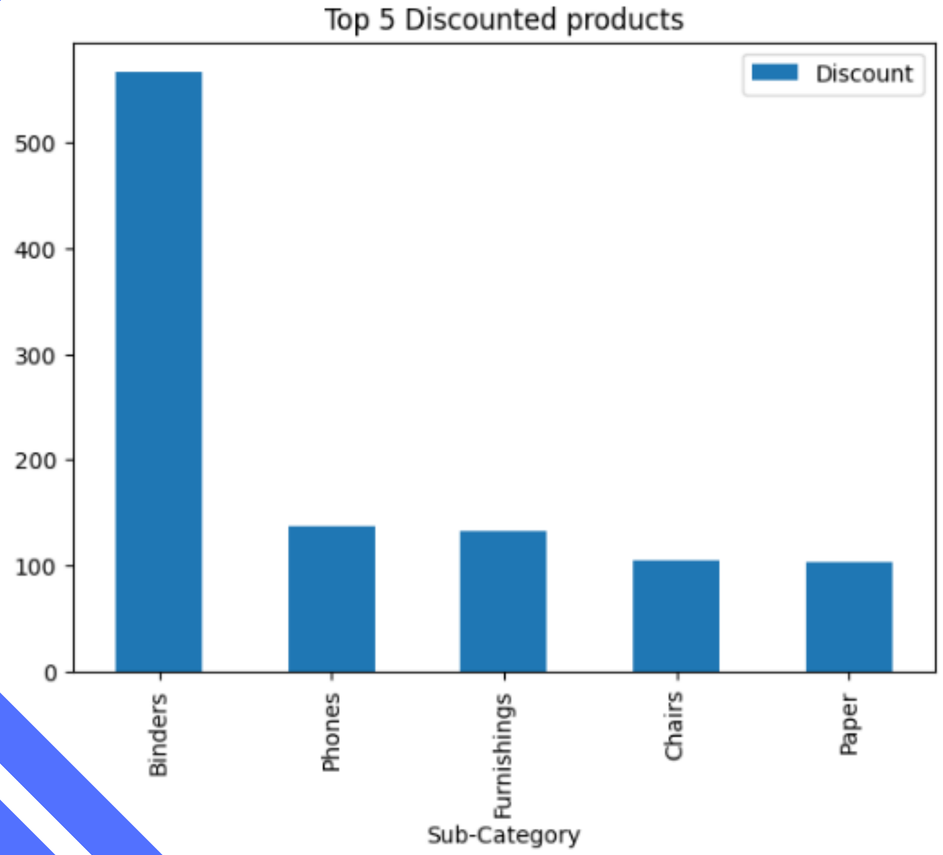

Top 5 Selling Product-type
Sales
300000
250000
200000
150000
100000
50000
0
Total Sales
Phones
Chairs
Storage
Tables
Binders
Product Type

RESULTS – 2

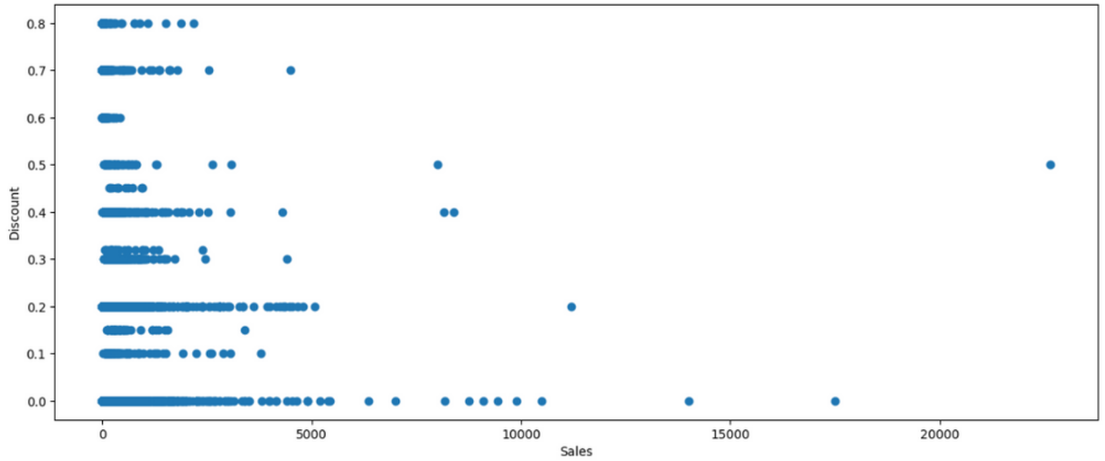

Sales
Profit
400
350
300
250
200
150
100
50
0
Central
East
South
West
Region


Top Cities by Sales
250000
200000
150000
100000
50000
0
Sales
New York City
Los Angeles
Seattle
San Francisco
Philadelphia
City

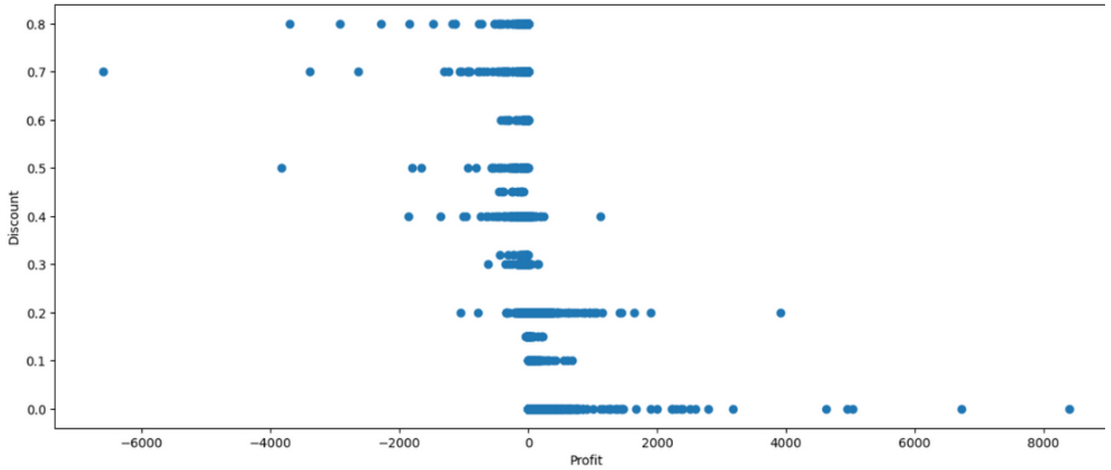
RESULTS – 3




```
# Create a scatter plot between sales and discounts
plt.figure(figsize=(15,6))
plt.scatter(df['Sales'],df['Discount'])
plt.ylabel("Discount")
plt.xlabel("Sales")
plt.show()
```




```
# Group the data by the discount and profit
plt.figure(figsize=(15,6))
plt.scatter(df['Profit'],df['Discount'])
plt.ylabel("Discount")
plt.xlabel("Profit")
plt.show()
```



CONCLUSION

The primary contributing factor to the decline in sales is attributed to discounts. While certain areas experience losses due to excessive discounts, others witness reduced sales because of fewer promotional offers. As a result, it necessitates improvement in this area.

Enhancing sales during festival seasons can be achieved by offering higher discounts, leading to increased customer engagement and greater revenue.

In certain cities, sales are relatively low, possibly due to a lack of awareness. To address this, targeted advertising in those regions could potentially boost sales and raise brand visibility.

LINKS & REFERENCES

Github:
https://github.com/nundika/Analysis_of_SuperStore_Dataset-Data-Analytics-Project

Article Help:
https://medium.com/analytics-vidhya/exploratory-data-analysis-super-store-cb91c37bcb06

THANK YOU!

IBM

edunet foundation

Warm Regards,

Lalnundika Darlong