***Accredian Telecom – Data Analysis report***



*Release: V1*

Group : Group 1007

Authors: Chethan Ananthan, Naveen Kumar B., Rupali Joshi

Reviewers: Vijaya S, Meghdeep Chada.

Contributors: Nirupama N., Pramod Johnson

Approvers: Team Accredian



*Revision History:*

|  |  |
| --- | --- |
| Version Date | Change Summary |
| 24th July - -2023 | Initial draft created |
| 6th August 2023 | Final Version |

Contents

[*Introduction:* 3](#_Toc142297131)

[*Project Description:* 3](#_Toc142297132)

[*Problem Analysis:* 3](#_Toc142297133)

[*Sources of Data* 4](#_Toc142297134)

[*Summary of Data Mining* 4](#_Toc142297135)

[*Proposed Solution for Customers* 5](#_Toc142297136)

[*Summary:* 14](#_Toc142297137)

# *Introduction:*

Accredian Telecom, one of the leading telecom players, understands that customizing offerings is very important for its business to stay competitive. Currently, Accredian Telecom is seeking to leverage behavioral data from more than 60% of the 50 million mobile devices active daily in India. They are doing this to help their clients better understand and interact with their audiences.

# *Project Description:*

The Acceredian team should provide the actionable insights on following points

* Number of Records analyzed
* Distribution of Customers State-wise
* Customer distribution across India
* Frequency of Events with Respect to Time
* Distribution of Brand of Mobile Phones among Customers
* Mobile Phone users based on Gender
* Distribution of different Age Groups

*Problem Statement:*

In this consulting assignment, Accredianians are expected to build a dashboard. This dashboard will help us to understand a user's demographic characteristics based on their mobile usage, geolocation, and mobile device properties Doing so will help millions of developers and brand advertisers around the world pursue data-driven marketing efforts. These efforts are relevant to their users and cater to their preferences.

# *Problem Analysis:*

We reviewed all the three given datasets in detail and tried to figure out the impact of each parameter in leveraging the customer base. As part of this strategy, we evaluated various columns in the dataset like age, gender, state, typical call duration, user count across various states etc.

# *Sources of Data*

There are three datasets used for Accredian Telecom analysis project. All the three datasets are static and provided by Accredian team. The datasets are not validated for its correctness.

# *Summary of Data Mining*

We analyzed the 3 major Data sets and performed the following actions:

1. **Age gender Train data** 
   1. This data set is of size 74645, 4. This dataset has device\_id as unique parameter (to be used as primary key).
   2. The gender is marked as M or F, which indicates Male / Female
   3. The Age Group has been divided in a very arbitrary way. Hence we create a NewGroup for the proper division of the age of the people

0 to 19 - Teenagers; 20 to 39 - Youth 30 to 69 - Mid Aged 60 and above - SeniorsCitizens.

1. **Mobile Brand Data**
   1. This data set is of size 87726, 3. This data has device\_id as the unique parameter. (to be used as the primary key).
   2. The device Names are in Chinese and English.
   3. The Mobile Brands are also in Chinese and English.
   4. We used a translator function to convert the Mobile brands from Chinese to English. Here we found that applying Chinese as a language of source is insufficient and we had to use “Simplified Chinese” as the source language for translations.
2. **Events Data**
   1. This data set is of size 3252950, 7. In the initial check there are no duplicate values.
   2. Event\_Id represents an event made by a device and there are 3252950 events.
   3. These events have been captured during the time span ranging from April 30th 2016 to May 8th 2016. 1 Week of data is available for us for analysis. There are very less observations on April 30th 2016 and May 8th 2016. We have not removed these observations and our analysis includes the data on these days.
   4. Out of 325290 entries, we observed Null data on the following data sets.
      1. State - 377 Entries are Null
      2. Device\_id - 453 Entries are Null
      3. Latitude and Longitude Data - We have 423 Entries which have Null
   5. We observed that a particular device is hooked into the same latitude and longitude *(except for few observations where latitude and longitude is outside of India).* Hence by taking the reference of latitude longitude and state of the device id, Missing Null values of the state , latitude and longitude are completed.
   6. For the missing data of device id, we scan through the device id’s present in the same latitude and longitude and fill the Null Value with the respective device id.
   7. We observed that about 42 entries have the latitude and longitude based out of India. These entries are corrected with the latitude and longitude of the respective device id.
   8. We observed that device\_id is the common link across 3 data sets. To perform proper operation we need to have device\_id as a common data type. Hence we convert the data type of device\_id in every dataset as float.

# *Proposed Solution for Customers*

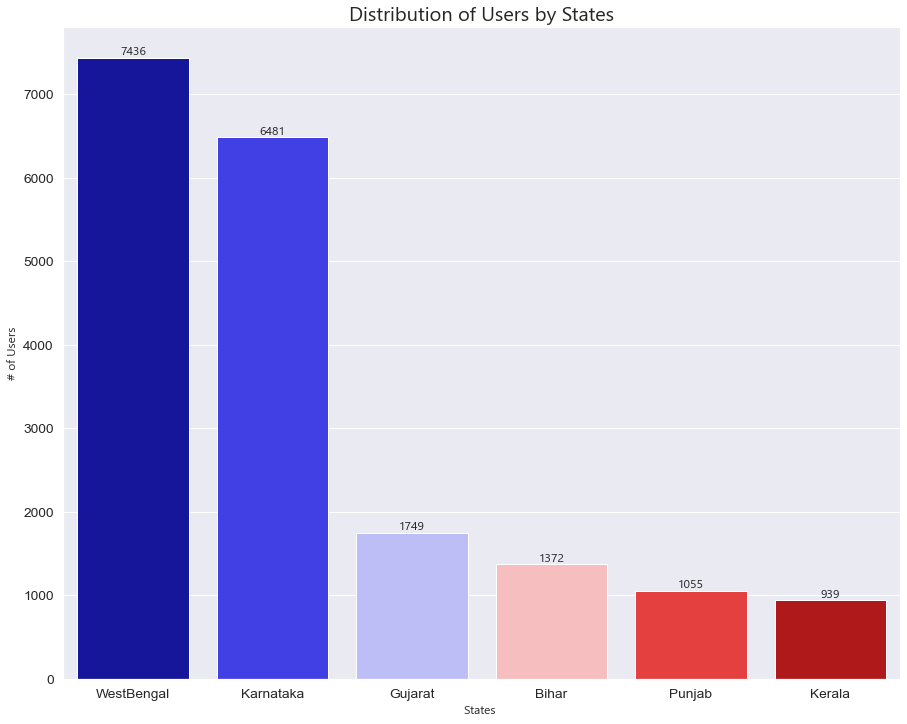
1. **Number of Records analyzed:**

The team was presented with three datasets and was asked to analyze only for “WestBengal”, “Karnataka”,” Gujarat”,” Punjab”,” Bihar” and “Kerala” states.

|  |  |
| --- | --- |
| event\_data | Analyzed 422971 rows of data out of 3252950 based on selected states |
| gender\_age\_train | Analyzed 74645 rows of data |
| phone\_brand\_device\_model | Analyzed 87726 rows of data |

1. **Distribution of Customers State-wise**

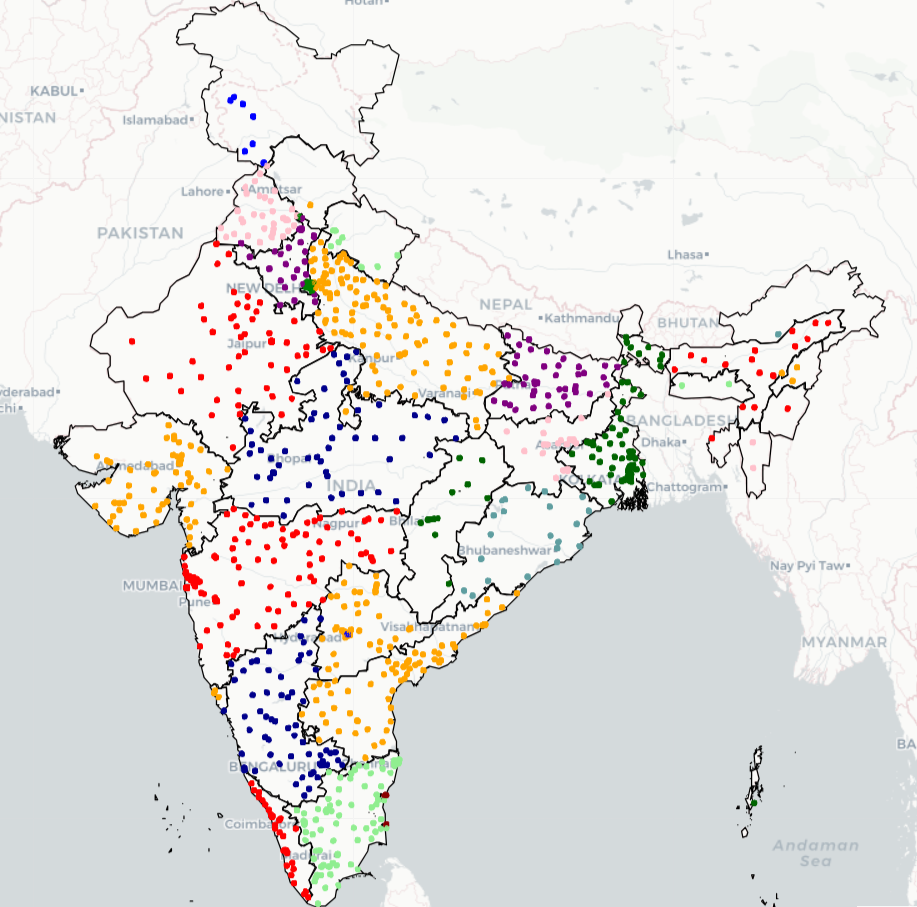
Maximum number of customers are in WestBengal while least number of customers are in Kerala



1. Customer distribution across India

Accredian Telecom has strong footprints in Maharashtra whereas the customer base in Manipur is weak.

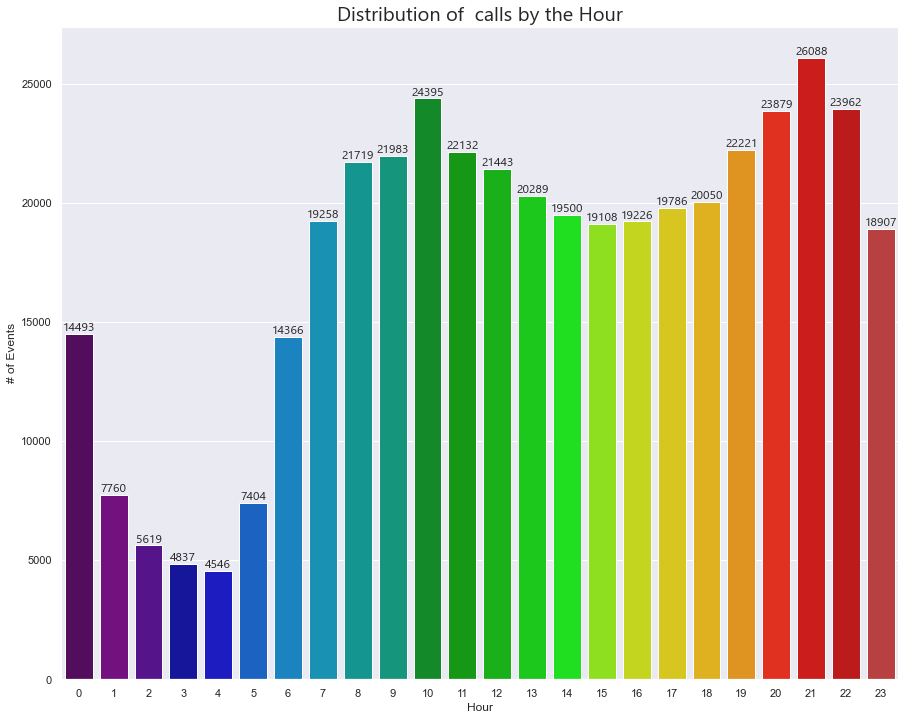
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **State** | **No Of Users** |  | **State** | **No Of Users** |
| Maharashtra | 9010 |  | Orissa | 655 |
| WestBengal | 7436 |  | Assam | 483 |
| Karnataka | 6481 |  | Chhattisgarh | 400 |
| TamilNadu | 5770 |  | Uttaranchal | 258 |
| AndhraPradesh | 4929 |  | JammuandKashmir | 236 |
| Delhi | 4909 |  | Nagaland | 107 |
| UttarPradesh | 3660 |  | Pondicherry | 96 |
| MadhyaPradesh | 3220 |  | Goa | 76 |
| Rajasthan | 3093 |  | Mizoram | 70 |
| Telangana | 3045 |  | Meghalaya | 56 |
| Gujarat | 1749 |  | AndamanandNicobarIslands | 41 |
| Bihar | 1372 |  | Tripura | 36 |
| Punjab | 1055 |  | Chandigarh | 35 |
| Kerala | 939 |  | ArunachalPradesh | 32 |
| Haryana | 816 |  | HimachalPradesh | 31 |
| Jharkhand | 739 |  | Manipur | 30 |

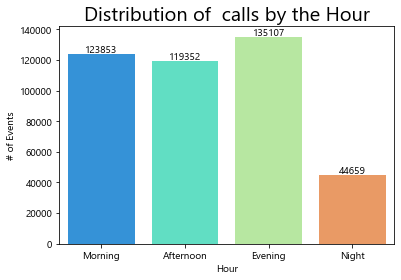


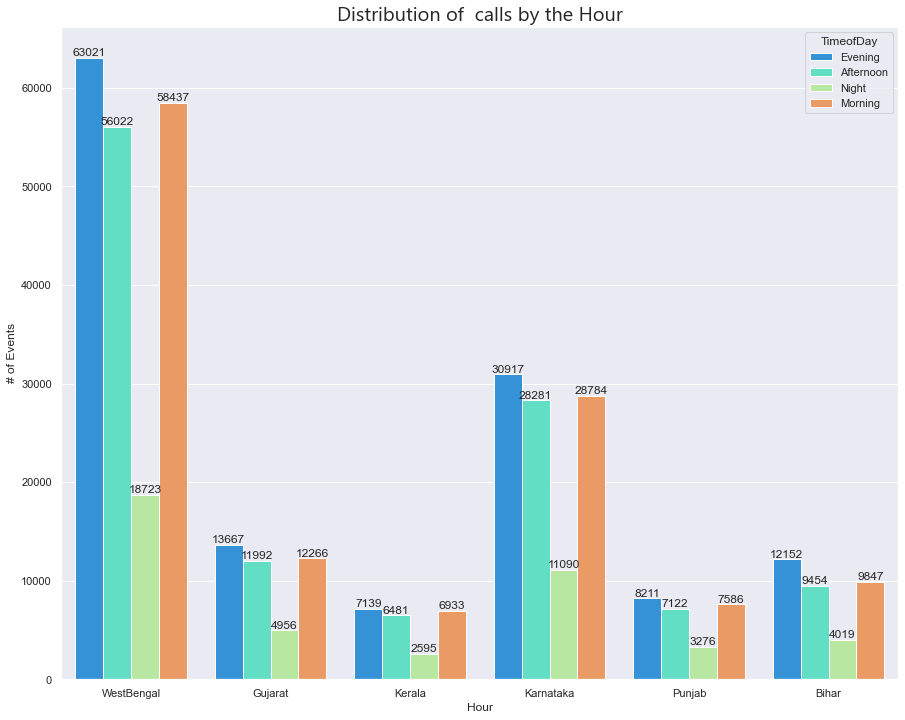
4. Frequency of Events with Respect to Time

The customers are most active at 10 AM in the morning and 9 PM in the evening.\

Least number of calls are made during 12- 4 AM

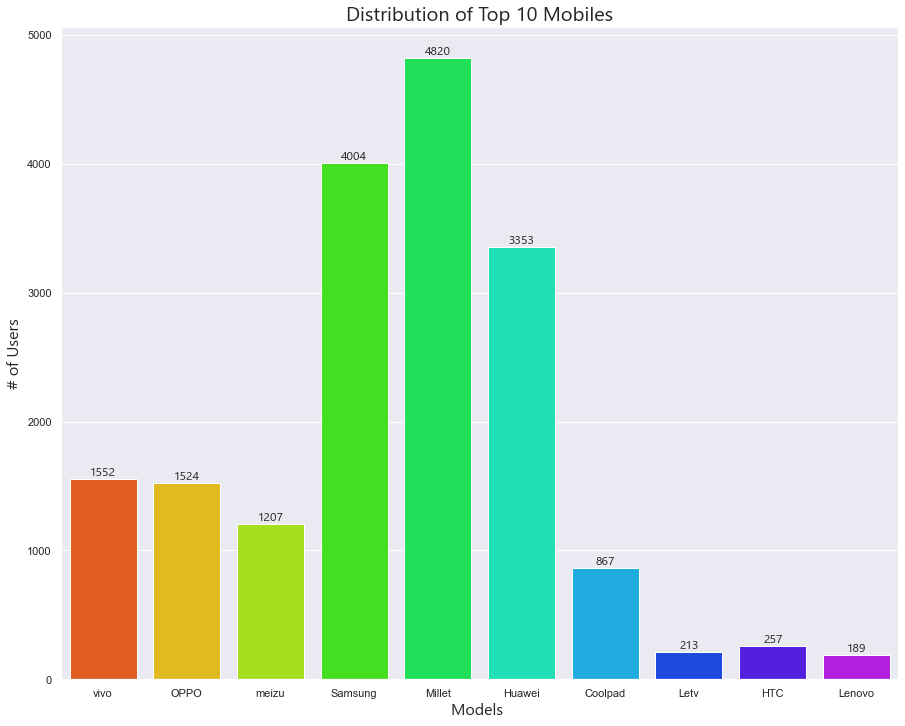


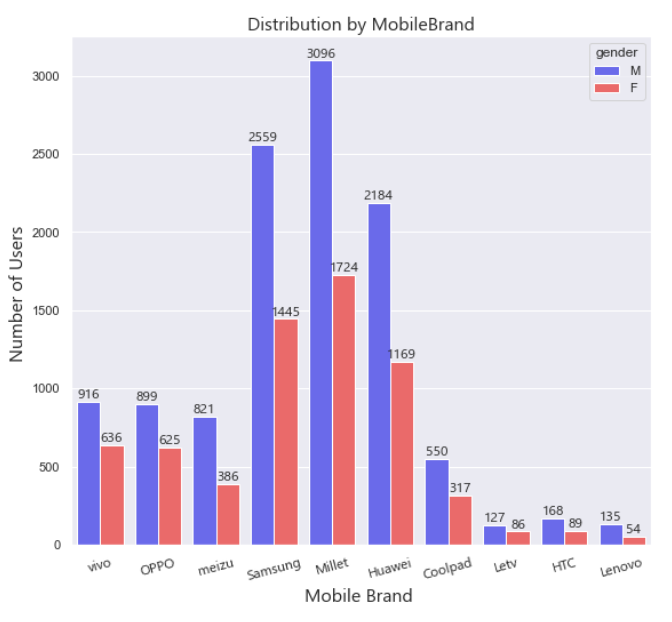


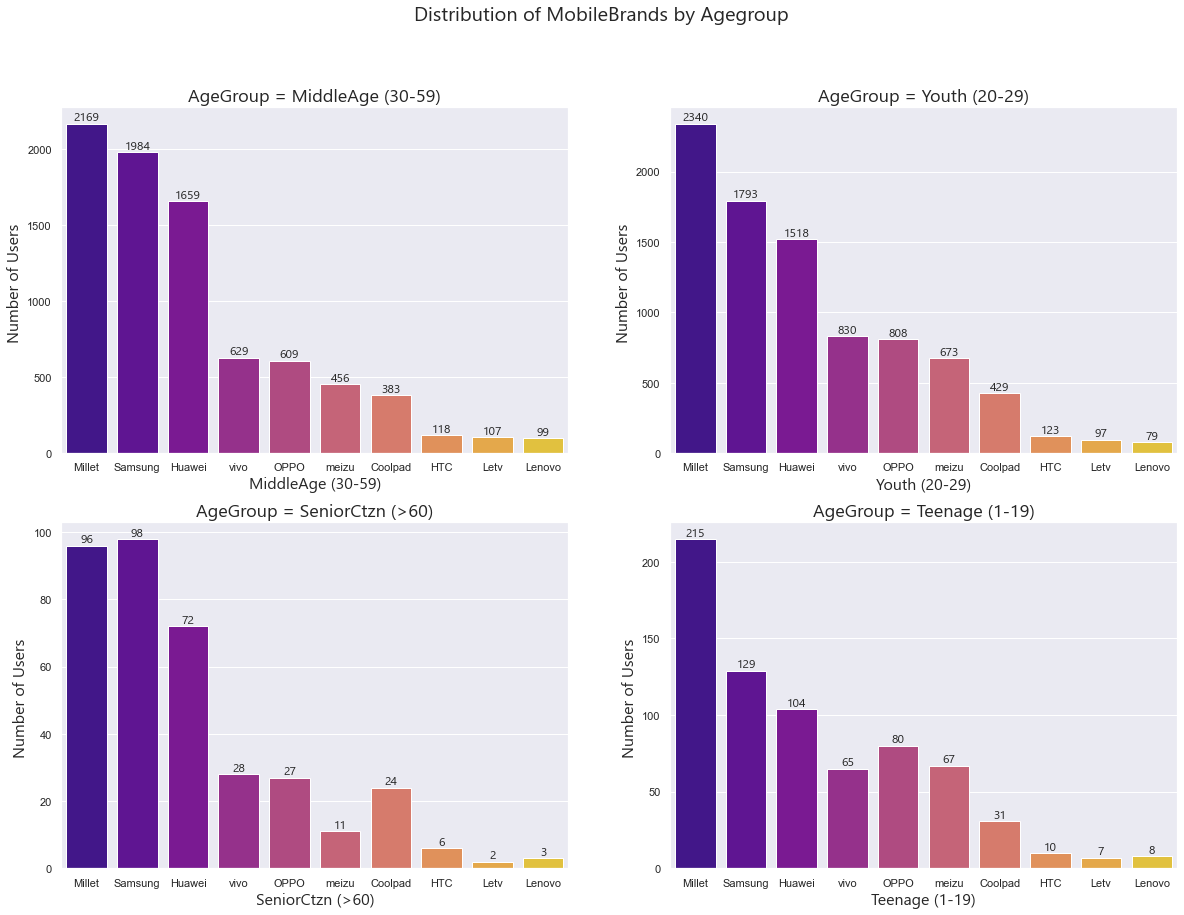


5.Distribution of Brand of Mobile Phones among Customers

Mobile brand “Millet” appears to be the first choice among customers.

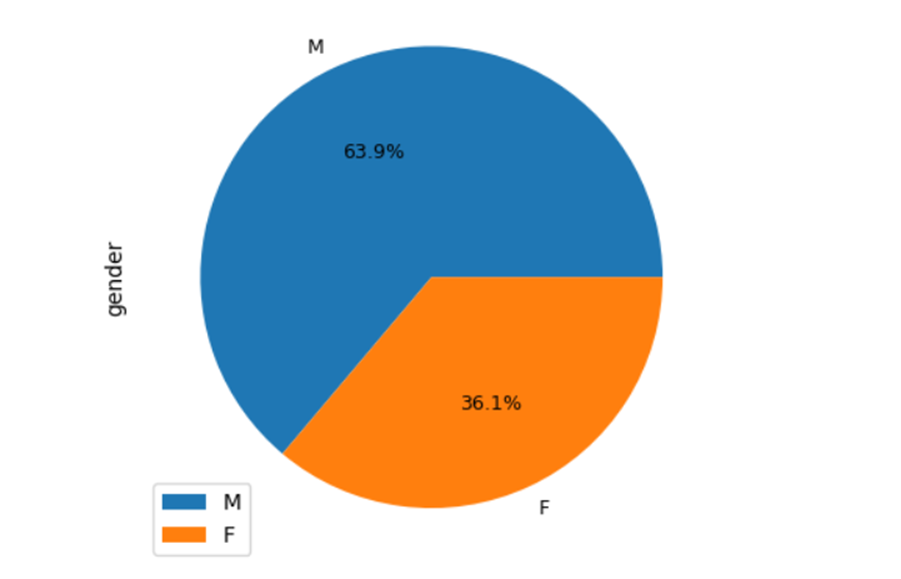


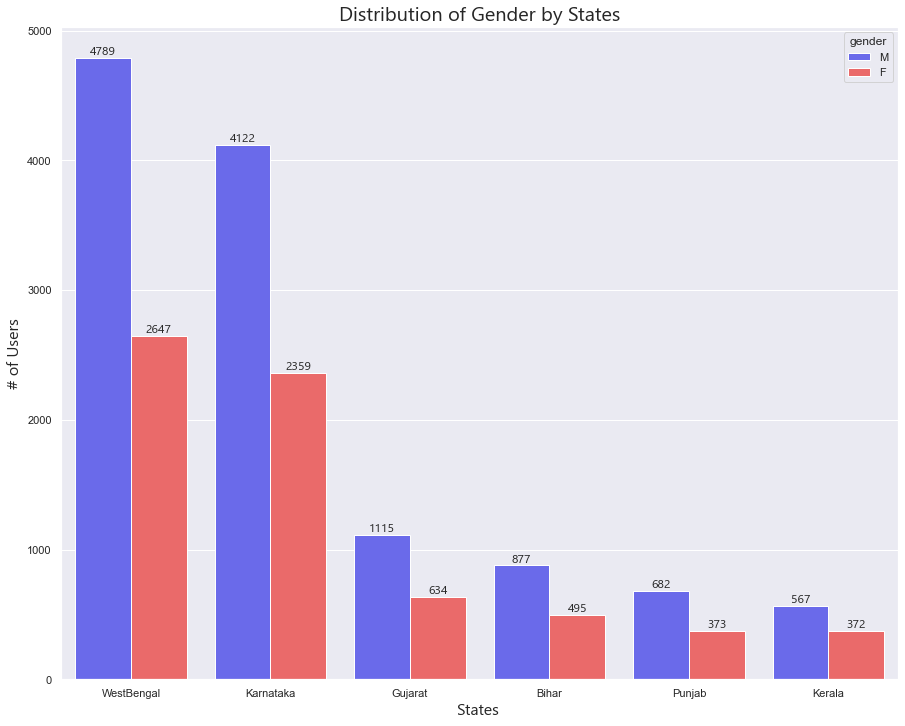
**

**

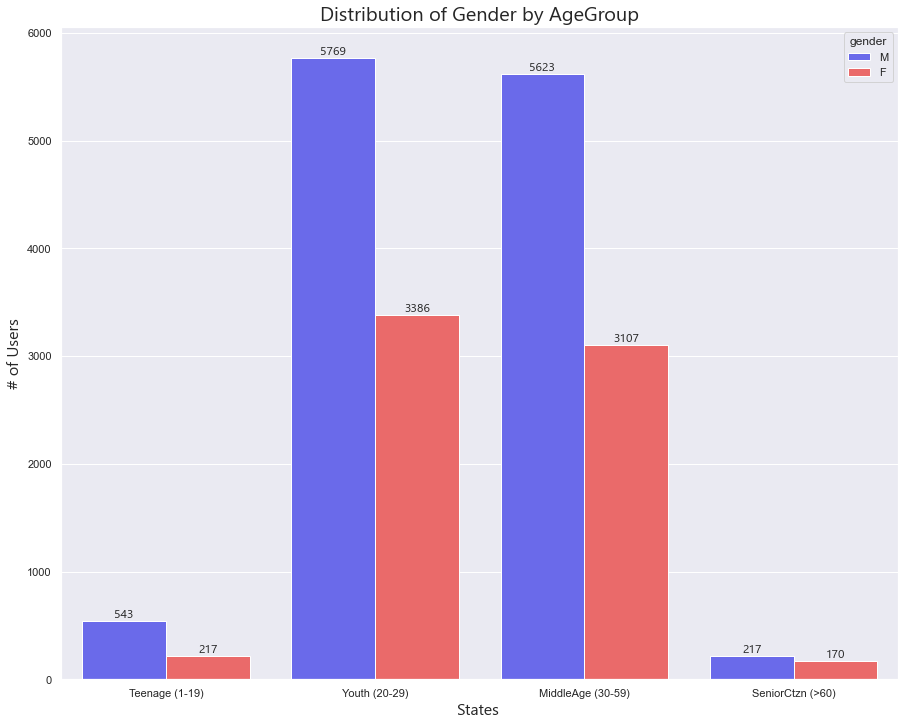
6. Mobile Phone users based on Gender

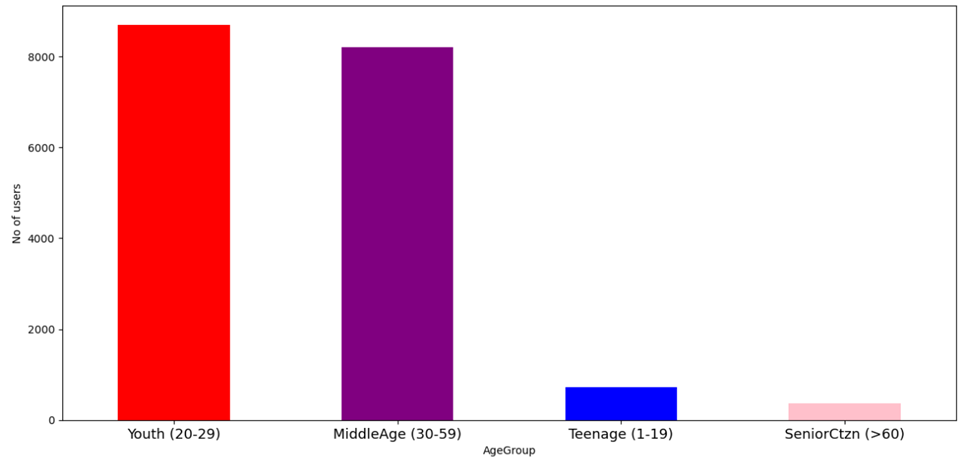
Male users are more in number than that of female users across all states





7. Distribution of different Age Groups





# *Summary:*

Following insights are drawn after analysing the given datasets.

1. There is huge gap between the number of male and female users. Female users are 57% less than male users and this provides a great opportunity to increase the female user count.
2. The maximum number of users are youngsters in age group 20-29 where as the senior citizens show tremendous decline in the usage. There is a great amount of potential in increasing the senior citizen user count by providing the services based on their needs
3. “Millet” is the most favorite mobile brand across users of all age group and gender whereas Lenovo ranks as the least favorite brand.
4. Most of the calls are made at 10AM and 9 PM. The off-peak offers during 12 - 4 AM time period can boost the call count throughout day.
5. Maharashtra tops in terms of number of users and Manipur is at the bottom of the list. The states like Orissa, Assam, Chhattisgarh and Uttaranchal shows scope for improving user base.