|  |
| --- |
| Project Title : PhonePe Transaction Insights  Domain : Finance / Payment Systems  Name : Deepika R,  +91 8675486892 |

Summary:

This project analyzes PhonePe Pulse data by extracting JSON files, loading them into a SQL database, and performing in-depth analysis on aggregated transactions, insurance data, and user statistics across states, districts, and pin codes. Insights are visualized using Python and presented in an interactive Streamlit dashboard, enabling dynamic exploration of payment trends, regional patterns, and user behavior. The workflow showcases skills in data extraction, SQL, visualization, and analytical thinking for actionable business insights in the digital payments domain..

Problem Statement:

With the rapid growth of digital payments, there is a need to **understand transaction patterns, user activity, and insurance adoption** to make informed business decisions.

PhonePe’s publicly available Pulse data is in **unstructured JSON format**, making it difficult to query and analyze without proper data transformation.

Organizations require **efficient visualization tools** to explore geographical and category-wise trends in transactions at state, district, and pin code levels.

Identifying **top-performing regions, popular payment categories, and engagement patterns** is essential for targeted marketing and service improvement.

Objectives:

PhonePe handles millions of transactions daily across India, involving various payment categories, insurance products, and user activities.  
To **improve business strategy and decision-making**, the company needs:

1. **Identify High-Value Regions** – Which states, districts, and pin codes contribute the most to total transactions and insurance sales.
2. **Detect Low-Performance Areas** – Highlight underperforming regions to target with promotional offers or awareness campaigns.
3. **Track Insurance Growth** – Understand where insurance adoption is strong or weak to optimize product offerings.
4. **Monitor User Base Trends** – Determine which regions have the highest registered users and app engagement.
5. **Measure Brand Popularity** – Identify which brands are most popular in transactions.
6. **Support Targeted Marketing** – Use these insights to run region-specific campaigns that maximize ROI.

Approach:

GitHub JSON

Python (ETL)

MySQL DB

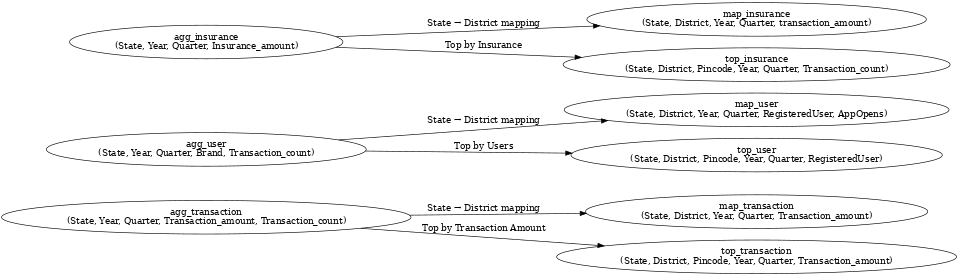
SQL Analysis

Python Visualization

Streamlit Dashboard

Database Schema Explanation:

The project database is structured into **three main categories**: **Aggregated**, **Map**, and **Top** tables — each storing different levels of detail about transactions, insurance, and users.

Relationships

* Aggregated tables give state-level summaries.
* Map tables provide district-level granularity.
* Top tables store only top-ranked records, making leaderboard queries faster.

### **ETL Pipeline Explanation**

#### **EXTRACT**

* Use to OS traverse the directory hierarchy.
* Use json to load nested data structures from json files.

TRANSFORM

* Normalize nested JSON into Python dictionaries/lists.
* Clean and format:
  + Extract quarter numbers from filenames.
  + Handle missing values with try-except.

#### **Load**

* Establish MySQL connection using pymysql.
* Create database (CREATE DATABASE IF NOT EXISTS phone\_pe).
* Create tables with appropriate data types.
* Insert transformed data using INSERT statements inside a loop.

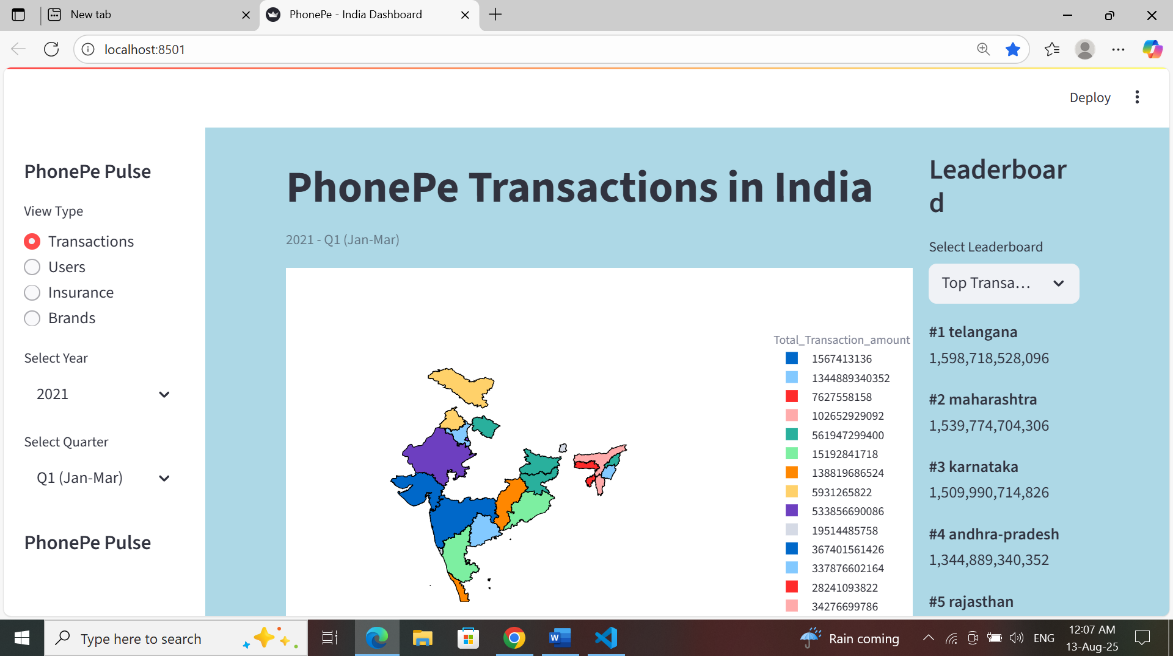
Technology used:

 **Programming Language:** Python 3.x

 **Database:** MySQL

 **Libraries:**

* **Json -** To parse and handle the JSON data from PhonePe Pulse datasets
* **Pandas** - Data manipulation and analysis library
* **Pymysql** - Python connector for MySQL database
* **Plotty.express** - Interactive plotting library for Python
* **Request** - To send HTTP requests in Python.
* **Streamlit** - To create interactive dashboards for data visualization

****

Business Finding:

* **High Transaction States:** Maharashtra, Karnataka, Tamil Nadu lead in value & volume.
* **Low Transaction States:** Northeastern states & UTs — potential growth markets.
* **Insurance Trends:** Urban states have high premiums; rural states show low adoption.
* **User Base:** Tier-1 cities have most users; tier-2 & tier-3 towns growing fastest.
* **App Opens:** Some states have many registered users but low activity — retention needed.
* **Brand Usage:** Xiaomi & Samsung dominate; brand insights useful for promotions.
* **Geographic Insights:** Urban pincodes dominate, but rural areas have rising adoption.

Business Impact:

The analysis of PhonePe Pulse data provides significant business value by enabling data-driven decision-making across multiple strategic areas. By identifying the top-performing and low-performing states, districts, and pincodes in terms of transaction amounts, insurance adoption, and user registrations, businesses can tailor their marketing campaigns to target high-value regions and design engagement strategies for underperforming areas.

The insights from insurance leaderboards highlight regions with high penetration, offering opportunities to introduce advanced or premium products, while low-penetration areas can be prioritized for awareness campaigns. User data, including registered user counts, app opens, and brand preferences, allows for refined customer segmentation, enabling personalized offers and promotions.

Furthermore, brand-level transaction analysis supports collaboration and co-marketing with popular device manufacturers to boost adoption. These findings also empower operational teams to allocate resources efficiently, focusing on regions with the highest growth potential. Overall, the project supports enhanced customer retention, improved product-market fit, and optimized revenue strategies, making it a valuable tool for competitive advantage in the digital payments and fintech landscape.

**Strategic Recommendations:**

1. **Targeted Marketing Expansion**
   * Focus ad campaigns in **emerging states** showing steady year-on-year growth to accelerate adoption.
   * Create **localized offers** (language-specific and festival-based) for districts with high potential but low engagement.
2. **Insurance Product Growth**
   * Offer **bundled digital payment + micro-insurance packages** in areas with high transaction activity but low insurance adoption.
   * Partner with **local influencers and community organizations** to improve trust in insurance products.
3. **User Retention & Loyalty**
   * Implement a **tiered rewards program** for frequent users to encourage repeat transactions.
   * Send **personalized push notifications** with cashback offers or bill payment reminders.
4. **Operational Optimization**
   * Increase **server infrastructure** during festival seasons and salary credit periods when transaction spikes are likely.
   * Optimize **regional payment processing hubs** to reduce latency in high-demand areas.
5. **Partnerships & Cross-Promotions**
   * Collaborate with **popular e-commerce and food delivery apps** for co-branded offers in high-growth districts.
   * Partner with **device manufacturers** in states where specific phone brands dominate to offer instant onboarding benefits.
6. **Data-Driven Product Innovation**
   * Use transaction trend data to **introduce new bill payment categories or loan products.**
   * Develop **predictive models** to anticipate demand surges and fraud attempts.

Conclusion:

The PhonePe Pulse project successfully extracted, processed, and analyzed large-scale digital payment data to uncover key trends in transactions, user behavior, and insurance adoption across India. By leveraging JSON data extraction, SQL-based analytics, and interactive Streamlit visualizations, the project delivered actionable insights into top-performing states, districts, and pin codes, while identifying regions with untapped growth potential.

These findings enable data-driven decision-making for marketing, infrastructure scaling, user engagement, and product development, ultimately contributing to better financial inclusion and enhanced customer satisfaction.