**Consulting Report**

**Table of Contents Page No.**

1. Introduction 02
2. Project Description 02
3. Problem Statement 02
4. Problem Analysis 03
5. Sources of Data 03
6. Summary of Data Mining 05
7. Proposed Solution for Customers 06
8. Tools 20

* DS Tools
* Web UI Tools
  + PHP
  + JavaScript

1. Conclusion 20
2. **INTRODUCTION**

Insaid Telecom, one of the leading telecom players in India. The Company understands that customizing offering is very important for its business to stay competitive. Currently, Insaid Telecom is seeking to leverage behavioral data from more than 60% of the 50 million mobile devices active daily in India. The project resolves to help the Company to better understand and interact with their audiences.

The Company seeks direction in terms of understanding its customer segments, demographics, behavior, brand affinity and engagement with the company. The Company is looking for directions to form strategies for customer acquisition, retention and market expansion.

The project aims to provide actionable insights to the Company.

1. **PROJECT DESCRIPTION**

In this consulting assignment, we are expected to build an interactive dashboard for the client to understand user's demographic characteristics based on their mobile usage, geolocation, and mobile device properties in terms of preferred mobile brands and device models. The assignment also seeks to understand customer segment behavior by gender and age groups across geo locations.

Doing so will help millions of developers and brand advertisers around the world pursue data-driven marketing efforts which are relevant to their users and cater to their preferences.

**Consulting Goals:**

We have been assigned 6 states for consulting i.e. Madhya Pradesh, Chhattisgarh, Uttaranchal, Jammu and Kashmir, Nagaland and Goa. Our goal is to:

* Analyze User behavior to help the company design its customer offering.
* Study Customer demographics to help company make future strategies.
* Do help the company understand what is the right way forward.
* Suggest actionable insights for marketing and product teams.

1. **PROBLEM STATEMENT**

Insaid Telecom is one of the Telecom players in India. Currently, the company want to leverage on the User data of active users on their network. The Company has 60% of the 50 million mobile devices active daily in India. Company wants to understand the customer demographics for better interaction and engagement with subscribers for 6 states - Madhya Pradesh, Chhattisgarh, Uttaranchal, Jammu and Kashmir, Nagaland and Goa.

1. **PROBLEM ANALYSIS**

We have deployed the following strategies to understand the data and the problem statement:

1. Importing right packages
2. Data preprocessing
3. Data cleaning
4. Imputing missing values as no deletion of data is allowed.
5. Data translation to English text from Chinese
6. Merging of multiple files into a single master file.
7. Filtering data to focus states. Our assignment is to study Madhya Pradesh, Chhattisgarh, Uttaranchal, Jammu and Kashmir, Nagaland and Goa
8. Check for Outliers in the dataset, if any
9. Exploratory Data Analysis
   1. Distribution of Users across Sates
   2. Top phone brands preferred by Users
   3. Distribution of users across gender
   4. Distribution of users across age group
   5. Preferred phone brand across age group
   6. Preferred phone brand across states
   7. Preferred phone brands by gender
   8. Most preferred device models
   9. States with highest no. of calls
   10. Average Calls Per User (ACPU) state-wise
10. Actionable insights from the EDA
11. **SOURCES OF DATA**

In this assignment, we are going to study the demographics of users (gender and age) based on their app download and usage behavior. The Data is collected from mobile apps that use Insaid Telecom services. Full recognition and consent from individual user of those apps have been obtained, and appropriate anonymization have been performed to protect privacy. The data is accurate ground truth for prediction.

The data schema can be represented in the following table:

1. gender\_age\_train - Devices and their respective user gender, age and age\_group

|  |  |  |
| --- | --- | --- |
| **Column** | **Dtype** | **Description** |
| device\_id | int64 | Unique device\_id of users of Insaid Telecom network |
| gender | object | Gender of the Users- male or female |
| Age | int64 | age of the users |
| Group | object | gender age group of the users |

1. phone\_brand\_device\_model - device ids, brand, and models phone\_brand: note that few brands are in Chinese

|  |  |
| --- | --- |
| **Brand Name** | **Brand English Mapping** |
| '华为' | 'Huawei' |
| '小米' | 'Xiaomi' |
| '三星' | 'Samsung' |
| 'vivo' | 'vivo' |
| 'OPPO' | 'OPPO' |
| '魅族' | 'Meizu' |
| '酷派' | 'Coolpad' |
| '乐视' | 'LeEco' |
| '联想 ' | 'Lenovo' |
| 'HTC' | 'HTC' |

|  |  |  |
| --- | --- | --- |
| **Column** | **Dtype** | **Description** |
| device\_id | int64 | Unique device id of users of Insaid Telecom network |
| phone\_brand | object | phone brands used by users |
| device\_model | object | different device models of various brands being used by users |

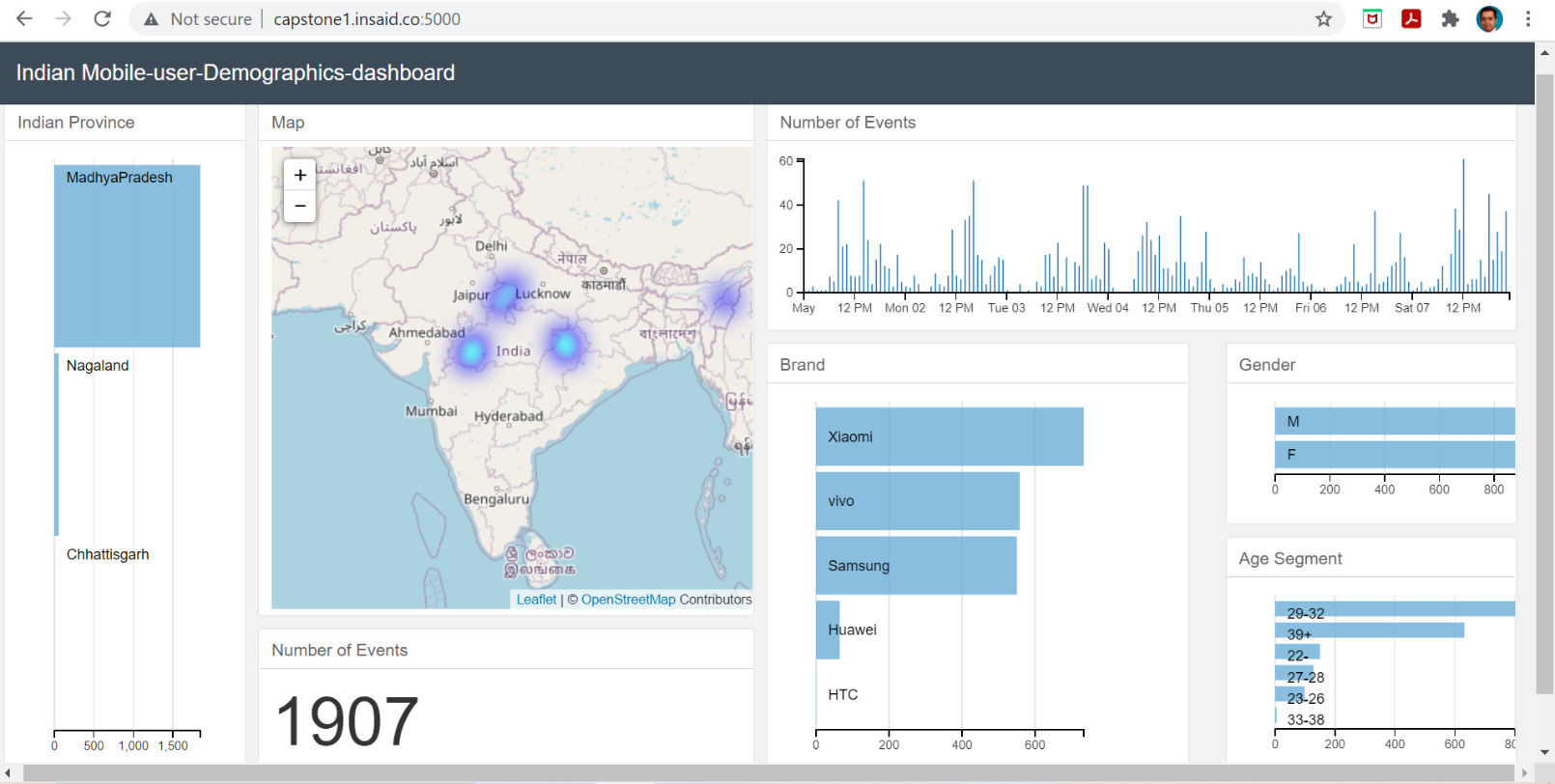
1. events\_data - when a user uses mobile on INSAID Telecom network, the event gets logged in this data. Each event has an event id, location (lat/long), and the event corresponds to frequency of mobile usage, timestamp: when the user is using the mobile.

|  |  |  |
| --- | --- | --- |
| **Column** | **Dtype** | **Description** |
| device\_id | float64 | Unique device id of users of Insaid Telecom network |
| Timestamp | object | date and time when the events occurred |
| Longitude | float64 | longitude location of the device id |
| Latitude | float64 | latitude location of the device id |
| City | object | city location of the user |
| State | object | state location of the user |

* Events data being 293 Mb, was downloaded to local system for fast fetching of the data while processing on notebook.
* Phone and Gender data was in SQL database. We applied SQL connector tools to fetch the data.

1. **SUMMARY OF DATA MINING**
   1. **Challenges faced with the data**

* Events data was of 293 Mb and we were not able to fetch the data from the cloud server while executing the python program on local system. Every time we faced either server time out or too long wait time to fetch data. Finally, it made sense to download the data to local server and use it for processing.
* Though events data had 3252950 rows of data, it contained only 60865 unique device\_id.
* Data type of device id was different in the three data sets float64 vs int64. Conversion to float64 was required for merge operation. We struggled merging of data on device\_id when it was kept as int64 data type. This was creating duplication of data. Converting device\_id to float64 and keeping it as an absolute number solved a big challenge.
* In events data, device\_id, longitude, latitude and state have missing data in 453, 423, 423 and 377 fields respectively.
* Few Device ids were found to be negative integer. “-ve” sign was removed to avoid error in merging or mapping of unique device\_id
* Phone data set had 87754 and Gender data set had 74645 rows of data. Merging this data created many 13109 NaN rows for gender, age and group. We found additional device\_id in this dataset which effectively means non-active or churned users.
* Since using Google translation API was creating errors while translating and behaving differently each time codes were executed, we chose to create a dictionary of brand map and device model map for the purpose of translation of chinese text to english text.
* When we filtered data to focused states, the dataset was found to be too small containing only 4297 unique device\_id
* The Interactive dashboard is not capable of populating the events data of the focus 6 states, which has only 264118 rows of data.
* device\_id in the sample CSV file was in int64 format. When we were uploading data with device\_id as float64, it was showing error. So, we changed the events\_time\_df CSV's device\_id to int64 to resolve the issue.
* Dashboard has some technical issue. We could upload the events CSV successfully for visulaization, but the visualization only shows 1907 events instead of 261148 events.
* We also noticed that our events CSV file doesn't contain the data regarding the phone\_brand, gender, group. But the dashboard reflects this information. We can easily conclude that there is some internal issue in dashboard's database referencing to fetch this information which we are not uploading.
* Ideally the dashboard should have been designed in such a way that it allows loading of the complete master dataset containing events, phone and gender data.



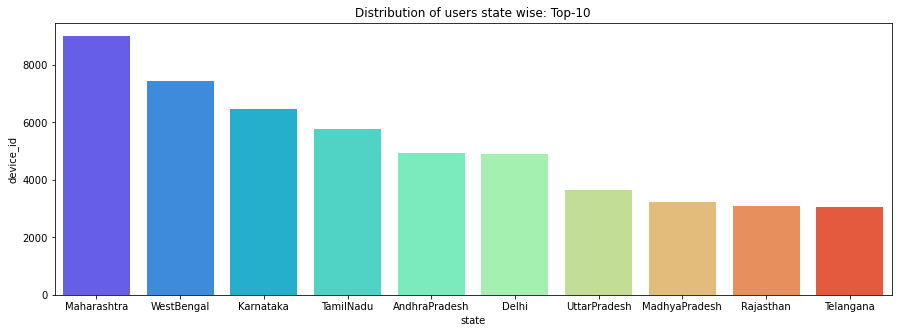
* 1. **Summary of analysis**

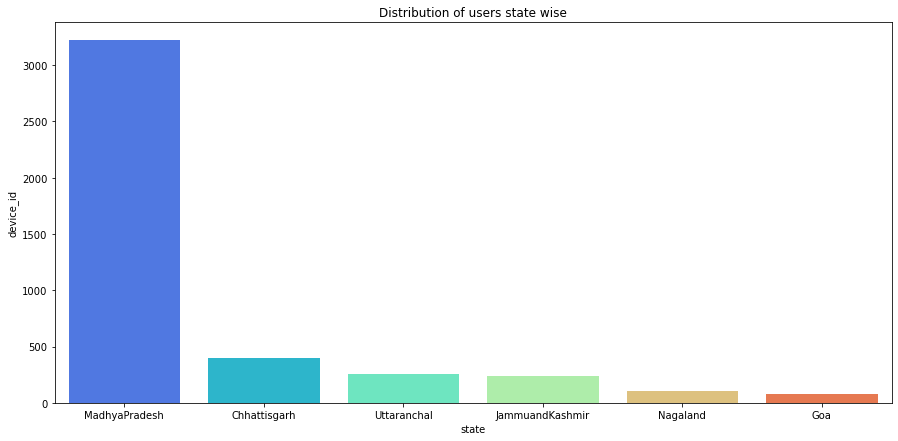
The data set threw multi-pronged challenges in front of us while data mining and pre-processing. Starting with the size of data, dealing with cloud server, using of different database CSV vs SQL, language translation, cleaning of data, imputing of missing values, detecting outliers, handling different data types for same data column, challenges in merge operations and erratic behavior of Google translation API.

1. **PROPOSED SOLUTION FOR CUSTOMERS**

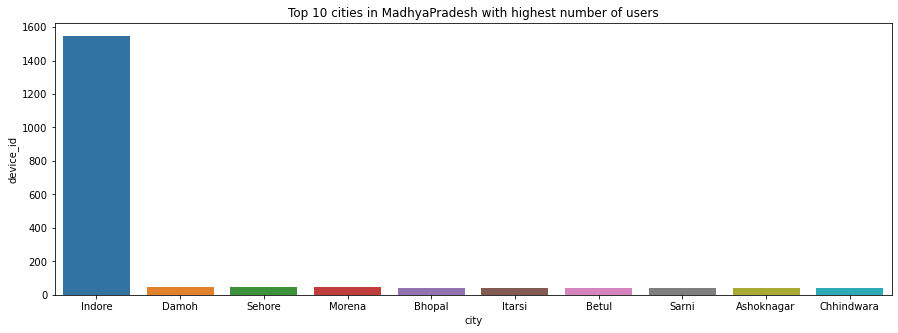
EDA generated lot of observations and insights pertaining to user behavior and demographics. Based on the EDA, we have found the following observations:

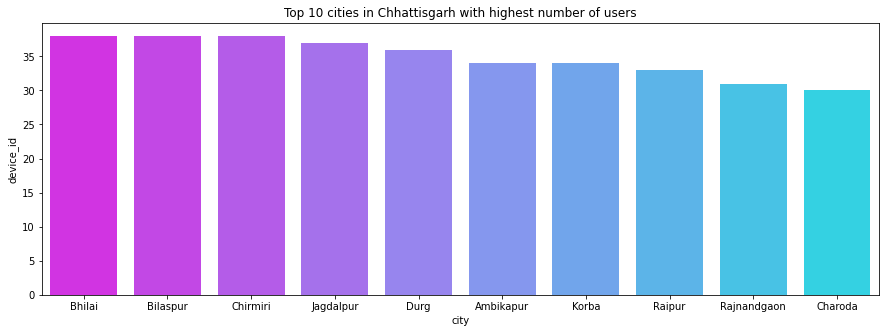
* 1. **Distribution of Users across States**

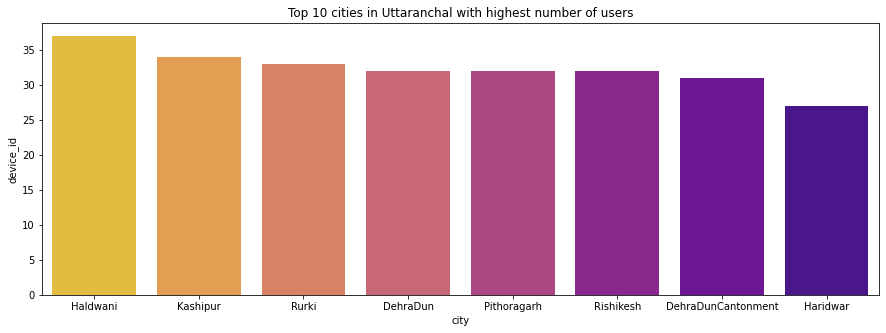
****

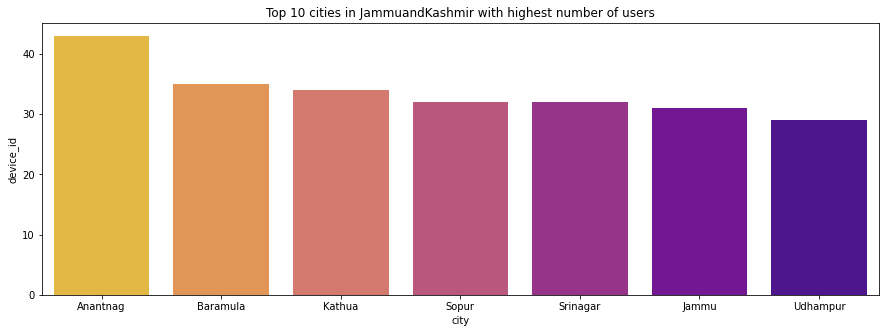
****

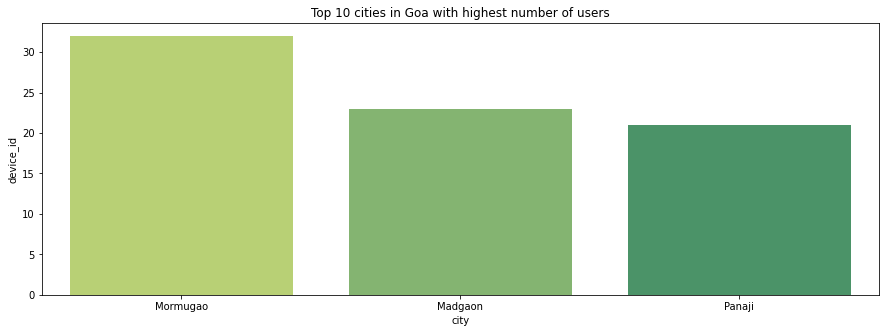
* 1. **Top 10 cities State-wise in focus States**

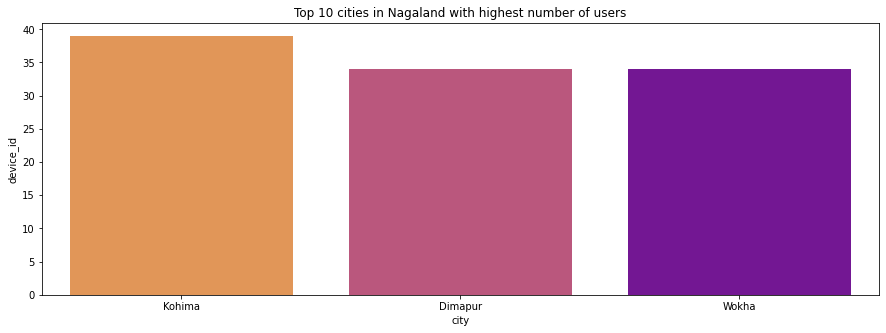
****

****

****

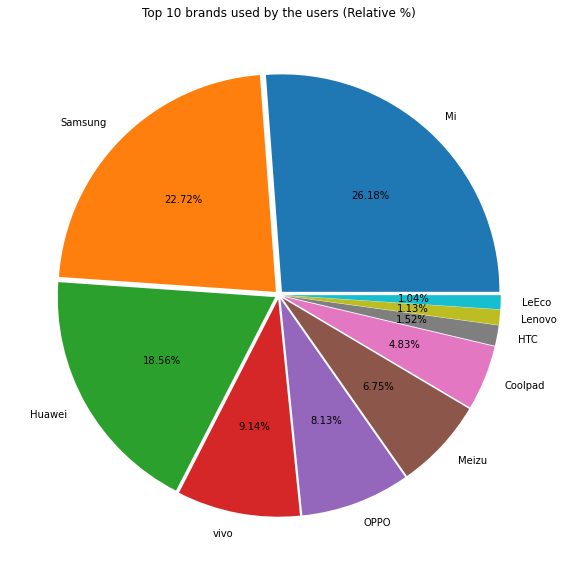
****

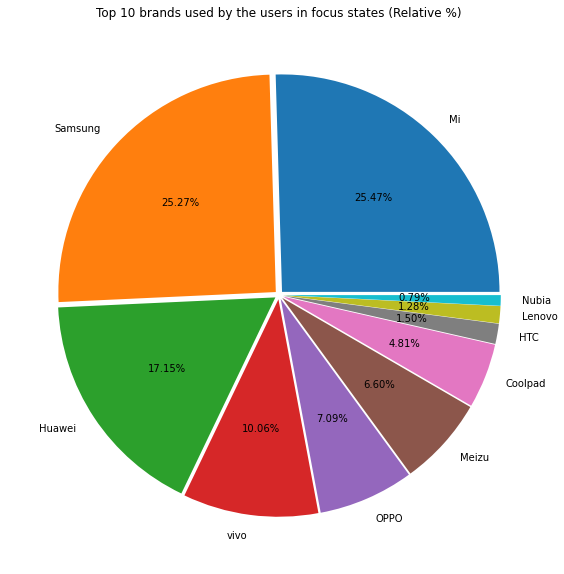
****

****

**Observations:**

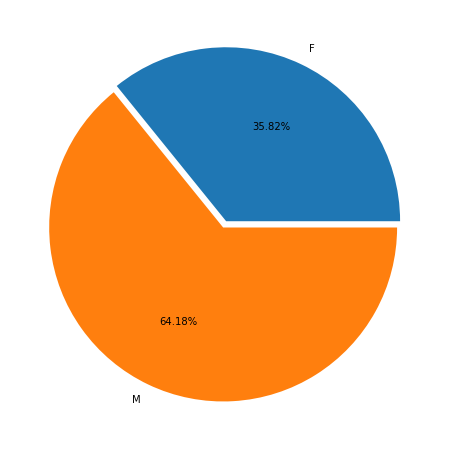
* Among the top 10 states by Users, Maharastra have the maximum connections at 9010 followed by West Bengal (7436) and Karnataka (6481).
* Out of our focus states, only MadhyaPradesh figured at 8th position by no. of users at 3220.
* In the focus states category, maximum Users found in Madya Pradesh (3220) followed by Chhattisgarh (400), Uttaranchal (258), Jammu and Kashmir (236), Nagaland (107) and Goa (76).
  + Leaving Madhya Pradesh, it can be inferred that INSIAD Telecom has poor customer base.
  + In Madhya Pradesh, maximum subscriber base is in Indore. Rest cities have very poor penetration.
  + Chhattisgarh hardly have any subscriber base, but the spread across cities is even.
  + Uttaranchal also have poor penetration with presence only limited to 8 cities.
  + J&K has low penetration limited to only 7 cities. The subscriber base is too small.
  + Goa has the lowest penetration in terms of subscribers as well as cities (only 3).
  + Nagaland has the second lowest subscriber base after Goa with presence limited to three cities- Kohima, Dimapur, Wokha.
  1. **Top 10 Phone Brands by Users**

****

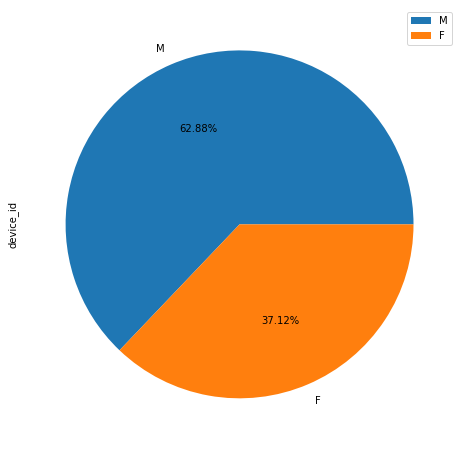
****

**Observations:**

* Top 10 brands by Users across Indian subcontinent are Mi, Samsung, Huawei, Vivo, Oppo, Meizu, Coolpad, HTC, Lenovo and LeEco.
* Almost 75% market is being dominated by the chinese handset manufacturers.
* It can be inffered that maximum Users use Android devices.
* Top 10 brands in the focus states are Mi, Samsung, Huawei, Vivo, Oppo, Meizu, Coolpad, HTC, Lenovo and Nubia. We could see that the User preference of Brands in focus states is similar to that of India as a whole.
  1. **Distribution of Users across Gender**

****Gender-wise users across India

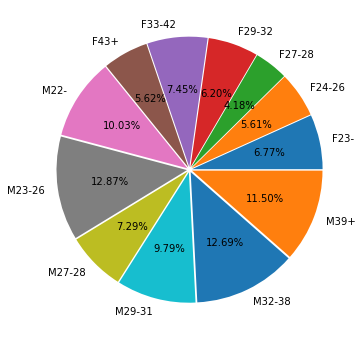
Gender-wise users across Focus States

****

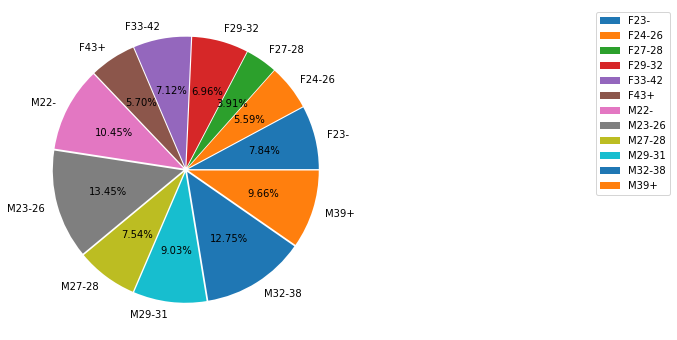
**Observations:**

* Across India, the company have one female subscriber for each two male subscribers.
* Though focus states have similar gender trends, it has better female subscribers ratio at 37.12% vs 35.82% national average.
  1. **Distribution of Users across Age Group**

Users across India by Age group



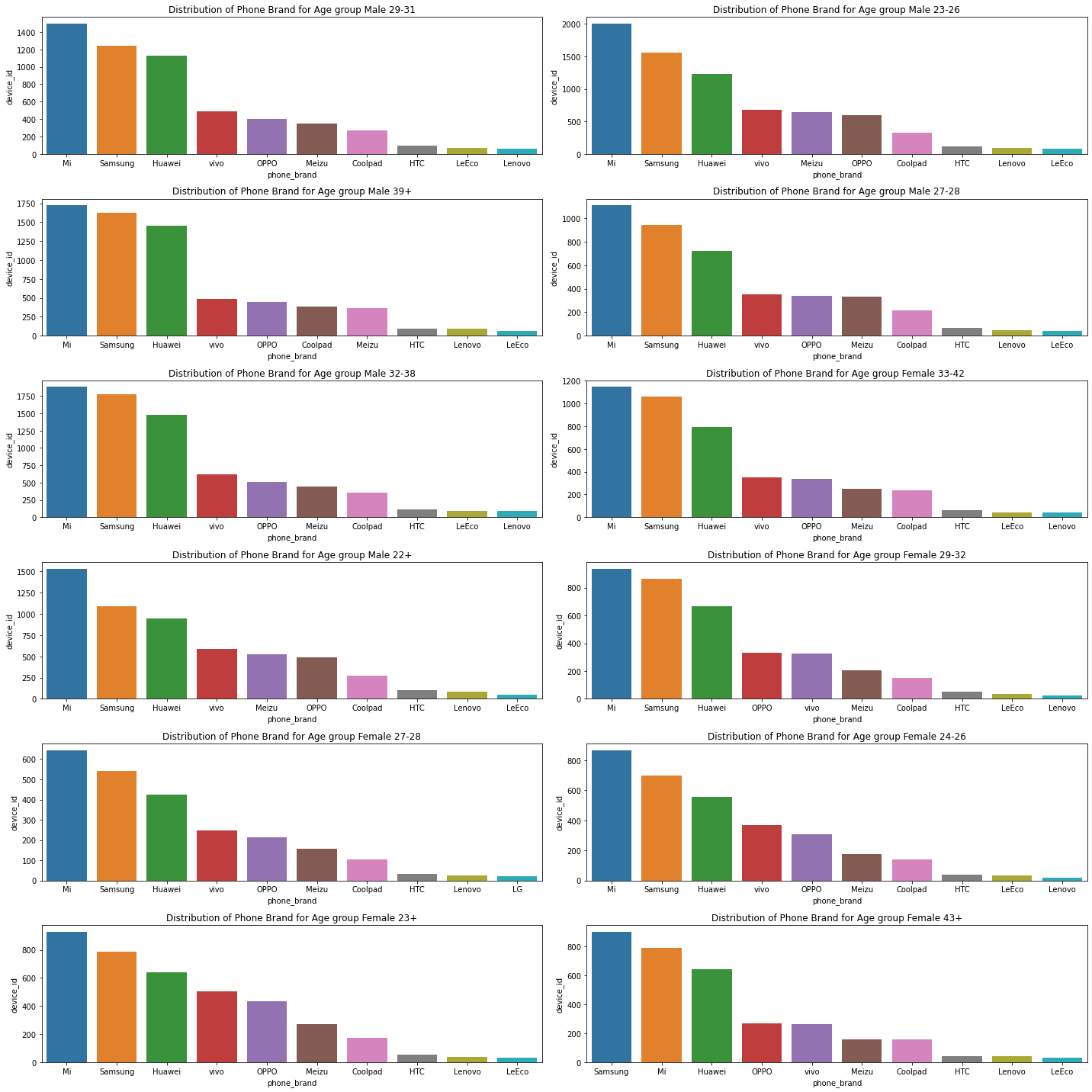
Users across focus states by Age group



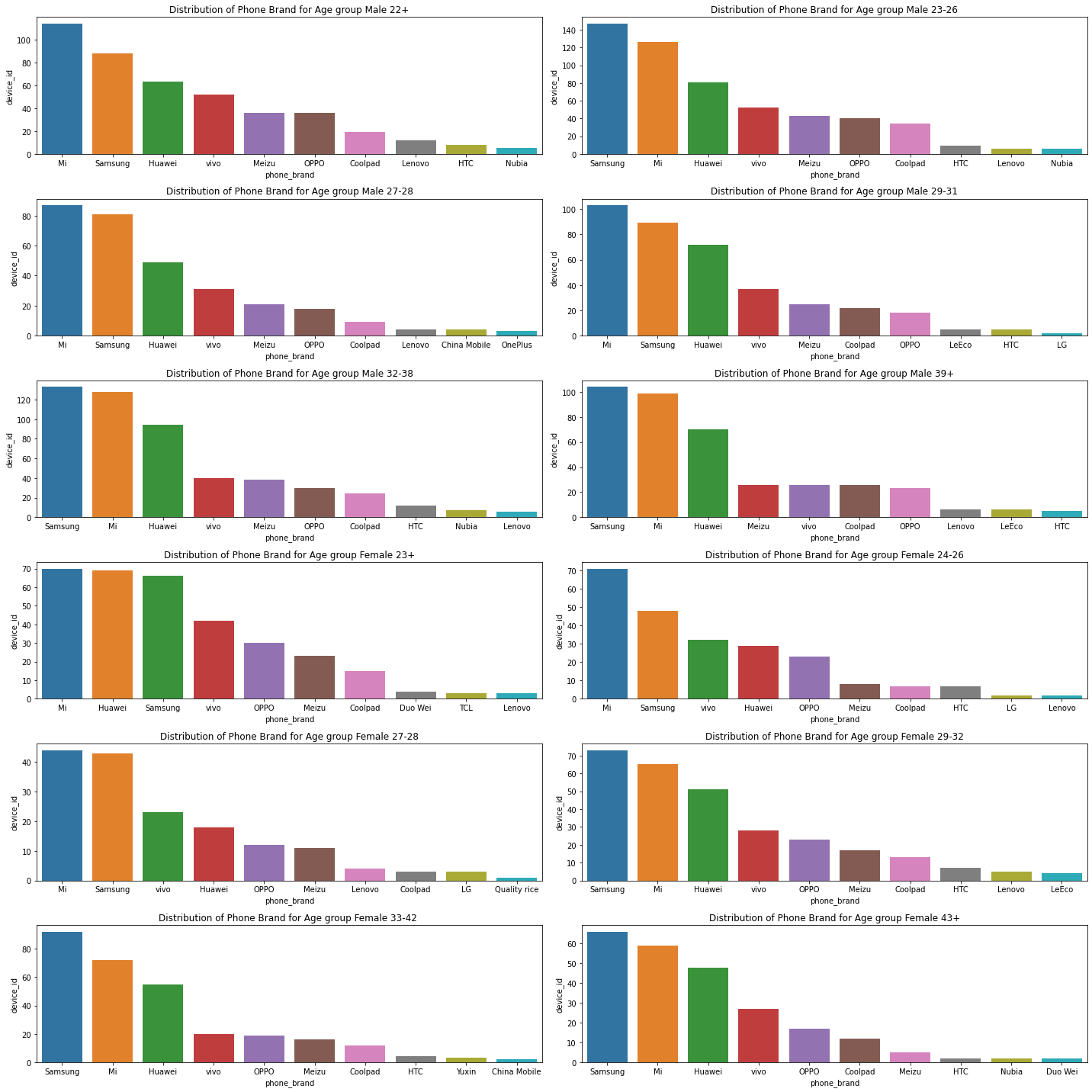
**Observations:**

* More than 80% subscribers are youth both in male and female population.
* Young and mid-age customers are main subscribers to INSAID Telecom i.e. age group 23-43 yrs. They represent the young working population.
  1. **Preferred Phone Brand across Age Group**

Nationwide



Focus States



**Observations:**

* Mi, Samsung and Huawei are top-3 brands across all gender age groups
* Female 24-28 age group likes Mi, Samsung followed by Vivo as preferred top-3 brands
* Nationwide, Female 43+ age group like Samsung over Mi. This means older age group females has higher brand affinity.
* In focus States, Female age group 33-42 and 43+ prefer Samsung over Mi unlike other age groups and Male counter-part. This means Female are more brand conscious.

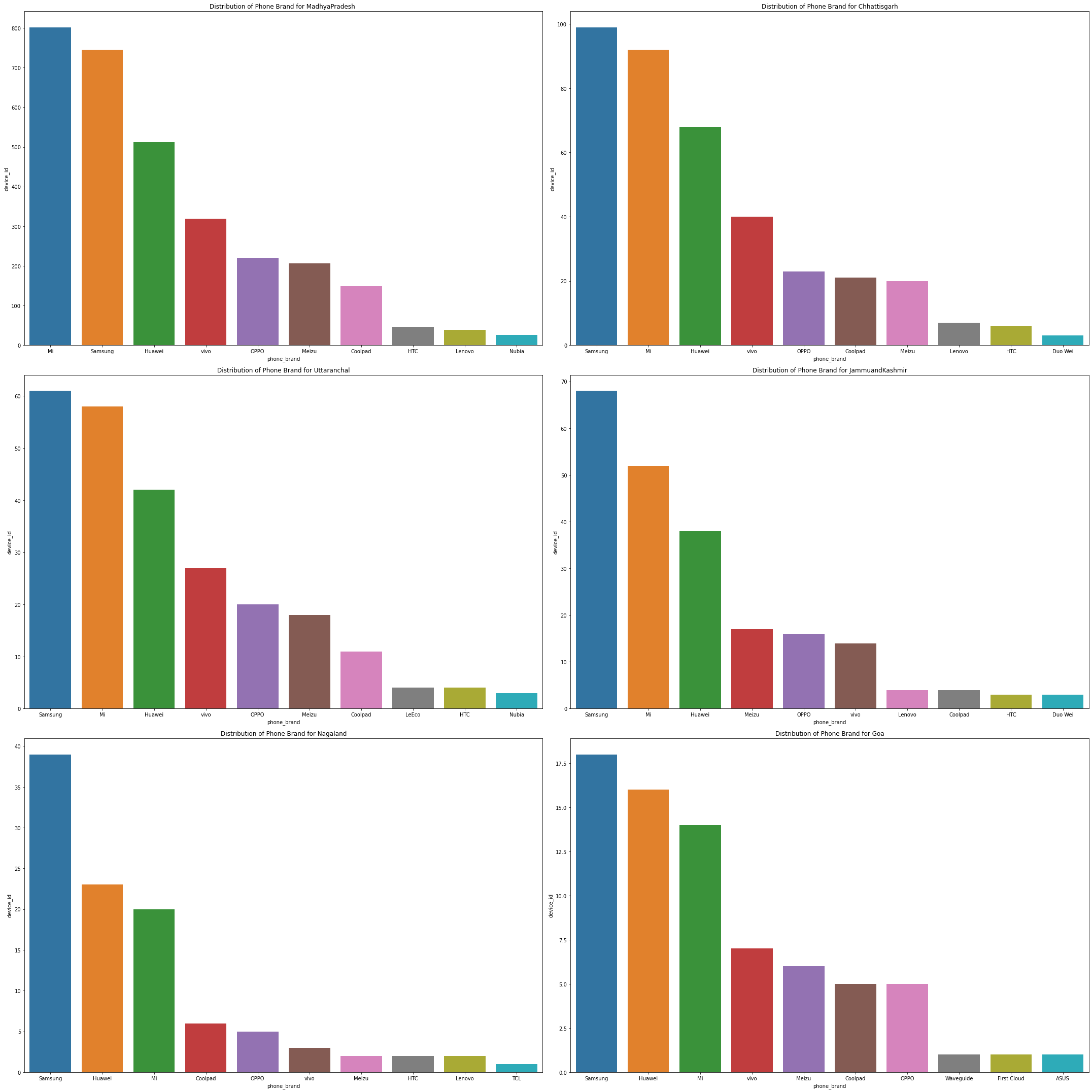
## Preferred Phone Brand across States

## Nation wide

## 

## 

Focus States



**Observations:**

* Nationwide Mi and Samsung are top consumer brands across all states except Delhi where Huawei is no.1 brand.
* Except Madhya Pradesh all other focus states have Samsung as the top brand of choice. Madhya Pradesh has Mi followed by Samsung as the top choice.

## Preferred Phone Brands Gender wise

## Nation wide

## 

## Focus States

## 

**Observations:**

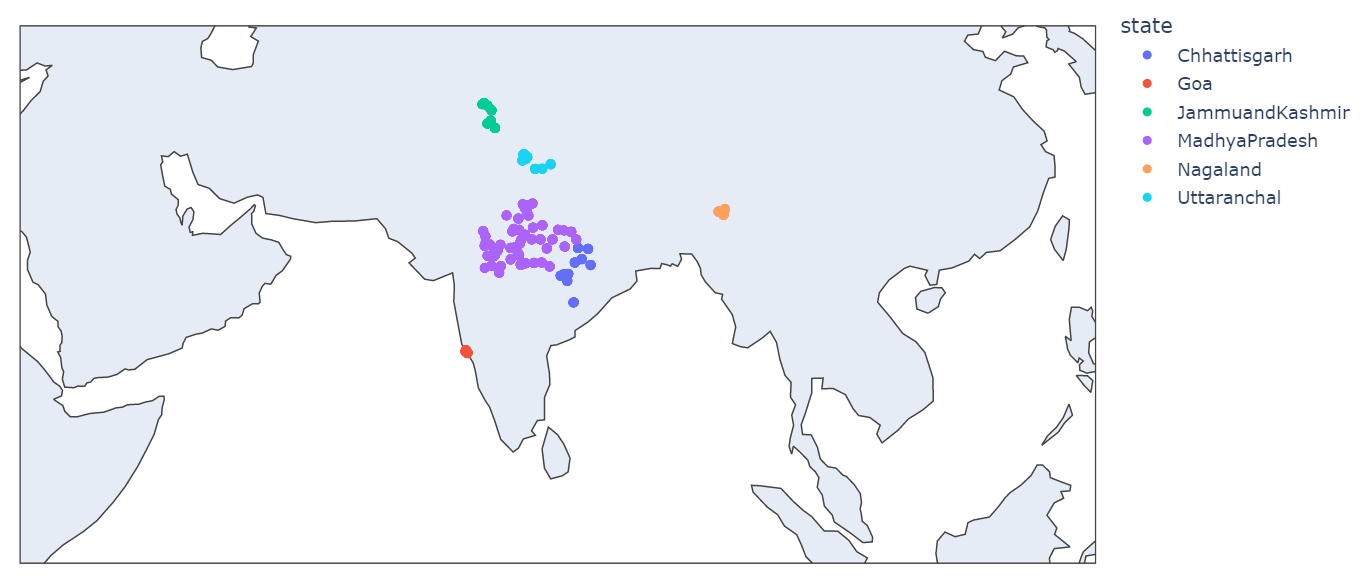
* Nationwide Mi and Samsung are top brands liked by Male as well as Female.
* In Focus States Mi is the top brand in Males and Samsung is the top brand in females.

## Most Preferred Device Model

## 

## 

**7.10 Geographical coverage in focus States**

****

**Observations:**

* Mi-Redmi note, MI 3, MI 2S are top 3 device models preferred by Subscribers across India.
* In Samsung-Galaxy Note 3, Galaxy S4 and Galaxy Note 2 are the most preferred models.
* In focus States Mi-Redmi Note, MI 3 and Samsung Galaxy S4 are the top three device models.

1. **TOOLS**

* DS Tools:
  + Python: Jupyter Notebook
  + Packages / libraries: matplotlib, seaborn, pandas, numpy, plotly, chart-studio, pymysql
* Web UI Tools
  + PHP
  + JavaScript

1. **CONCLUSION**

After careful analysis of the data, the team proposes the following actionable insights for Insaid Telecom, as mention below. It will be in the interest of the company to build strategies around it. Implementation of the said observations will definitely improve revenue and profitability of the company along with improvement in customer services leading to more satisfied subscribers to the business.

Following are the actionable insights proposed to the company:

* Leaving Madhya Pradesh, it can be inferred that INSAID Telecom has poor customer base in our focus states of study (MP, Chhattisgarh, Uttaranchal, J&K, Goa, Nagaland) and need to drive acquisition of new subscribers.
  1. Need to launch lucrative Talk-time plans for customers so that they switch from competition to INSAID Telecom.
  2. Lunch Enterprise plans to attract corporate customers.
  3. Help Customers for easy Mobile Number Portability to INSAID Telecom.
  4. Address current service issues, if any, like infrastructure, network coverage, data service, customer services, POS, recharge vouchers, payments, new connection services etc
  5. Launch Customer acquisition BTL activities-Retail POS marketing activities, corporate tie-ups, canopy campaigns etc.
  6. Launch ATL campaigns – Digital marketing initiatives, encourage own app downloads, POS branding, Print and TV commercial to grab mind share in important markets.
* In Madhya Pradesh, maximum subscribers base is in Indore. Rest cities have very poor penetration. INSAID Telecom has to drive customer acquisition beyond Indore. Cities like Gwalior, Mandu, Orchha, Ujjain are not even in the list of top 10 cities of Madhya Pradesh. Bhopal being one of the important cities in MP has very low penetration.
* Chhattisgarh hardly have any subscriber base, but the spread across cities is even. INSAID has to focus on activation starting with major cities like Raipur, Bhilai, Bilaspur, Durg, Jagdalpur, Ambikapur, Korba and so on.
* Uttaranchal also have poor penetration with presence only limited to 8 cities. Though evenly spread subscriber base is too low. The focus should be on expanding subscriber base. On the same time, INSAID Telecom should plan to launch in more cities like Roorkee, Pantnagar, Almora, Joshimath, Kotdwar, Mussoorie, Nainital, Pauri, Ramnagar, Raniket, Rishikesh, Uttarkashi etc.
* J&K has low penetration limited to only 7 cities. The subscriber base is too small. Need to drive customer acquisition. Cities like Pulwama, Budgam, Ramban, Bandipora, Kargil, Reasi, Kupwara, Ganderbal, Kishtwar, Leh, Ladakh, Doda, Samba etc.
* Goa has the lowest penetration in terms of subscribers as well as cities restricted to Mormugao, Madgaon and Panaji. INSAID Telecom should penetrate other cities like Margao, Mapusa, Ponda, Curchorem etc.
* Nagaland has the second lowest subscriber base after Goa with presence limited to three cities- Kohima, Dimapur, Wokha. Other cities to focus expansion is Mokokchung, Mon, Mopungchuket, Peren, Tuensang and Sumi.
* Going by the Handset brands used by the users which is mostly Chinese, it can be safely concluded that current subscribers are price sensitive customers. Hence designing of right talk-time and validity recharge vouchers / subscription plan is critical to customer acquisition and retention.
* As most users are Android users, focus should be to develop app eco-system and services around it. iOS users are non-existent as per the dataset provided. Company needs to launch premium services to acquire this high value customers which might be currently at different networks. Such premium services can be around plans bundled with broadband, data, mobile combination. Plan which caters to the need of family e.g. family plans can be a good starting point. Other high value plans could be bundled around OTT free subscriptions to Netflix / Amazon Prime / Zee / Hotstar etc.
* Company can also evaluate launching latest branded handset bundled offers, though it is not so successful in India as dominated by prepaid customers. Company can use this as a tool to shift prepaid customers to post-paid customers. Post-paid customers are most loyal with lowest churn rate and higher ARPU. So, it is prudent that Company should focus on expanding the pool of post-paid customers. Providing EMI option will ensure these customers to stick to the network.
* Focus states has better female subscribers’ ratio than national average. INSAID Telecom should leverage on this trend and design acquisition offers, validity plans, recharge plans keeping women in mind. Women subscribers are more loyal if given better value. OTT bundled offers can do wonders.
* Almost 80% of the INSAID subscribers belong to young and mid-age customers. These are younger working population who are price and benefit sensitive. They switch easily. The older working population are high paying customers i.e. customers who are 43+. This population has a sizeable representation of 27% in India by size but under representation in terms of subscription. Company should device subscription plans keeping this population in mind. This age group of customers are loyal, less demanding and sticky customers.
* Female customers in mid-age group 32-43 and older age group 43+ are found to be more brand conscious. They prefer Samsung brand over Mi which is not true for male counterpart. We can infer that this are high value customers and can be targeted to increase ARPU (Average Revenue Per User) from this segment.
* INSAID telecom should focus on increasing ACPU (Average Calls Per User). This can be achieved by introducing unlimited talk-time plans. This will also help in improving ARPU.

\*\*The End\*\*