**📊 PhonePe Pulse Data Analysis & Dashboard**

**🔖 Project Title:**

**PhonePe Pulse Data Visualization and Business Insights Using Python, SQL, and Streamlit**

**📌 Overview:**

**This project focuses on extracting and analyzing transaction data from the official PhonePe Pulse GitHub repository. The key objective is to transform raw JSON data into meaningful insights through SQL queries, Python-based analysis, and an interactive Streamlit dashboard.**

**The project also covers real business use cases such as:**

* **Customer segmentation**
* **Fraud detection**
* **Insurance insights**
* **State-wise performance comparisons**
* **Growth analysis**

**📂 Data Source:**

* **Official GitHub Source:** [**PhonePe Pulse GitHub**](https://github.com/PhonePe/pulse)
* **Format: JSON**
* **Data Categories:**
  + **Aggregated Transactions (by year, quarter, state)**
  + **Aggregated Users (App opens, user registrations)**
  + **Aggregated Insurance (Coverage statistics)**
  + **Top performing states/districts**

**🛠️ Tech Stack:**

* **Programming Language: Python, SQL**
* **Database: MySQL**
* **Dashboard Framework: Streamlit**
* **Libraries Used:**
  + **pandas, plotly, matplotlib, seaborn**
  + **mysql-connector, sqlalchemy**
  + **json, os, glob**

**📈 ETL Process (Extract - Transform - Load):**

1. **Extract: Raw JSON files were parsed using Python to collect and flatten the nested structure.**
2. **Transform: Cleaned and standardized data into tabular format.**
3. **Load: Data was stored into structured MySQL tables with proper schema (9 tables total).**

**🧮 SQL Tables Created:**

* **aggregated\_transaction**
* **aggregated\_user**
* **aggregated\_insurance**
* **map\_transaction**
* **map\_user**
* **map\_insurance**
* **top\_transaction**
* **top\_user**
* **top\_insurance**

**Each table contains structured info with dimensions like year, quarter, state, district, pincode, and metrics like transaction count, app opens, insurance count, etc.**

**📊 Data Analysis (Python + SQL):**

**We performed the following analysis:**

* **Top performing states and districts**
* **Year-over-year transaction growth**
* **User adoption trends**
* **Insurance penetration across regions**
* **Seasonal trends (by quarter)**
* **Heatmaps for insurance and transaction volume**

**💼 Business Use Cases:**

1. **Customer Segmentation**
   * **Identified high-engagement regions based on app opens and transaction counts.**
2. **Fraud Detection (Theoretical)**
   * **Detected anomalies in low-population districts with unusually high transaction volume.**
3. **Insurance Insights**
   * **Evaluated insurance coverage growth state-wise and quarterly trends.**
4. **Regional Performance**
   * **Compared growth rates across states and districts to highlight top contributors.**

**📎 Future Scope:**

* **Add forecasting models using machine learning**
* **Integrate real-time API (if client want to provides us)**
* **Expand business use case coverage**
* **Deploy the dashboard online (Streamlit Cloud, Render, or Heroku)**

**✅ Final Plan to Build STREAMLIT Dashboard**

**An interactive web dashboard was built using Streamlit where:**

* **Users can filter data by year, state, or quarter**
* **Visualizations are rendered using Plotly and Seaborn**
* **Multiple pages like Transactions, Users, Insurance, and Heatmaps**
* **Supports business queries and real-time analytics views**

**🧩 Based On:**

* 📒 **CheckDataExtraction.ipynb** logic we have taken according to this file.
* 🧱 **streamlitcode.py** style, functions, UI for Showing interactive web dashboard.
* 🧮 MySQL data (we will convert it properly) to find our insights.

**📦 Structure of Your Final App:**

**✅ 1. Sidebar Navigation:**

* Home
* Data Insights
* Customer Segmentation
* Fraud Detection
* Insurance Analysis
* Trend Analysis
* Competitive Benchmarking
* Geographical Insights
* Top Charts

**✅ 2. Features from Notebook:**

All use cases will be included:

| **Use Case** | **Feature Included in App** |
| --- | --- |
| ✅ Customer Segmentation | KMeans Clustering with state-level plot |
| ✅ Fraud Detection | Z-score anomaly map |
| ✅ Geographical Insights | Choropleth Map from GeoJSON |
| ✅ Insurance Insights | Choropleth by insurance type |
| ✅ Payment Performance | Transaction amount analysis |
| ✅ Trend Analysis | Quarter-wise time trends |
| ✅ Competitive Benchmarking | Top 10/Bottom 10 states |
| ✅ User Engagement | AppOpens vs Registered Users line graph |
| ✅ Marketing Optimization | Coming from clustering + transactions |
| ✅ Product Development | Based on brand usage trends |

**🚀 🖥️ How to Run the Project:**

1. **Clone the repository**
2. **Install dependencies:**

**Bash -- pip install -r requirements.txt**

1. **Set up MySQL and import the structured tables**
2. **Run the Streamlit app:**

**Bash -- streamlit run phonepe\_dashboard.py**

**✅ Project Status:**

**✔️ ETL completed  
✔️ SQL schema created  
✔️ Full analysis done in Jupyter  
✔️ Streamlit dashboard developed and tested  
✔️ Business use cases documented**