Walmart Sales Data Analysis - Project Report

Author: Vikas Kumar

Tools Used: MySQL, Python (Pandas, Matplotlib, Seaborn), Jupyter Notebook

Date: [Add Date]

# 1. Introduction

Retail giants like Walmart generate a vast amount of transactional data every day. Efficient analysis of this data provides valuable insights into business performance, customer preferences, and operational efficiency. This project focuses on extracting actionable insights from Walmart’s sales data using SQL and Python.

# 2. Objective

- To explore and analyze sales data from Walmart across different dimensions.

- To identify trends in payment methods, product categories, customer ratings, and branch-wise performance.

- To derive key performance metrics and visualize them for decision-making support.

# 3. Tools & Technologies Used

- SQL (MySQL): For data extraction, aggregation, and transformation.

- Python: Used for advanced data visualization and EDA.

- Libraries: Pandas, Matplotlib, Seaborn

- Jupyter Notebook: Interactive analysis environment.

- CSV File: orders.csv used as the main dataset.

# 4. Dataset Overview

The dataset contains transactional-level information from Walmart including:

- Branch

- City

- Payment Method

- Date & Time

- Product Category

- Quantity Sold

- Total Sales

- Customer Rating

# 5. Methodology

- Data was first loaded into MySQL under the table name 'walmart'.

- SQL queries were used to derive key metrics like:

\* Total transactions by payment method

\* Product category-wise sales

\* Branch-wise trends across time

\* Revenue and quantity metrics

- Python was used to:

\* Load and preprocess the orders.csv file

\* Generate plots and graphs to visualize trends

\* Complement SQL findings with charts and visual interpretation

# 6. Key SQL Queries & Analysis

- Payment Trend: Credit Card emerged as the most preferred payment method.

- Branch Performance: Each branch has distinct busiest days – valuable for staffing and inventory planning.

- Category Ratings: Ratings vary across city and product category, helping identify customer satisfaction levels.

- Revenue Drop Analysis: Year-over-year revenue drop identified for specific branches in 2023.

- Time-of-Day Impact: Most sales transactions happen in the Afternoon, followed by Evening.

- Profitability: Category-wise analysis helps in identifying top profit-generating segments.

# 7. Python-Based Exploratory Data Analysis

- Used Pandas to clean and prepare the data.

- Created visualizations:

\* Pie charts showing payment method shares

\* Bar plots for category-wise revenue

\* Heatmaps for sales across weekdays and branches

- Derived correlation between quantity, rating, and total amount.

# 8. Insights & Findings

- Electronic Accessories and Fashion Accessories are high-revenue segments.

- Branch A is the top performer in both quantity and total sales.

- Customer ratings remain mostly positive across categories and cities.

- Afternoon is the peak business time, suggesting marketing focus should align accordingly.

# 9. Conclusion

The analysis successfully identified key business insights from the Walmart dataset. By leveraging SQL for slicing and dicing the data and Python for visual representation, a comprehensive understanding of store performance, customer behavior, and sales trends was achieved.

# 10. Future Scope

- Incorporate forecasting models to predict future sales.

- Build interactive dashboards using Power BI or Tableau.

- Perform market basket analysis for product recommendation systems.

- Analyze seasonal trends by incorporating external calendar data.

# 11. Appendix

- SQL Queries File: sdfs.sql

- Python Notebook: SQL + Python.ipynb

- Raw Data: orders.csv