

Supermart Grocery Sales Analysis and Prediction

Exploratory Data Analysis

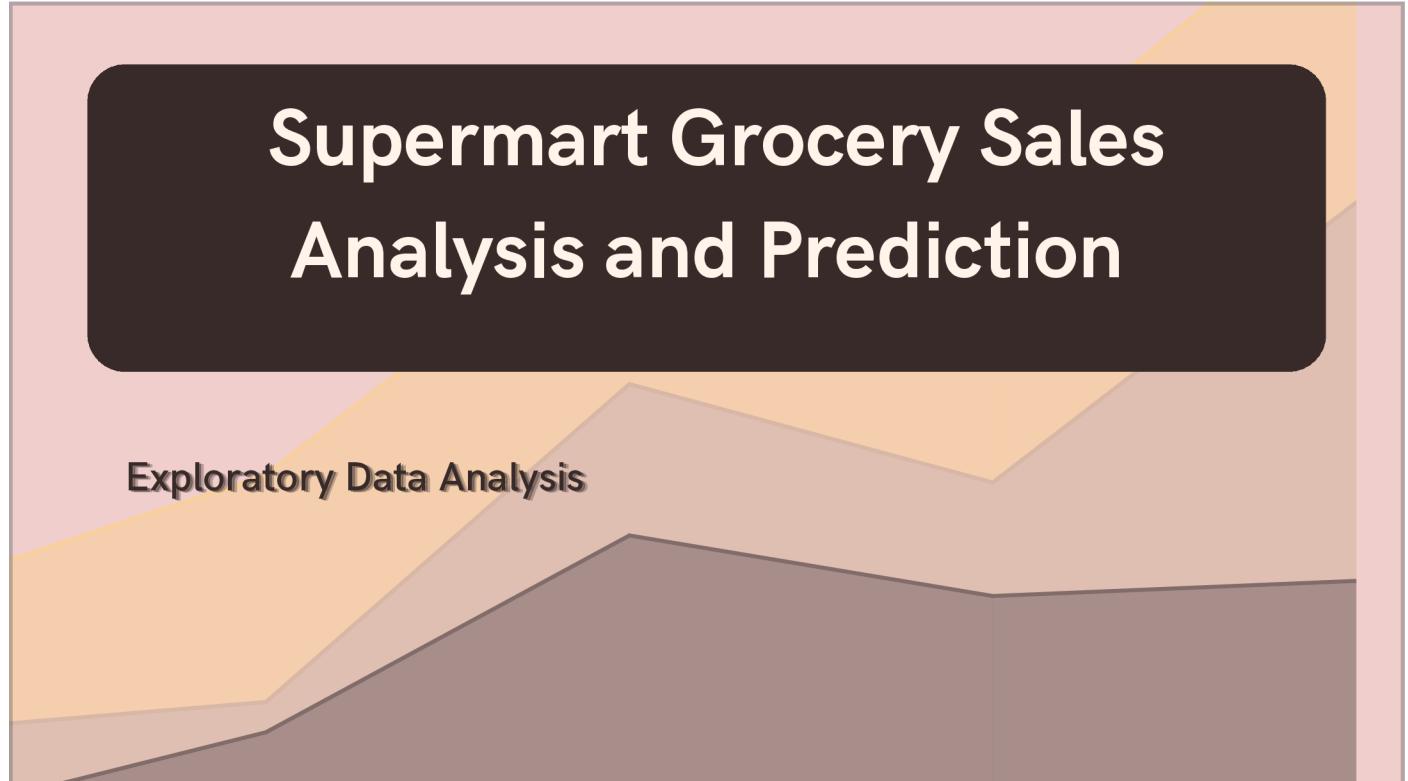
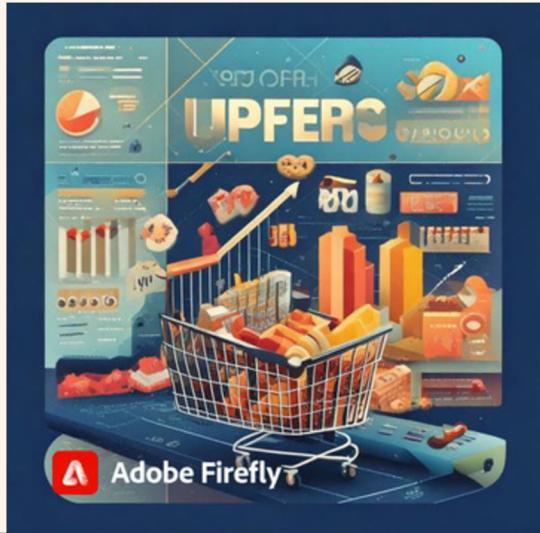


Table of Contents

- INTRODUCTION
- DATA PREPROCESSING AND CLEANING
- EXPLORATORY DATA ANALYSIS (EDA)
- FEATURE ENGINEERING
- MACHINE LEARNING MODELING
- VISUALIZATIONS
- CONCLUSION AND KEY TAKEAWAYS

Introduction

The purpose of this project is to analyze the Supermart Grocery Sales - Retail Analytics Dataset and to predict profits using machine learning techniques. The grocery sales industry has become increasingly competitive, making data-driven decision-making essential for maximizing profitability. This analysis focuses on understanding sales trends, identifying key performance metrics, and applying machine learning models to predict future profits based on historical data.



Data Preprocessing and Cleaning

The dataset was initially loaded using Pandas, and an exploratory review was conducted using the `head()`, `shape`, `info()`, and `isnull().sum()` methods. Key steps in data preprocessing included:

- Converting the 'Order Date' column to a datetime format to facilitate time-series analysis.
- Dropping unnecessary columns, including 'Order ID', 'Customer Name', and 'State', which were not essential for the analysis.
- Identifying and removing outliers in the 'Profit', 'Discount', and 'Sales' columns to ensure the quality of the data for subsequent analyses.

Exploratory Data Analysis (EDA)

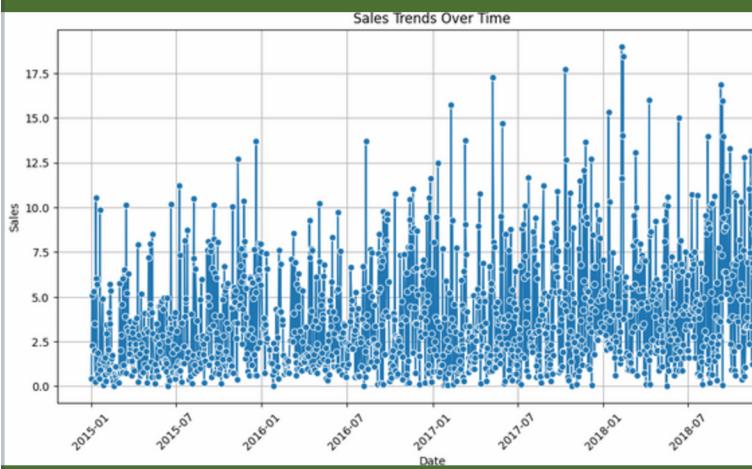
Exploratory Data Analysis was performed to gain insights into the dataset. Key visualizations included:

- A bar chart showing total sales by category, which revealed the categories contributing most significantly to overall sales.
- A line plot illustrating sales trends over time, highlighting seasonal patterns and identifying peak sales periods.
- A correlation heatmap that displayed the relationships between numerical variables, aiding in feature selection for machine learning modeling.



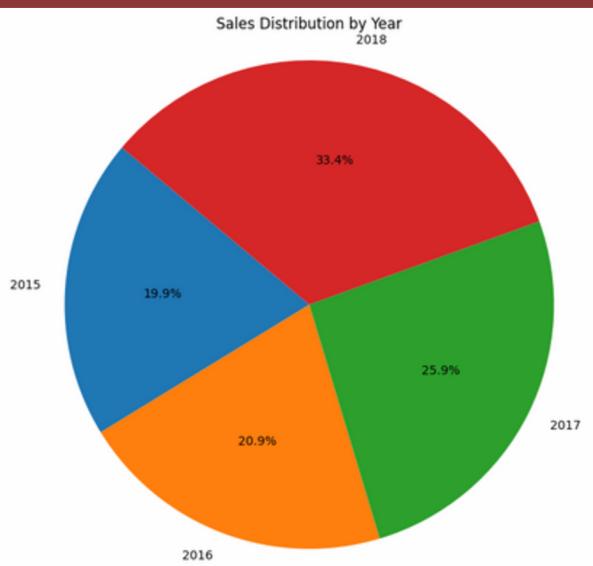
**Let's explore different ways
we can represent data!**

LINE GRAPH



The "Sales Trends Over Time" visualization is best represented by a Line Graph or Time Series Plot, which displays data points connected by a line to show the progression of sales over time. This type of graph is commonly used to track changes across continuous intervals, making it ideal for identifying trends and fluctuations in sales.

PIE CHART

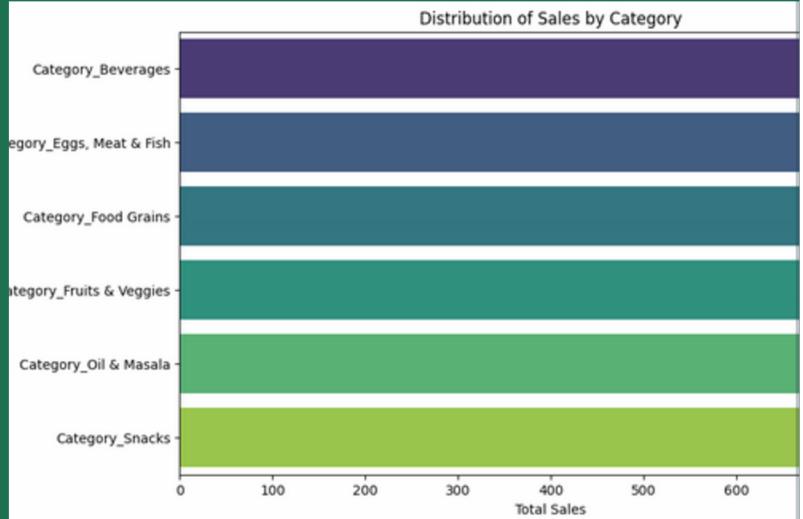


The pie chart illustrates the proportion of total sales contributed by each year. By visualizing the sales distribution this way, we can easily identify which years were most profitable for the grocery store and which ones saw relatively lower sales. This helps to pinpoint trends over time and assess the impact of any significant events or changes in business strategy on annual sales.

BAR CHART

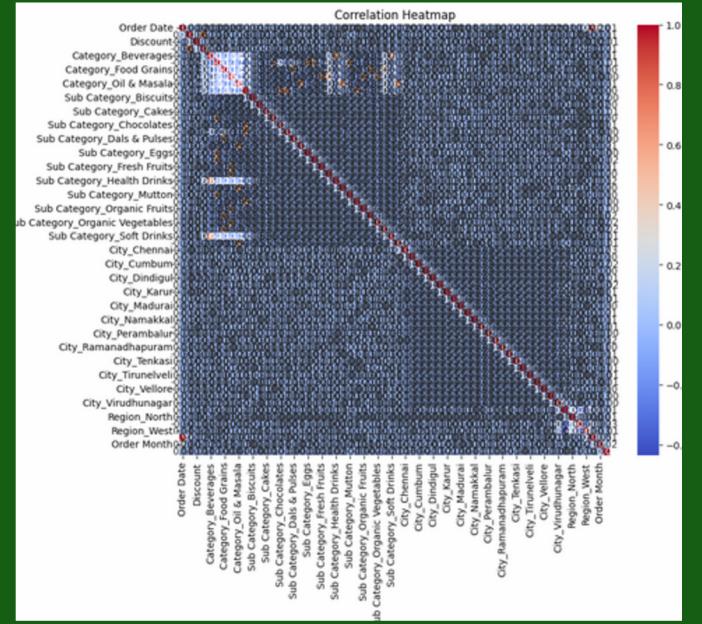
This bar chart visualizes the total sales contribution of each product category within the dataset. By comparing categories, we gain insights into which product types drive the highest sales and which may require strategic improvement or marketing.

Category Comparison
Identifying Key Contributors
Market Trends
Operational Decisions



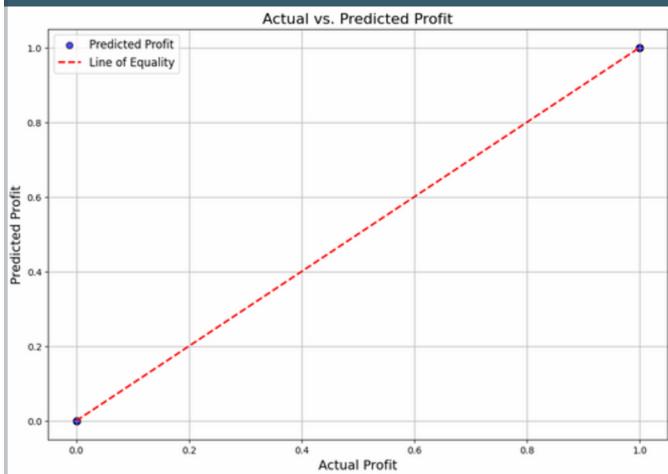
Correlation Heatmap

The correlation heatmap is a visual representation of the relationships between different variables in the dataset. By displaying correlations as a color-coded grid, we can easily identify which variables are positively or negatively associated and the strength of these associations.



LINE GRAPH

This scatter plot compares the actual profit values with the model's predicted profits. By visualizing this comparison, we can evaluate the model's performance and identify how closely the predictions align with the true values.



Conclusion

The analysis of the Supermart Grocery Sales Dataset reveals valuable insights into sales distribution, category performance, and sales trends over time. Visualizations like the correlation heatmap and actual vs. predicted profits provide an understanding of key relationships among variables, while trend analysis highlights seasonal sales patterns. These findings support data-driven decision-making and offer a foundation for further exploration to refine sales strategies and improve profitability.



Thank You.