# PhonePe Pulse Data Analysis and Visualization Project

## Objective / Goal

The goal of this project is to analyze the digital payments landscape in India using PhonePe Pulse data. The objective is to visualize and extract insights about transaction volume, transaction amount, user growth, and insurance coverage across states and districts using Python, SQL, and Streamlit.

## Business Problem

PhonePe handles billions of transactions every year, and the dataset contains vast information about transaction patterns, user adoption, and digital payment penetration across regions. The challenge is to efficiently clean, transform, and visualize this data to identify growth trends, high-performing regions, and areas with low adoption.

## Dataset Description

The dataset contains multiple JSON files from the PhonePe Pulse GitHub repository covering:   
- Aggregated Transactions (year, quarter, type, count, amount)  
- Aggregated Users (year, quarter, brand, registered users, app opens)  
- Aggregated Insurance Data  
- Map-level Data (state, count, amount)  
- Top-level Data (district, registered users, transaction count)

## Steps Involved

1. Data Extraction: Parsed JSON files into structured data using Python.  
2. Data Cleaning: Handled null values, standardized column names, and merged datasets.  
3. Data Transformation: Created unified master datasets for transactions, users, and insurance.  
4. Data Storage: Stored processed data in MysqL for efficient querying.  
5. Visualization: Built an interactive dashboard using Streamlit for year-wise, state-wise, and category-wise analysis.

## Visualization Summary

1. Transaction Trends: Bar and line charts showing total transaction count and amount by year and state.  
2. User Growth: Brand-wise user registration and app opening analysis.  
3. Insurance Analysis: State-wise insurance adoption and amount trends.  
4. Top Districts: Map and table visualizations showing top-performing regions.

## Key Insights

- Maharashtra, Karnataka, and Uttar Pradesh dominate in transaction volume and value.  
- Digital adoption in Northeast states and Union Territories is increasing but still below national average.  
- Insurance penetration shows strong correlation with digital payment maturity.  
- User engagement patterns reveal high activity during festival seasons and year-end quarters.

## Business Impact

The project helps PhonePe and policymakers understand digital payment adoption patterns across India. It supports strategic decisions for targeted marketing, infrastructure planning, and user engagement campaigns.

## Tools & Technologies Used

Python, Pandas, NumPy, Matplotlib, Plotly, Streamlit, Mysql workbench, Visual Studio Code, Git.

## Challenges Faced

- Parsing large nested JSON files into relational format.  
- Handling inconsistent naming conventions across states and districts.  
- Managing performance issues during data merging and visualization.  
- Ensuring clean integration between Python, Mysql, and Streamlit components.

## Conclusion

The PhonePe Pulse Data Analysis project successfully delivered a data-driven dashboard that visualizes India's digital payment ecosystem. The interactive Streamlit app allows users to explore trends across time, regions, and transaction categories, offering actionable insights for digital finance expansion.