IBM SKILLSBUILD INTERNSHIP PROJECT

TOPIC: ANALYSIS OF SUPERSTORE DATASET

STUDENT DETAILS:

NAME: Roshan Parida

SKILLSBUILDS EMAIL: roshanparida277@gmail.com

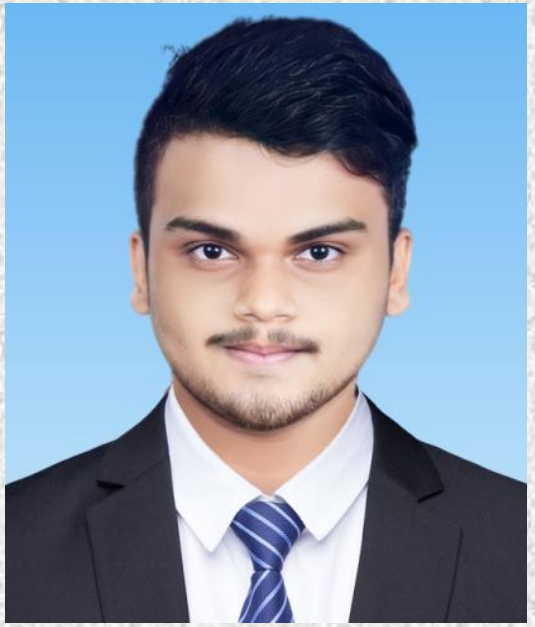
COLLEGE NAME: C.V. Raman Global University

STATE: Odisha

INTERNSHIP DOMAIN: Data Analytics (DA)

INTERNSHIP START & END DATE: 12/06/2023 – 24/07/2023

ORGANIZATION: DGT



ANALYSIS OF SUPERSTORE DATASET

A **Superstore dataset**, a kind of a simulation where you perform extensive data analysis to deliver insights on how the company can increase its profits while minimizing the losses.

This project focuses on analyzing the Sample Superstore dataset, which contains sales data from a fictional retail store. The dataset was obtained from **Kaggle** - an online community of data scientists and machine learning practitioners plus a repository of datasets that the users can upload. This data set was in the form of a csv file and could be mainly used for data pre-processing, exploratory data analysis, visualization.

The goal of the analysis is to gain insights into the store's performance and identify areas for improvement.

AGENDA

DATA ANALYSIS:

Loading and Importing the Dataset

Discovering and statistical description the dataset

Finding unique value and cleaning the duplicate data

1. Exploratory Data Analysis & Visualization

- Top Sales and Profits of company
- Top 10 States by Sales and Profit
- Sales and Profit Analysis based on State, Region, Segment, Category
- Quantity Analysis
- Discount Analysis

2. Data Visualization & Insights based on Region, Category, Segment, States

3. Ship Modes Analysis & Visualization

PROJECT OVERVIEW

This project aims to provide valuable insights and recommendations to help optimize business operations and drive growth for the Superstore. The analysis is based on thorough data exploration, statistical analysis, and visualization techniques, enabling stakeholders to make informed decisions to improve overall performance.

- **Objectives:-**

1. Performing Data Inspection and cleaning.
2. To Perform Exploratory Data Analysis.

   a) Sales & Profit Analysis b) Discount & Quantity Analysis c) Ship Mode Analysis

3. Maintains the details of Sales by Region/State/Category/Segment
3. Data Visualization to get key insights and problems.
4. Identify key areas for improving profits in different fields

WHO ARE THE END USERS OF THIS PROJECT?

Mainly end users in this sample superstore dataset is the ‘Consumer’. But it is divided into 3 segments such as ‘Consumer’, ‘Corporate’ and ‘Home Office’.

So, to improve the business we should focus on ‘Consumer’ and ‘Corporate’ Segment that make up more than 70% of customer base.

So, the target is specially upon customers from the East and West region in the Top 10 cities with **Highest Sales** by introducing special promotions and bundles for mass Consumer and Home Offices to increase the quantity sales and profit.

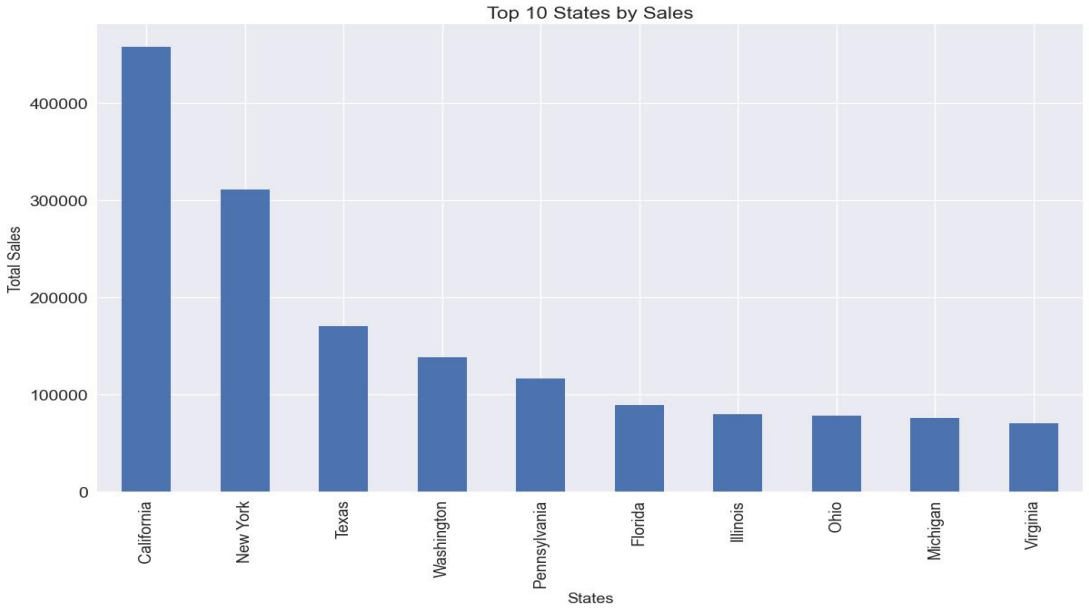
YOUR SOLUTION AND ITS VALUE PROPOSITION

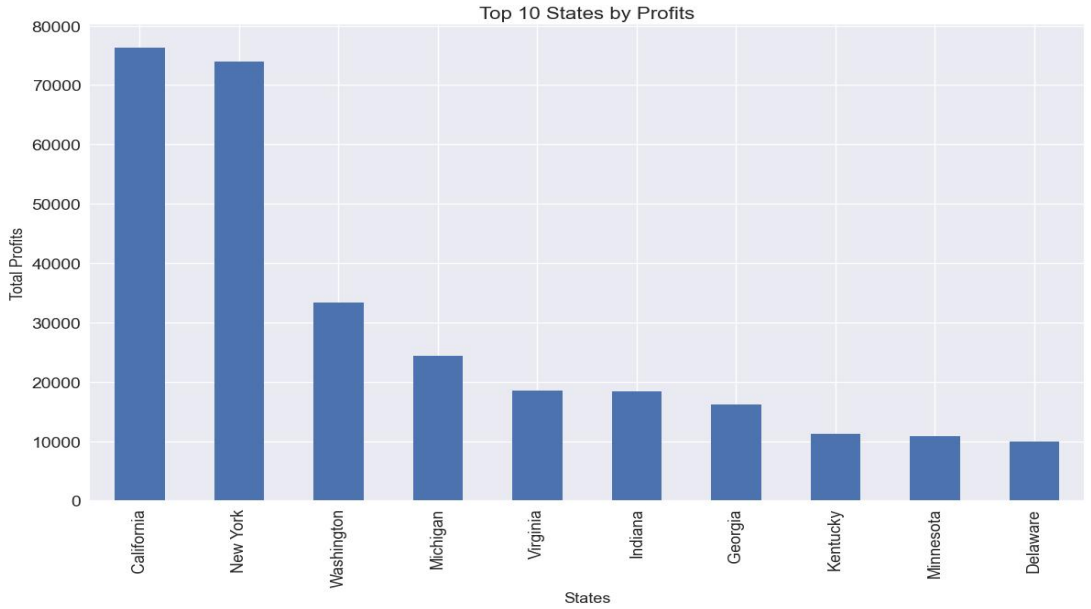
- Checked the description statistics to identify any inconsistencies or outliers in the data.
- Removed any duplicate records to achieve a high level of accuracy and avoid error calculation.
- Correlation Matrix shows relationship between Profit & Sales, Quantity & Sales to give analytics and ideas.
- Plotting **Visualization Cluster graphs** of the dataset to capture additional information and insights.
- **Sales** = 2,296,195.59 and **Profit** = 286,41.42 of the company. Most sales and profit were made in **'West' Region** followed by **'East'**.
- Our customer base is made up of **'Consumer'** & **'Corporate'** segments, which make up over 70% of the customers.
- Company offering more Discount on **'Office Suppliers'** & **'Furniture'** Category. To reduce loss in other category, we could provide discount in festival season to increase the sales and reduce the loss.
- The **'Ship Mode'** doesn't affect much to the sales but most profit was made in **'First'** and **'Second'** Class

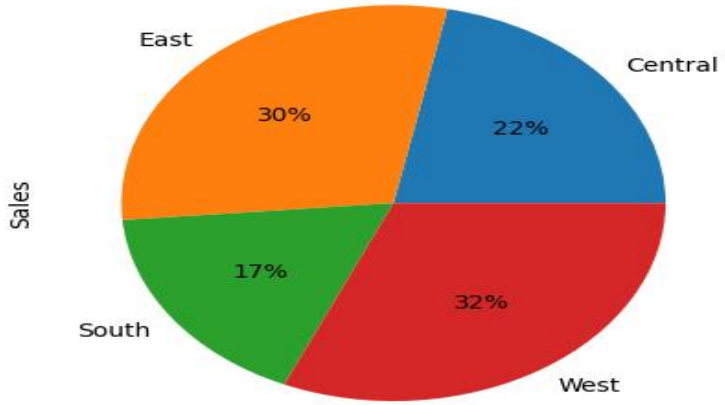
HOW DID YOU CUSTOMIZE THE PROJECT AND MAKE IT YOUR OWN

- As we have seen that every project has these 3 basic steps: 1)Loading and Importing the Dataset. 2) Descriptive statistics of the dataset. 3)Finding unique value and cleaning the duplicate data.
  I have also followed the same and performed the EDA by using a creative idea that is by finding the **Correlation Matrix** to perform the required analysis such as sales, quantity, discount and profit.
- Other project usually has sales & profit analysis. But in this we have also included discount & quantity analysis to get overall analysis and find the business problem. Which helps us to key insights and shows the areas of improving the quantity sales and profit based on Region, State, Category & Segments.
- We also included data visualization on different fields using the cluster graphs to get proper insights, ideas and to find the business problems/ profits. There is additional ‘Ship Mode’ analysis done which is one of the key transport mode for sales & profit of the company.

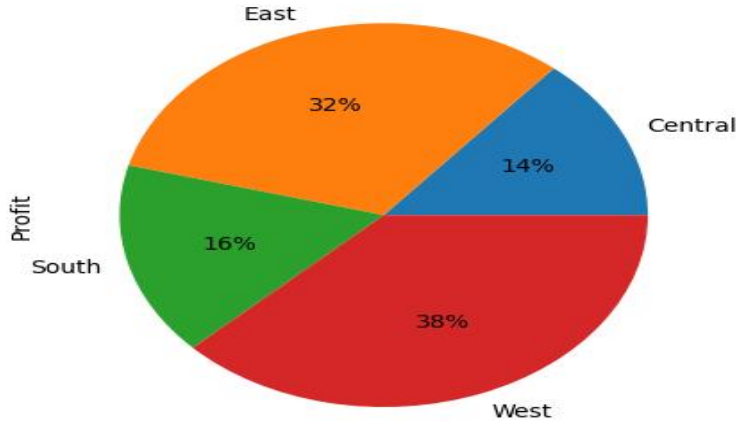
MODELLING





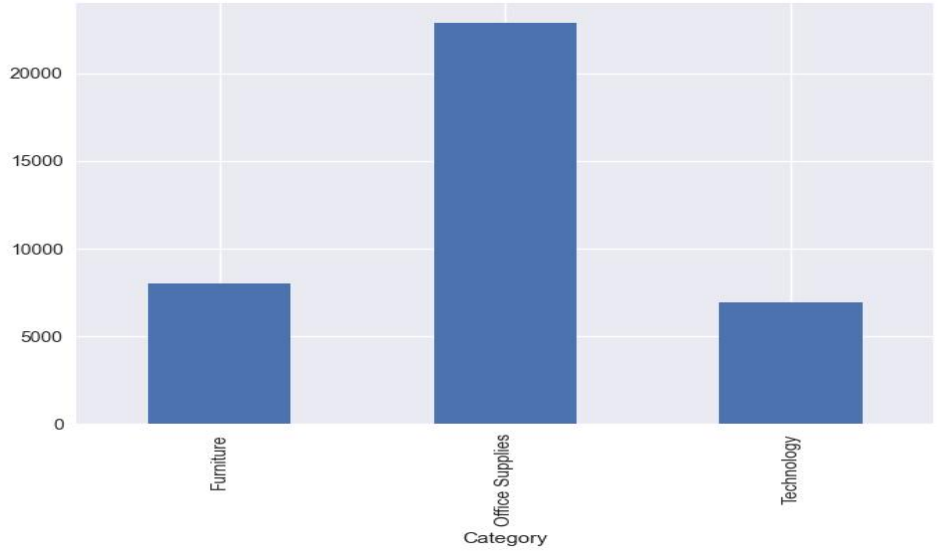


Sales Visualization By Region

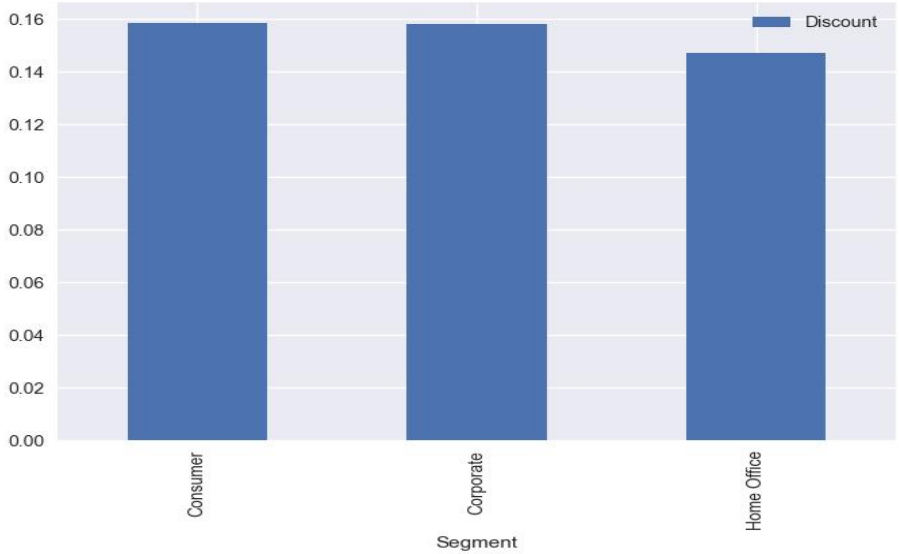


Profit Visualization By Region

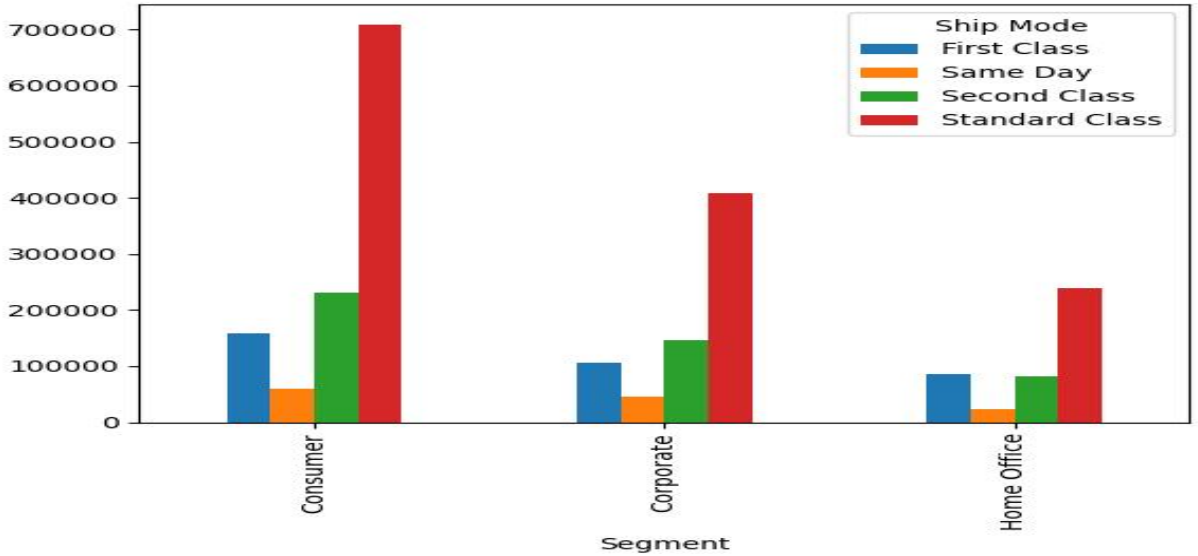
Quantity Analysis based on Category



Discount Analysis based on Segment



Ship Mode Analysis



RESULTS

- **Sales = 2,296,195.59** and **Profit = 286,41.42** of the company.
- Out of **Top 10 States**. Most sales & profit were done in **'California'** followed by **'New York'**. So, the company should focus on above 2 states to make most profit and most quantity sales are done there only.
- Most sales and profit were made in **'West' Region** followed by **'East'.**
- **'Consumer'** segment made the most sales and profit as compared to 'Corporate' and 'Home Office' segments.
- In **'Technology'** category business we get more Profit as compared to other two business. These is because of less Discount. So, we should shift focus on 'Furniture' category due to less profit and more discount
- We must concentrate on the Sales & Profit of **'Arizona', 'Colorado', 'North Carolina', 'Ohio', 'Texas'** states.
- **Visualization cluster graphs** are plotted for analytical insights on category, region, segment, states.
- In **'Ship Mode' Analysis**, most value was made on 'Standard Class' but the **Most Sales** were made on **'Same Day'** and **'Second Class'. Most Profit** was made on **'First Class'** followed by **'Second Class'**
- Similarly, **Most Quantity** was transported in 'Second Class' & 'Standard Class' and **Most Discount** were given in 'First Class' followed by 'Standard Class'
- Hence, To get good profit in any business you must focus on increasing sales but not giving more discount.

LINKS

- **GitHub Link:**
- https://github.com/roshp21/Analysis_of_SuperStore_Dataset-IBM_Internship_DataAnalysis

- **Reference:**
- https://medium.com/analytics-vidhya/exploratory-data-analysis-super-store-cb91c37bcb06
- https://github.com/alairdata/Superstore_Analysis
- https://github.com/Sachinnavgale/-The-Spark-Foundation-Tasks