PHONEPE TRANSACTION INSIGHTS

01

Introduction

This dataset shows how PhonePe is used across India, including data on users, insurance, and transactions. It is divided into three types:

Aggregated: Gives an overall summary of how many transactions or users there were and the amounts involved.

Map: Shows the data linked to specific places, including their location coordinates, to see where the activity happened.

Top: Lists the top-performing states, districts, or pincodes with the highest numbers.

The insurance data covers the years 2020 to 2024, and the user and transaction data starts from 2018. The data originally came in separate folders and JSON files, which were cleaned and converted to CSV files for easier analysis.

02 | Objective

✅ To extract, clean, and transform the provided data from its folder and JSON structures into structured CSV datasets.

✅ To analyze and draw insights from the data regarding PhonePe's growth, user distribution, insurance adoption, and transaction patterns across India.

✅ To build an interactive Streamlit dashboard showcasing key findings and visualizations to make the data easily explorable and actionable.
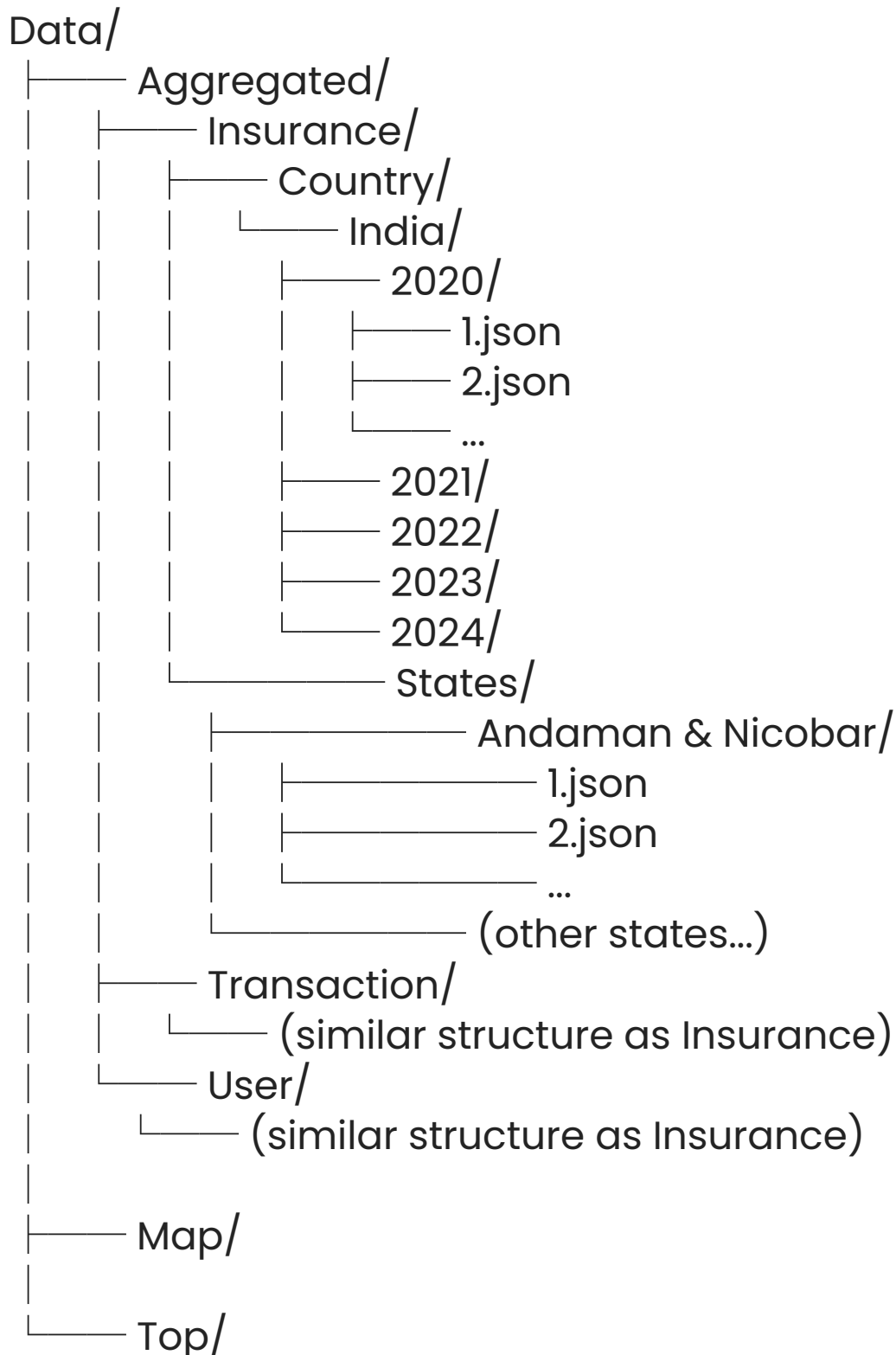
03 | Sources

`🔗Github: [Click Here]

`🔗Case Study: [Click Here]

`🔗Dataset: [Click Here]

04 | **Data Structure**

```
Data/
├────── Aggregated/
│    ├────── Insurance/
│    │    ├────── Country/
│    │    │    └────── India/
│    │    │         ├────── 2020/
│    │    │         │    ├────── 1.json
│    │    │         │    ├────── 2.json
│    │    │         │    └────── ...
│    │    │         ├────── 2021/
│    │    │         ├────── 2022/
│    │    │         ├────── 2023/
│    │    │         └────── 2024/
│    │    └────── States/
│    │         ├────────────── Andaman & Nicobar/
│    │         │    ├────────── 1.json
│    │         │    ├────────── 2.json
│    │         │    └────────── ...
│    │         └────────────── (other states...)
│    ├────── Transaction/
│    │    └────── (similar structure as Insurance)
│    └────── User/
│         └────── (similar structure as Insurance)
│
├────── Map/
│
└────── Top/
```

04 | New Data Structure

```
Data/
├────── Insurance/
│     ├────── Aggregated/
│     │     ├────── Country/
│     │     │     └────── 2020/
│     │     │           ├────── 1.json
│     │     │           ├────── 2.json
│     │     │           └────── ...
│     │     │     ├────── 2021/
│     │     │     ├────── 2022/
│     │     │     ├────── 2023/
│     │     │     └────── 2024/
│     │     └────── States/
│     │           ├────── Andaman & Nicobar/
│     │           │     ├────── 1.json
│     │           │     ├────── 2.json
│     │           │     └────── ...
│     │           └────── (other states...)
│     ├────── Map/
│     │     └────── (similar structure as aggregated)
│     └────── Top/
│           └────── (similar structure as aggregated)
│
├────── Transaction/
│
└────── User/
```
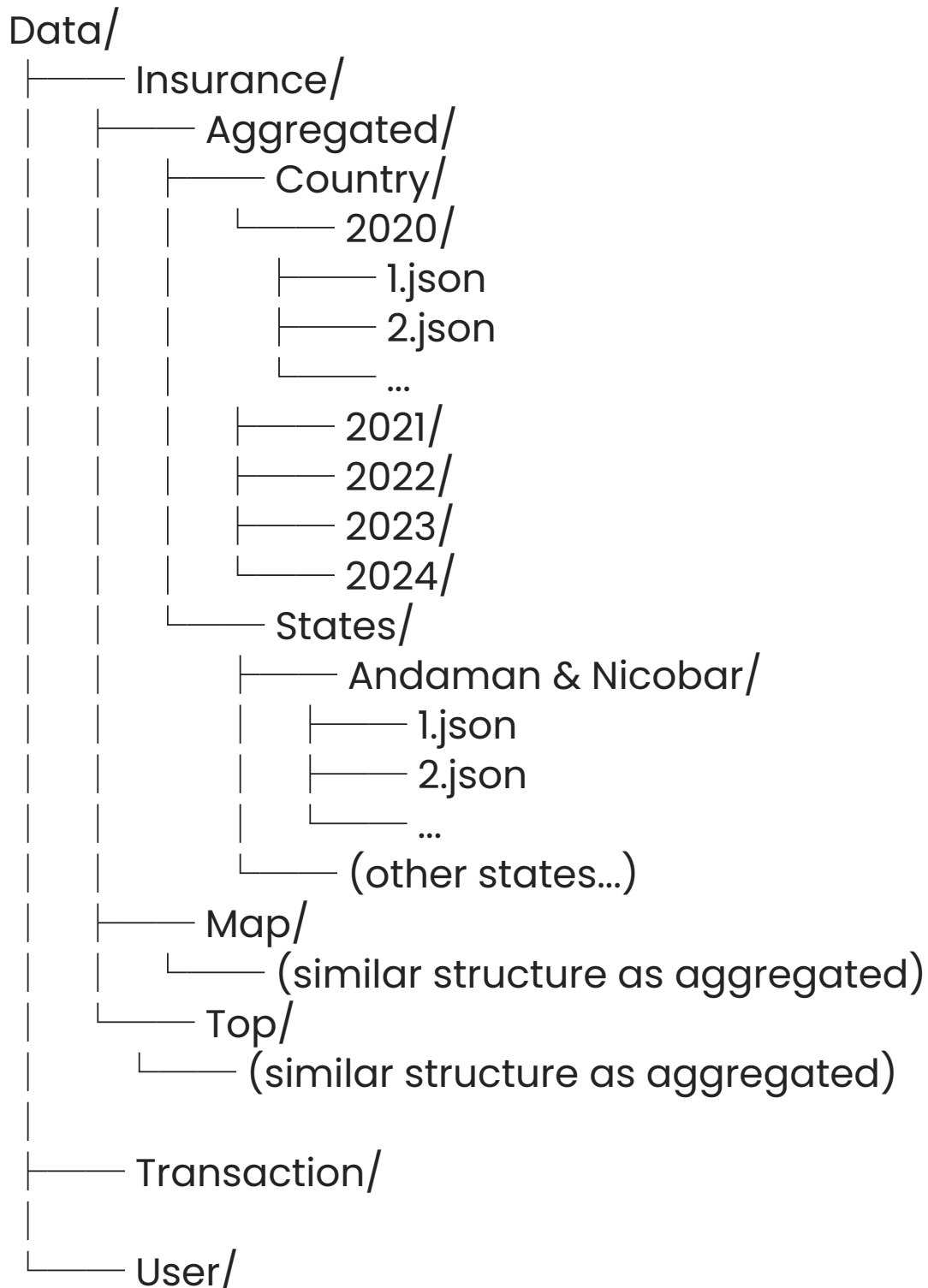
05 | Data Conversion

The entire dataset was converted from nested JSON files into structured CSV format using Python libraries such as json, os, datetime, and pandas. During this process, relevant folder and file names were also extracted to capture important contextual information like year, state, or level, and were added as new columns to the final datasets. This approach ensured that all metadata embedded in the directory structure was preserved for accurate analysis.

06 | Links

`🔗New Data: [Click Here](#)

`🔗Python Notebook: [Click Here](#)

`🔗SQL Table Query: [Click Here](#)

01 | About

The Insurance dataset consists of three main subfolders:

- Aggregated
- Map
- Top.

Each of these contains data at both country and state levels. The original data was provided in nested JSON files organized by year and region. During preprocessing, these files were converted into CSV format, and columns like state and year were extracted directly from file paths and names to enrich the dataset. Overall, this dataset captures the insurance activity conducted through PhonePe from 2020 to 2024, providing insights into regional trends, user behavior, and high-performing segments across India.

Links

`🔗Dataset: Click Here

`🔗Python NoteBook: Click Here

02 | Aggregated Charts
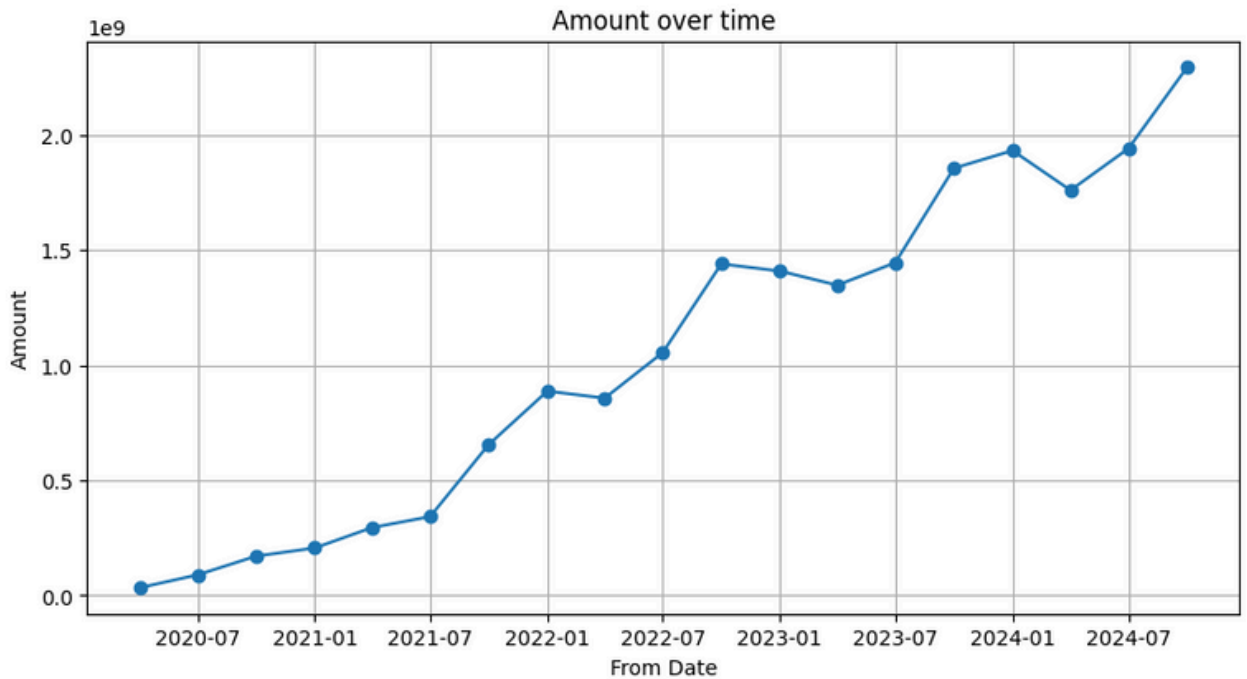
- This time series graph displays the aggregated transaction amounts reported quarterly over the period from 2020 to 2024.
- The plot highlights a consistent upward trend, suggesting strong growth in these financial transactions.

02 | Aggregated Charts

Average Amount per Transaction by Year
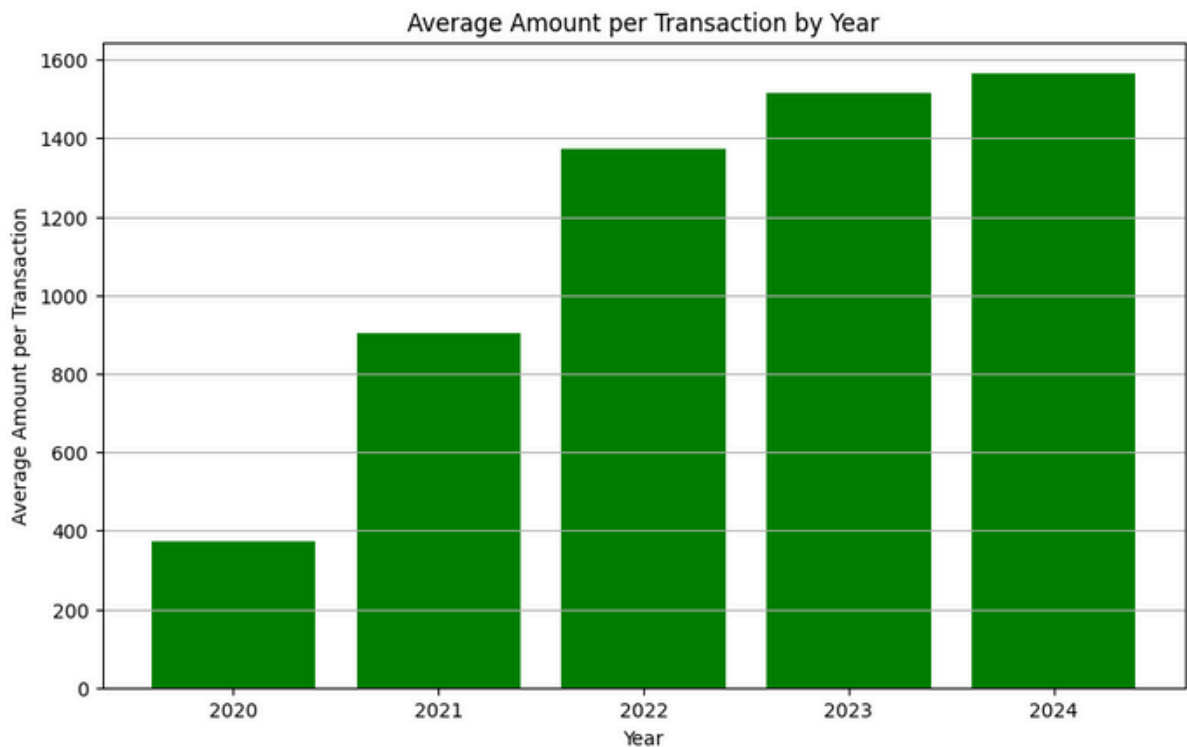
- This bar plot shows the average amount per insurance transaction at the country level from 2020 to 2024.
- The average amount increased steadily from just below ₹400 in 2020 to nearly ₹1600 in 2024.
- This trend highlights clear growth in the value of insurance transactions over these years.

02 Aggregated Charts

Total Insurance Transaction Amount by State (2020-2024)
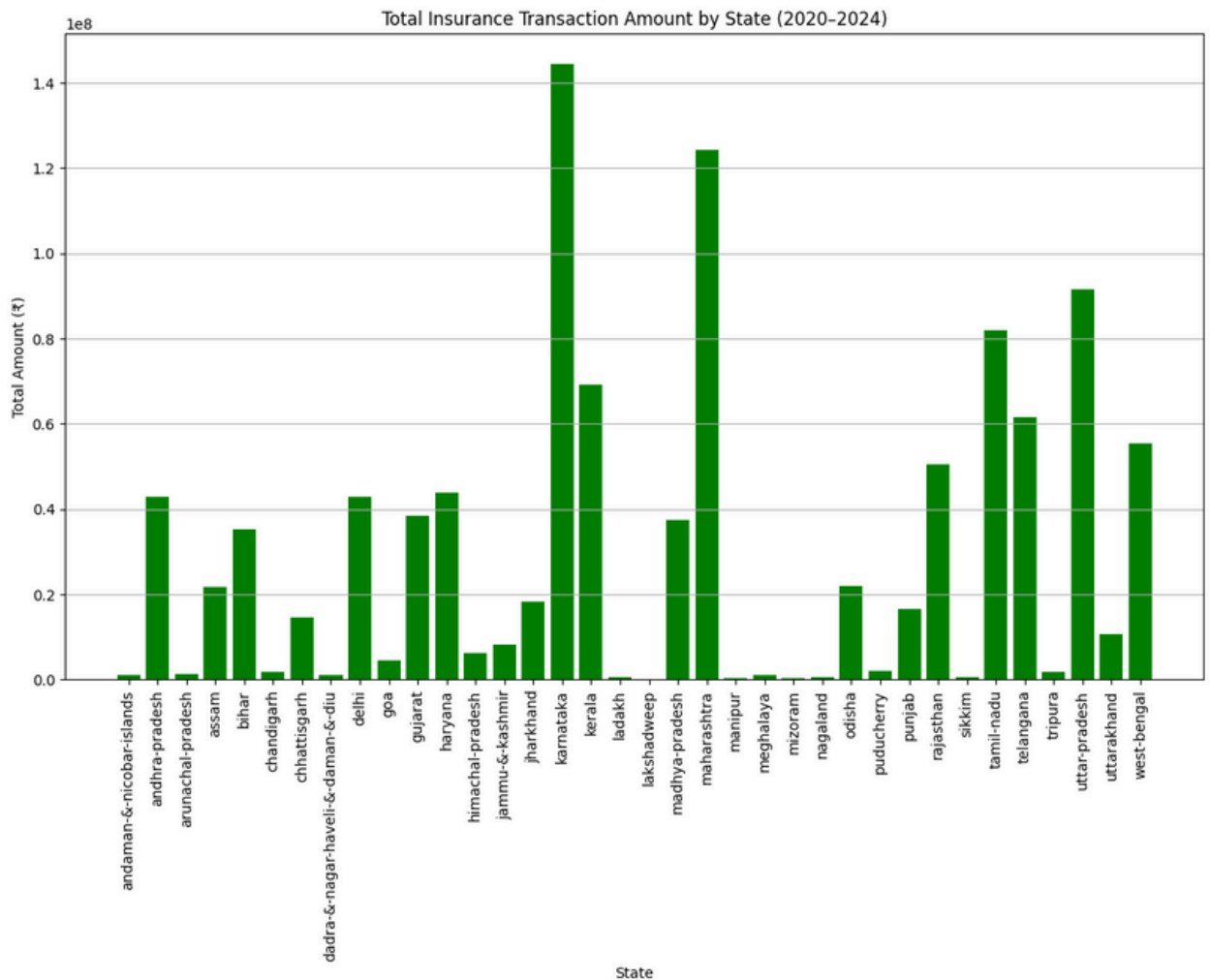
- Karnataka and Maharashtra not only had more transactions but also had higher average insurance amounts per transaction.
- This means people there are choosing bigger insurance plans.
- In contrast, states like Ladakh, Mizoram, and Nagaland had much smaller average amounts — almost negligible in comparison.

02 | Aggregated Charts
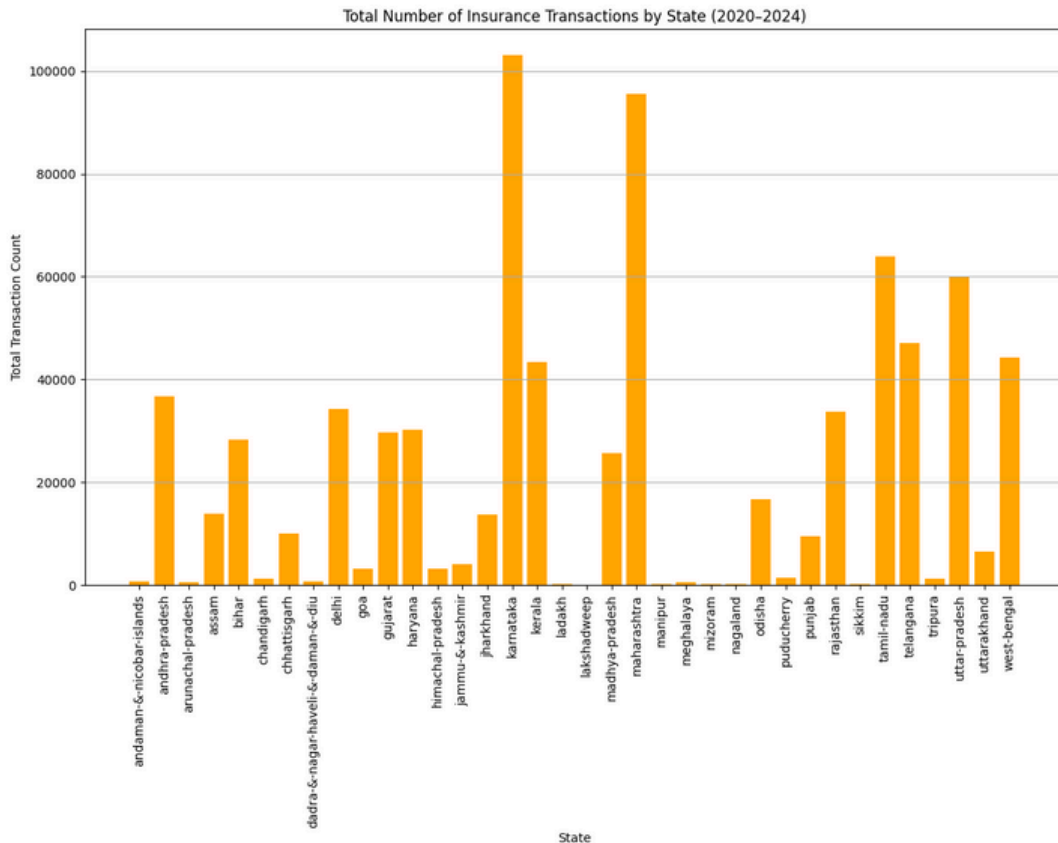
- Karnataka and Maharashtra had the highest number of insurance transactions.
- These states clearly saw more people using PhonePe for insurance.
- Tamil Nadu and Uttar Pradesh had a moderate count, while smaller states like Ladakh and Sikkim had very few — barely visible on the chart.

03 | Map Charts

State with Highest Insurance Amount by Year
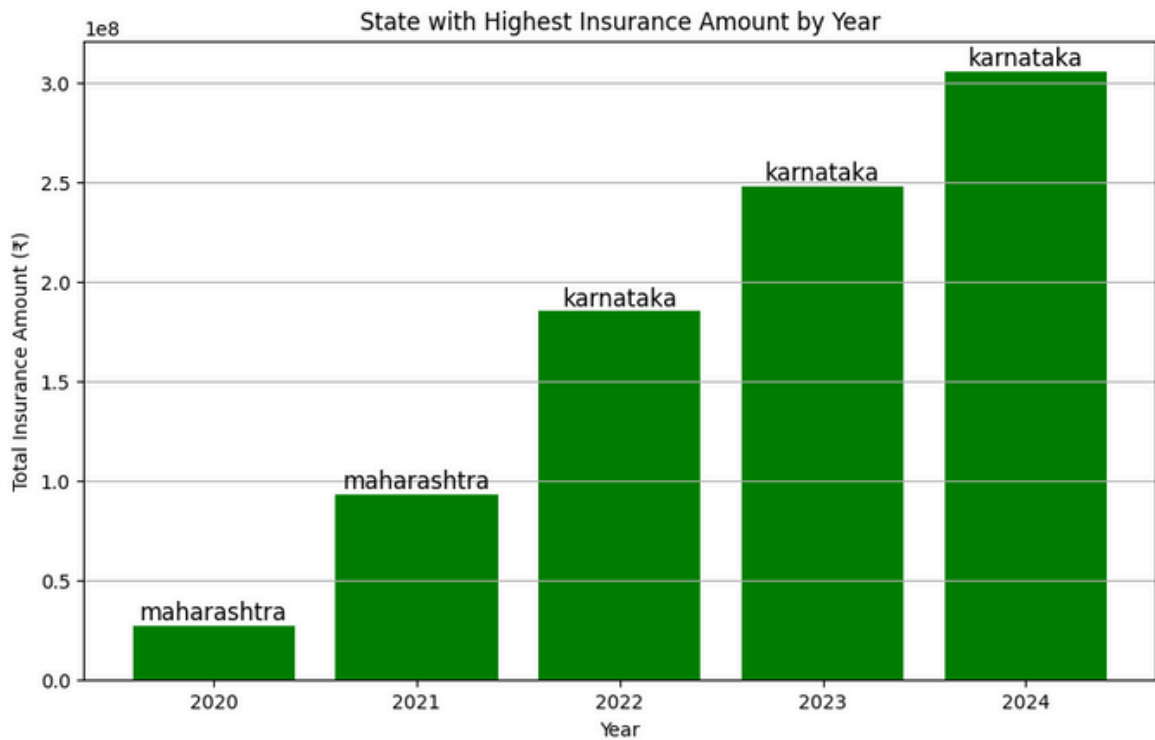
- Karnataka and Maharashtra led in total insurance amount.
- Karnataka started rising in 2022 and overtook Maharashtra in the later years.

INSURANCE

03 | Map Charts

State with Lowest Insurance Amount by Year
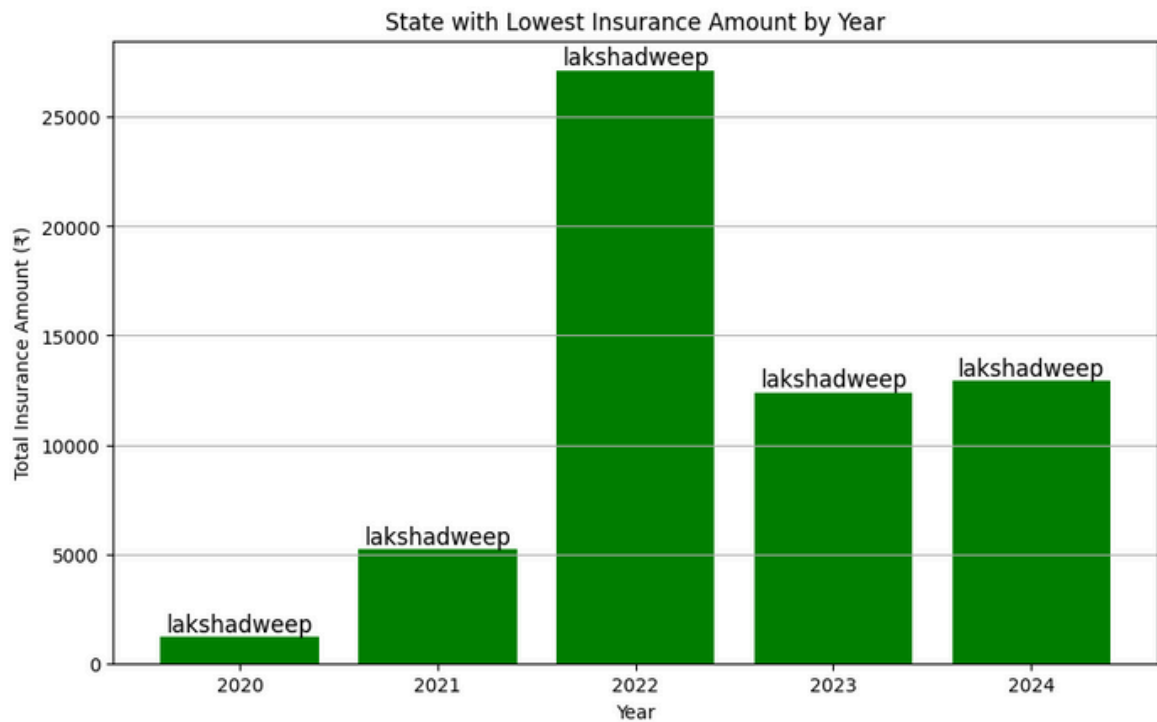
- Lakshadweep has been the last since 2020
- Its Peak is just above 25k & didn't achieve it after 2022
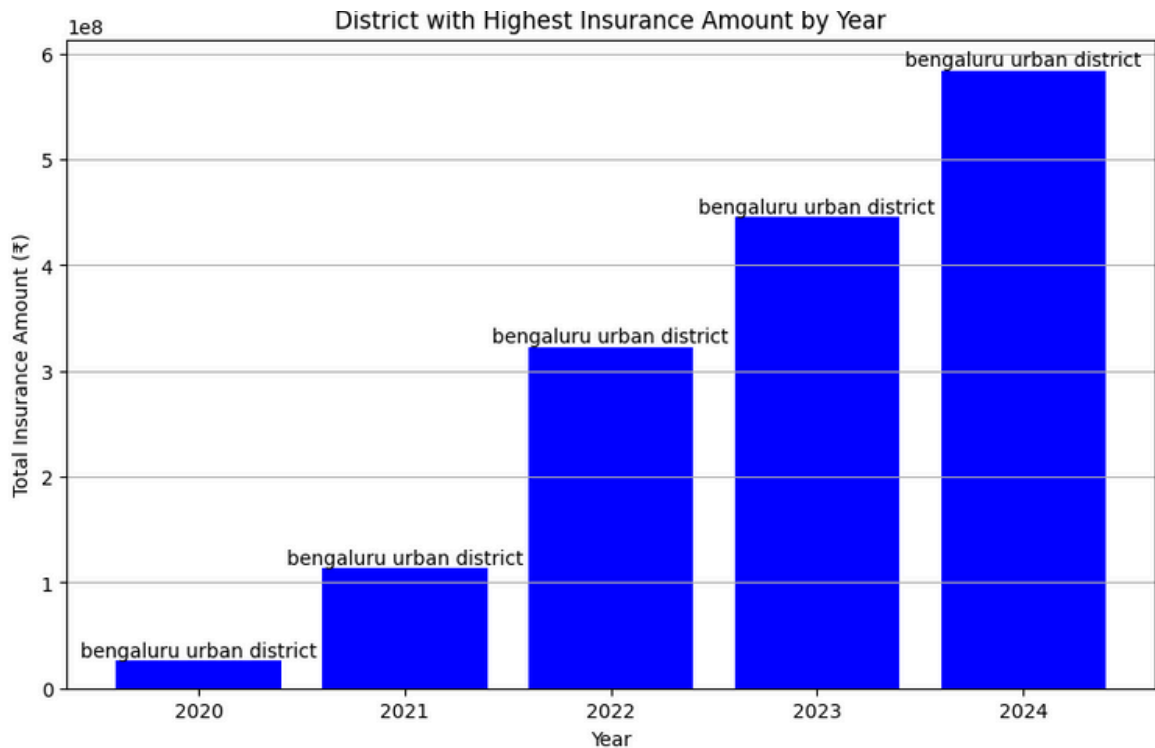
03 | Map Charts

- Bengaluru Urban district held the top spot for insurance amount in every year.
- It's consistently the highest-performing district in insurance.

03 | Map Charts

District with Lowest Insurance Amount by Year
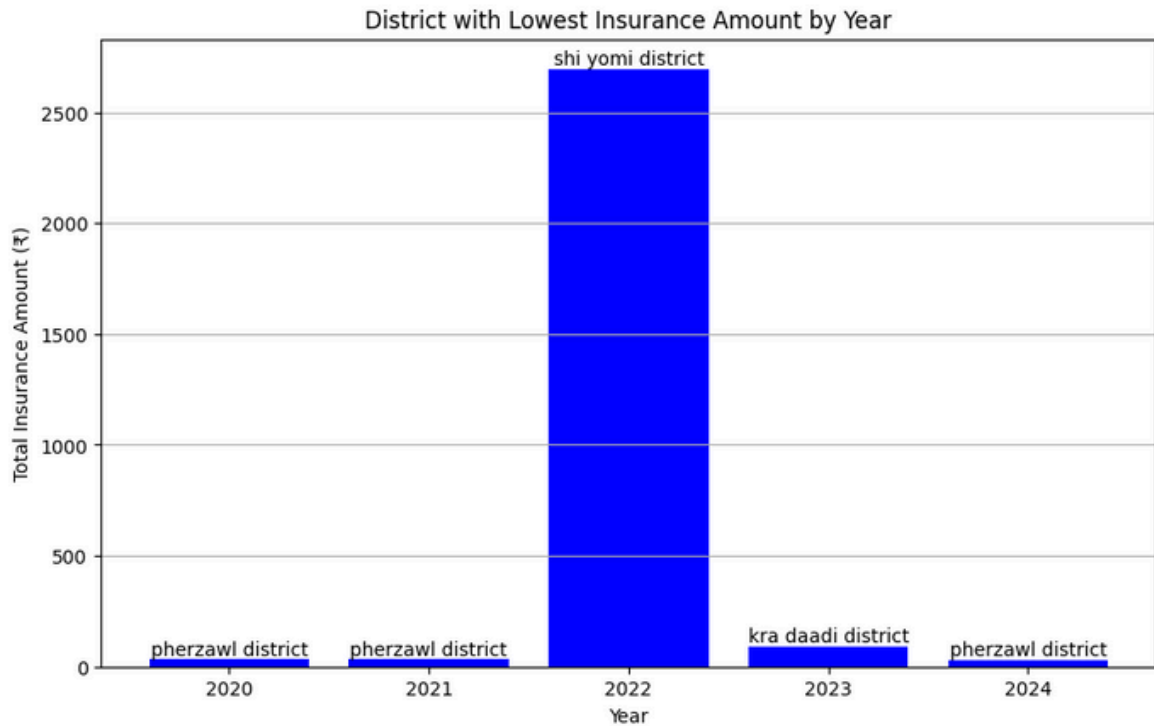
- `Pherzawl district` is the most frequently occurring lowest amount district.
- `Shi yomi district` had a peak among lowest in 2022
- 2022 Seems to be the peak age of the lowest-amount states but not for the Highest

04 | Top Charts

Top States by Total Insurance Amount
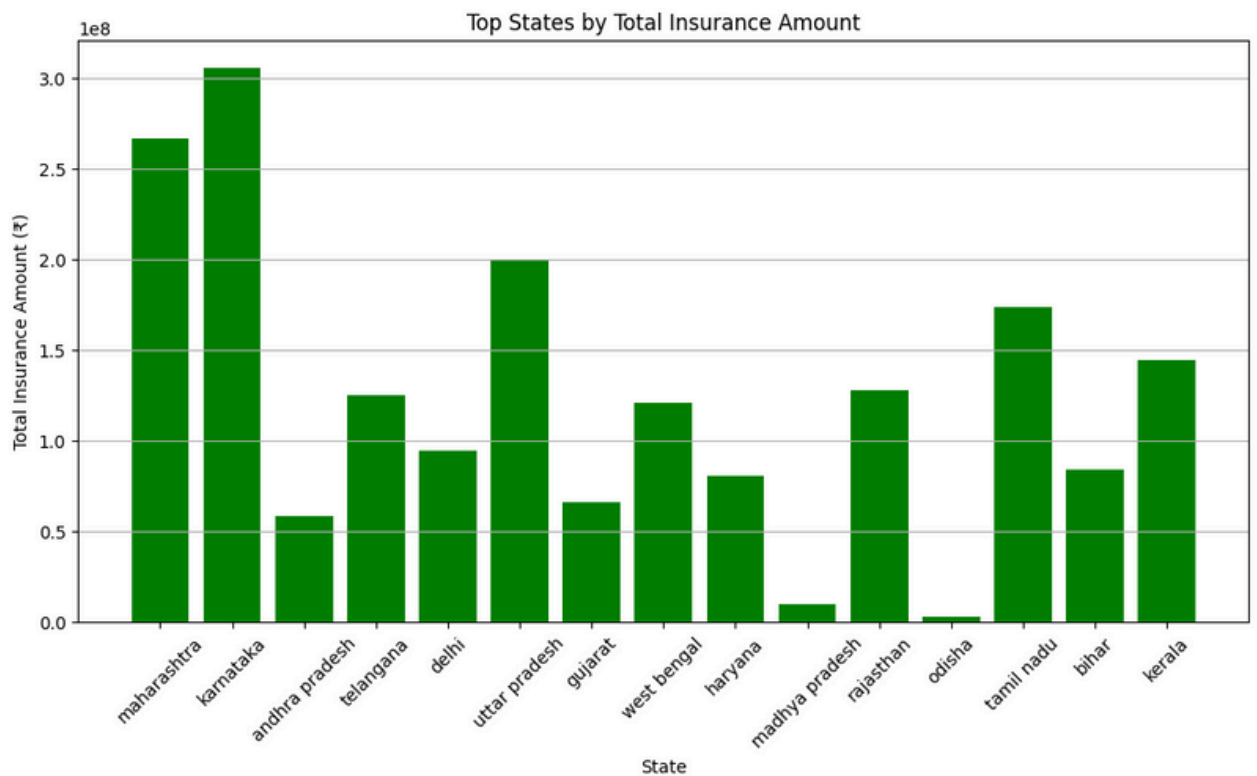
- Karnataka & Maharashtra are almost similar in state level.
- UP & Tamil Nadu are average.
- The rest are below averages.
- Odisha & Madya Pradesh are very bottom.

04 | Top Charts
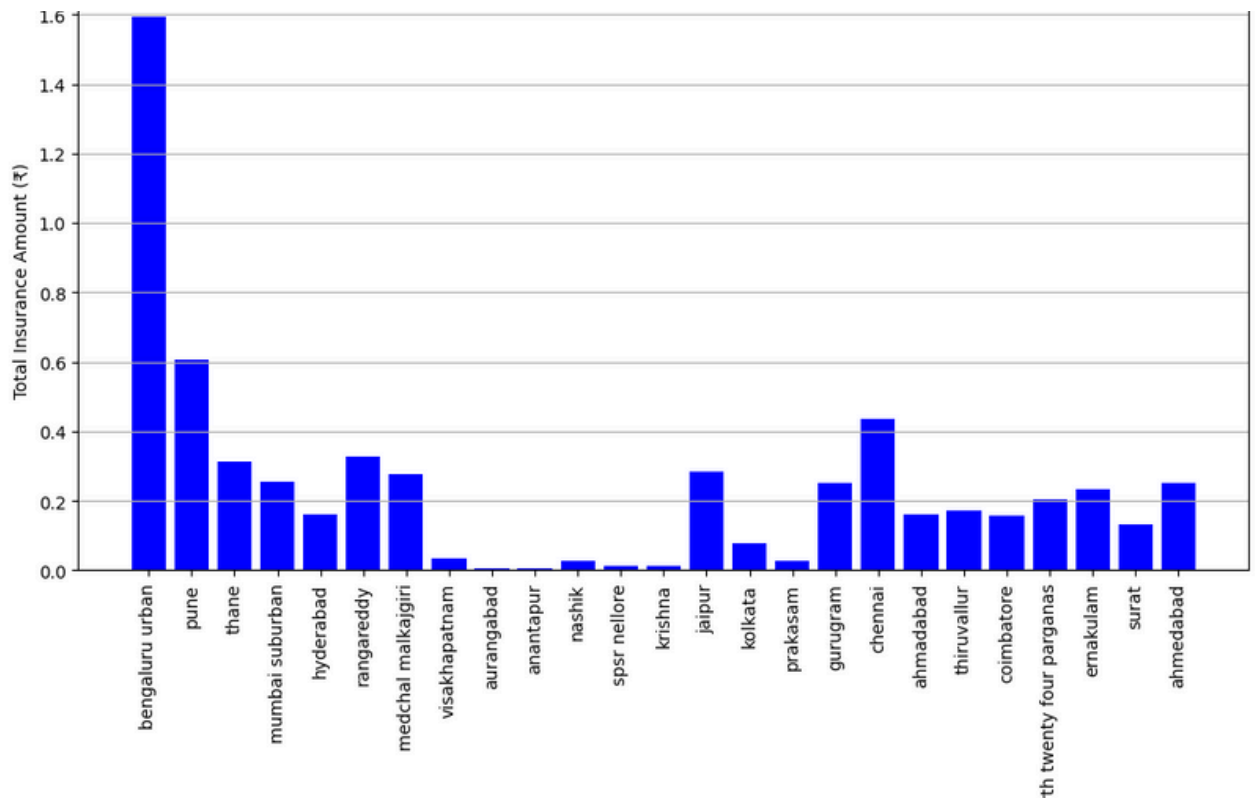
- Bengaluru urban district has a huge lead compared to others.
- Statewise, most are similar.
- Second best is Pune, which didn't even reach half of the Bengaluru urban district.

01 | About

The Transaction dataset is structured similarly to the Insurance dataset, with three main subfolders:

- Aggregated
- Map
- Top

Each subfolder contains data at both the country and state levels. The original data was provided in JSON format and was converted into separate CSV files using Python. This dataset covers PhonePe transaction activity from 2018 to 2024, offering a comprehensive view of digital payment trends. The Aggregated data provides an overall summary, the Map data highlights the geographic distribution of transactions, and the Top data identifies the highest-performing regions and categories.

Links

`🔗Dataset: [Click Here](#)

`🔗Python NoteBook: [Click Here](#)

02 | Aggregated Charts

Total Transaction Amount by Type (2020–2024)
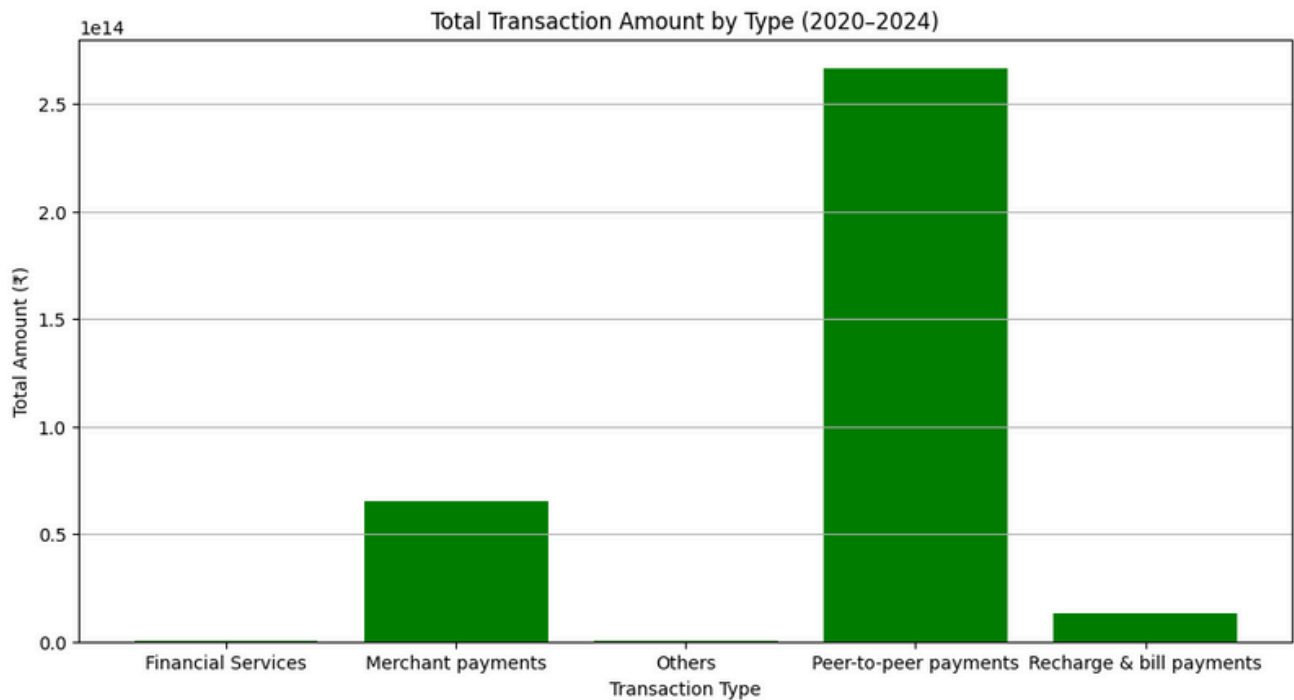
- Peer-to-peer payments were used for the highest amounts that were transferred.
- Merchant Payment is not even half of the peer-to-peer payments.
- This shows people transfer more money in person than to merchants.
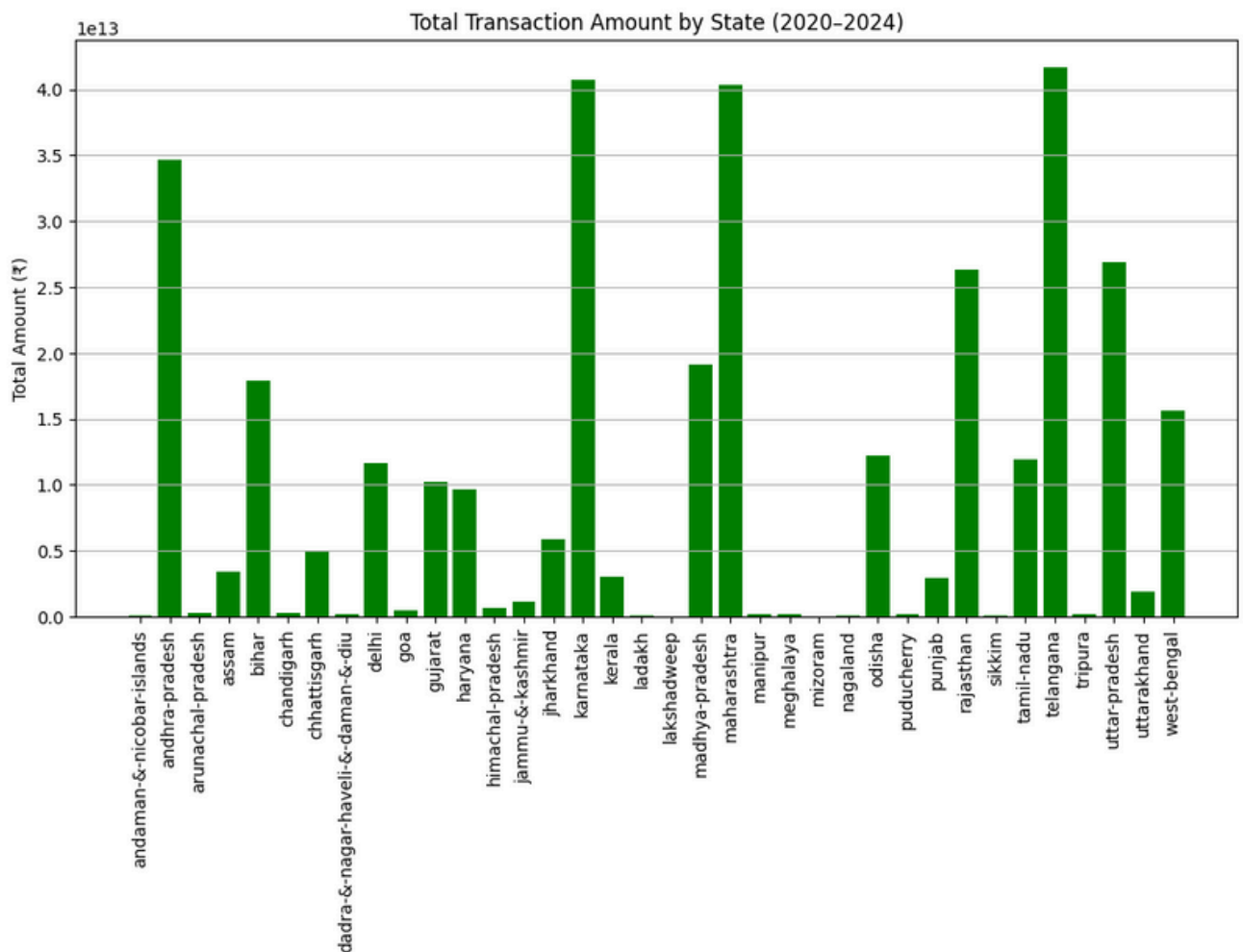
02 | Aggregated Charts

Total Transaction Count by Type (2018–2024)
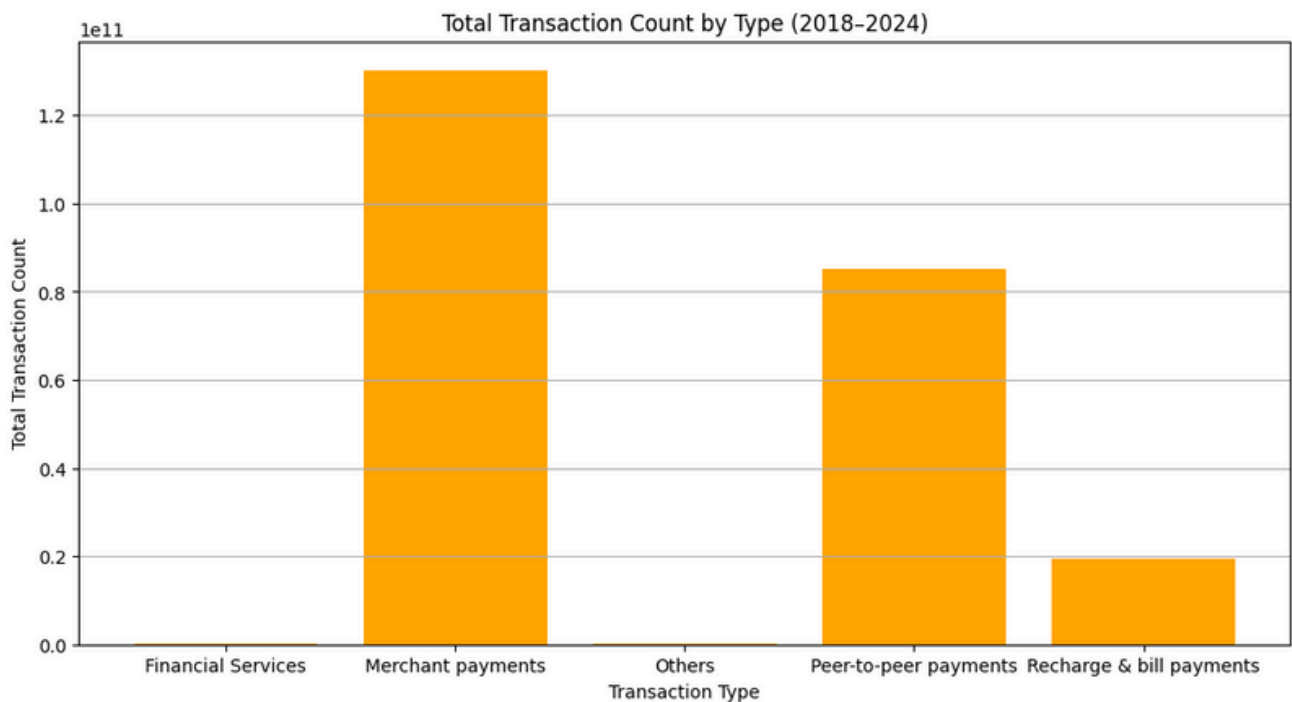
- Merchant payments count far exceeds that of peer-to-peer payments.
- This means merchant payments are done heavily but only in small amounts compared to peer-to-peer.
- This also shows people started opting for UPI payments.

02 | Aggregated Charts

Total Transaction Amount by State (2020–2024)

- Telangana has the highest Transactional Amount.
- Its wasn't high in the insurance purchase but high in transactions.
- AP has raised to the highest also.
- this reveals people of Telugana & AP spent less on Insurance.
- Maharashtra & Karnataka has maintained their position as the highest.
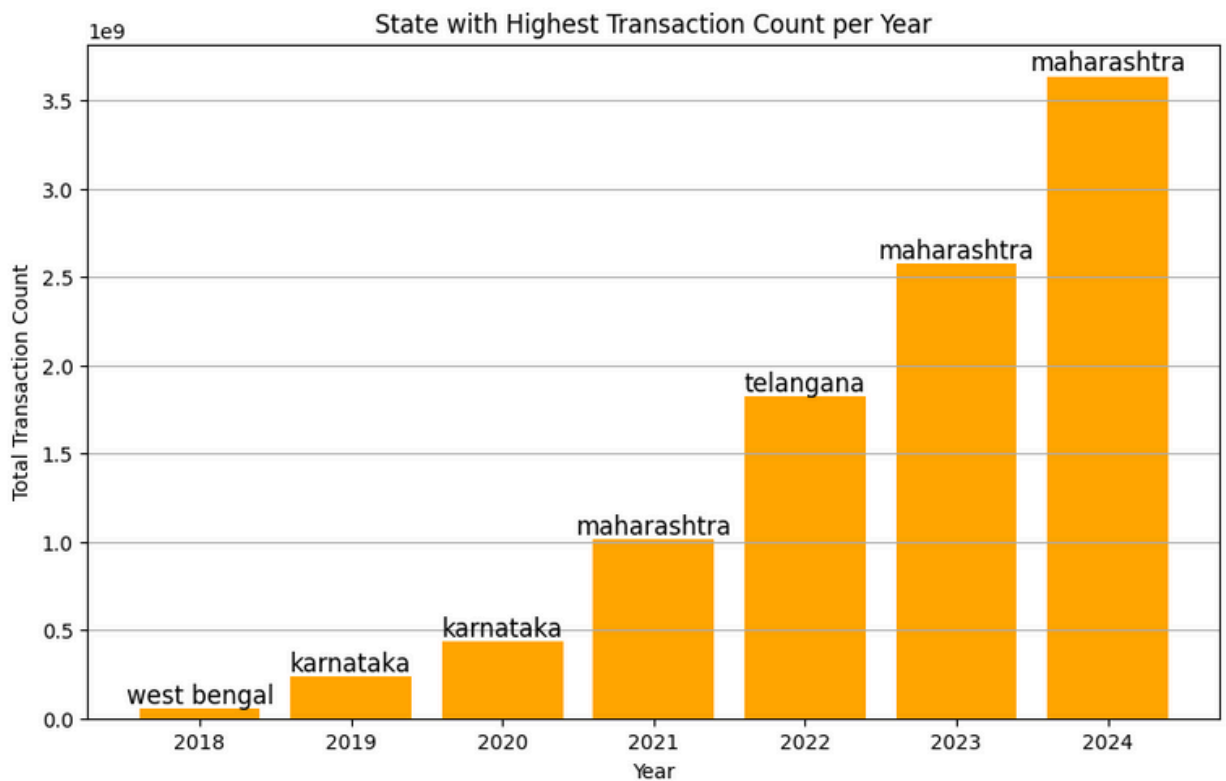- Kerala is down bottom in Transactions but held a good position in Insurance.
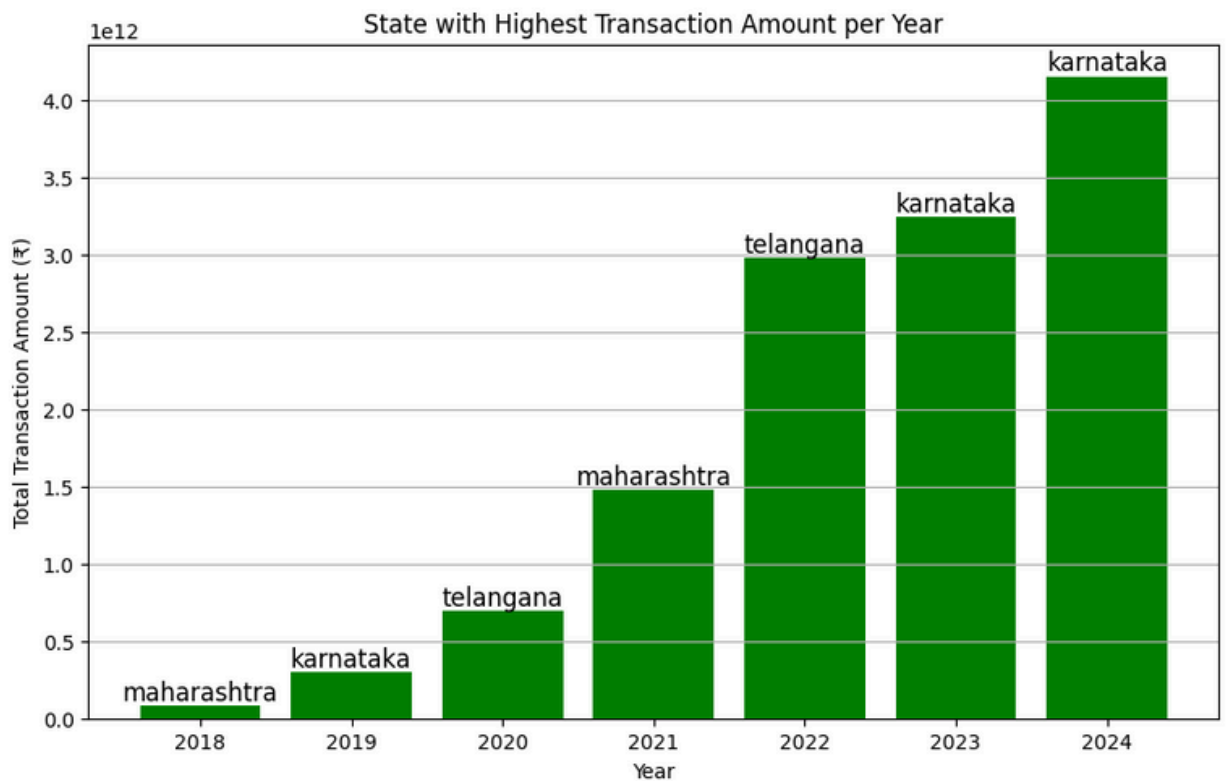
03 | Map Charts

- Maharashtra was leading at the start of 2018.
- Telangana made an appearance in 2020 & 2022.
- Karnataka overtook all from 2023 onwards.

03 | Map Charts
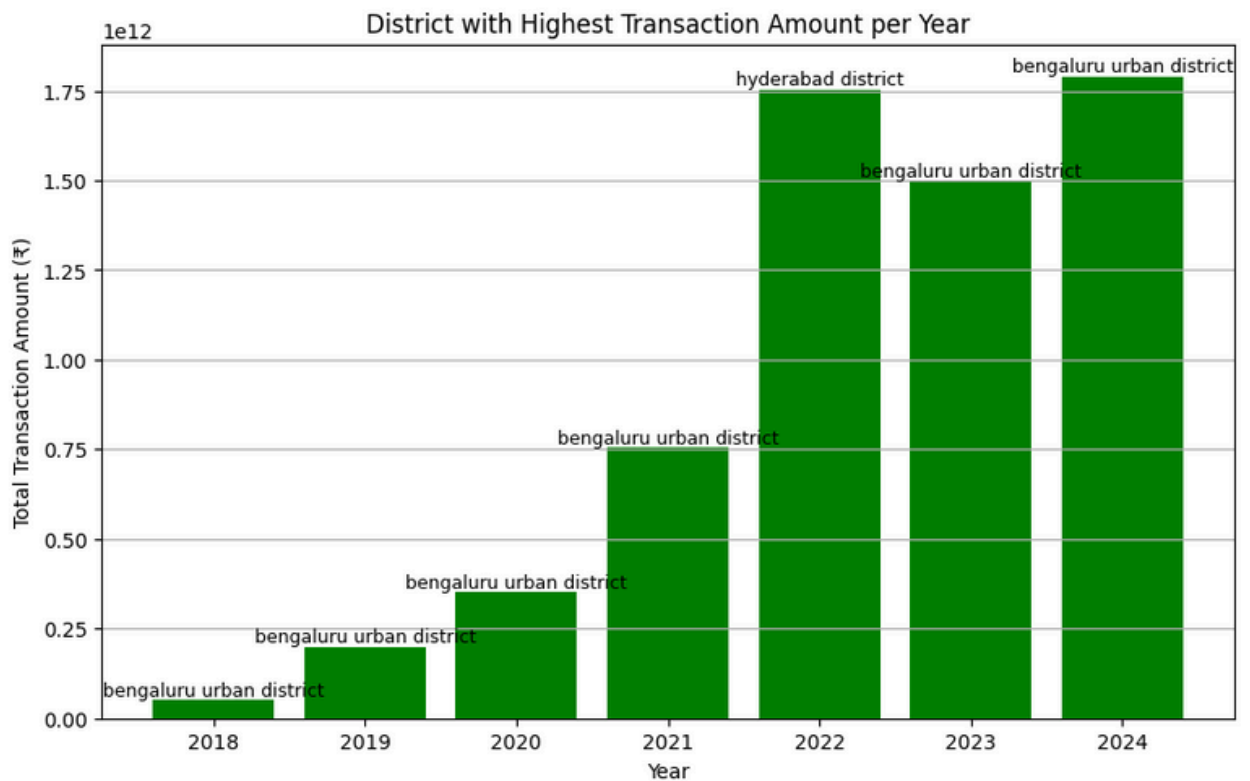

State with Highest Transaction Count per Year

- Maharashtra has the highest count record but lost to Karnataka in terms of amount in 2024.
- West Bengal showed an appearance in count in 2018 but didn't reach the amount mark.
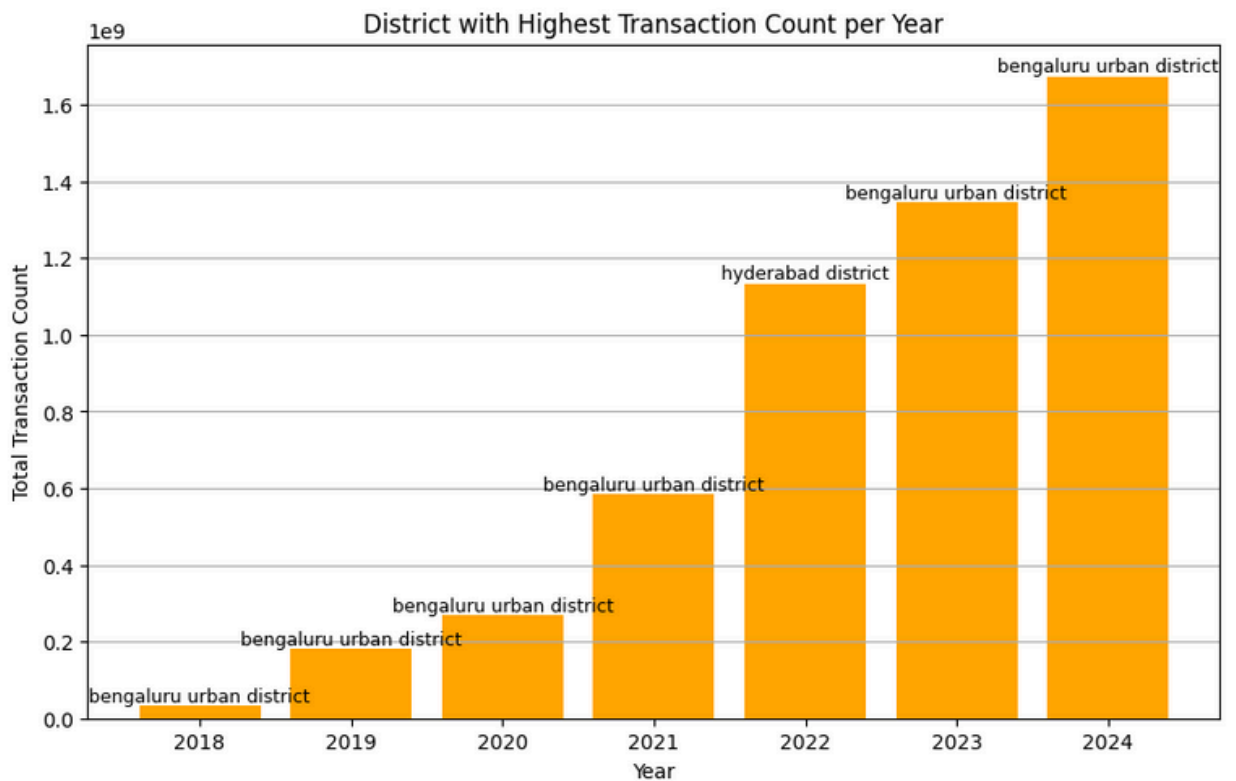- Telangana's Peak was 2022.

03 | Map Charts

- The Bengaluru urban district has held the largest amounts since 2018.
- Hyderabad district took the lead in 2022 on amounts.
- That lead was equivalent to the 2024 lead of the Bengaluru urban district.
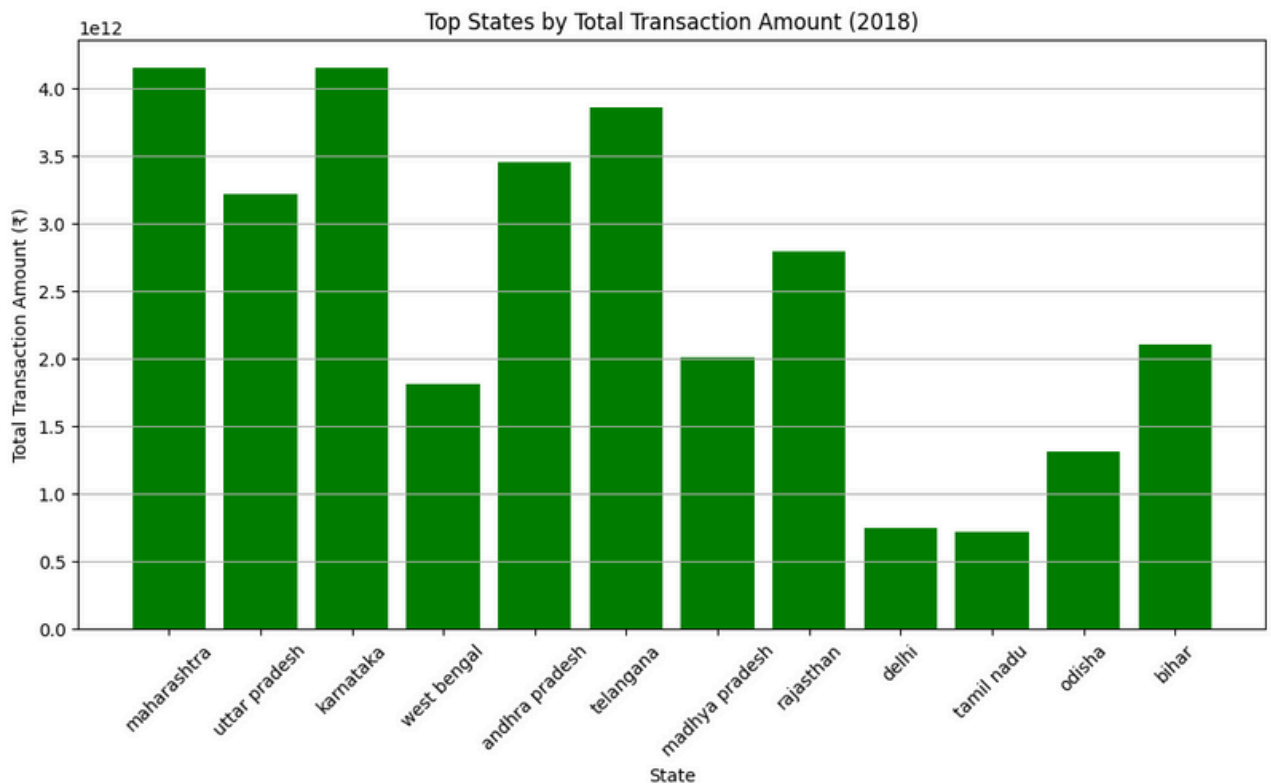- 2022 was Hyderabad district's peak

03 | **Map Charts**

District with Highest Transaction Count per Year

- Bengaluru urban district also held the largest count since 2018.
- Only lost once in 2022 to the Hyderabad district.
- The following year it took back its position on top.

04 | Top Charts
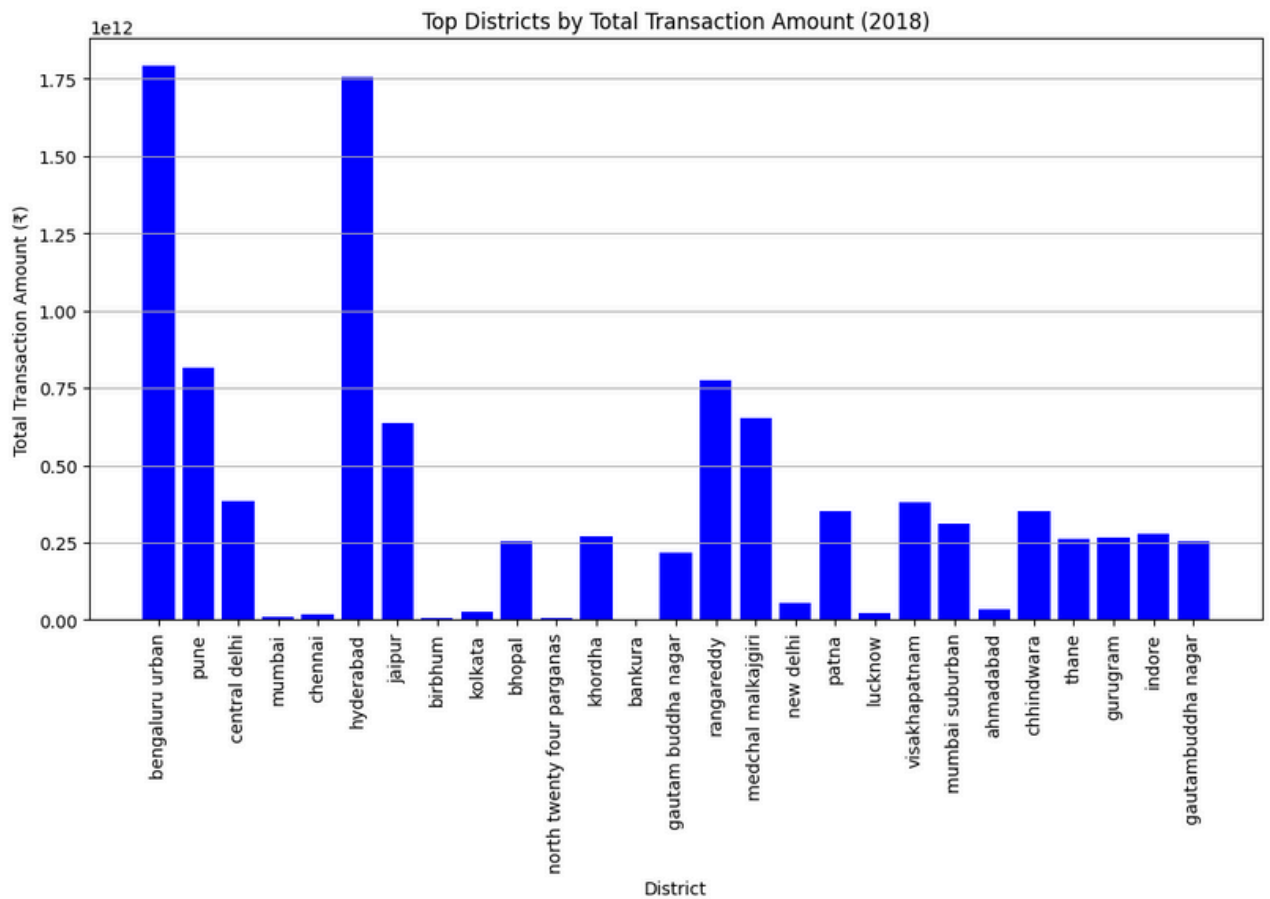

Top States by Total Transaction Amount (2018)

- Maharashtra & Karnataka are on top and on the same level.
- Telangana is right below them.
- AP & UP are on the rise.
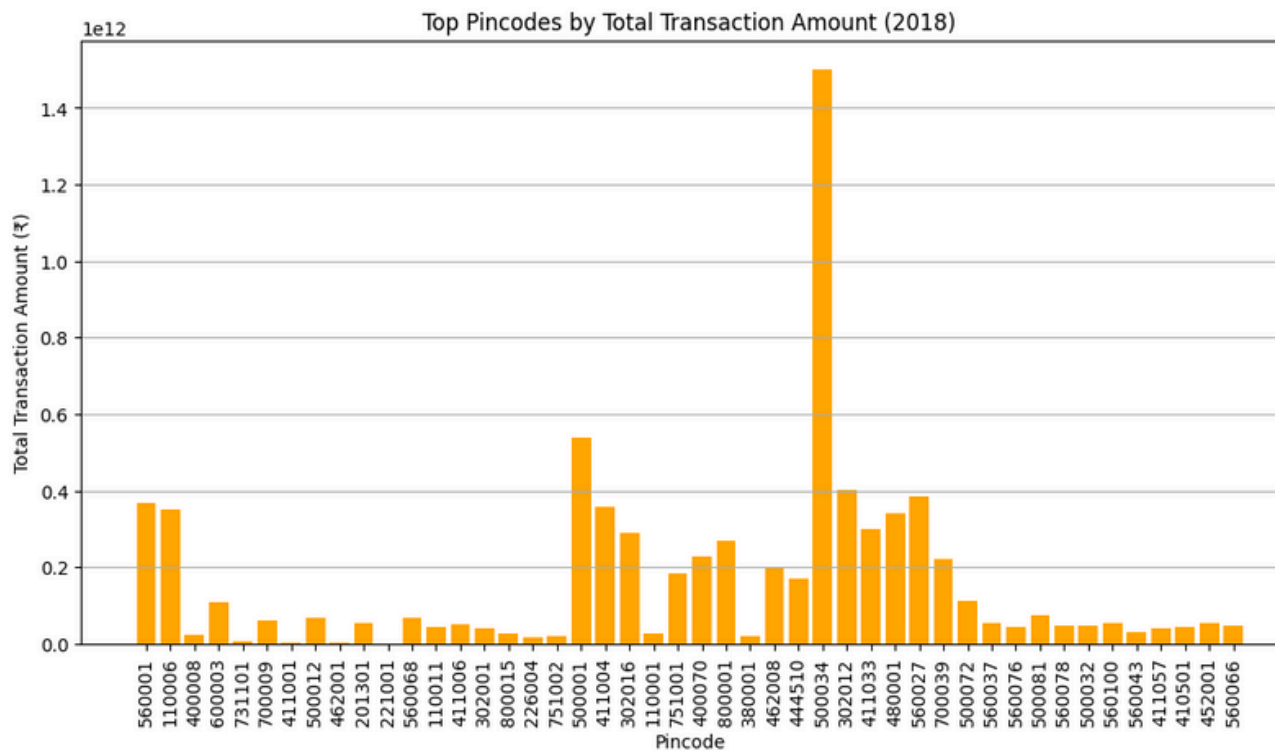- Delhi & Tamil Nadu seem to be last.

04 | Top Charts

Top Districts by Total Transaction Amount (2018)

- On the district level, Bengaluru Urban & Hyderabad are nearly on the same level.
- The rest are below their half level.

04 | Top Charts

Top Pincodes by Total Transaction Amount (2018)

- The pincode `500034` has a huge lead in transactions compared to others.
- The second-best pincode `500001`, couldn't even reach half of the first pincode.

01 | About

The User dataset also follows the same structure with three key subfolders:

- Aggregated
- Map
- Top.

Each of these contains data at both the country and state levels, organized originally in JSON format and later converted into structured CSVs using Python. This dataset captures user-related metrics on PhonePe from 2018 onward, including the number of registered users, app opens, and device brand distribution. The Aggregated section summarizes user activity over time, the Map section shows user presence across different locations, and the Top section highlights states, districts, or pincodes with the highest user counts.
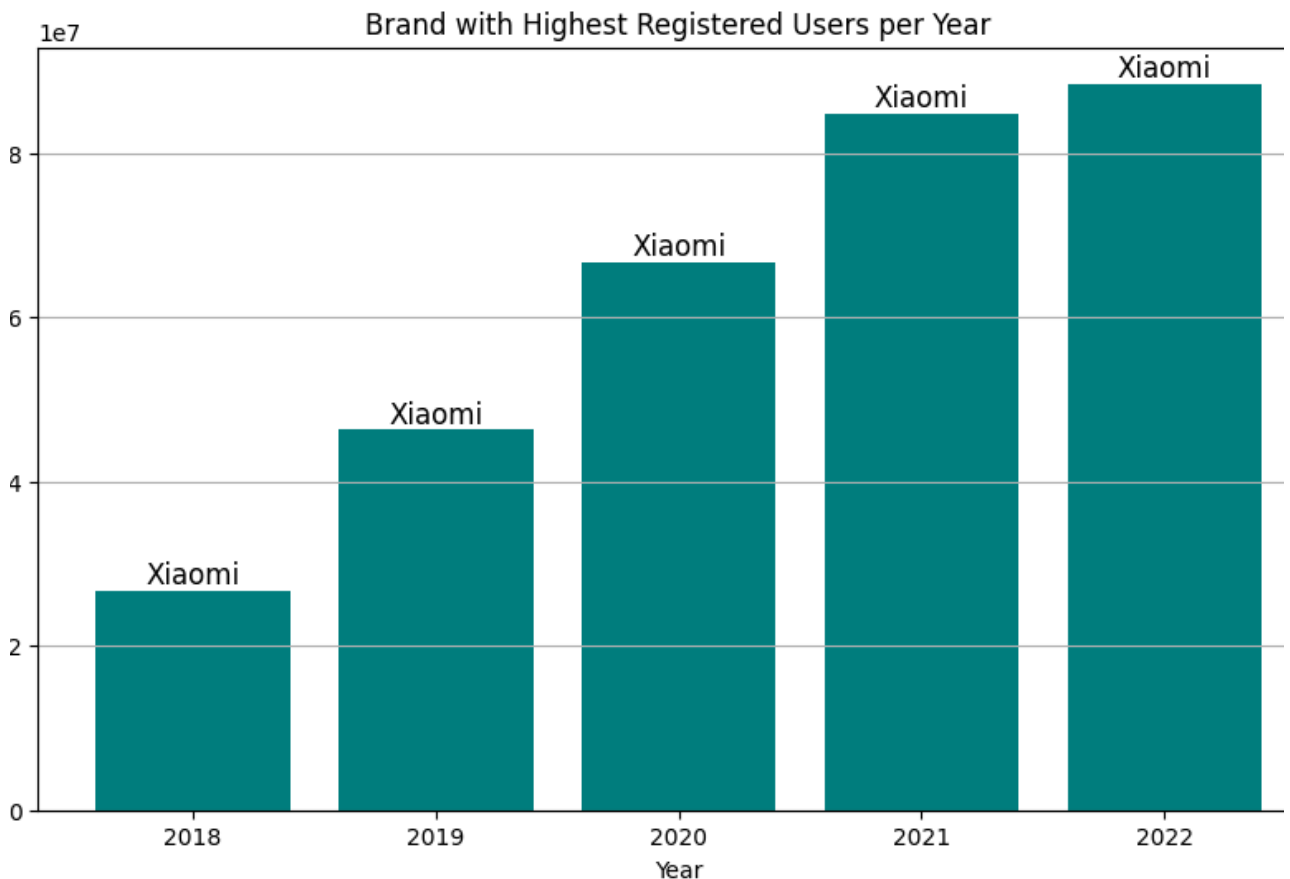
Links

`🔗Dataset: [Click Here](#)

`🔗Python NoteBook: [Click Here](#)

02 Aggregated Charts
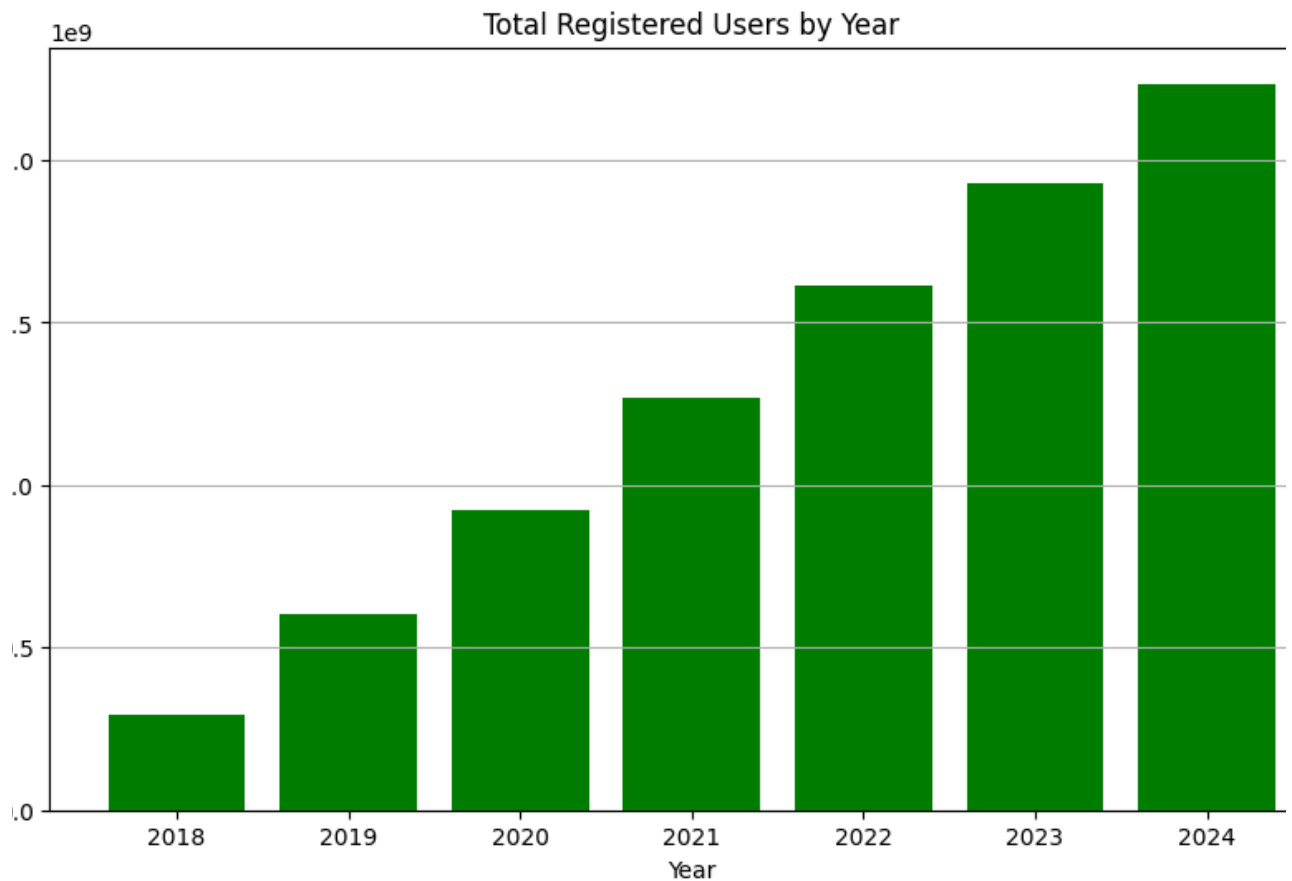

Brand with Highest Registered Users per Year

- The brand with the highest number of registered users every year is `xiaomi`.
- No other brands have been on the list at the country level till 2022.
- The Brand Specifications are only given till 2022
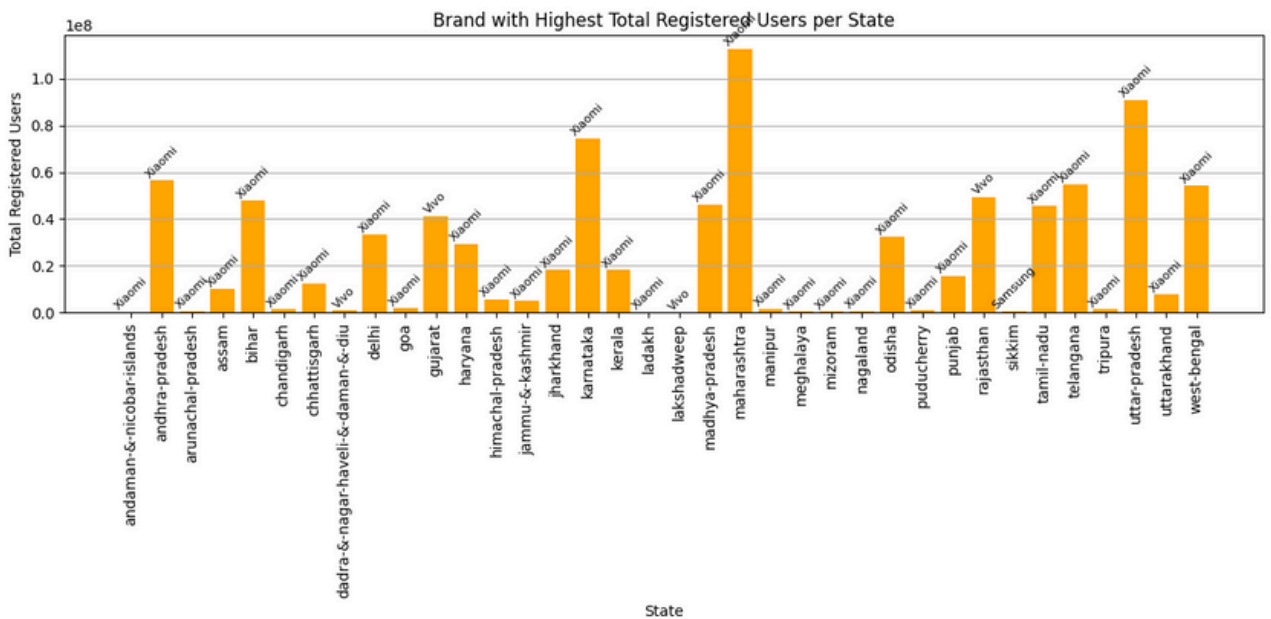
02 | Aggregated Charts

- The registered users are increasing steadily from 2018 to 2024.
- It grew from `8,08,774` to `5,86,756,452`

02 | Aggregated Charts
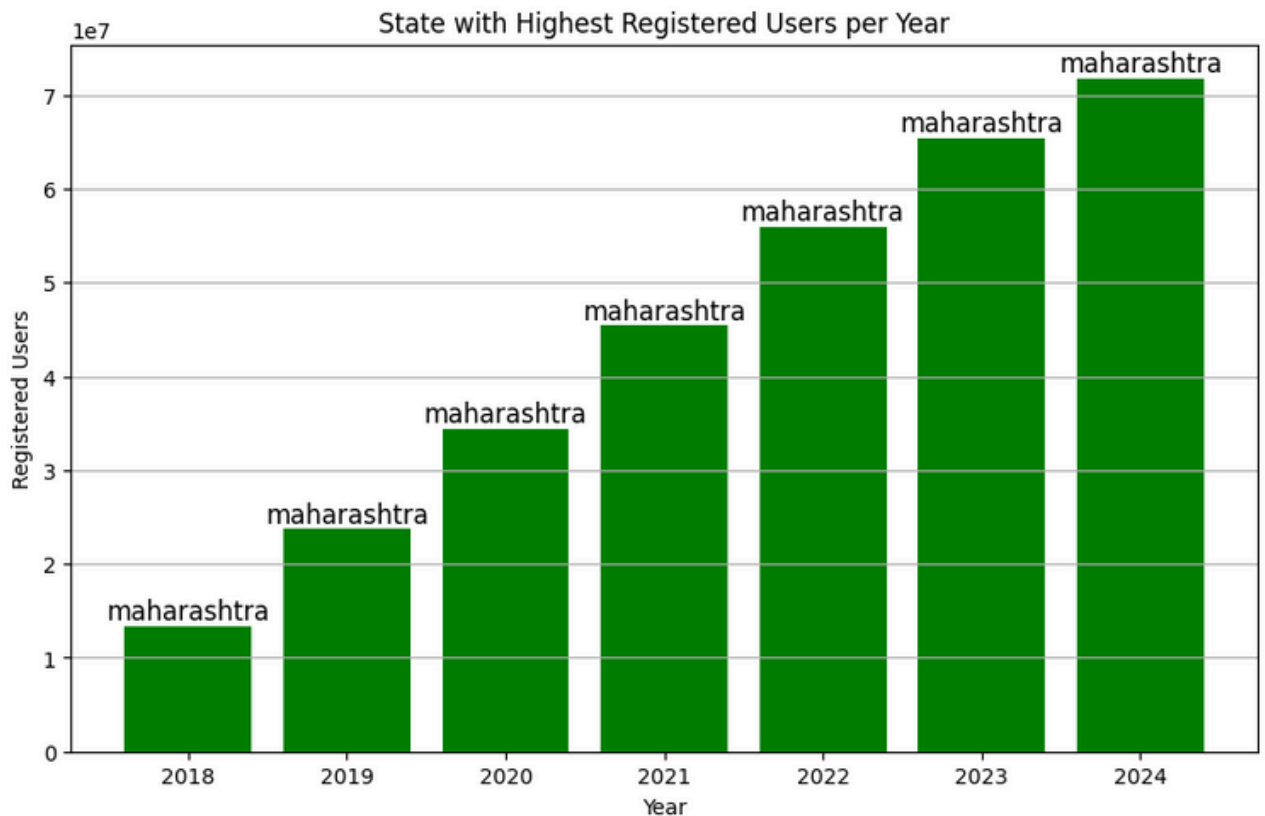

Brand with Highest Total Registered Users per State

- `Xiaomi` is the biggest brand in user registrations in India.
- Then there is `Vivo` with only 4 and a small occurrence at the state level.
- Then `Samsung` with only 1 and the lowest occurrence in Sikkim.
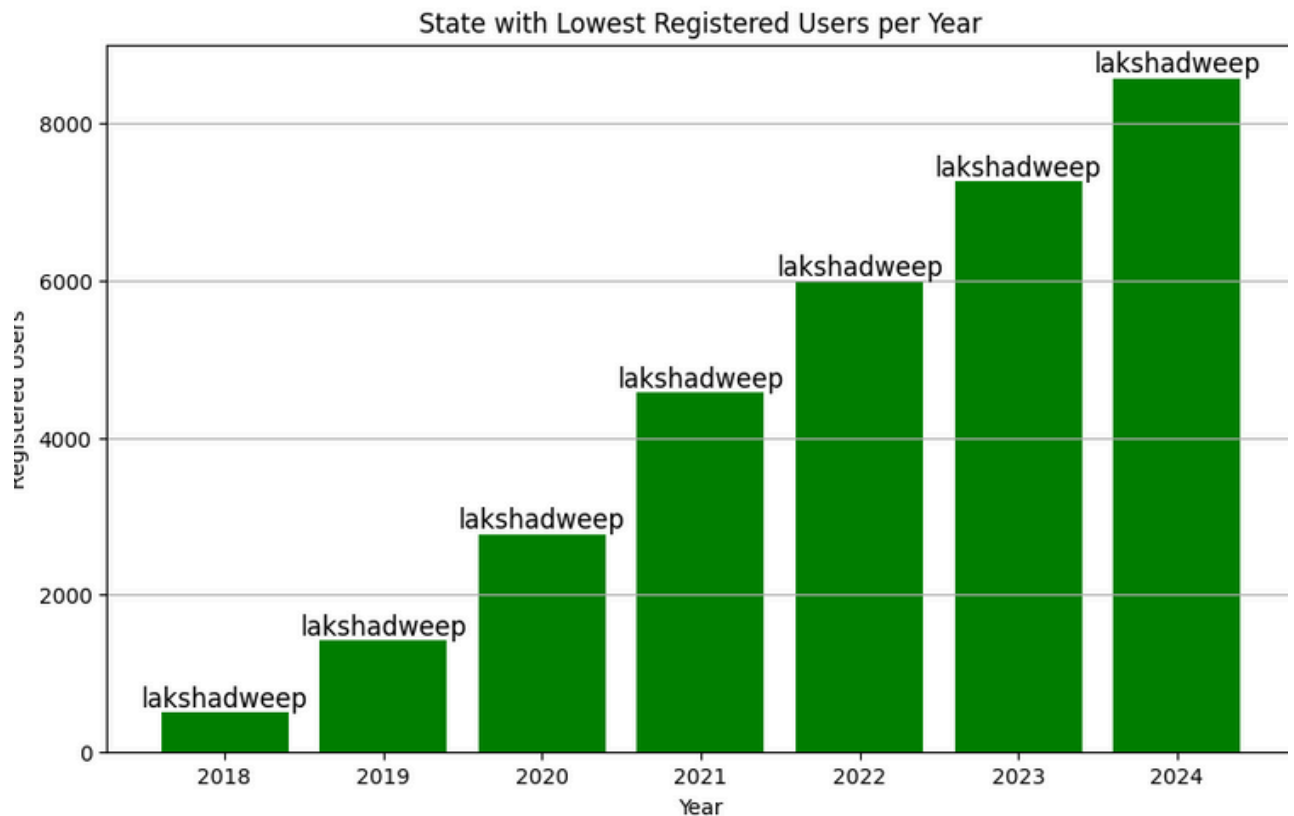
USER

03 | Map Charts

- `Maharashtra` is the state with the highest registration every year.
- Other big states haven't occurred once.

03 | Map Charts

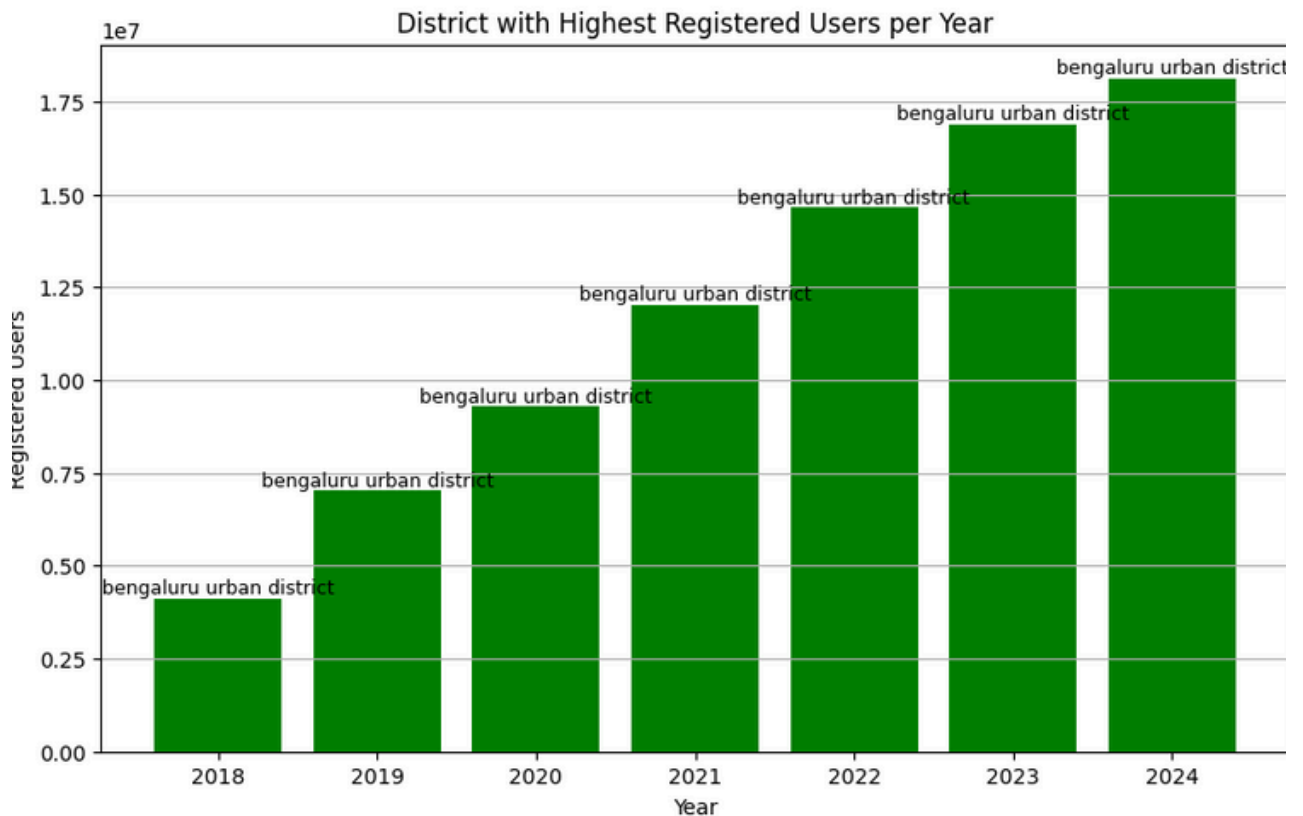State with Lowest Registered Users per Year

- `Lakshadweep` is the state with the lowest registration every year.
- No other states have occurred in this list.

03 | Map Charts

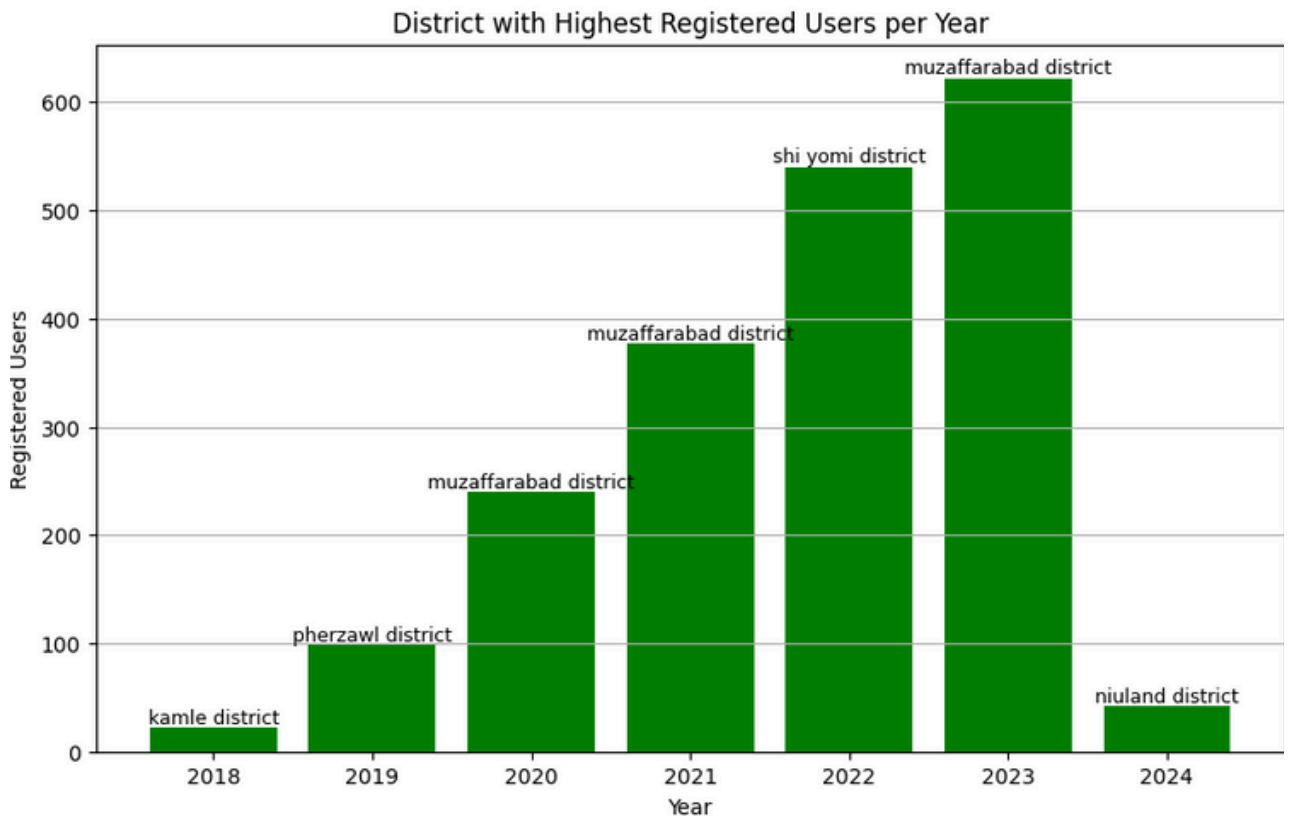
- `Bengaluru urban district` Held the Largest in Registration since 2018.
- No other districts have occurred in this list.

03 | Map Charts

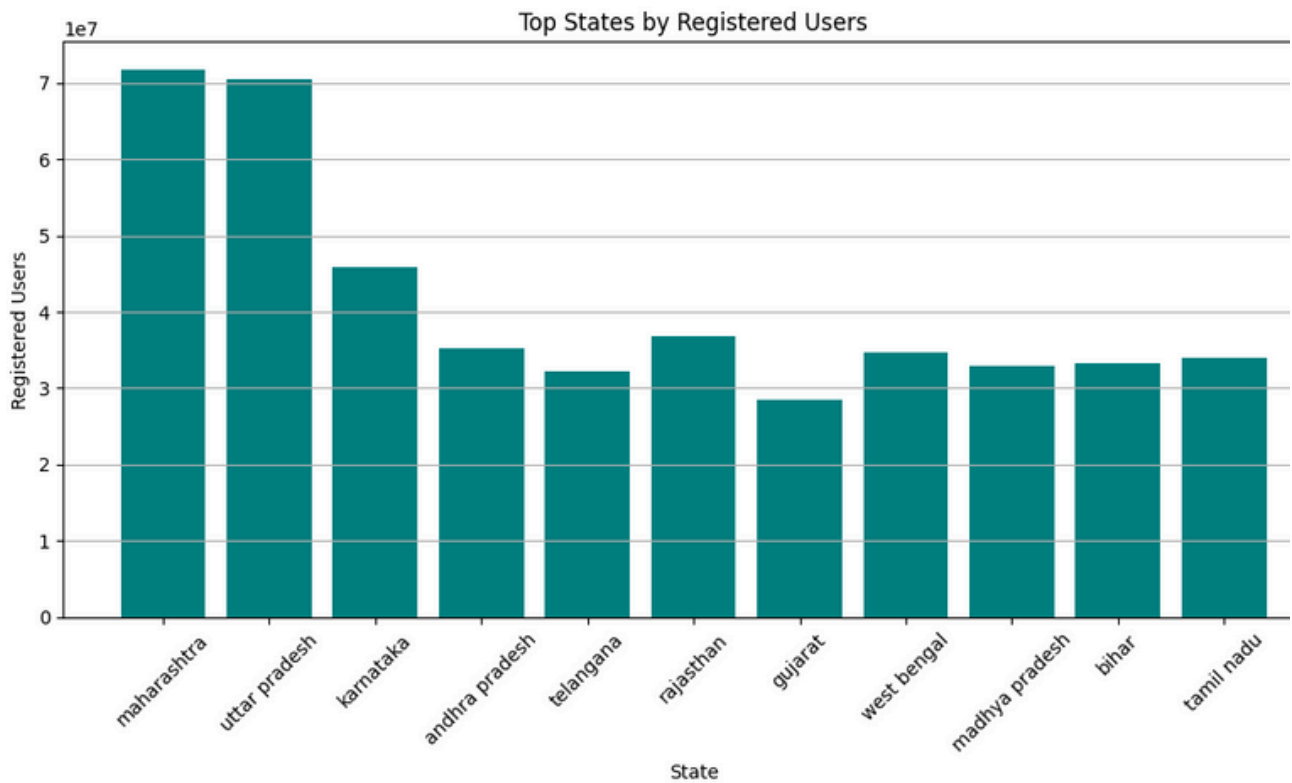
District with Highest Registered Users per Year

- `Muzaffarabad district` occurred more in the Lowest Registration list.
- `Kamle district` was the lowest in 2018.
- `niuland district` was the lowest in 2024.
- The shocking fact is that both are almost similar in registrations despite the year gap.

04 | Top Charts

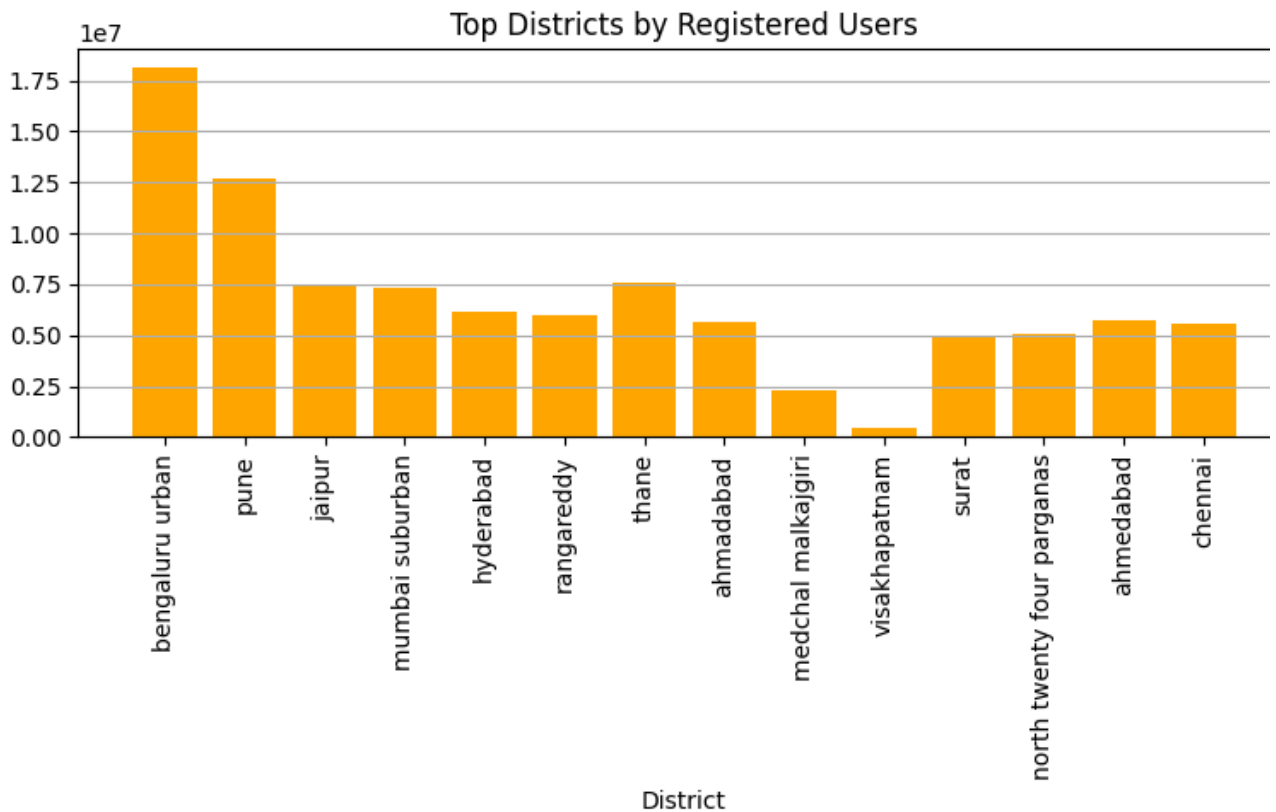
Top States by Registered Users

- `Maharashtra` & `UP` are shown as the top states in registration.
- Karnataka isn't on top this time.
- Most other states are the same on average level.

04 | Top Charts
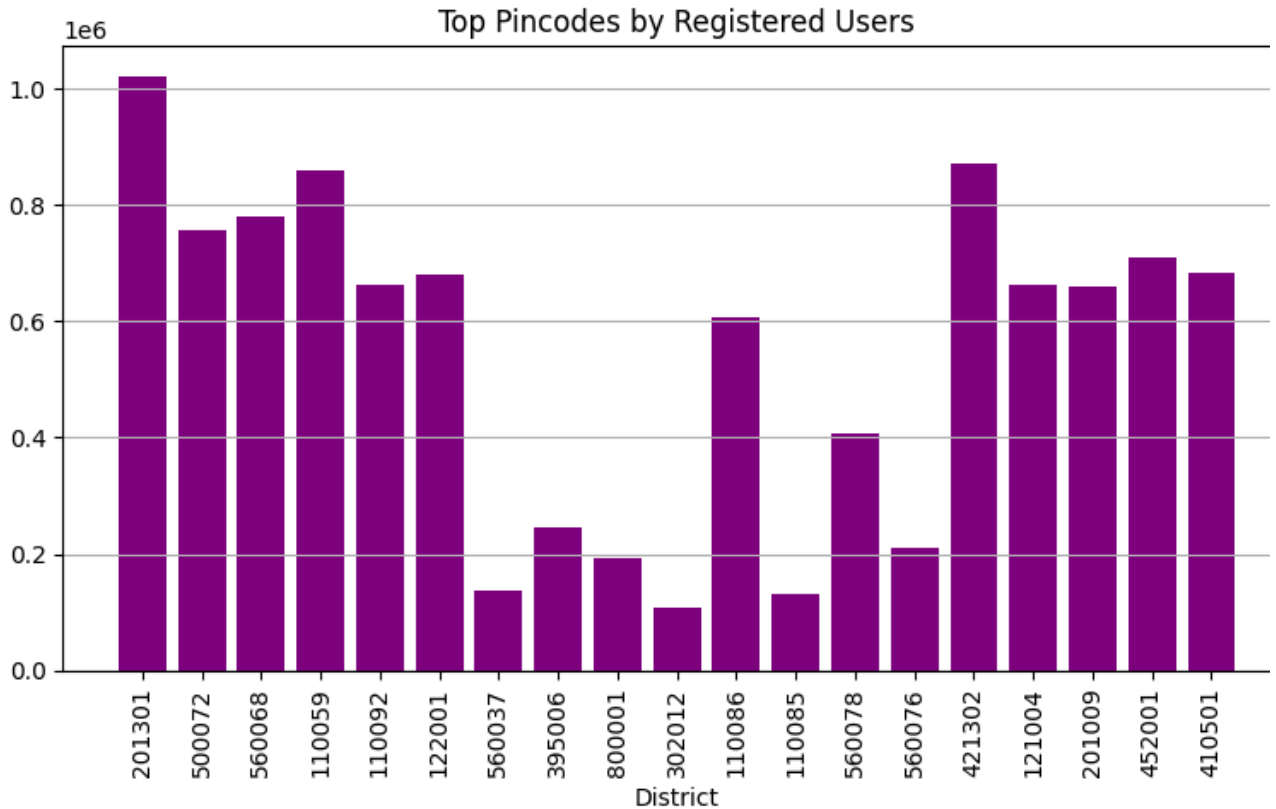

Top Districts by Registered Users

- `Bengaluru Urban` is the top district in registration.
- `Pune` is on the second.
- `Visakhapatnam` is at the bottom level.
- Rest are all the same in the average level.

04 | Top Charts

Top Pincodes by Registered Users

- `201301` is the most registered user pincode.
- `560037`, `302012` & `11085` are on the bottom,
- `11086` seems to be above average when `11086` is at the bottom

01 | Conclusion

- The PhonePe dataset reveals strong digital growth across insurance, transactions, and user registrations.
- Insurance values rose steadily from 2020 to 2024, with Karnataka overtaking Maharashtra as the top state from 2022.
- Bengaluru Urban consistently led in insurance and user counts, while districts like Pherzawl and Shi Yomi featured at the lower end.
- Peer-to-peer payments had the highest transaction amounts, but merchant payments dominated by volume.
- Telangana saw high transactions but low insurance activity, while Kerala showed the opposite.
- Xiaomi was the top brand in registrations (till 2022), and Maharashtra remained the highest-registered state.
- Lakshadweep and districts like Muzaffarabad and Kamle ranked lowest.
- Notably, pincode 500034 led in transactions, and 201301 topped user registrations.

SOCIALS

01 **GitHub: [Click Here](#)**

02 **Linkedin: [Click Here](#)**

03 **Project Repo: [Click Here](#)**

04 **Streamlit: [Click Here](#)**