**PhonePe Transaction Insights**

**Table of Contents**

1. Introduction

2. Data Extraction

3. SQL Database and Table Creation

4. Data Analysis

5. Dashboard Creation Using Streamlit

6. Data Insights

7. Results

**1. Introduction**

**1.1 Objective**

The objective of this project is to analyze PhonePe transaction data to extract valuable insights on user behavior, transaction patterns, and brand popularity.

**1.2 Why PhonePe Transaction Insights is Needed**

Understanding PhonePe transaction data helps uncover trends in digital payments, detect anomalies or fraud, and assist businesses and governments in making data-driven decisions.

**1.3 Tools and Technologies Used**

- Python  
- SQL (PostgreSQL)  
- Streamlit  
- Plotly  
- Pandas  
- SQLAlchemy  
- GitHub

**1.4 Dataset Overview**

The dataset is derived from PhonePe GitHub repositories, containing transaction details like amount, count, user demographics, device brands, and geographical distribution across states, districts, and years.

**2. Data Extraction**

Data was extracted from the PhonePe Pulse GitHub repository. JSON files were parsed and converted into structured DataFrames for analysis.

**3. SQL Database and Table Creation**

A PostgreSQL database was set up, and relevant tables were created to store aggregated and map-based transaction and user data. Python's SQLAlchemy was used to handle the connection and queries.

**4. Data Analysis**

Multiple SQL queries were performed to analyze transaction count, transaction amount, user count, and percentage by year, quarter, state, district, and brand.

**AGGREGATED TRANSACTION ANALYSIS**

* The transaction data is analysed to find the top payment method in every year and quater
* The state with highest transaction
* The quater with the highest payment are analysed

**INSURANCE ANALYSIS**

* The total insurance transaction by each state is analysed
* The yearlly transaction growth in insurance is analysed.
* The top districts in insurance transactions are found.
* The nation wide yearly growth is analysed.

**MAP TRANSACTION ANALYSIS**

* The top districts in the transaction is found
* The total transaction by each state is analysed.
* The yearly transaction trend is anlaysed.
* The transaction heatmap is used to analyse how transaction has changed over the years.

**DEVICE DOMINANCE**

* The geographical view of the registered user is visualised using streamlit that dynamically changes with the selection of year.
* The top brand device that is used widely is analysed.
* The device dominance is analysed to find the user behaviour by device which helps in marketing
* The most dominant device in each state is analysed.
* The heat map is used to analysed the device dominance over the phonepe.

**FRAUD DETECTION**

* The unusal transactions are analysed.
* The transacations with low and high count is analysed.
* The district with high transactions but low registered user are analysed.
* The sudden transaction spike is analysed.

The transaction count deviation in each state is anlaysed.

**5. Dashboard Creation Using Streamlit**

An interactive dashboard was built using Streamlit and Plotly to visualize heatmaps, bar graphs, line charts, and maps showing trends in transaction and user behavior.

**6. Data Insights**

- States like Maharashtra, Karnataka, and Tamil Nadu consistently showed high transaction volumes.  
- Quarter 4 of each year typically recorded higher transaction activity.  
- Certain brands like Xiaomi and Samsung were dominant in terms of user count.  
- Seasonal spikes and anomalies were detected using heatmaps and line charts.

**7. Results**

The analysis of user distribution based on device brands across different states, years, and quarters revealed the following:

- The most popular brands across all states were Samsung, Xiaomi, Vivo, Realme, and Oppo, based on total user count.

- There was a consistent growth in the number of users across all top brands from 2018 to 2023.

- Samsung had a strong presence in southern states such as Tamil Nadu and Kerala, while Xiaomi dominated Bihar, Uttar Pradesh, and other northern states.

- The average user percentage for each brand showed clear brand preference trends, sometimes exceeding 50% in certain regions.

- Seasonal shifts and spikes in usage were observed, particularly during promotional periods or festivals.