WALMART DATA ANALYSIS

Presented by Md Tarif

PROJECT OVERVIEW

- **Objective:** Extract actionable insights from Walmart sales data.
- **Technologies Used:** Python, SQL (PostgreSQL/MySQL), Kaggle API.
- **Focus Areas:** Data cleaning, analysis, and business problem-solving.
- **Outcome:** Optimized strategies for sales, customer behavior, and profitability.

TOOLS AND TECHNOLOGIES

Python Libraries

- Pandas
- SQLAlchemy
- Psycopg2

Databases

- PostgreSQL

Other Tools

- Kaggle API
- Jupyter Notebooks
- Visual Studio Code

PROJECT ARCHITECTURE

kaggle

Extract Data

API

Process DATA
CLEAN DATA
LOAD into RDBMS

PostgreSQL

SQL TOPICS
GROUP BY
AGGRIGATION FUNC
WINDOW FUNCTION
DATE FUNCTIONS

STEP 1

Set Up Kaggle API

- Obtain Kaggle API token from profile settings.
- Configure the API: Place kaggle.json in the .kaggle folder.
- Download dataset using kaggle datasets download command.

STEP 2

Install Required Libraries

- Install Python libraries using pip install commands.
- Load the dataset into Pandas for initial analysis.

STEP 3

Explore the Data

- Initial Analysis:
  - Check column names, data types, and missing values.
  - Use .info(), .describe(), and .head() functions.
- Identify potential issues in data structure.

STEP 4

Data Cleaning

- Remove duplicates and handle missing values.
- Ensure consistent data types (e.g., dates as datetime).
- Format currency values for analysis.

STEP 5

Feature Engineering

- Add calculated fields like Total Amount = (Unit Price * Quantity).
- Enhance dataset for SQL analysis.

STEP 6

Load Data into SQL

- Use SQLAlchemy to connect Python with PostgreSQL.
- Automate table creation and data insertion.
- Verify data using initial SQL queries.

SQL ANALYSIS

1- Find different **payment methods**, number of transactions, and quantity sold by payment method

2- Identify the highest-rated **category** in each branch, Display the branch, category, and average rating

3- Identify the busiest day for each branch based on the number of **transactions**.

4- Calculate the total **quantity** of items sold per **payment method**.

SQL ANALYSIS

5- Determine the average, minimum, and maximum **rating** of the category for each **city**. List the city, average_rating, min_rating, and max_rating.

6- Calculate the **total profit** for each category by considering total_profit

7- Determine the most common **payment method** for each Branch.

8- Categorize sales into 3 groups MORNING, AFTERNOON, EVENING. Find out each of the shift and number of invoices.

INSIGHTS

Among credit cards, wallets, and cash, people preferred **credit cards** for their purchases.

Business Decisions Based on Credit Card Preference:

1- Promote Credit Card Offers.
2- Enhance Credit Card Payment Infrastructure.
3- Targeted Marketing Campaigns

	payment_method text	no_of_payment bigint	quantity_sold double precision
1	Credit card	4256	9567
2	Ewallet	3881	8932
3	Cash	1832	4984

INSIGHTS

Among all 6 categories **Fashion Accessories** generated the highest total revenue.

Business Decisions Based on this:

1- Prioritize Stocking Fashion Accessories.
2- Expand Product Range.
3- Focus Marketing Efforts,

	category text	total_revenue double precision	profit double precision
1	Fashion accessories	489480.89999999997	192314.89320000037
2	Electronic accessories	78175.02999999998	30772.489499999978
3	Health and beauty	46851.17999999998	18671.7345
4	Food and beverages	53471.28000000006	21552.862200000003
5	Sports and travel	52497.93000000002	20613.808199999996
6	Home and lifestyle	489250.06	192213.6380999999

FUTURE ENHANCEMENTS

Integrate dashboard tools (e.g., Power BI, Tableau) for visualization.

Add more datasets for deeper analysis.

Automate real-time data ingestion and analysis.

MICROSOFT FABRIC

THANK YOU