Retail Data Analysis Documentation

Click Here for the Dashboard

1. Problem Definition and Objectives Problem Definition

The goal of this project is to analyze retail sales data to understand customer behavior, product sales, and store performance. Businesses need this information to improve their sales, manage inventory, and make better decisions.

Objectives

- Find out who the best customers are and how they shop.
- Identify which products sell the most and which ones don't.
- Analyze store performance and find which stores are doing well or struggling.
- Understand how holidays and different times of the year affect sales.
- Predict future sales trends to help with planning and decision-making.
- Provide suggestions on how to improve sales and customer satisfaction.

2. Description of the Datasets Used and Their Relevance

Retail_DataSet.csv

This file contains all the sales data from the business. It includes customer details, product information, store locations, sales amounts, purchase dates, and more. This data is important for analyzing sales trends, customer behavior, and store performance.

3. Summary of Key Findings and Recommendations

Key Findings

- Some customers spend a lot and buy frequently. These customers can be targeted for loyalty programs and personalized offers.
- Some products sell very well, while others do not. Low-performing products might need discounts or better marketing.
- Some stores have high sales, while others perform poorly. Poorly performing stores may need better marketing or operational improvements.
- Sales tend to drop slightly on holidays, possibly due to store closures or fewer shoppers. Planning holiday discounts or promotions could help improve sales.

- Some months have high sales, meaning customers shop more during those times. Businesses can stock up and prepare for these busy periods.
- A machine learning model was used to predict future sales. It has around 85% accuracy, helping businesses plan inventory and staffing.

Assumptions

- Data is assumed to be accurate and complete, though some missing values were handled with mean or mode imputation.
- External factors like holidays have an impact on sales, though weather and economic conditions were not included.
- Customer segmentation is based on past purchase behavior without real-time updates.

Challenges

- Weather API Integration Problem: Historical weather data is not publicly available and requires a paid service, making it difficult to incorporate weather-based sales analysis.
- Data Quality Issues: Missing values and inconsistent entries required significant preprocessing.
- Limited External Factors: Only holidays were considered; economic factors and customer sentiment were not included.
- Prediction Accuracy: The forecasting model has an accuracy of around 85%, but real-world variables may cause deviations.
- Store-Level Variability: Some stores may have external factors affecting sales that were not captured in the dataset.

Recommendations

- Offer discounts, coupons, or membership benefits to high-spending customers.
- Stock more of the popular products and reduce stock of low-selling items.
- Focus on improving sales in underperforming stores by using better marketing.
- Try promotions, discounts, or repositioning low-selling products to increase sales.
- Increase stock and staff in peak sales months to meet customer demand.
- Use the forecasting model to prepare for future demand and optimize operations.
- Consider including economic indicators, weather conditions, and customer feedback to enhance analysis.