Supermarket Sales Data Analysis

Context

Supermarkets have gained immense popularity in recent years due to their unmatched convenience, offering a diverse range of products all under one roof, from groceries to electronics and clothing. Shoppers are drawn to the abundant variety and choices available, enabling them to select from numerous brands and sizes. Furthermore, supermarkets leverage their purchasing power to provide competitive prices and frequent discounts, making them an attractive option for budget-conscious consumers. With extended operating hours, including late evenings and weekends, they cater to busy schedules. Additionally, their commitment to offering fresh produce, seamless technology integration, exceptional customer experiences, and community engagement initiatives have solidified their appeal. Supermarkets continuously innovate to meet evolving consumer preferences, including the demand for organic and eco-friendly products.

About this project

The growth of supermarkets in most populated cities are increasing and market competitions are also high. The dataset is one of the historical sales of supermarket company which has recorded for 4 years data. Target is to present Consumer Behaviour and Retail Trends through Supermarket Sales Data.

About this dataset

This dataset provides a comprehensive overview of supermarket sales, offering insights into consumer purchasing behaviours, product trends, and retail dynamics. Analysing this data enables retailers, marketers, and analysts to optimize strategies, improve inventory management, and enhance customer experiences. Explore sales volumes, seasonal patterns, and promotional impacts for informed decision-making in the competitive retail sector.

Here's a brief description of each of the attributes or labels in the dataset:
Rows: Around 10k.
Timeframe: Data covers the years 2014 to 2017.

- **Row ID:** Every record in this dataset starts with a Row ID which helps identifying rows uniquely
- **Order ID:** Unique identifier of the order placed

- **Order Date:** The date when the order was placed
- **Ship Date:** The date on which goods are sent out to a customer
- **Ship Mode:** This column has 4 categories - First Class, Second Class, Same Day, Standard Class
- **Customer ID:** Unique identifier of the customer
- **Customer Name:** Name of the customer
- **Segment:** The Segment column has 3 categories - Consumer, Home Office, Corporate
- **Country:** A column in a table that contains the country name of the customer
- **City:** Location of supermarkets / the city of customer
- **State:** State name of the customer
- **Postal Code:** The postal code contains numbers that helps to identify a location
- **Region:** This column has 4 categories - South, East, West, Central
- **Product ID:** Unique identifier of a product
- **Category:** Name of the product
- **Sub-Category:** Category of the product
- **Product Name:** Sub-category of the product
- **Sales:** The total sales amount earned from the transaction
- **Quantity:** Number of products purchased by customer
- **Discount:** Discount applied
- **Profit:** Profit tells whether a particular combination of order and product is making or losing money and how much

Data Cleaning & Analysis

Tools Used: MySQL Workbench
The dataset underwent thorough cleaning and analysis to prepare it for further exploration.

Tasks performed

- Imported the csv data into a MySQL database for analysis
- Performed the basic data cleaning activities
- Performed general data exploration and validation checks on the supermarket sales dataset
- Generated table summary statistics counted total number of rows and total number of columns

- Analysed distinct values and distributions for Ship Mode, Segment, Region, Product Category / Sub-Category, Country/City/State etc
- Summarized total profit/loss by transactions across the regions
- Ranked models by sales, discounts and profits
- Used SQL techniques like WINDOW functions, JOINs for advanced analysis
- Presented findings in a clear format for consumption by leadership and stakeholders

Key Findings

- The dataset consists of order dates ranging from 2014 upto 2017.
- The customers are mostly choosing Standard class (37%) as their preferred mode of shipping.
- The product category "Office Supplies" is mostly ordered for all the segments Consumer, Corporate and Home Office
- Customers from West region have placed maximum number of orders and stands on top in terms of maximum sales and profits.
- Maximum number of states (14) from the East region have been placed the orders in this super market.
- There are 305 products which made loss in this store.
- There are 39 states which made profit in this store. California is the topper and Wyoming caused the least profit.
- Texas, Florida and Illinois are the top three states in terms of number of cities the orders came from for negative total profit.
- Profit percentage of District of Columbia is highest i.e. 36.98%.
- There are 5 cities wherein the profit is at least 30%.
- California, New York and Texas are the cities where the most of the customers placed orders in this super store.

- Shipping modes the customers choose are Same Day, First Class, Second Class and Standard Class. The average days of product shipment from the date of order are 0.04, 2.18, 3.24, 5.01 respectively

- The segment "Consumer" has maximum number of customers and made maximum profit.

- Staples, Staple envelope and Easy-staple paper are the top 3 most selling product throughout the year.

- There are 95 customers who have placed orders in all the shipping modes.

- 123 cities are reported from which the orders are placed in all the years from 2014 to 2016.

- The maximum profitable orders were received in the month of December.

- Furnishings, Appliances, Paper and Copiers are the product sub-categories where the profit has been gradually increased from 2014 to 2017.

- There are 3 customers who have only ordered in "First Class" shipping mode.

- Lansing is the city from which the customers order from all shipping mode except "Standard Class".

- The profit has turned down into loss above 0.3% of discount offered, also the sales are downwards.

Purpose

This dataset can be used for predictive data analytics purpose.

Conclusion

The main point of doing this project are -

- Creative thinking
- Understanding the Dataset
- Use the best tools for working
- Financial performance and operation
- Analysing customer behaviour
- Deriving data-driven insights from supermarket sales data to support key business decisions and strategy.