Case Study: Walmart Sales Data Analysis

# Project Title

Sales Insights & Business Intelligence from Walmart Transaction Data

# Domain

Retail Analytics

# Tools Used

SQL (MySQL), Python (Pandas, Matplotlib, Seaborn), Jupyter Notebook

# Author

Vikas Kumar

# 1. Background

Walmart, being one of the world’s largest retail corporations, processes massive volumes of sales transactions across its numerous branches daily. Efficient analysis of such data can provide strategic insights for optimizing inventory, improving customer satisfaction, and enhancing revenue streams.

# 2. Problem Statement

Despite having large datasets, Walmart faces the challenge of converting raw transactional data into actionable insights. Questions like:

- Which branches or categories are underperforming?

- What are the preferred customer payment methods?

- Are there any seasonal trends in sales?

- How do customer ratings vary across branches or product categories?

remained unanswered without a proper data analytics framework.

# 3. Objective

The primary goal of this project was to analyze Walmart’s transactional data using SQL for data querying and Python for visualization to uncover:

- Customer behavior trends

- Branch and category performance

- Time-based sales patterns

- Revenue drop areas and possible causes

- Correlation between quantity, total sales, and customer feedback

# 4. Approach

🧾 Data Collection & Cleaning:

- The dataset (orders.csv) contains over 1,000+ transaction records with details like branch, city, payment method, category, rating, total, quantity, and datetime.

🧮 Data Analysis (SQL):

- Loaded into MySQL (walmart table) for structured querying.

- SQL was used to answer specific business queries like revenue, quantity, payment preferences, and revenue drops.

📊 Visualization (Python):

- Data imported into Pandas DataFrame.

- Charts created using Matplotlib and Seaborn: bar, pie, heatmap, and line graphs.

# 5. Insights & Key Findings

- ✅ Credit Card was the most used payment method across all cities.

- ✅ Branch A showed the highest sales and customer volume, particularly on weekends.

- ✅ Electronic Accessories and Fashion Accessories yielded the highest revenue.

- ✅ A noticeable drop in revenue in 2023 vs. 2022 for some branches was identified and quantified.

- ✅ Afternoon hours emerged as the time of peak transactions.

- ✅ Ratings remained consistently high across most categories, with few outliers.

# 6. Business Impact

These insights help Walmart:

- 📦 Optimize Inventory: Align stock with most-purchased categories and peak hours.

- 🧑‍💼 Enhance Staffing: Staff planning for peak hours and busiest branch days.

- 💳 Refine Payment Strategy: Focus on popular payment gateways.

- 📉 Track Declining Revenue: Identify underperforming branches and take corrective action.

- 🎯 Improve Customer Engagement: Monitor feedback by ratings and focus on low-rated segments.

# 7. Challenges

- The dataset contained string-formatted dates and times, requiring preprocessing.

- Some calculated metrics (e.g., year-wise revenue comparison) required custom SQL logic using STR\_TO\_DATE and YEAR().

# 8. Conclusion

This project successfully translated raw retail sales data into a strategic business asset. Using SQL and Python, comprehensive insights were derived to guide Walmart in improving sales, marketing, and operational strategies.

# 9. Next Steps

- Integrate a BI Dashboard using Power BI/Tableau for real-time KPI tracking.

- Deploy forecasting models to predict future sales and seasonal performance.

- Expand analysis with demographic data for more detailed customer profiling.