TELECOM INDUSTRY INVESTMENT

Market Analysis

CLIENT: INSAIDTELECOM | PREPARED BY: INSAIDIANS 1006

Table of Contents

INTRODUCTION

India is currently the second largest telecommunications market in the world with a subscriber base of 1.16 billion as of March 2021. The telecom economy is growing rapidly in the last decade and is expected to contribute substantially to India's Gross Domestic Product (GDP) as per a report prepared by GSM Association (GSMA) in collaboration with Boston Consulting Group (BCG). India overtook the US to become the second largest market in terms of the number of app downloads in 2019.

Government of India's policies and reforms such as the deregulation of Foreign Direct Investment (FDI) norms along with the increasing consumer demand has helped telecom emerge as the leading sector in the Indian market. Easy access to telecom equipment and an agile regulatory framework by the Government has ensured the availability of the telecom services to consumers at affordable prices.

MARKET SIZE

India has the world's second largest market of total internet users. The number of internet subscribers in the country increased at a CAGR of 21.36% from FY16 to FY20 to reach 743.19 million in FY20. The total subscriber base in the country stood at 1,173.83 million, as of December 31, 2020.

Gross revenue of the telecom sector stood at Rs. 68,228 crore (US$ 9.35 billion) in the third quarter of FY21.

Over the next five years, rise in mobile-phone penetration and decline in data costs will add over **650 million new internet users in India**, creating opportunities for new businesses.

INVESTMENT POTENTIAL

It is estimated that by 2023, there would be over 650 million internet users in the country. With a daily increasing subscriber base, there is a lot of investment and development potential in the sector. FDI inflow into the telecom sector during April 2000 – December 2020 totaled US$ 37.62 billion according to the data released by Department for Promotion of Industry and Internal Trade (DPIIT).

These factors are a basis for the client INSAIDTELECOM, one of the leading telecom players in the country to have a closer look at this market and accordingly plan their strategy for the coming 5 years to capture this market effectively and also have a data led approach to understand the consumers and their mobile usage pattern to help in investing towards the right things.

Project Description

INSAIDTELECOM, one of the leading telecom players, understands that customizing offerings is very important for its business to stay competitive. Hence, INSAIDTELECOM approached INSAIDIANS 1006 Consulting Pvt. Ltd. to better understand and leverage from the behavioral data of over 30 million mobile devices active daily in India to help its clients better understand and interact with their audiences.

PROJECT TEAM MEMBERS

Name	Email-id
Sridhar Emmadi	sridhar.emmadi@gmail.com
Hemanth Kasturi	kasturihemanth3392@gmail.com
Sanjay Dubey	sanjay261289@gmail.com
Sonal Kulkarni	sonal.kulk@gmail.com
Srirama Pavan Kasibhatla	mr.kpavan@gmail.com
Debanjana Basu	bdebanjana@gmail.com
Kiran Dhanwada	dhanwadakiran@gmail.com
Rahul Thakare	rahulthakare26jan@gmail.com
Prasanna Venkatesan Narayanan	venkatesan84@yahoo.co.in
Ambrish Choudhary	ambrish.choudhary@gmail.com
Gautam Gahlaut	gautam.gahlaut77@gmail.com
Neha Awasthi	ultimatenehu@gmail.com
Piyush	piyush.mishra90@gmail.com
T C P ELAVARASU	tcp.elavarasu@gmail.com
Sarika Kushwaha	sarika.kushwaha0511@gmail.com
Renuka Sachdev	r3nuka.sachdev@gmail.com
Karthik Sheggoju	karthik.sheggoju@gmail.com
Salil Shaikh	shaikh.salil@gmail.com
Gaurav Virmani	virmani_gaurav@yahoo.com
Ashin Sasidharan	sasidharanashin@gmail.com
Ravi Gopalkrishna	ravig.pmp@gmail.com
Sarita Singh	singhsarita02@gmail.com
Keith DSouza	keith.dsouza.80@gmail.com
Rohit Pratapwar	pratapwarrohit13@live.com

PROBLEM STATEMENT

Understanding over 30 million mobile device user's demographic characteristics based on their mobile usage, geolocation, and mobile device properties to help developers and brand advertisers to pursue data driven marketing efforts to increase its consumer base.

PROBLEM ANALYSIS

A thorough analysis of the available data of the mobile usage patterns by location of around 30 million mobile users in India was carried out and insights are drawn based on the understanding of the available robust data. A dashboard is built based on the findings, to reflect various demographic characteristics of users which would facilitate developers and brand advertisers to pursue their marketing efforts catering to user preferences.

The team at INSAIDIANS 1006, followed the following strategy in deriving insights from this large set of data:

1. Studying the data available and understanding the challenges in the data
2. Pre-processing and cleaning the data to resolve the challenges observed
3. Preparing analytics for the pre-processed data
4. Preparing analytics dashboard
5. Preparing consulting report with strategies from the analytics

SOURCES OF DATA

INSAIDTELECOM provided three major datasets consisting of location wise distributed data of 30 million users including their demography (age group, gender) as well as data on the phone brands used by them and the usage patterns for the year 2016. A total of three datasets were provided separately to the INSAIDIANS 1006 team:

1. **events_data** from Github repository which talks about the number of events involved per device which provides us with the usage pattern of the 30 million customers
2. **gender_age_train** from SQL database which provides the demographic data of the customers
3. **phone_brand_device_model** from SQL database which provides the devices used by these customers.

The "mysql.connector" package was used to extract data from SQL database.
The data is collected from mobile apps that use INSAID Telecom services. Full recognition and consent from individual users of these apps have been obtained and appropriate anonymization have been performed to protect privacy.

Below is brief description of each of the datasets:

1. **events_data dataset:** When a user uses mobile on INSAID Telecom network, the event gets logged in this data. Each event has an event id, location (latitude/longitude), and the event corresponds to frequency of mobile usage. The timestamp is the moment when the user is using the mobile. Sample data points is as below:

	event_id	device_id	timestamp	longitude	latitude	city	state
0	2765368	2.973348e+18	2016-05-07 22:52:05	77.225676	28.730140	Delhi	Delhi
1	2955066	4.734221e+18	2016-05-01 20:44:16	88.388361	22.660325	Calcutta	WestBengal
2	605968	-3.264500e+18	2016-05-02 14:23:04	77.256809	28.757906	Delhi	Delhi
3	448114	5.731369e+18	2016-05-03 13:21:16	80.343613	13.153332	Chennai	TamilNadu
4	665740	3.388880e+17	2016-05-06 03:51:05	85.997745	23.842609	Bokaro	Jharkhand

2. gender_age_train dataset: Device IDs and their respective user gender, age and age_group. Sample data points is as below:

	device_id	phone_brand	device_model
0	1877775838486905855	vivo	Y13
1	-3766087376657242966	小米	V183
2	-6238937574958215831	OPPO	R7s
3	8973197758510677470	三星	A368t
4	-2015528097870762664	小米	红米Note2

3. phone_brand_device_model dataset: Device ids, brand, and models phone_brand. Sample data points is as below:

2	-8260683887967679142	M	35	M32-38
3	-4938849341048082022	M	30	M29-31
4	245133531816851882	M	30	M29-31

SUMMARY OF DATA MINING CHALLENGES

Following were the discrepancies found in the datasets and the solution to it:

Sr. No.	Challenge	Resolution
1	81 nos. of missing values in device_id in events_data	The missing values were found to be from Jaipur city. Hence, the missing value was replaced with mode of device id for Jaipur city
2	66 nos. of missing values in location (latitude/longitude) in events_data	The missing values were found to be from Jaipur city. Hence, the missing value was replaced with mode of respective latitude & longitude values for Jaipur city
3	6 nos. of data points were outside India as observed from folium plot	The city for these points was mentioned as Jaipur and accordingly the latitude & longitude of these points were replaced by mode of respective values for Jaipur city
4	Some of the phone brand names were in Chinese	google_translator package was used to translate these Chinese names to English
5	Timestamp data was of object datatype instead of date-time	Pandas function "to_datetime" was used to convert the data type from object to date time format. Furthermore, the hour in which the event was created was extracted from this information which will be used for analysis

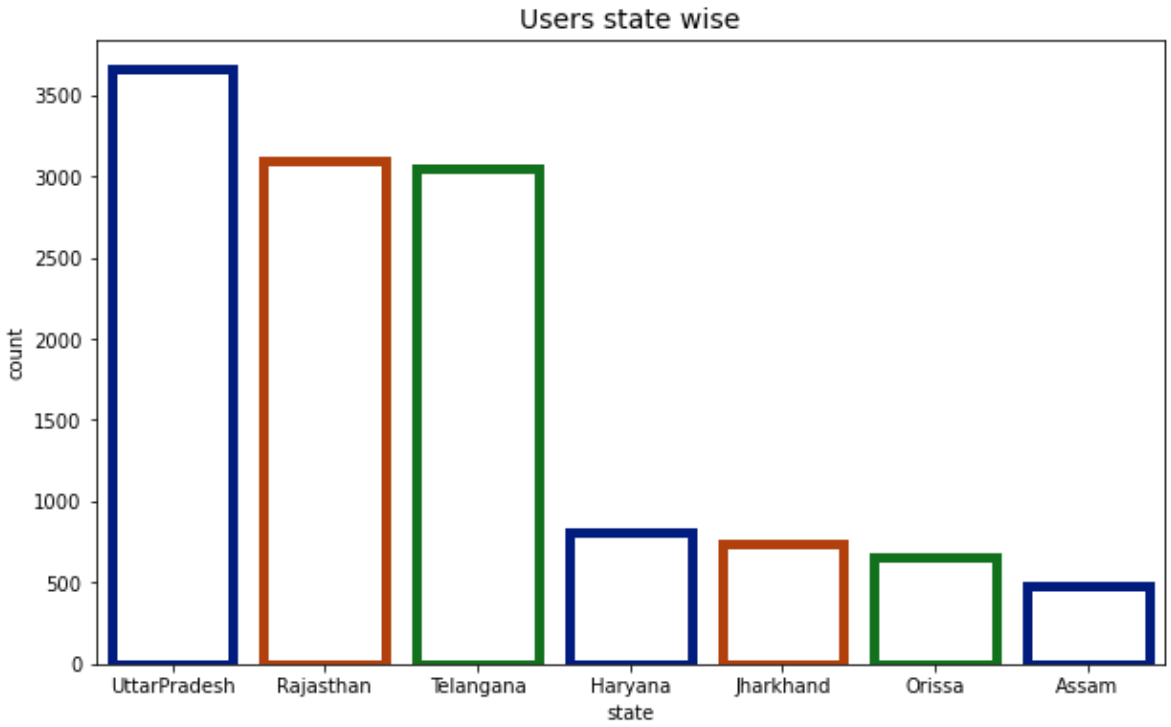
Following credentials were used to extract data from SQL database:

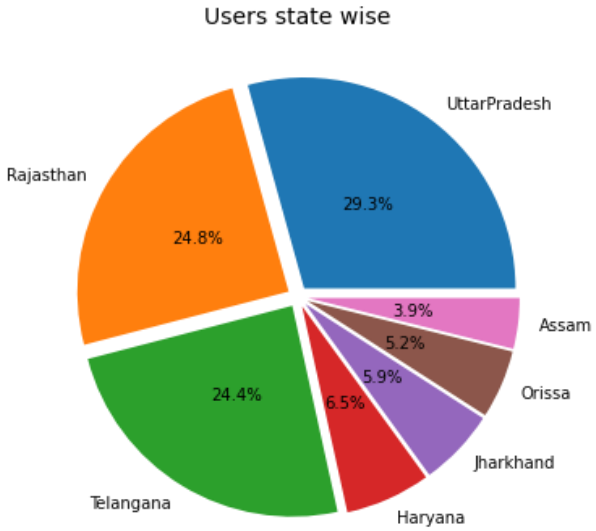
host	'cpanel.insaid.co'
user	'student'
password	'student'
database	'Capstone 1'

PROPOSED SOLUTIONS

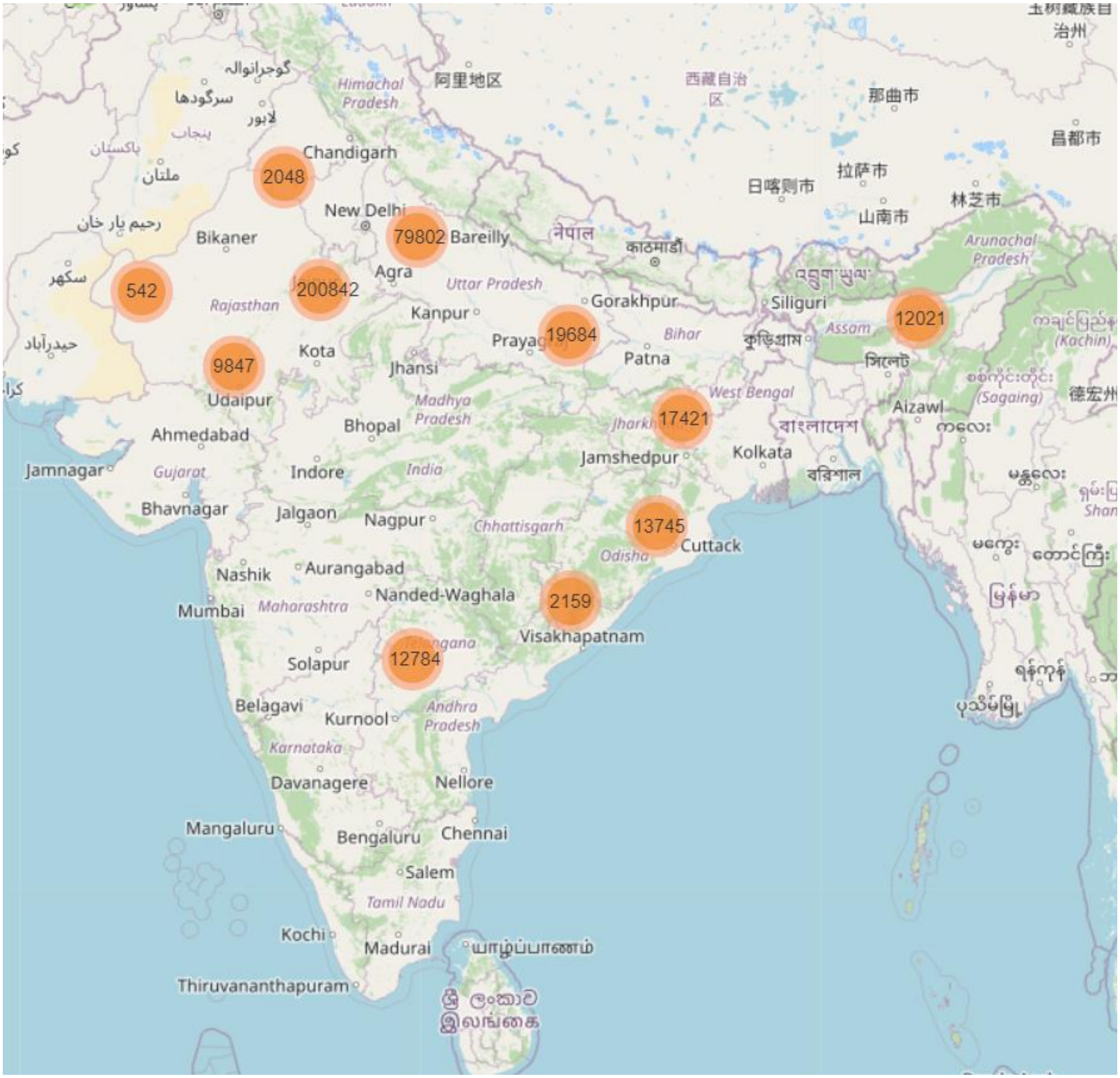
DATA ANALYTICS

1. State wise number of customers:



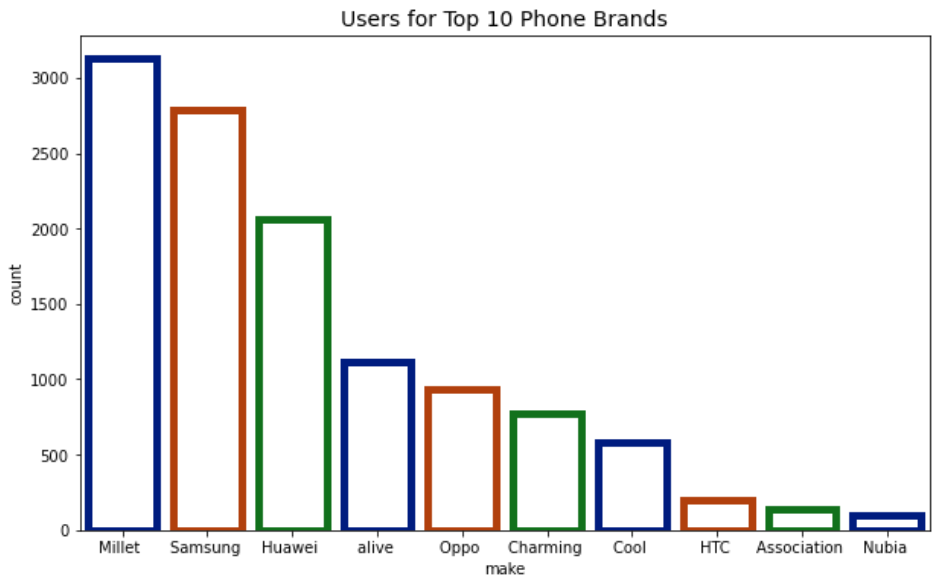


The three states of **Uttar Pradesh, Rajasthan & Telangana** comprises around **80% of the users**, the topmost being in Uttar Pradesh. This is mostly due to higher population in these states



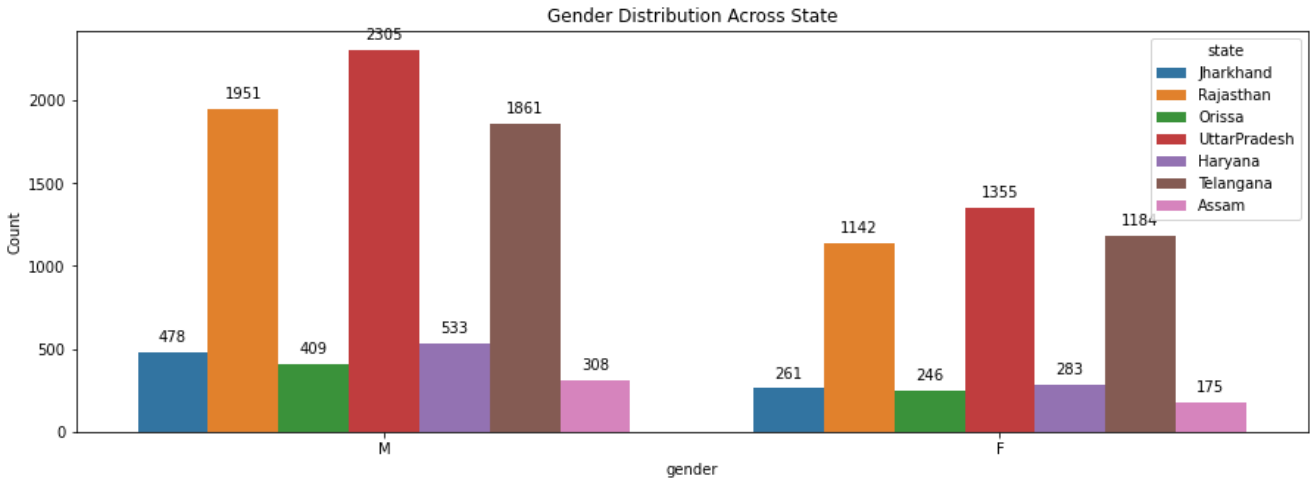
2048
79802
542
200842
9847
19684
12021
17421
13745
2159
12784
Chandigarh
New Delhi
Bareilly
Bikaner
Agra
Rajasthan
Uttar Pradesh
Kanpur
Gorakhpur
Kota
Bihar
Patna
Udaipur
Jhansi
Madhya Pradesh
Bhopal
West Bengal
Ahmedabad
Jamshedpur
Kolkata
Jamnagar
Gujarat
Indore
India
Bhavnagar
Jalgaon
Nagpur
Chhattisgarh
Cuttack
Odisha
Nashik
Aurangabad
Nanded-Waghala
Mumbai
Maharashtra
Visakhapatnam
Solapur
Belagavi
Kurnool
Andhra Pradesh
Karnataka
Davanagere
Nellore
Mangaluru
Bengaluru
Chennai
Salem
Tamil Nadu
Kochi
Madurai
Thiruvananthapuram
Himachal Pradesh
Siliguri
Assam
Arunachal Pradesh
Aizawl
(Kachin)
(Sagaing)
阿里地区
西藏自治区
那曲市
昌都市
拉萨市
日喀则市
山南市
林芝市
नेपाल
काठमाडौं

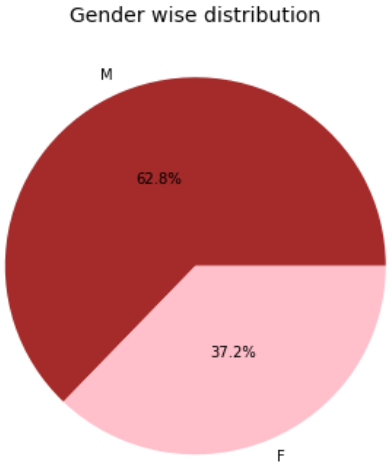
2. Top 10 Phone brands:



Millet, Samsung & Huawei are the most preferred phone brands among the users.

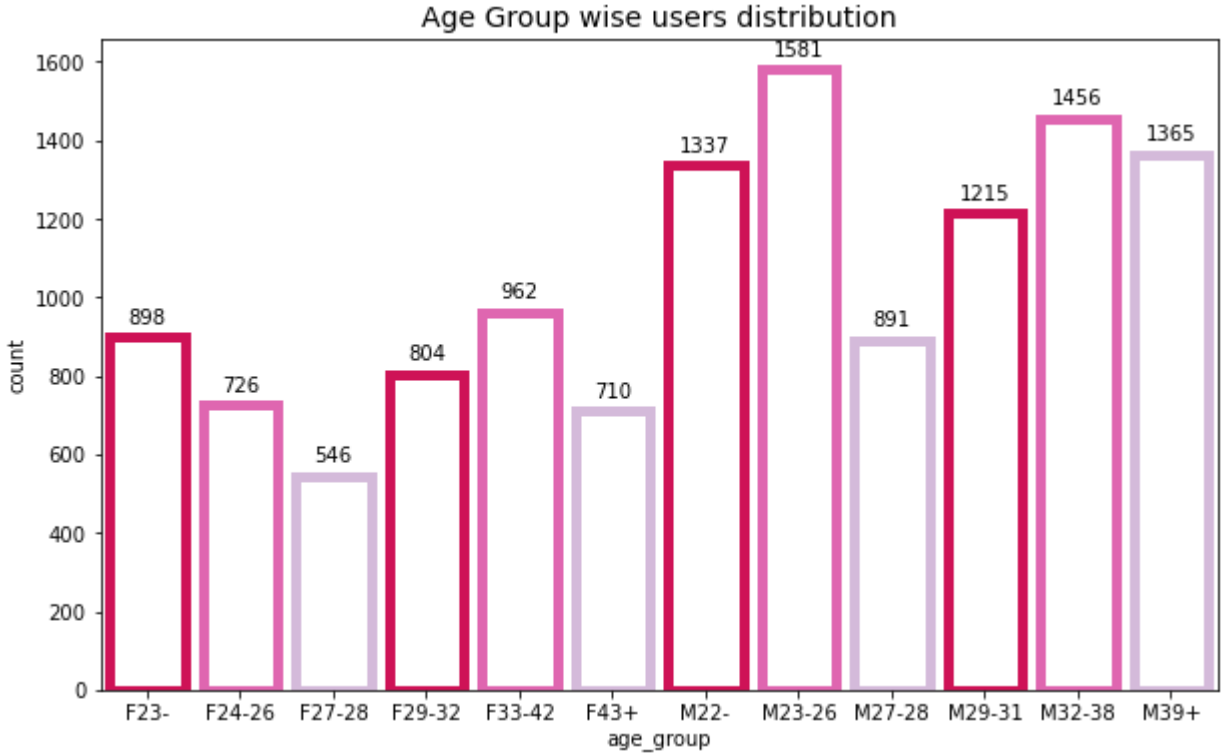
3. Gender-wise Distribution:





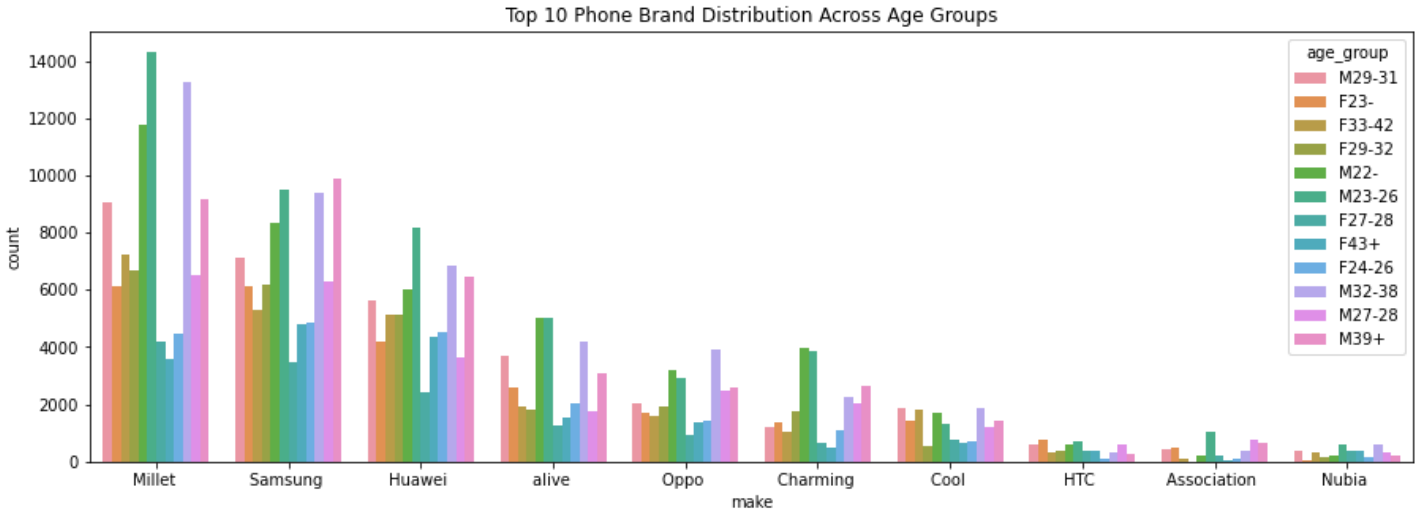
There are almost **double the number of male users** compared to female users. the users.

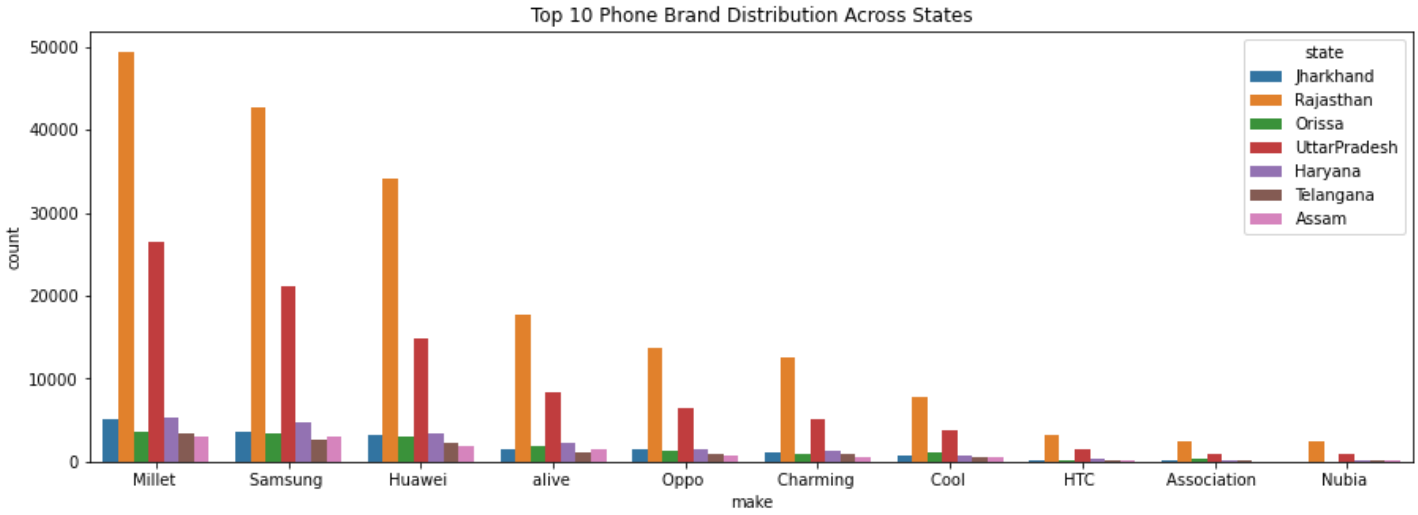
4. Age group-wise distribution:



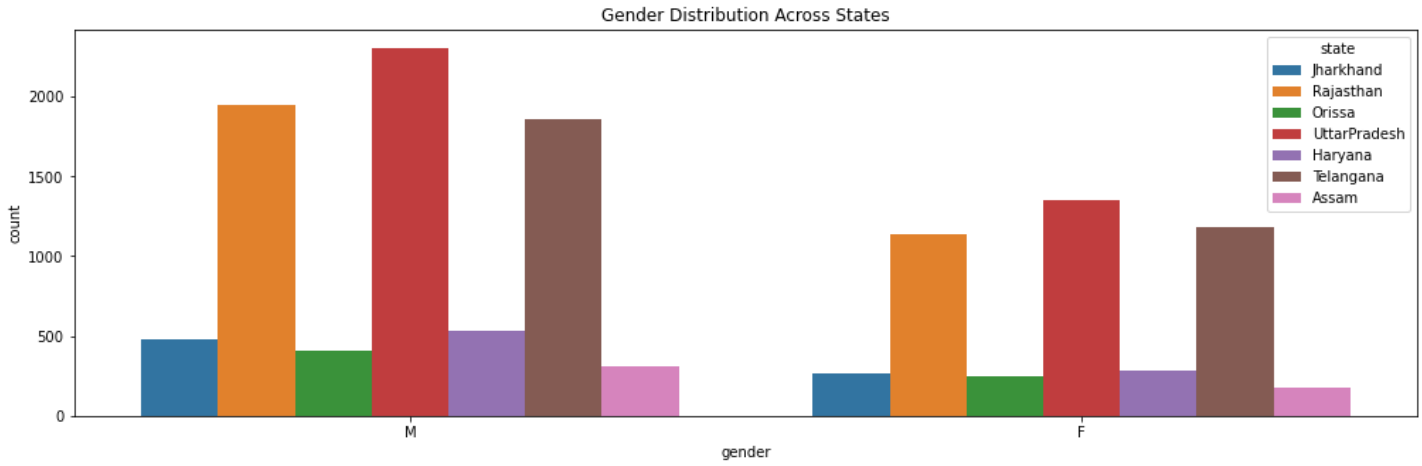
Large number of users are between age of **23-30 years**.
We have a significant number of users under the age of 23 years for both genders.
We have close to **2000 users above the age of 39 years combining both genders** which is ~16% of total population in these 7 states.

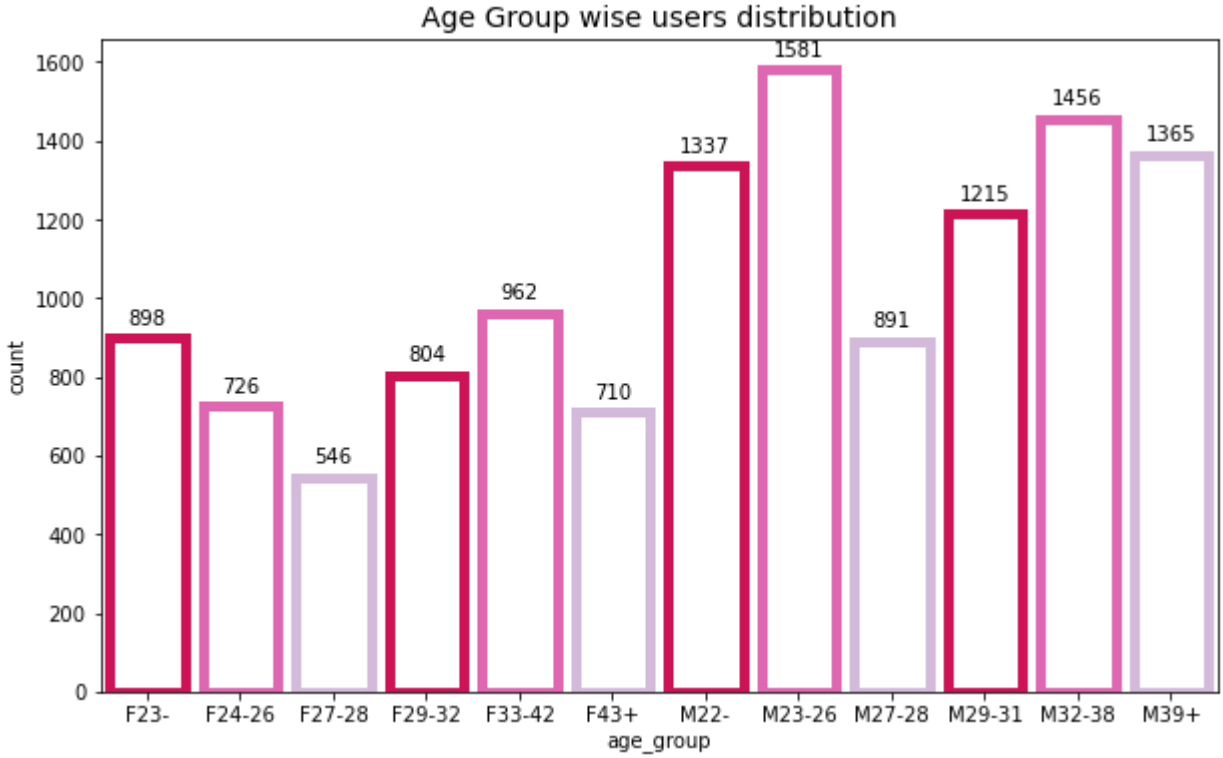
5. Distribution of Phone brands across Age segments / State / Gender:

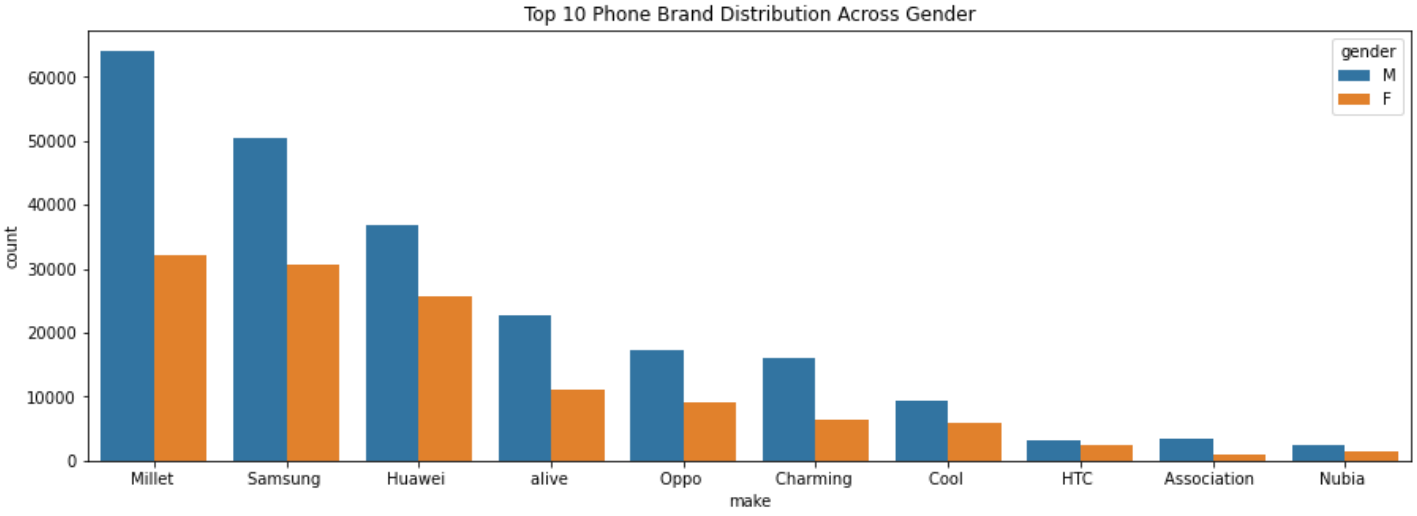




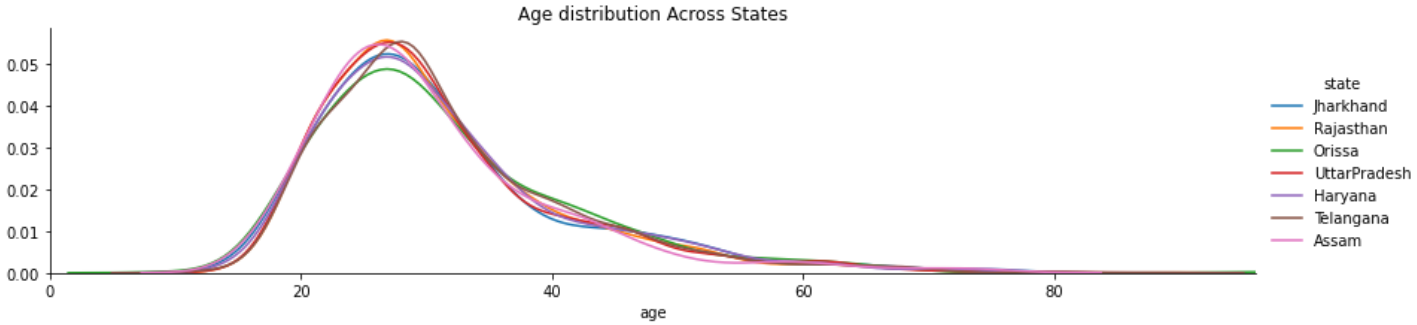
6. Gender distribution for State / Age group & Top 10 Phone Brands:

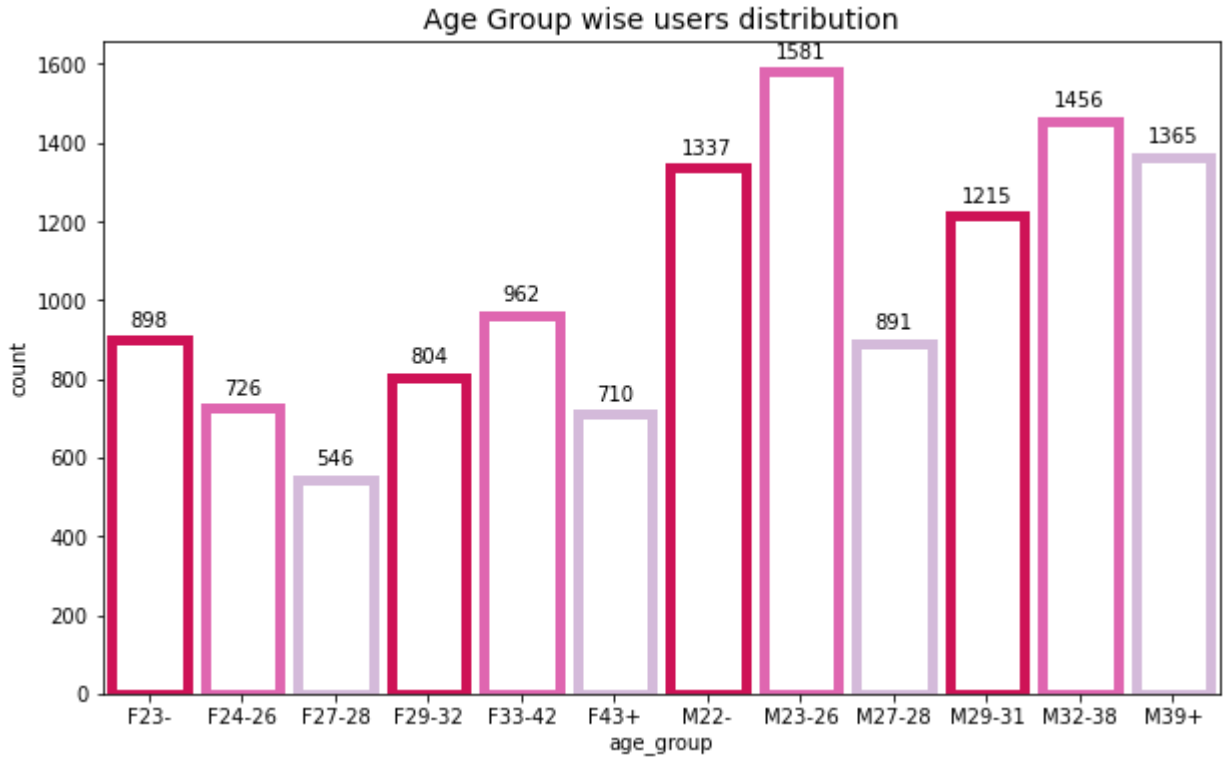


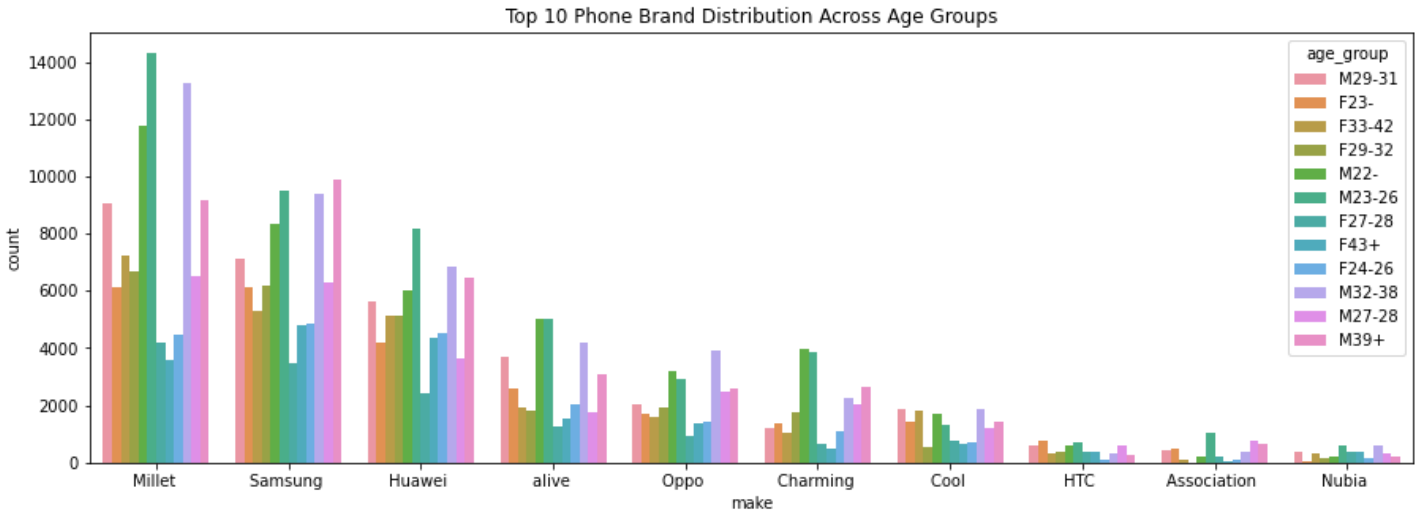




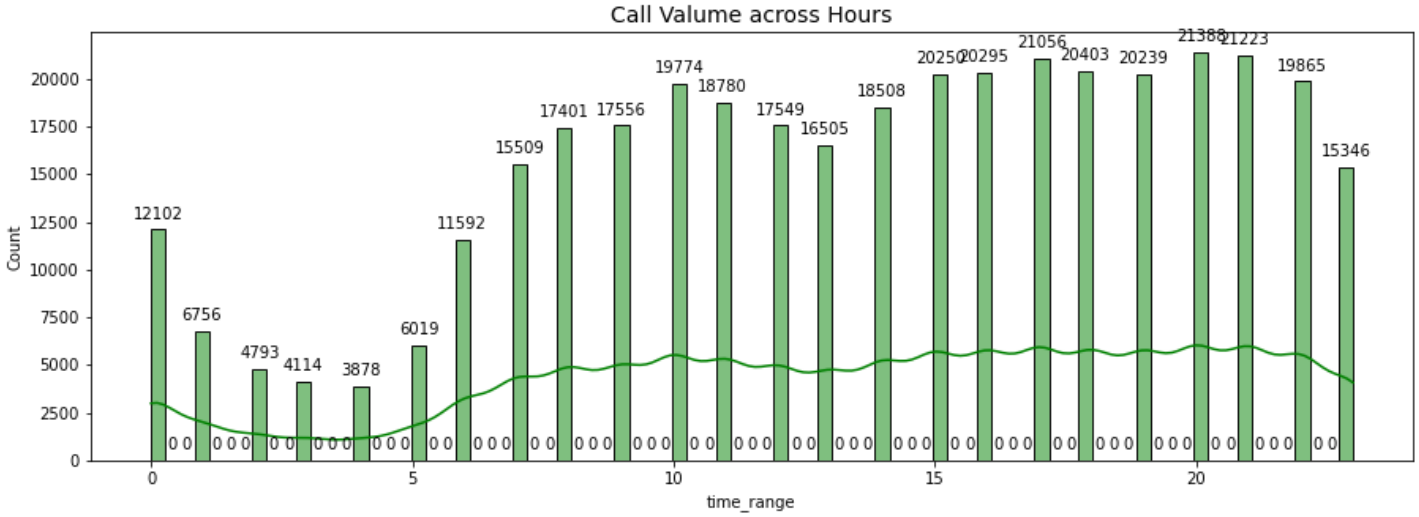
7. Age distribution across State / Gender & Top 10 Phone Brands:







8. Hourly distribution of Phone Calls:



There is a **peak period between 8.00am to 10.00am** and then again from **3.00pm to 10.00pm**.
There is **less demand during nighttime (10.00pm to 7.00am)** and during afternoon time **(12.00pm to 2.00pm)**

S T R A T E G I E S

- Refer a Friend scheme can be deployed in the states having lower number of users i.e. Haryana, Jharkhand, Orissa & Assam. The person who referred will get a bonus data pack or talk-time worth one month or so.
- The telecom company can partner with Top 10 Phone brands like Millet, Samsung, Huawei, Alive, Oppo, Charming and provide offers like free data for a few months in the beginning as well as full year talk-time at a discounted price.
- Offers can be provided to stream TV daily soaps, regional drama, movies on mobile. This should increase the female number of users as well as users over 40 years of age.
- The telecom company can announce happy hours between 11.00pm to 7.00am and 12.00pm to 3.00pm during which the calling rates will be significantly low. This will result in reduced load during peak hours improving system maintenance
- To increase user interaction for users above the age of 40 years in these 7 states, offers to be sent to users in their regional language.

TOOLS USED

Insaidians 1006 have created this consulting report using various data science and Web UI tools which helped in not only warehousing, cleaning and preparing data analytics but also in visualizing this data through bars and charts as well as a comprehensive dashboard for better understanding. The various tools used by Insaidians 1006 are:

1. **Data Science Tools:** Python, MySQL, Numpy, Pandas, Seaborn, Folium packages for mapping, etc.
2. **Web UI Tools:** This was used to create the dashboard
    a. PHP
    b. JavaScript

CONCLUSION

INSAIDTelecom has a strong user base from the highly populated Indian states. Also, among these the data shows that the company has a good foothold among the youth in these states between the age of 20-30 years.
While the reach is good, to further improve the sales, INSAIDTelecom should adopt various target-based strategies to capture some of the user base that currently do not exist in its customer base like the 40 years and above age group.
It should also concentrate on increasing its foothold amongst the women in these states using the strategies provided in this report.

In conclusion, a targeted approach to create different schemes can be looked upon. Additionally, INSAIDIANS 1006 will be able to prepare a further detailed clustering model to help INSAIDTelecom with identifying schemes.