

# Data science

## Mini Project – 2

### The Project Report On PhonePe Transaction Insights

Batch code: DS-C-WD-E-B39

Name: Kavya P P

# 1. Problem Statement

Understanding customer behavior is essential for optimizing business strategies and improving customer satisfaction. ShopEasy aims to analyze customer interactions, purchase patterns, sentiment trends, and marketing effectiveness to gain actionable insights. The objective is to enhance engagement, optimize marketing spend, and boost conversions by leveraging data-driven decisions.

---

# 2. Introduction

The purpose of this project is to analyze digital transaction data from PhonePe Pulse, a platform that aggregates extensive financial interaction statistics across India. The analysis focuses on understanding user engagement, transaction patterns, and insurance growth across various states and quarters. Using this data, key business insights and trends are derived to inform decision-making and potential market strategies.

---

# 3. Objectives

- To visualize state-wise and district-wise transaction volumes and values.
  - To examine user registration and app usage trends.
  - To assess the penetration and growth of PhonePe's insurance services.
  - To identify top regions by transaction activity using interactive visualizations.
- 

# 3. Tools and Technologies Used

| Tool/Tech | Purpose                          |
|-----------|----------------------------------|
| Python    | Data processing and analysis     |
| MySQL     | Data storage and querying        |
| Streamlit | Web-based data dashboard         |
| Plotly    | Data visualization               |
| Pandas    | Data manipulation                |
| GeoJSON   | ,State-level map rendering       |
| GitHub    | Data source (PhonePe Pulse Repo) |

---

# 4. Data Source and Preprocessing

The data was cloned from the PhonePe Pulse GitHub repository. It includes JSON files representing transaction data by state, district, and pincode. The data was parsed using Python and loaded into MySQL for structured querying. State names were standardized, and key fields like transaction type, count, and amount were extracted.

---

## 5. Data Preprocessing

- JSON files were parsed and converted into Pandas DataFrames.
- Necessary fields such as State, Year, Quarter, Transaction Type, Count, and Amount were extracted.
- State and district names were cleaned for consistency.
- Tables were created in MySQL and data was inserted for use with SQL queries.

---

## 6. Database Tables Created

- aggregate\_transaction
- aggregate\_user
- aggregated\_insurance
- map\_transaction
- map\_user
- map\_insurance
- top\_transaction
- top\_user
- top\_insurance

Each table contains fields such as States, Years, Quarters, Registered Users, Transaction Count, and Transaction Amount.

---

## 7. Functional Modules and Visualizations

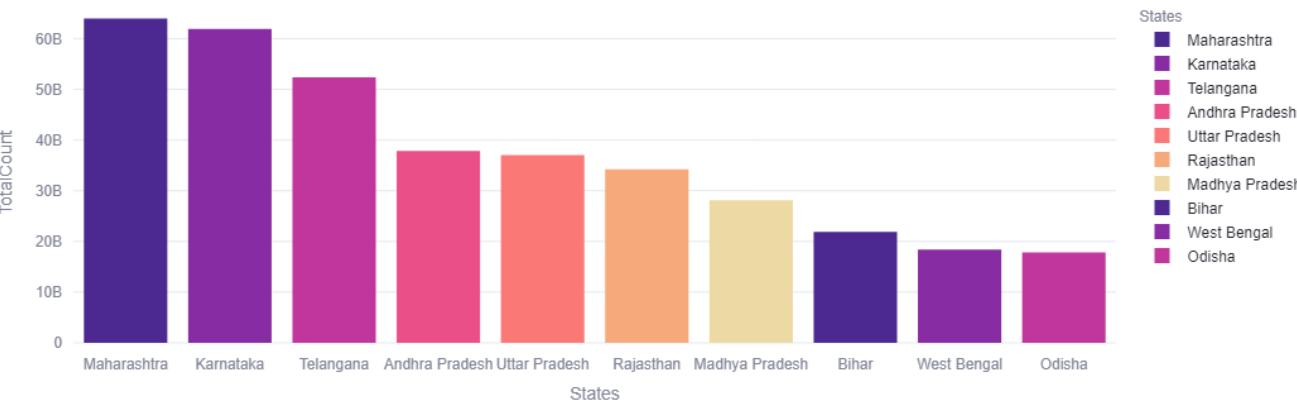
### 7.1 Decoding Transaction Dynamics

Analyzes transaction type distribution, state-wise transaction counts, and overall growth trends over the years.

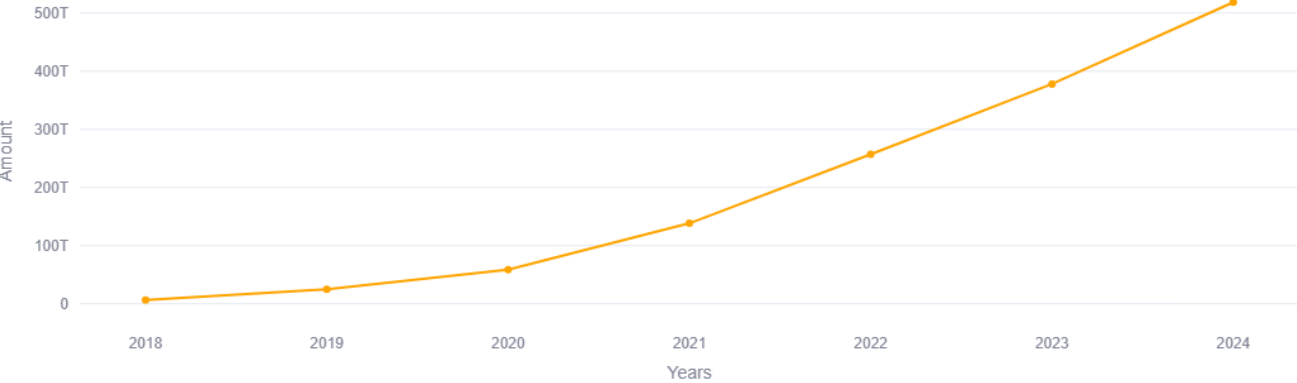
Transactions by Type



Top 10 States by Transactions



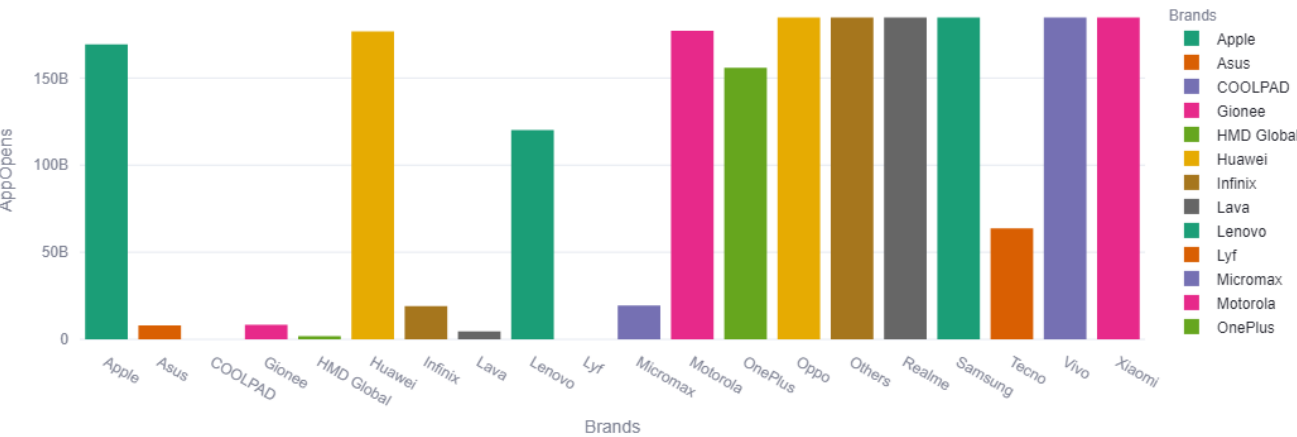
Transaction Trend Over Years



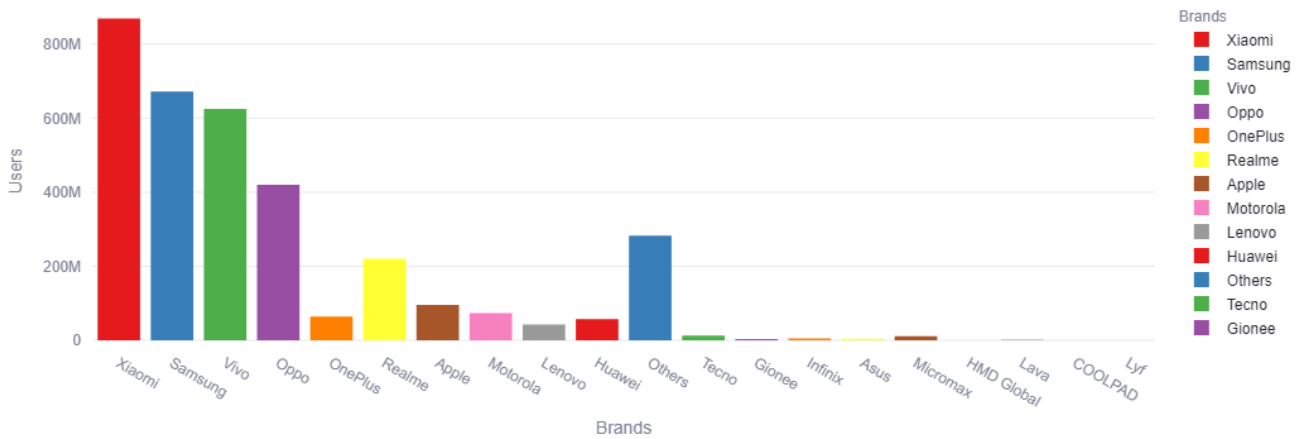
7.2 Device Dominance & Engagement

Explores user engagement across mobile brands, focusing on app opens and user registrations.

App Opens by Device Brand



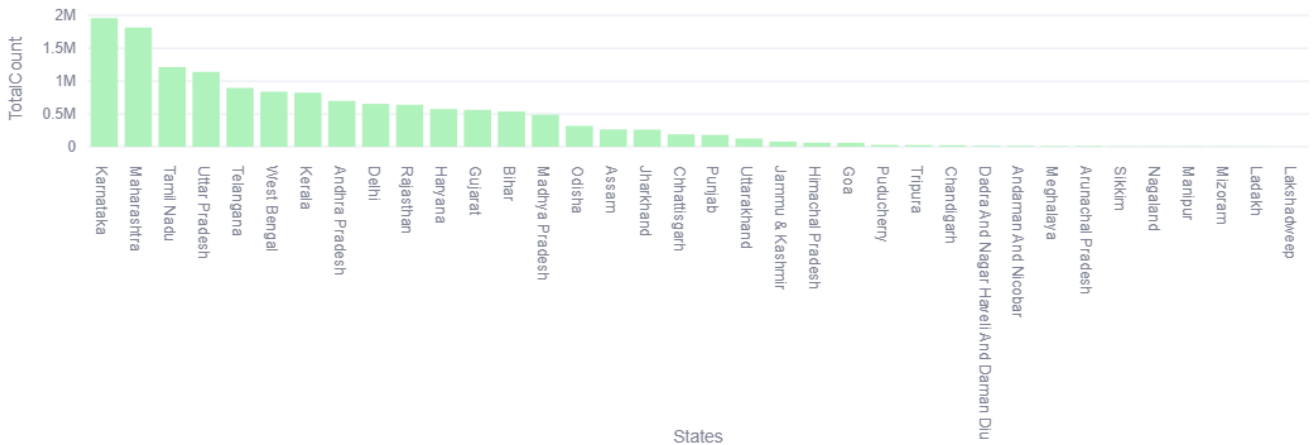
Users by Device Brand



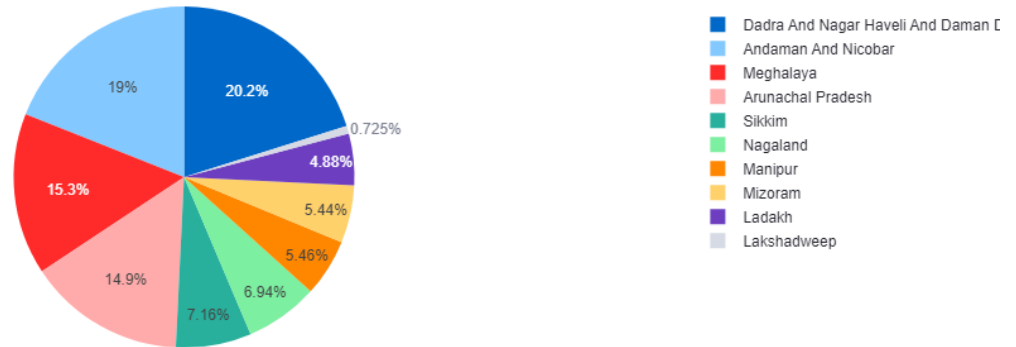
### 7.3 Insurance Growth Analysis

Tracks insurance transaction volume over time and ranks states by adoption rate and growth.

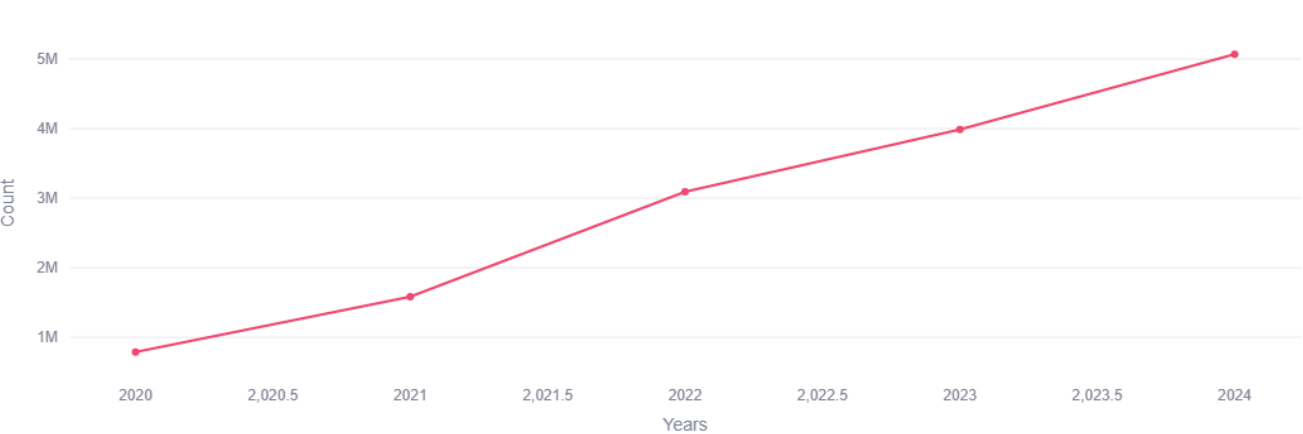
Top States by Insurance Transactions



Least Penetrated States (Pie View)



Insurance Growth Over Time



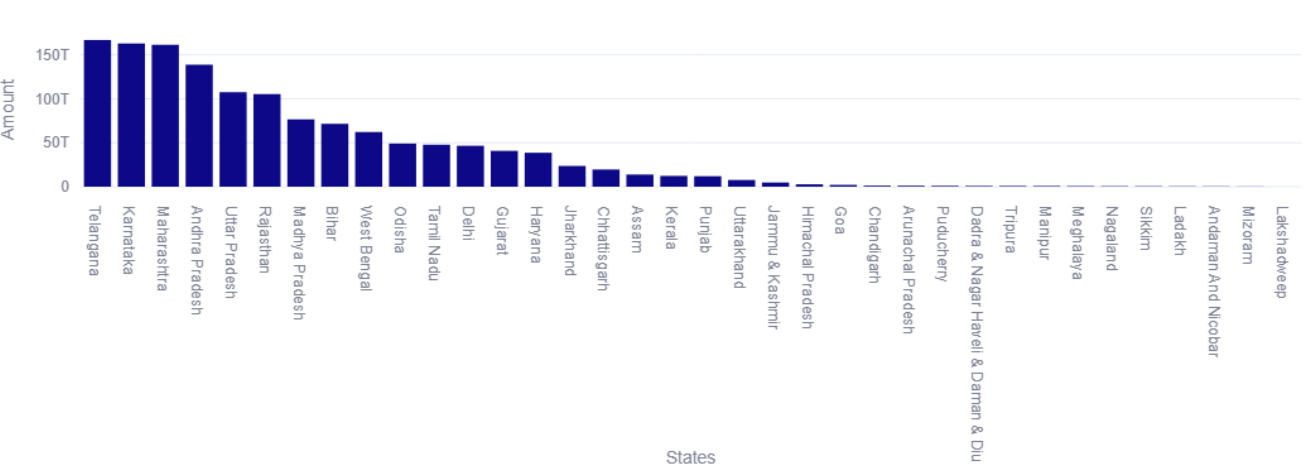
7.4 Market Expansion via Transactions

Highlights states with high transaction values and trends, useful for identifying expansion opportunities.  
states with high transaction values and trends, useful for identifying expansion opportunities.

State-wise Market Value Map



States by Transaction Value



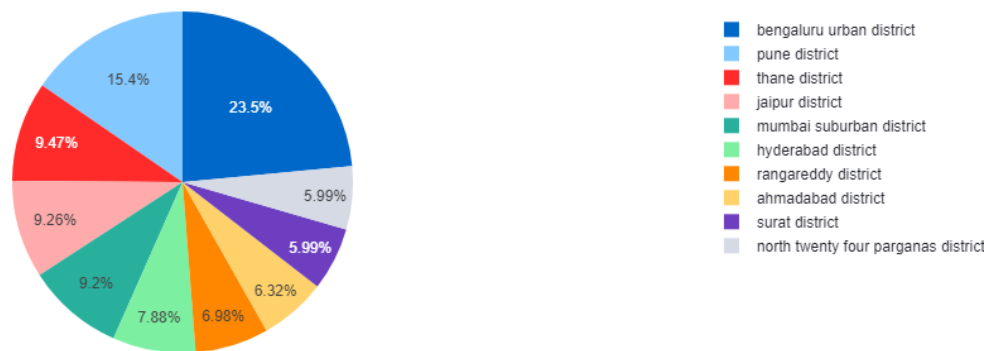
## 7.5 User Engagement and Growth Strategy

Analyzes app opens and user registrations by region and over time, showcasing top-performing districts.

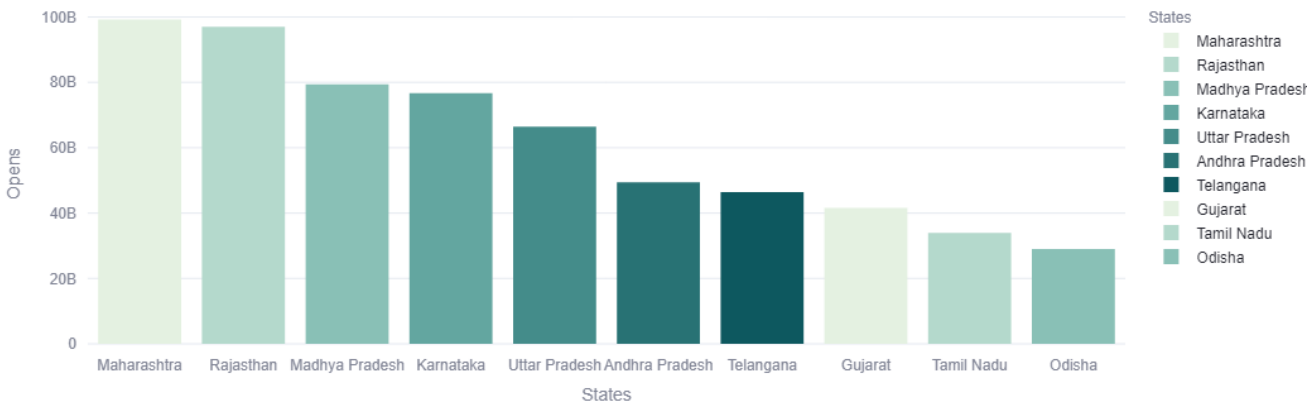
User Registrations Over Time



Top Districts by Registrations Share



Top States by App Opens



## 7.6 Insurance Engagement by District

Provides district-wise breakdown of insurance transactions and value distribution.

Insurance Transactions by District



Insurance Value Distribution by District



## 8. Conclusion

This project successfully utilized the PhonePe Pulse dataset to develop a fully functional dashboard capable of providing key insights into transaction trends, user behavior, and insurance penetration. The integration of MySQL, Streamlit, and Plotly enabled real-time querying and visualization, empowering decision-makers to interactively explore the data.