DataVista: Sales Data Analysis and Visualization
SUPERMARKET SALES

~Gagankumar Seth

PROJECT OVERVIEW: DATASET

OBJECTIVE

To analyze and visualize sales data, uncovering patterns and insights to support business decisions. The project involves exploring customer behavior, identifying key sales trends, and examining factors affecting revenue, which will aid in optimizing future sales strategies.

TOOLS & LIBRARIES

Python:
For Data Processing, Analysis, and Visualization.
Libraries:
Pandas, NumPy, Matplotlib, Seaborn for Analysis and Visualization
Scikit-learn for Data Splitting, Normalization and Regression Analysis.

KEY FEATURES

Data Preprocessing:
Handling Missing Values, Converting Data Types, and Scaling Numerical features for Analysis.

Exploratory Data Analysis (EDA):
Visualizations such as Bar plots, Pie charts, and Line graphs to understand metrics like popular Products, Payment methods, and Customer demographics. Trend analysis to observe sales patterns over time.

Statistical Analysis:
Correlation analysis to find relationships between variables.
Regression analysis to predict sales based on factors like unit price and quantity sold.

Data Splitting and Scaling:
Splitting data into training and testing sets to evaluate model performance.
Applying Min-Max Scaling, Z-score normalization, and Decimal Scaling to standardize the data.

EDA: DATA VISUALIZATION

1. SALES INSIGHTS
   - <u>Total sales by Product type:</u> Shows total sales across different product lines, highlighting which products contribute most to revenue.
   - <u>Sales distribution by City:</u> Visualizes sales distributed across cities, helping identify regions with high or low sales.
   - <u>Customer type sales Comparison:</u> Compares sales generated by different customer types (e.g., member vs. Non-member).
   - <u>Sales by Payment method:</u> Shows the preferred payment methods based on sales volume and proportion.
   - <u>Branch-wise Sales distribution:</u> Compares total sales across branches, useful for regional performance analysis.
   - <u>Months-wise Sales distribution:</u> Compares total sales in each month, useful for monthly performance analysis.
   - <u>Total sales by Hour of day:</u> Examines sales volume by the hour, identifying peak sales times.

2. REVENUE & INCOME INSIGHTS
   - <u>Gross income Distribution:</u> Shows the distribution of gross income, highlighting variability and profitability.
   - <u>Gross income by Customer type and Gender:</u> Compares gross income by customer type and gender, offering insights into demographic profitability.

3. CUSTOMER SATISFACTION & RATING INSIGHTS
   - <u>Customer rating Distribution:</u> Examines the distribution of customer ratings, highlighting satisfaction levels.
   - <u>Average rating of Products:</u> Displays average rating of each product type.

4. QUANTITY & PRODUCT LINE INSIGHTS
   - <u>Quantity sold by Product type:</u> Examines the distribution of quantities sold across product lines, useful for identifying popular product categories.

5. FEATURE RELATIONSHIPS AND INTERACTION INSIGHTS
   - <u>Heatmap of Correlation matrix:</u> Displays correlations among numerical features, useful for identifying dependencies.

OUTCOME

A comprehensive overview of the Sales data through various Visualizations and Analyses. The project highlights Actionable insights, such as Top-selling products, Effective payment methods, and Seasonal sales trends, supporting data-driven business decisions.

PROJECT FILE