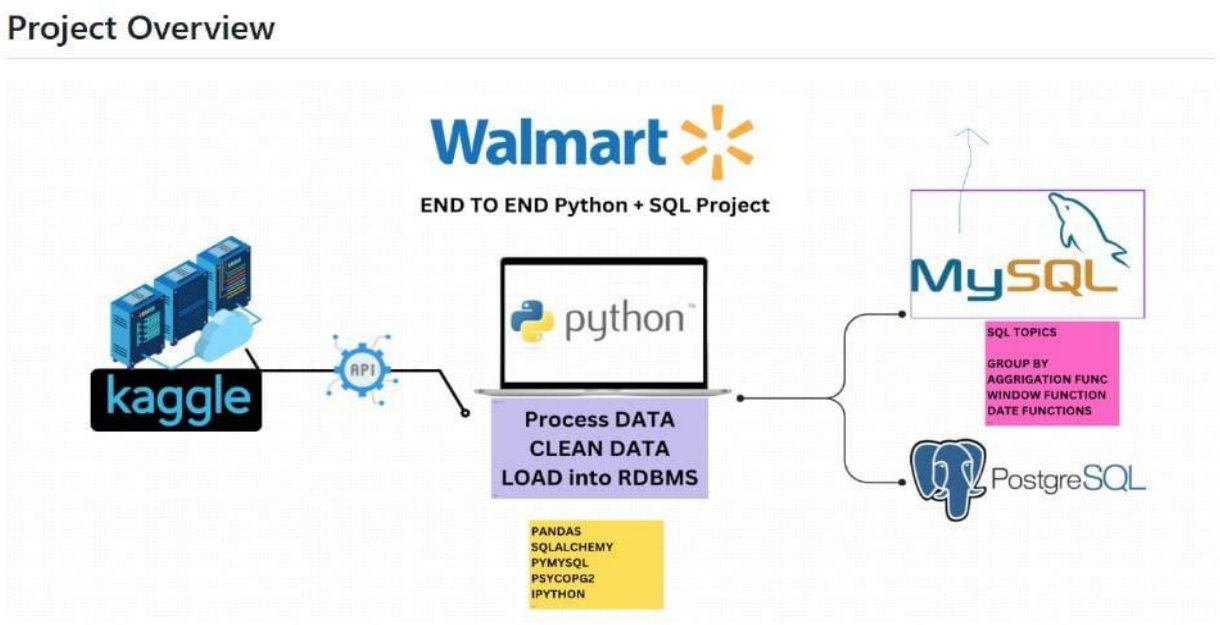
Detailed Analytical Report

Title: Walmart Sales Data Analysis – End-to-End SQL & Python Project

Project Overview



This project focuses on analysing Walmart sales data to derive actionable business insights. The goal is to identify revenue trends, customer behaviour, and sales patterns using SQL and Python.
The project is beginner-friendly yet covers an end-to-end data analysis workflow—from data extraction and cleaning to visualization**.**

Tools & Technologies

- Languages: SQL (MySQL, PostgreSQL), Python
- Libraries: Pandas, Requests, Matplotlib
- Databases: MySQL, PostgreSQL
- Data Sources: Walmart dataset from Kaggle and APIs
- Skills Practiced: Data cleaning, SQL querying, date formatting, window functions, and visual storytelling.

Key Objectives

- Import and clean Walmart sales data.
- Perform advanced SQL queries to analyse branch performance, sales trends, and customer behaviour.
- Work with dates to identify time-based trends (morning/noon/evening sales, weekdays vs weekends).
- Visualize sales insights using Python for clear storytelling.

Key Insights

1. Highest Profit Categories

    - Home and Lifestyle (E-wallet, Total Profit ≈ 79,757)
    - Fashion Accessories (E-wallet, Total Profit ≈ 79,141)
2. Most Common Payment Method (per branch)
    - E-wallet was most preferred in the majority of branches.
3. Sales by Time of Day
    - Evening Sales had the highest number of invoices in most branches.
4. Branches with Highest Revenue Decrease (2022 → 2023)
    - Top 3: WALM045, WALM047, WALM089
    - Decrease Ratio: 62.6% → 57.9%
5. Highest Rated Category per Branch
    - Example: WALM002 → Food & Beverages (8.25 Avg Rating)
6. Busiest Day (per branch)
    - Example: WALM001 → Saturday, WALM002 → Friday

Methodology

1. Data Extraction & Cleaning: Pulled data from Kaggle & APIs, cleaned using Python (Pandas).
2. SQL Analysis:
    - Wrote 20+ optimized queries using CTEs, window functions, and aggregation.
    - Derived branch-wise and category-wise revenue & profit.

Major Analysis Performed (SQL Queries & Results)

Below are the key SQL queries, results, and business insights.

Total Profit per Category & Payment Method

Query:

```
sql
CopyEdit
SELECT
    category,
    payment_method,
    SUM(unit_price * quantity * profit_margin) AS total_profit
FROM walmart
GROUP BY category, payment_method
ORDER BY total_profit DESC;
```

Insight: Focus on **Home & Lifestyle** and **Fashion Accessories** categories; **E-wallet** drives the highest profit.

Most Common Payment Method Per Branch

Query:

```
sql
CopyEdit
SELECT * FROM (
  SELECT
    branch,
    payment_method,
    COUNT(*) AS transactions,
```

```
    RANK() OVER(PARTITION BY branch ORDER BY COUNT(*) DESC) AS "rank"
  FROM walmart
  GROUP BY branch, payment_method
) AS t
WHERE t.rank = 1;
```

Insight: Most branches prefer **E-wallet**, making it the dominant payment method.

Sales by Time of Day

Query:

```
sql
CopyEdit
SELECT
  branch,
  CASE
    WHEN TIME(`time`) BETWEEN '06:00:00' AND '12:00:00' THEN 'Morning'
    WHEN TIME(`time`) BETWEEN '12:00:01' AND '17:00:00' THEN 'Afternoon'
    ELSE 'Evening'
  END AS time_of_day,
  COUNT(*) AS total_invoices
FROM walmart
GROUP BY branch, time_of_day
ORDER BY branch, total_invoices DESC;
```

Insight: Most transactions occur in the **Evening**, suggesting promotions could be targeted during these hours.

Branches with Highest Revenue Decrease (2022 → 2023)

Query:

```
WITH revenue2022 AS (
  SELECT
    branch,
    SUM(total) AS revenue_2022
  FROM walmart
  WHERE EXTRACT(YEAR FROM STR_TO_DATE(`date`, '%d/%m/%Y')) = 2022
  GROUP BY branch
),
revenue2023 AS (
  SELECT
    branch,
    SUM(total) AS revenue_2023
  FROM walmart
  WHERE EXTRACT(YEAR FROM STR_TO_DATE(`date`, '%d/%m/%Y')) = 2023
  GROUP BY branch
)
SELECT
  r22.branch,
  r22.revenue_2022,
  r23.revenue_2023,
  ( ( r22.revenue_2022 -  r23.revenue_2023)/ r22.revenue_2022 )*100 as decreaseRatio
FROM revenue2022 r22
JOIN revenue2023 r23 ON r22.branch = r23.branch
WHERE  r22.revenue_2022  > r23.revenue_2023
ORDER BY 4 DESC
LIMIT 5;
```

Insight: Top declining branches: **WALM045, WALM047, WALM089** with ~60% revenue drop.

Highest Rated Categories Per Branch

Query:

```
sql
CopyEdit
SELECT * FROM (
  SELECT
    branch,
    category,
    AVG(rating) AS avg_rating,
    RANK() OVER(PARTITION BY branch ORDER BY AVG(rating) DESC) AS "rank"
  FROM walmart
  GROUP BY branch, category
) AS t
WHERE t.rank = 1;
```

Insight: Top categories include **Food & Beverages** and **Electronic Accessories** with highest average ratings.

Business Recommendations

- Boost E-wallet campaigns – it's the most preferred payment mode.
- Target Evening sales with promotions/discounts.
- Investigate revenue decline in top 5 branches to recover sales.\
- Push Home & Lifestyle products through high-profit channels.

Conclusion

This project demonstrates how **data-driven decisions** can improve operational efficiency by identifying the best-selling categories, busiest time slots, and most profitable branches. The workflow replicates real-world data analyst responsibilities.