# RetailPulse Analysis - Documentation

## Overview

This project analyzes retail sales data to identify trends and insights related to customer behavior, sales patterns, and profitability. The code processes, transforms, and visualizes sales data using Python, leveraging libraries such as Pandas, NumPy, Seaborn, and Matplotlib.

## Setup and Data Import

Libraries Used:

**- pandas:** For data manipulation and analysis.  
 **- numpy:** For numerical operations.  
 **- seaborn and matplotlib.pyplot:** For data visualization.

Data Import: The dataset is read from a CSV file named Different\_stores\_data.csv.  
I got the dataset from Kaggle.  
<https://www.kaggle.com/datasets/kzmontage/sales-from-different-stores?select=Different_stores_dataset.csv>

## Data Preparation and Transformation

1. Public Holidays and Weekend Tagging:

- A list of public holidays for 2016 (US-based) is defined.  
- The code creates a DataFrame (dates\_df) with columns:  
 - invoice\_date: Dates in 2016.  
 - Is\_Holiday: Indicates if the date is a holiday (weekend or public holiday).  
 - Reason: Reason for being a holiday (e.g., Weekend or Public Holiday).

2. Merging Holiday Information:

- The dates\_df is merged with the sales dataset to include holiday-related information for each invoice.

3. Column Cleaning:

- Unnecessary columns (Unnamed: 13, Unnamed: 14) are dropped.

4. Derived Metrics:

- Profit per Unit: selling\_price\_per\_unit - cost\_price\_per\_unit.  
 - Invoice Value: quantity \* selling\_price\_per\_unit.  
 - Invoice Profit: profit\_per\_unit \* quantity.

5. State to Region Mapping:

- A mapping dictionary is used to assign regions to states, ensuring consistent regional data.

## Exploratory Data Analysis (EDA)

1. General Dataset Insights:

- Shape and column data types of the dataset.  
 - Checking for missing values and duplicates.

2. Visualizations:

**- Sales Over Time:** A line plot of invoice values by date.  
 **- Correlation Heatmap:** Displays correlations between numeric variables.  
 **- Monthly and Daily Sales Trends:** Monthly sales visualized using bar plots and daily sales averages plotted as a line graph.  
 **- Sales and Profits by State and Region:** Bar plots to visualize state and region-wise sales and profits.  
 **- Category Analysis:** Sales and profits visualized by category and gender.  
 **- Holiday Impact:** Average sales on holidays vs. non-holidays, and category-wise sales trends.  
 **- Shopping Mall Insights:** Profits by shopping malls and a comparison of specific malls.  
 **- Customer Analysis:** Age distribution, spending patterns by age groups, and regional distribution.  
 **- Payment Method Analysis:** Total sales by each payment method.

## Key Findings

1. Seasonality:

- Holidays and weekends have higher sales compared to regular days.  
 - Monthly trends reveal peak sales in certain months.

2. Customer Behavior:

- Customers aged 20-50 contribute the most to sales, with distinct spending patterns by region.  
 - Gender-specific purchasing trends observed across categories.

3. Profitability:

- Certain states, regions, and shopping malls outperform others in profit generation.  
 - Categories like Technology and Food & Beverage generate significant revenue.

## Potential Next Steps

- Further refine insights by integrating external data (e.g., weather, promotions).  
- Predict future sales using machine learning models.  
- Automate reporting with dashboards for real-time insights.