

# Walmart Sales Data Analysis Project

This project analyzes Walmart sales data to uncover meaningful business insights, combining data preprocessing in Jupyter Notebook with business problem-solving using PostgreSQL. The dataset from Kaggle contains transaction-level details including payment methods, sales categories, customer ratings, and revenue.



# Project Workflow

## Data Collection

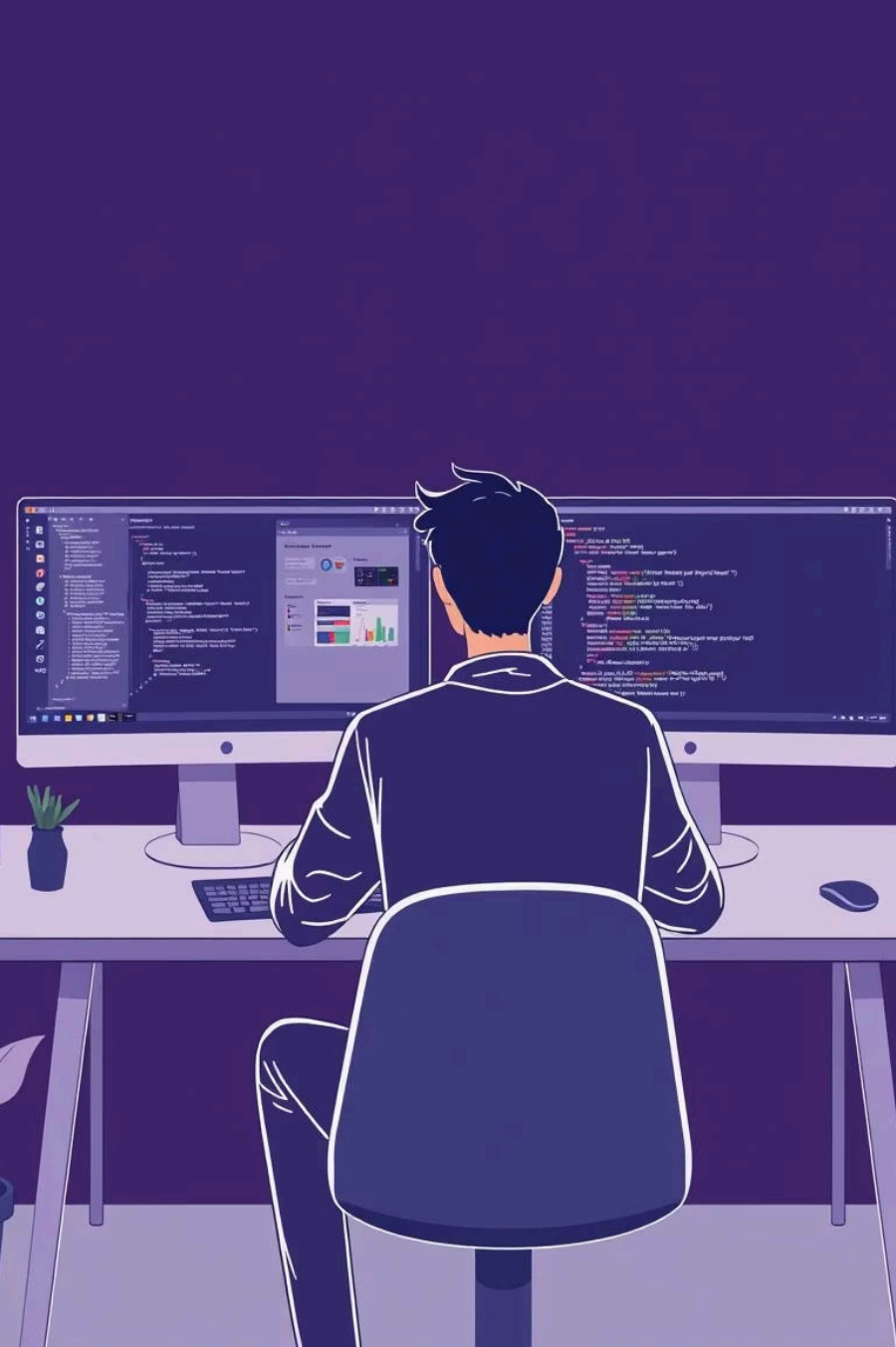
Imported dataset from Kaggle containing transaction-level details

## Data Preparation

Cleaned and processed data in Jupyter Notebook using Python

## SQL Analysis

Connected to PostgreSQL and performed business analysis with SQL queries



# Data Preparation Process

## Data Cleaning

- Removed rows with missing values
- Dropped duplicate records
- Standardized column names to lowercase

## Data Enhancement

- Added a new total column
- Connected cleaned dataset to PostgreSQL

All preparation was performed in Python using Jupyter Notebook before transferring to PostgreSQL for analysis.

# Business Questions Addressed

- Payment methods analysis (transactions and quantity)
- Highest-rated category per branch
- Busiest day identification for each branch
- Total quantity sold per payment method
- Rating statistics (avg/min/max) per city
- Total profit calculation per category
- Most common payment method by branch
- Sales distribution across daily shifts
- Revenue decline analysis (2023 vs 2022)

Each question was addressed using specific SQL queries designed to extract actionable insights.

# Sample SQL Query: Payment Analysis

```
SELECT payment_method,  
       COUNT(*) AS no_transactions,  
       SUM(quantity) AS total_quantity_sold  
FROM walmart  
GROUP BY payment_method;
```

This query identifies different payment methods used by customers, counts the number of transactions for each method, and calculates the total quantity of items sold through each payment type.

# Profit Calculation Methodology

$$\text{Total Profit} = \text{Unit Price} \times \text{Quantity} \times \text{Profit Margin}$$

This formula was implemented in SQL to calculate category-level profitability, allowing for comparison between product categories and identification of the most profitable segments.



# Time-Based Analysis

## Morning Shift

Sales transactions occurring  
before 12:00 PM

## Afternoon Shift

Sales transactions between 12:00  
PM and 5:00 PM

## Evening Shift

Sales transactions after 5:00 PM

Sales were categorized into these time periods to identify peak business hours and optimize staffing and inventory management accordingly.



# Year-over-Year Performance Analysis

The project identified the Top 5 branches with the highest revenue decrease ratio when comparing 2023 to 2022 performance.

⊗ This analysis highlighted branches requiring immediate attention due to significant revenue decline.

Understanding these trends allows management to investigate underlying causes and implement targeted improvement strategies.



# Key Insights Discovered



## Payment Preferences

Different branches show distinct preferences for payment methods



## Category Ratings

Some product categories consistently achieve higher customer ratings



## Business Hours

Peak business hours vary between morning, afternoon, and evening shifts



## Revenue Decline

Certain branches suffered major revenue decline in 2023 compared to 2022

# Project Conclusion

## Technical Achievement

This project demonstrates a complete data pipeline from cleaning raw data to solving real-world business problems using SQL.

It showcases the combined power of [Python for preprocessing](#) and [PostgreSQL for analytics](#).

## Business Value

The insights generated provide actionable intelligence for:

- Optimizing payment processing
- Improving product categories
- Adjusting staffing based on peak hours
- Addressing declining performance

