

# Phonepe Transaction Insights

**Skills :** Data Extraction, SQL Proficiency, Data Visualization, Analytical Thinking, Documentation, Streamlit ,Power BI

**Domain :** Finance/Payment System

**- Keerthika J S**

# Problem Statement

With the increasing reliance on digital payment systems like PhonePe, understanding the dynamics of transactions, user engagement, and insurance-related data is crucial for improving services and targeting users effectively. This project aims to analyze and visualize aggregated values of payment categories, create maps for total values at state and district levels, and identify top-performing states, districts, and pin codes.

## Data Extraction :

- **Clone the GitHub repository containing PhonePe transaction data and load it into a SQL database.**

## SQL Database and Table Cration:

**Setup a SQL database using a relational database management system (e.g., MySQL, PostgreSQL).**

## Create tables to store data from the different folders:

- **Aggregated Tables :** User, Transactions, Insurance
- **Map Tables :** User, Transactions, Insurance
- **Top Tables :** User, Transactions, Insurance



# Data Extraction

## To Extract the data using Python, and Store it into a MySQL database

1. Load Local Data (e.g., JSON, CSV)
2. Install MySQL Connector for Python - `pip install mysql-connector-python`
3. Connect Python to MySQL - `import mysql.connector`
4. Insert Data from DataFrame into MySQL
5. Close the Connection - `cursor.close()` and `conn.close()`

## Python Libraries we used in this project :

- **Pandas** – `import pandas as pd`
- **Plotly.express** – `import plotly.express as px`
- **Matplotlib.pyplot** – `import matplotlib.pyplot as plt`
- **Seaborn** – `import seaborn as sns`
- **Streamlit** – `import streamlit as st`

# Business Case Study



**Device  
Dominance and  
User  
Engagement  
Analysis**



**Insurance  
Penetration  
and Growth  
Potential  
Analysis**



**User  
Engagement  
and Growth  
Strategy**



**Transaction  
Analysis  
Across States  
and Districts**



**Insurance  
Transactions  
Analysis**

# Device Dominance and User Engagement Analysis

PhonePe aims to enhance user engagement and improve app performance by understanding user preferences across different device brands. The data reveals the number of registered users and app opens, segmented by device brands, regions, and time periods. However, trends in device usage vary significantly across regions, and some devices are disproportionately underutilized despite high registration numbers.

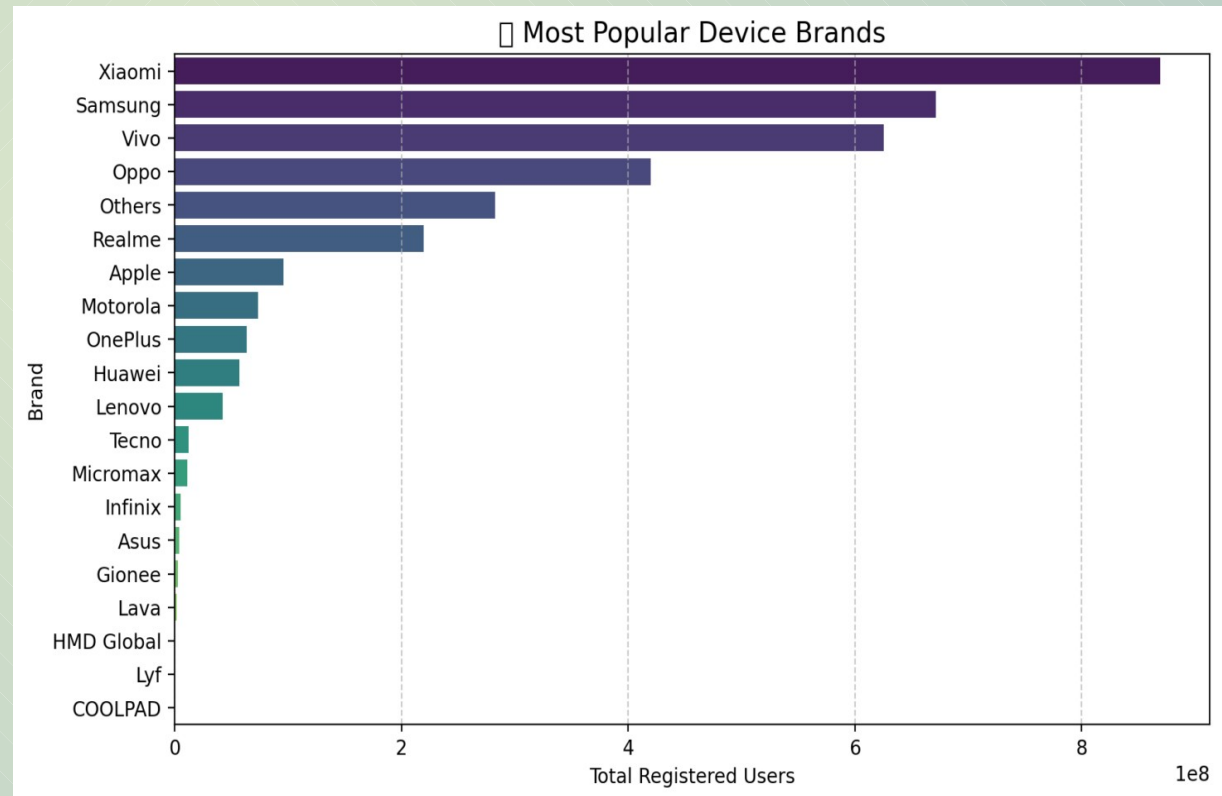
## Identify the Most Popular Device Brands

To analyze the most used devices, we calculate total registered users per brand.

### SQL query :

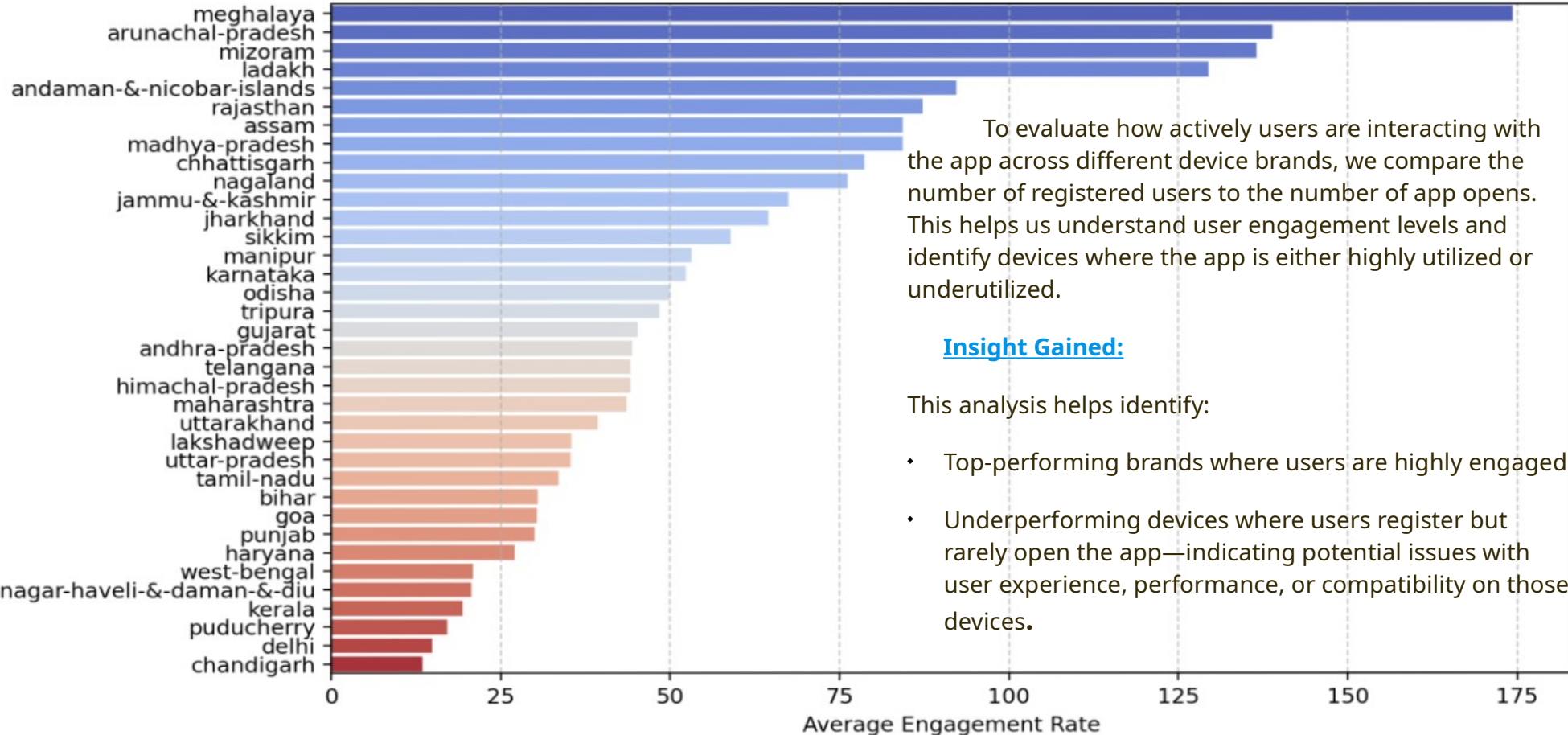
```
SELECT
  Brand,
  SUM(User_count) AS total_registered_users,
  ROUND(AVG(Percentage), 2) AS avg_market_share
FROM aggregated_user
GROUP BY Brand
ORDER BY total_registered_users DESC;
```

**This query finds which brands dominate the market based on total user registrations.**



# Compare App Engagement Across Devices

□ Average App Engagement Rate by State



To evaluate how actively users are interacting with the app across different device brands, we compare the number of registered users to the number of app opens. This helps us understand user engagement levels and identify devices where the app is either highly utilized or underutilized.

## Insight Gained:

This analysis helps identify:

- Top-performing brands where users are highly engaged.
- Underperforming devices where users register but rarely open the app—indicating potential issues with user experience, performance, or compatibility on those devices.

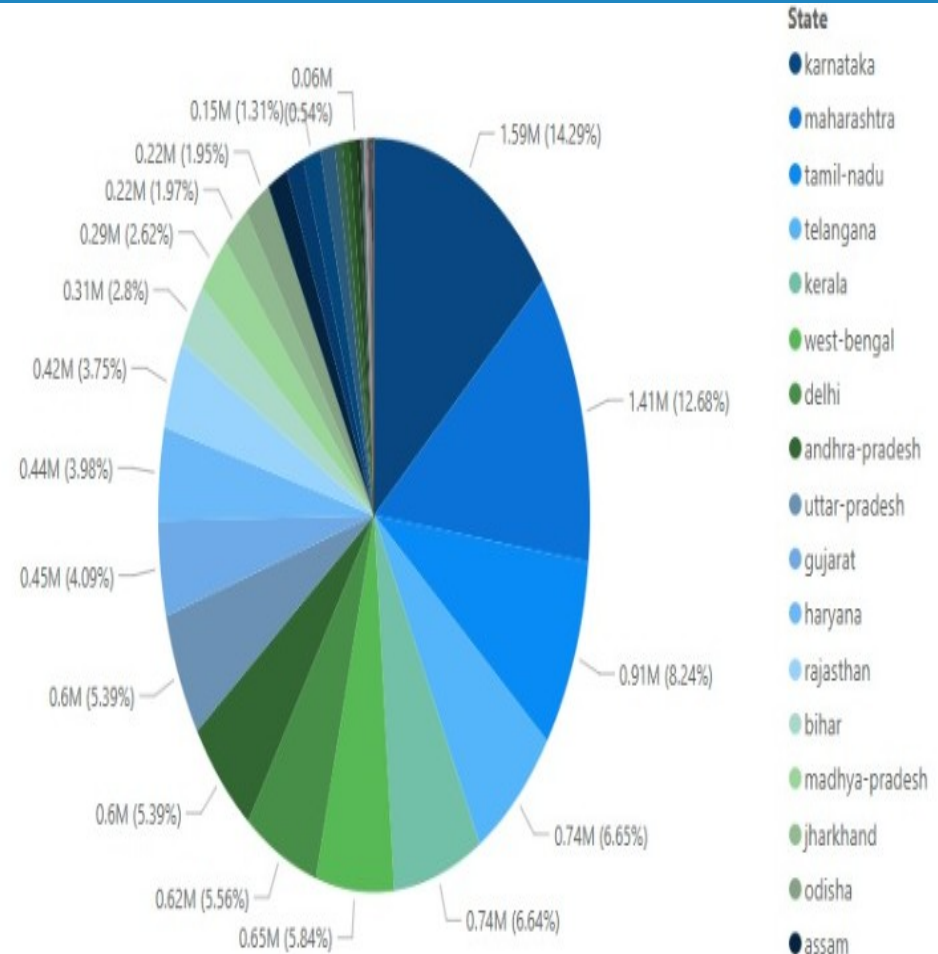
# Insurance Penetration and Growth Potential Analysis

## Identify the Top States for Insurance Adoption

This step focuses on analyzing insurance performance across different states by examining both the volume of policies sold and the total premium collected. It provides a deeper understanding of how well insurance products are being adopted geographically .

### Insight Gained:

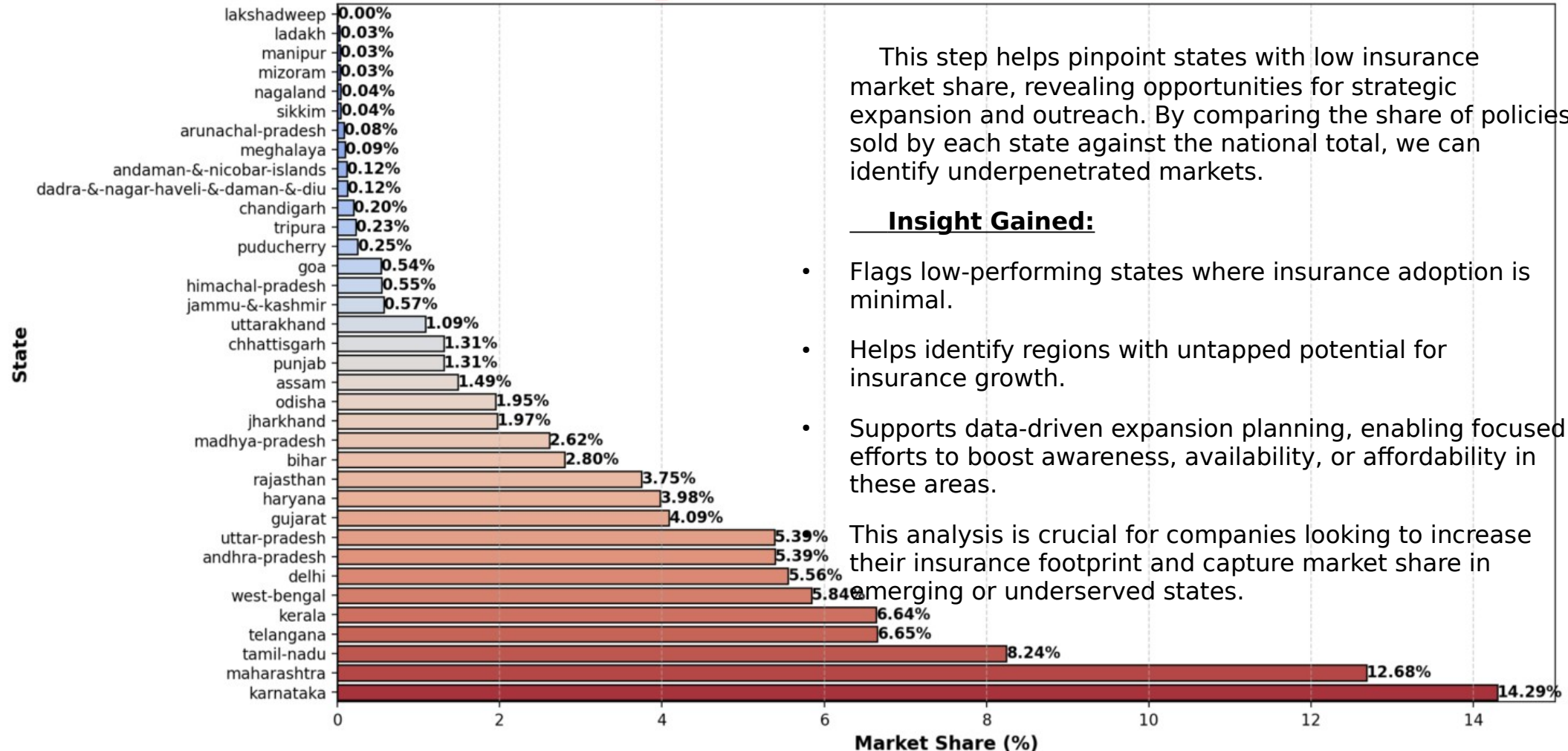
- Identifies top-performing states in terms of insurance adoption and revenue generation.
- Highlights regions with higher-value insurance policies, which could indicate strong market potential or higher-value customers.
- Helps prioritize states for targeted insurance marketing or expansion strategies.
- This analysis enables data-driven decisions to enhance insurance reach and performance across India.





# Insurance Underpenetrated States for Expansion

## Market Share of Insurance Policies Sold





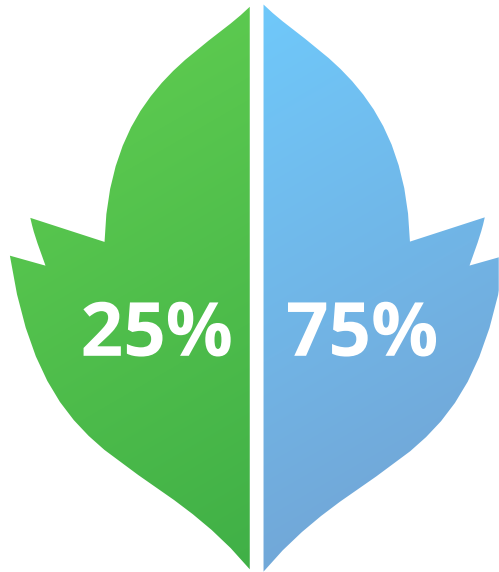
# States into Strong vs Weak Insurance Adoption

To better assess the distribution of insurance adoption across India, we categorize each state based on its market share of total insurance policies sold. This classification helps simplify the analysis by grouping states into two key categories: Strong and Weak adoption zones.

This classification helps insurance providers prioritize efforts, allocate resources wisely, and design region-specific campaigns to either maintain strongholds or boost adoption in weaker areas.

## **Strong Insurance Adoption**

- States that contribute more than 5% to the total insurance policies sold.
- Indicates high consumer trust, stronger awareness, better accessibility, and/or effective marketing.
- Represents mature or well-penetrated insurance markets.



## **Weak Insurance Adoption**

- States that contribute 5% or less to total insurance policy count.
- Signals underpenetrated markets with potential barriers like low awareness, affordability issues, or lack of access.
- Offers significant opportunity for targeted outreach and growth.

# User Engagement and Growth Strategy

## Identify Underperforming Regions with Growth & Market Share

### High Adoption (> 3% market share)

- Strong presence of registered users.
- Indicates robust brand awareness, market penetration, and possibly high user satisfaction.
- States in this category are likely already well-established markets.

### Moderate Adoption (1% - 3% market share)

- Balanced performance—states showing good engagement but with room for growth.
- These regions may benefit from targeted marketing or partnership strategies to push into the high adoption bracket.

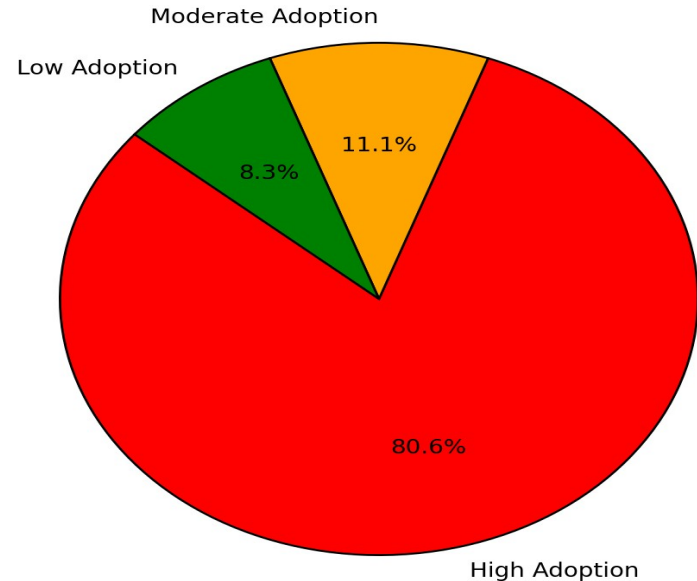
### Low Adoption (< 1% market share)

- Underperforming regions with limited user base.
- Could indicate challenges like low digital literacy, poor network coverage, or limited brand reach.
- Presents high potential for growth if addressed with the right regional strategies (e.g., local campaigns, regional language support, awareness drives).

Uses a CASE statement to assign each state into an adoption category based on its market share:

- > 3% → High Adoption
- 1%–3% → Moderate Adoption
- < 1% → Low Adoption

## □ User Adoption Categories



# Transaction Analysis Across States and Districts

**Analyze the Most Popular Transaction Types :** This analysis aims to identify which transaction types are the most popular and profitable by examining both volume and value metrics. It helps understand user preferences, revenue drivers, and high-value services offered on the platform.

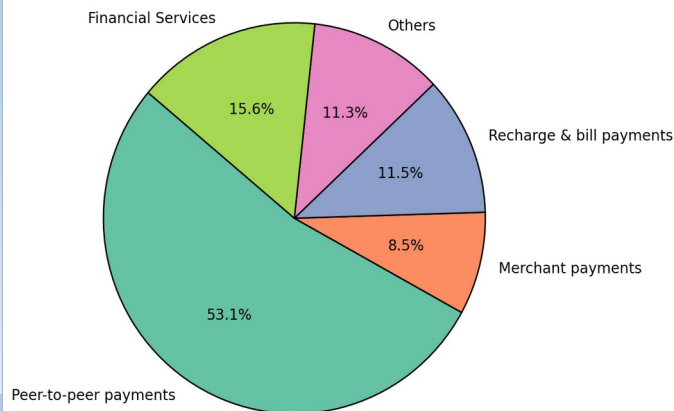
## Peer-to-Peer Payments

- Massive total revenue (₹266.5T) and high average value per transaction (~₹3,134).
- While not the most frequent, these transactions are large in value—likely personal or business transfers, rent, or bulk payments.
- High-value, low-frequency pattern.

## Merchant Payments

- Highest in transaction volume (130B+ transactions) but with a moderate average value (~₹501).
- Shows the dominance of daily, small-to-medium business payments—like groceries, cafes, and retail purchases.
- Critical for everyday usage and merchant ecosystem growth.

□ Total Revenue by Transaction Type



## Others

- Low in volume but maintains an average transaction value of ₹665.
- Could include miscellaneous services like donations, toll payments, subscriptions.
- Useful to watch for niche service expansion.

## Recharge & Bill Payments

- Moderate in volume and revenue, with average value around ₹681.
- Indicates utility-oriented transactions—frequent and essential, though not high in value.
- Great for retention and recurring engagement.

## Financial Services

- Lowest in transaction count, but with notably high average value (~₹921).
- These likely include loan repayments, insurance, investments—less frequent but financially significant.
- Ideal for targeted financial product campaigns.

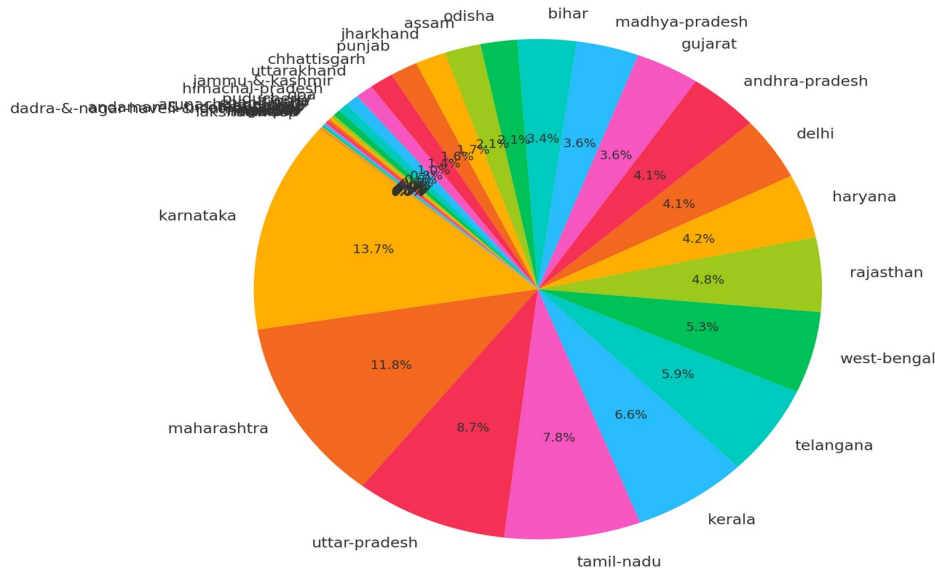
# Insurance Transactions Analysis (2020-2024)

To identify which states in India contributed the most to insurance transactions in terms of both:

Total policies sold

Total premium collected

Total Insurance Premium Collected by State (2020-2024)



## Top performers:

- Karnataka (~13.7%) and Maharashtra (~11.8%) are the leading contributors, showing strong adoption and higher-value policies.
- Uttar Pradesh, Tamil Nadu, Kerala, and Telangana also contribute significantly, each with a 5–9% share.

## Mid-range states:

- States like West Bengal, Rajasthan, and Haryana have a moderate share (3–5%), indicating decent insurance penetration.

## Underperformers:

- Smaller states/UTs like Dadra & Nagar Haveli, Sikkim, Mizoram, and Tripura have very low contributions, suggesting potential markets for future insurance growth and awareness campaigns.

# Conclusion

Over the course of this analysis, multiple insights emerged across user behavior, device engagement, insurance adoption, and transaction trends

## 1. App Engagement Across Devices

- Brands like Samsung and Xiaomi showed high engagement rates, indicating active user bases.
- Certain brands exhibited lower app open-to-user ratios, suggesting underutilization or poor app performance.

## 2. Insurance Transactions Analysis

- Karnataka and Maharashtra emerged as top states for insurance premium collection (13.7% and 11.8% respectively).
- Several smaller states and union territories showed very low participation, highlighting untapped markets.

## 3. Insurance Adoption Categorization

- States were effectively segmented into Strong and Weak Adoption zones.
- This categorization supports strategic planning for region-specific marketing and insurance product design.

## 4. Underpenetrated Markets for Expansion

- States with low market share and policy counts represent opportunities for growth.
- These findings aid in targeting new regions for insurance penetration and user acquisition.

## 5. Overall User Growth & Market Share

- States like Kerala, Telangana, and West Bengal showed a strong mix of growth rate and market share.
- A 3-tier adoption model (High, Moderate, Low) gave clarity on regional performance and potential.

## 6. Most Popular Transaction Types

- Peer-to-Peer payments dominated transaction value and volume, while Merchant payments had the highest transaction count.
- Financial services and bill payments showed solid average transaction values, suggesting trust in digital platforms for high-value services.



# Future Enhancements & Recommendations

## **Drill-Down Dashboards:**

- Develop interactive dashboards using Streamlit or Power BI with filters for year, quarter, state, and transaction type.
- Add heatmaps and trendlines to monitor seasonal shifts and growth hotspots.

## **Predictive Analytics:**

- Implement models to forecast user growth, insurance demand, or transaction volume based on historical trends.

## **Geo-Mapping with Insights:**

- Enhance visuals by integrating geo-spatial data to show adoption and transaction strength on Indian maps.

## **User Segmentation:**

- Combine app usage, transaction patterns, and device type to create customer personas for targeted marketing.

## **Insurance Awareness Campaigns:**

- Launch campaigns in low-adoption regions, especially North-Eastern states and UTs, with digital onboarding strategies.

## **Competitor Benchmarking:**

- Include comparative analysis with other apps to see how PhonePe stands in terms of market share and adoption.

## **Real-Time Monitoring:**

- Incorporate live APIs (if available) for real-time monitoring of usage metrics and fraud detection alerts.



**Thank You**

