CDF CAPSTONE PROJECT

INSAID TELECOM SERVICES

Presented by – Group 1007

INTRODUCTION

India is one of the biggest consumer of data worldwide. As per TRAI, average wireless data usage per wireless data subscriber was 11 GB per month in FY20. It is expected to reach to 18 GB by 2024.

Currently, India is the world's second-largest telecommunications market with a subscriber base of 1.16 billion and has registered strong growth in the last decade. The Indian mobile economy is growing rapidly and will contribute substantially to India's Gross Domestic Product (GDP) in the foreseeable future, according to a report prepared by GSM Association (GSMA) in collaboration with Boston Consulting Group (BCG). In 2019, India surpassed the US to become the second largest market in terms of number of app downloads.

***Insaid Telecom,** one of the leading telecom players, understands that customizing offering is very important for its business to stay competitive and it ensures customers feel attached to their services.*

INSAID TELECOM

PROJECT DESCRIPTION

- ***Insaid Telecom*** *is seeking to leverage behavioral data from more than 60% of the 50 million mobile devices active daily in India to help its clients better understand and interact with their audiences.*
- *This will help millions of developers and brand advertisers around the world pursue data-driven marketing efforts which are relevant to their users and catered to their preferences.*

INSAID TELECOM

PROBLEM STATEMENT

- *In this consulting assignment, we **team 1007** are expected to study and analyze the **user behavior and their demographics** for the 6 states i.e. **West Bengal, Karnataka, Bihar, Punjab, Gujarat and Kerala.***

- ***Build a dashboard** to understand users demographic characteristics based on their mobile usage, geolocation, and mobile device properties, and **get insights** for marketing and product terms.*

PROBLEM ANALYSIS

The team will analyze and answer following questions to understand the pattern in the data and provide insights:

- **How is the distribution of Users across States.**
- **How is the distribution of Users across Phone brands.**
- **How is the distribution of Users across Gender.**
- **How is the distribution of Users across Age segments.**
- **How is the distribution of Phone brands for each Age Segment, State, Gender.**
- **How is the distribution of Gender for each State, Age Segment and Phone Brand.**
- **How is the distribution of Age Segments for each State, Gender and Phone Brand.**
- **Hourly distribution of Phone Calls.**

EDA PROCESS

TOOLS USED

DS tools

- **Pandas:** *for data manipulation and analysis*
- **Numpy:** *for high-level mathematical functions*
- **Pandas profiling:** *perform data profiling*
- **Matplotlib:** *popular plotting library used along with pandas*
- **Seaborn:** *a library, built on matplotlib, to create beautiful plots*
- **Google translator:** *API for converting Chinese brand and model names to English names*
- **Folium package:** *to plot the users on map*

WEB UI tools

- *PHP*
- *Java script*

DATA SOURCE

Dataset Name	Source	Records	Features
Events data	CSV file	3252950 (total) / 422971 (for 6 states)	7
Gender age	MySQL database	74645	4
Phone brand - device model	MySQL database	87726	3

Events Dataset:

- **when a user uses mobile on INSAID Telecom network, the event gets logged in this data.**
- **the event corresponds to frequency of mobile usage.**

ID	Feature Name	Feature Description
1	Event ID	Unique event id for each event
2	Device ID	Device Id of the user
3	Timestamp	Date and time of the event
4	Longitude	Longitude
5	Latitude	Latitude
6	City	Name of the city
7	State	Name of the State

Gender Age Dataset:

- **The connection to MySQL database is established through mysql.connector python package.**
- **It contains demographic details of the user.**

ID	Feature Name	Feature Description
1	Device ID	Device Id of the user
2	Gender	Gender of the user
3	Age	Age of the user
4	Age group	Age segments eg: 0-20, 21-30, 31-40, 41-50, 51-60, 60+

Phone Brand Model Dataset:

- **The connection to MySQL database is established through mysql.connector python package.**
- **Contains brand and model details.**

ID	Feature Name	Feature Description
1	Device ID	Device Id of the user
2	Phone brand	Brand name of the device
3	Device model	Model name of the device

DATA CLEANING

Cleaning operations on the data is performed based on observations from **data description, data information and profiling report.**

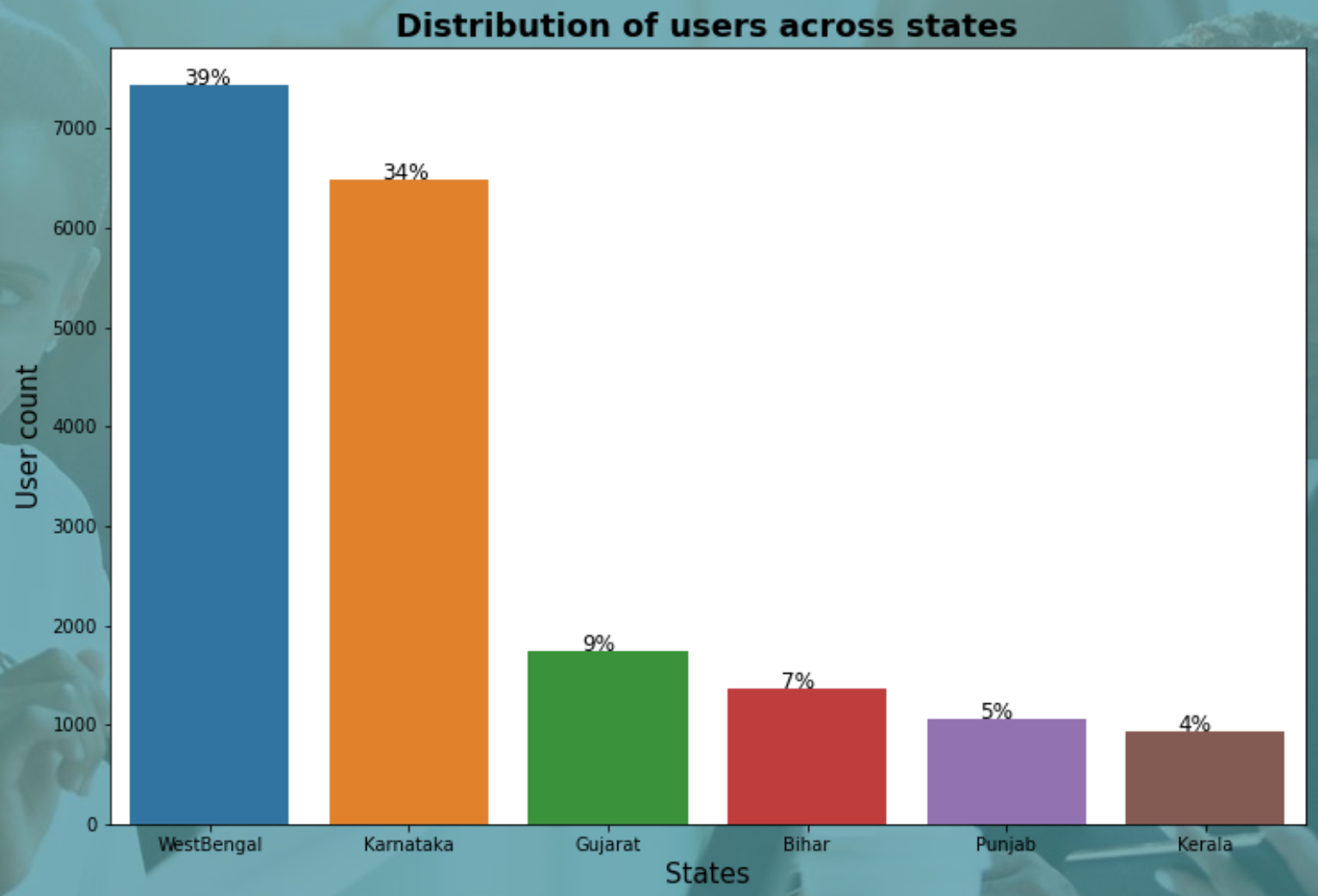
#	Observation	Resolution
1	Events dataset had 377 missing values in 'State' column	replaced missing state values w.r.t cities associated.
2	Events dataset had 453 missing values in 'Device Id' column	replaced missing device ids with device id values from its duplicate records using location information (longitude, latitude, city and state).
3	Events dataset had 423 missing values in 'latitude' and 'longitude' columns	replaced missing longitude and latitude values from its duplicate records using device id information.
4	Three records of Latitude and Longitude were corrupted	replaced the incorrect longitude and latitude values with the correct values from its duplicate records using device id information.
5	In Events dataset 'Timestamp' feature was stored as an object	converted timestamp to Datetime object for further analysis.
6	'In 'brand model' dataset phone brand and device model had names in Chinese characters	imported Google Translator from deep translator for translating Chinese brand and model names into corresponding English names.
7	The grouping of age segments were inconsistent among the two genders.	created a new categorical column ‘age group' and grouped the age into consistent groups between the two genders for further analysis.
8	Datatype of ‘device id’ was inconsistent between the 3 datasets	converted data type of ‘device Id' of ‘gender – age’ and ‘device – model’ datasets to same as ‘events’ dataset (from 'int64' to float64’).

DATA ANALYSIS

➢ How is the distribution of users across the 6 states?
(West Bengal, Karnataka, Punjab, Gujarat, Bihar and Kerala)

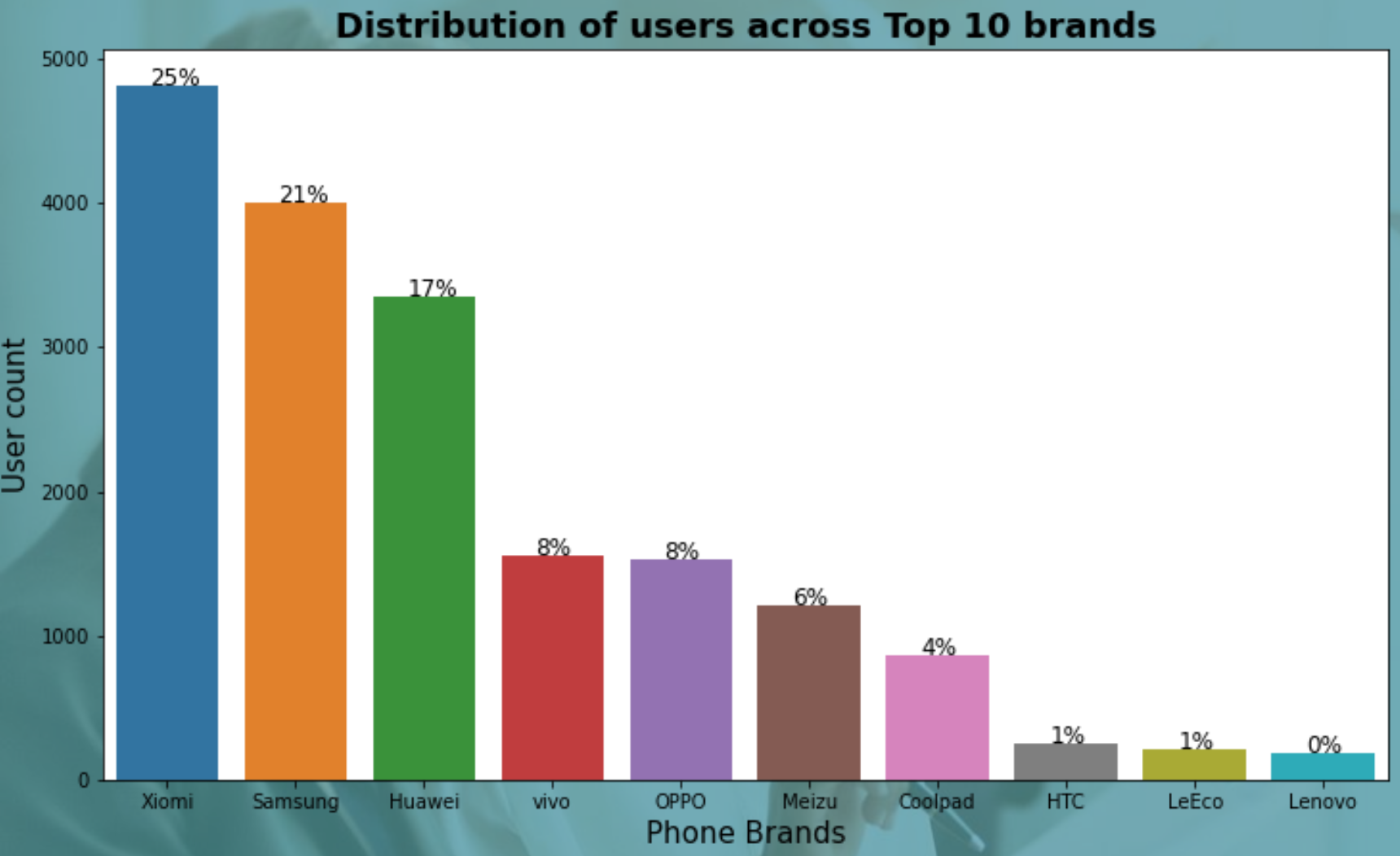
Observations:

- West Bengal has the highest number of users (39%) followed by Karnataka(34%).
- These 2 states contribute to around 73% of the total users.



DATA ANALYSIS

➢ How is the distribution of users across Phone brands (top 10)?

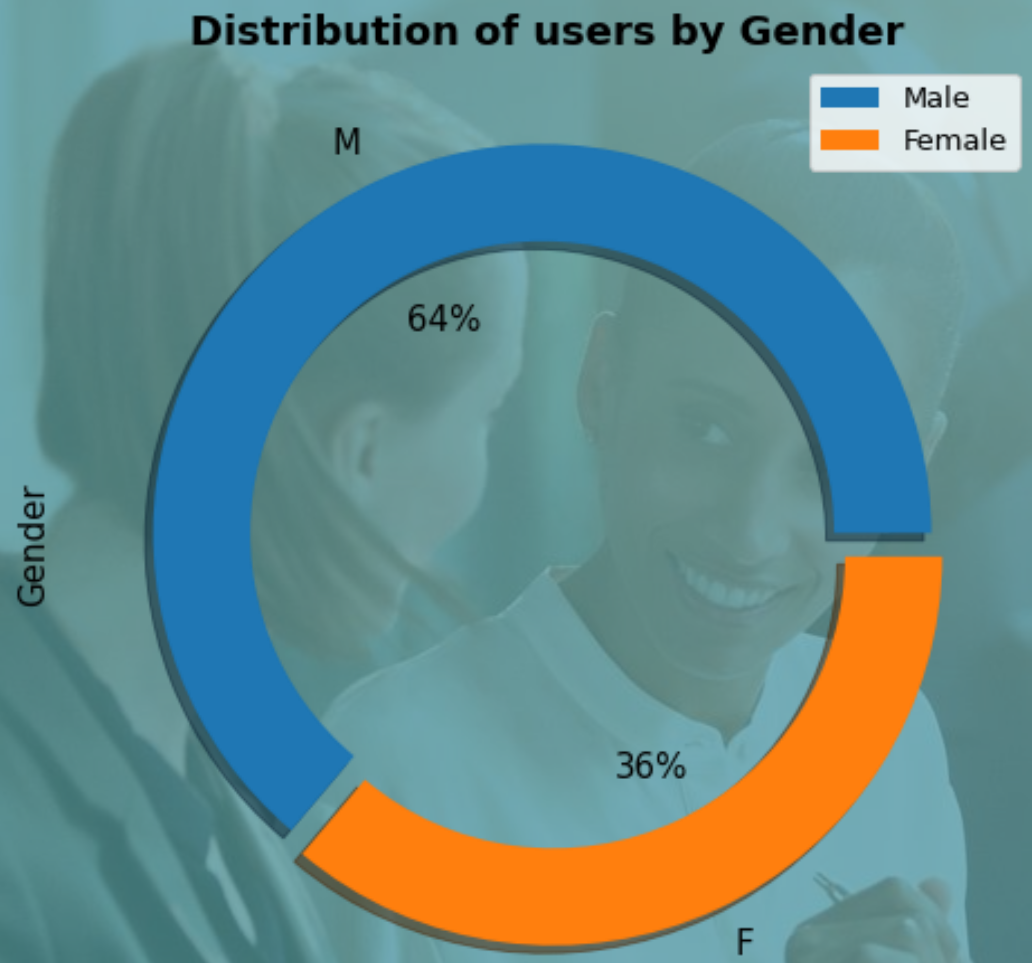


Observations:

- Xiaomi is the most popular phone brand (25%) across users followed by Samsung (21%) and Huawei (17%)

DATA ANALYSIS

➢ How is the distribution of users across Gender?



Observation:

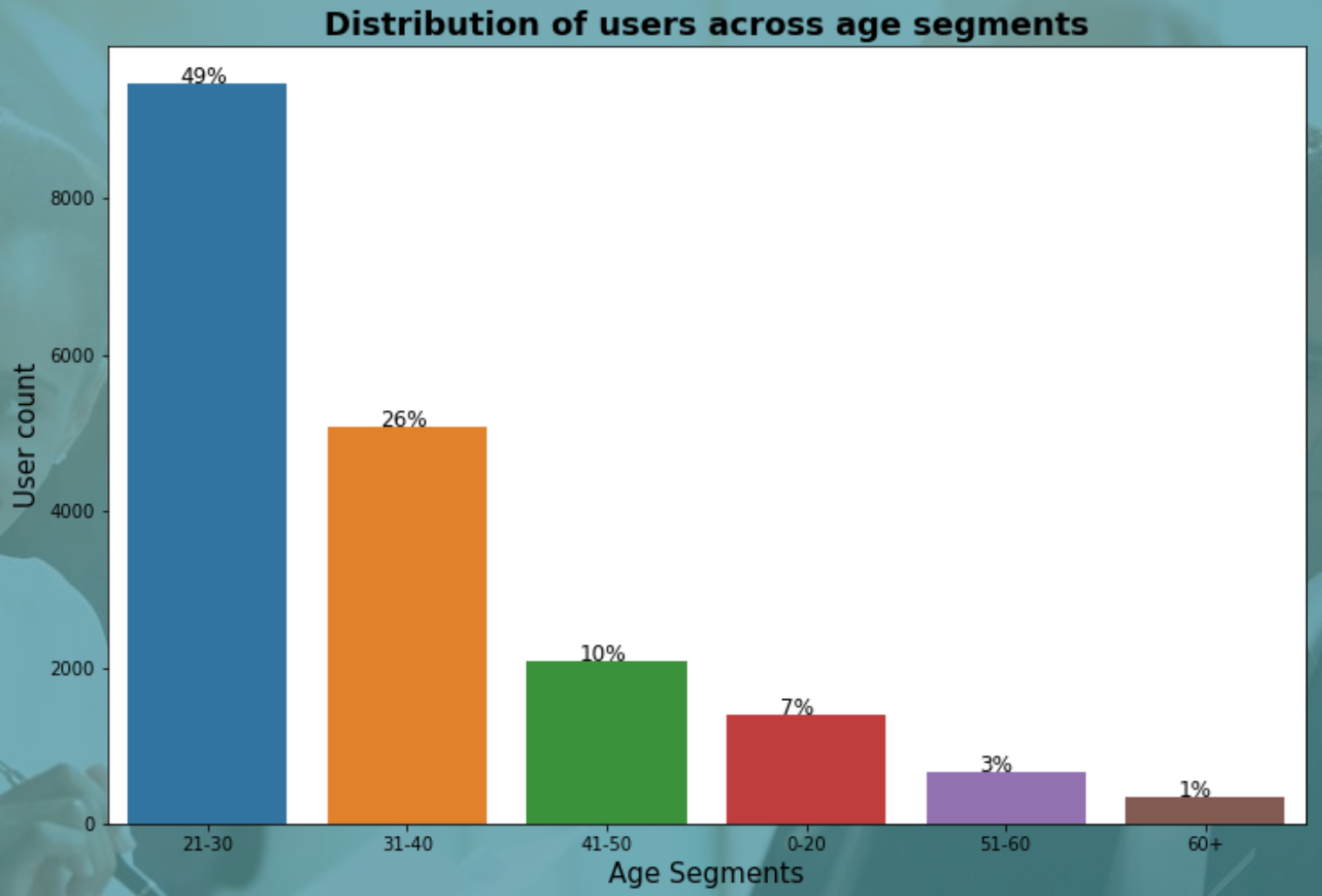
- Male users are more (64%) compared to female users.

DATA ANALYSIS

➢ How is the distribution of users across Age segments?

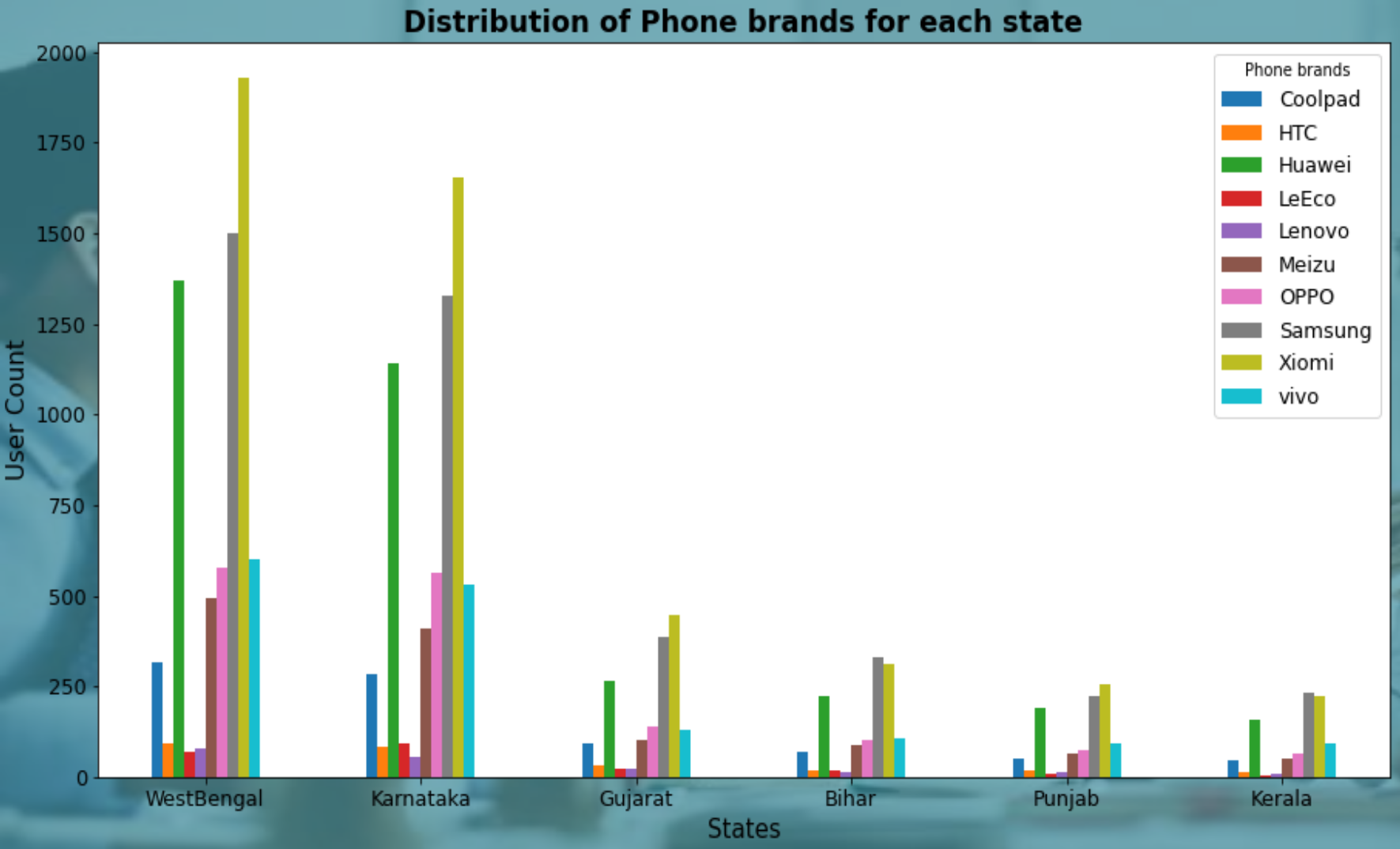
Observations:

- Around 50% of the users are in the age group 21-30 years followed by 25% users in the age group of 31-40 years.
- Teenage and elderly users are the least.



DATA ANALYSIS

➢ How is the distribution of Phone brands across each state?

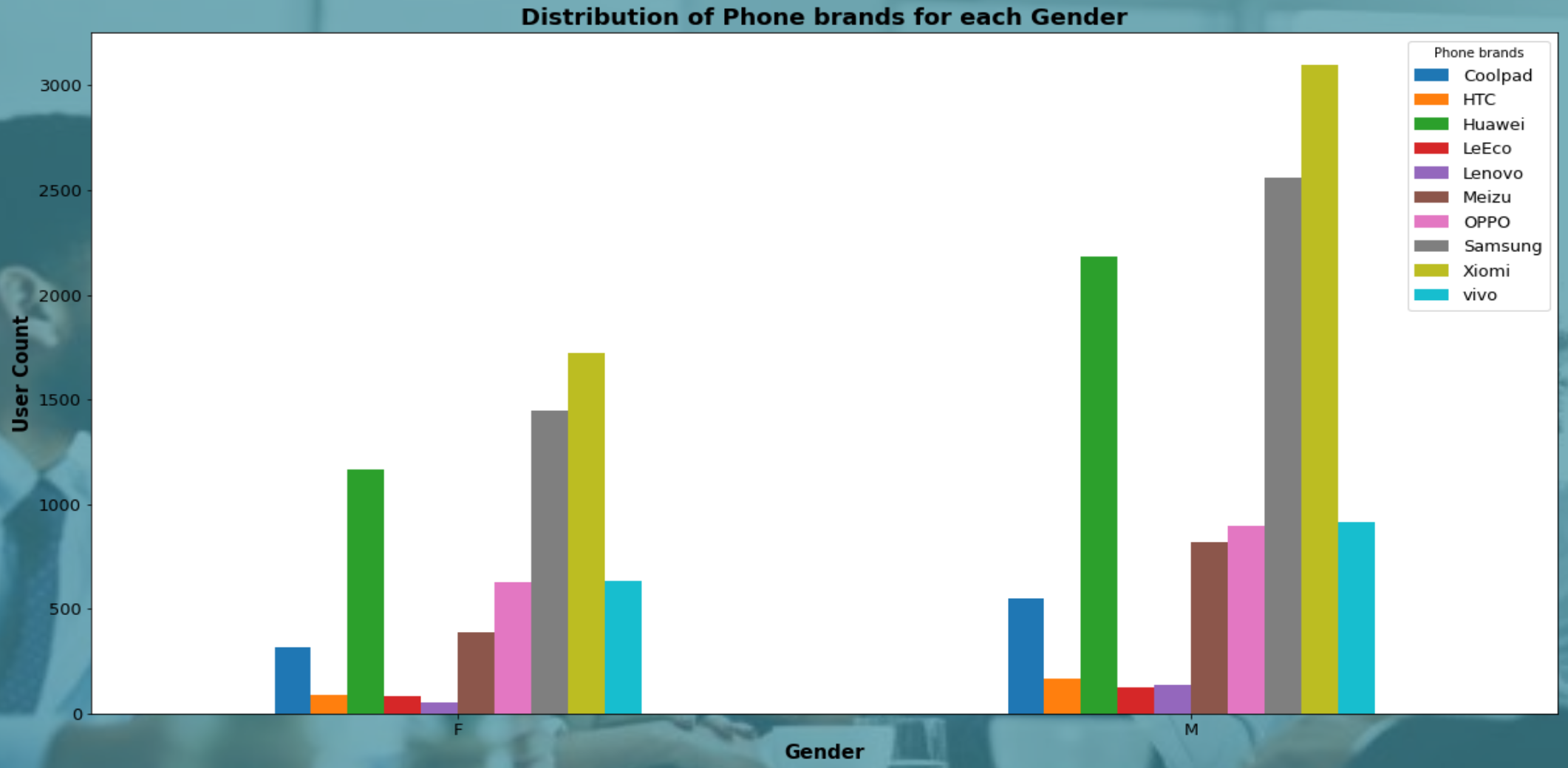


Observations:

- In West Bengal, Karnataka, Gujarat & Punjab, Xiaomi is the most popular brand followed by Samsung and Huawei.
- Whereas in Bihar and Kerala, Samsung is the top brand followed by Xiaomi and Huawei.

DATA ANALYSIS

➢ How is the distribution of Phone brands across gender?

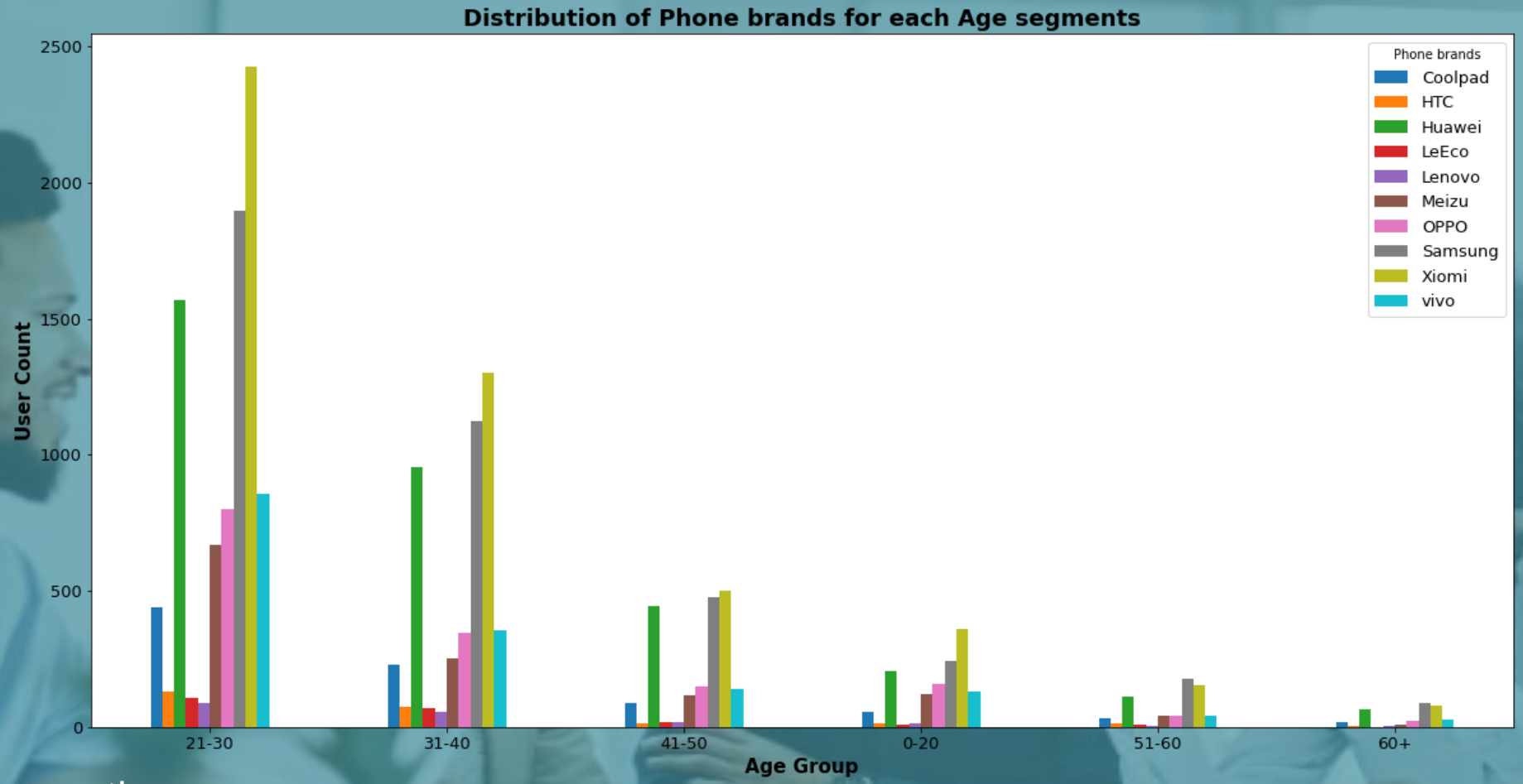


Observation:

- Xiaomi is the most popular phone brand followed by Samsung and Huawei among both males and females.

DATA ANALYSIS

➢ How is the distribution of Phone brands across each age group?

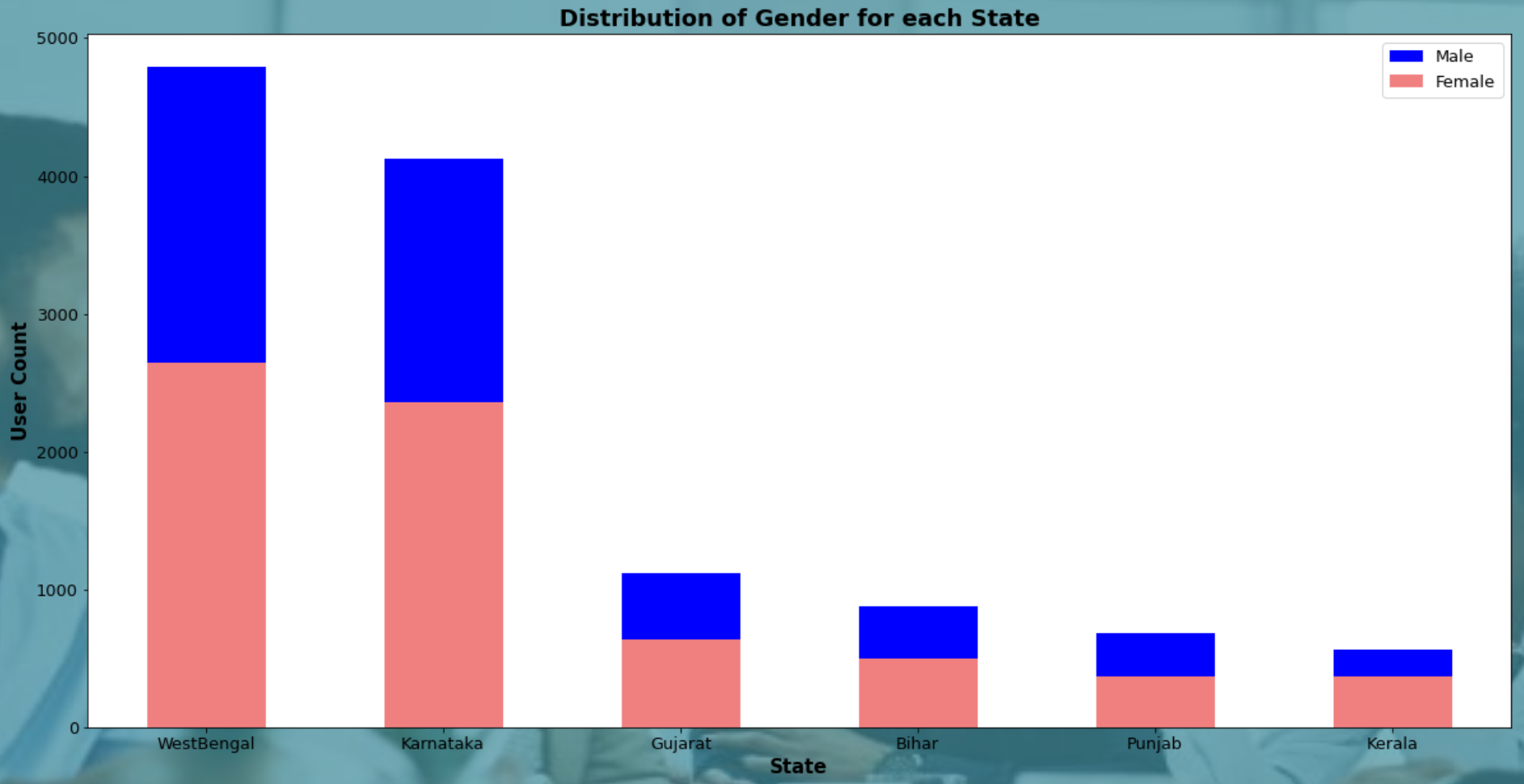


Observation:

- Among all the age group from 0-50 years, Xiaomi is the most popular phone brand followed by Samsung and Huawei
- But Samsung is more popular among elderly population i.e., 50+ age group.

DATA ANALYSIS

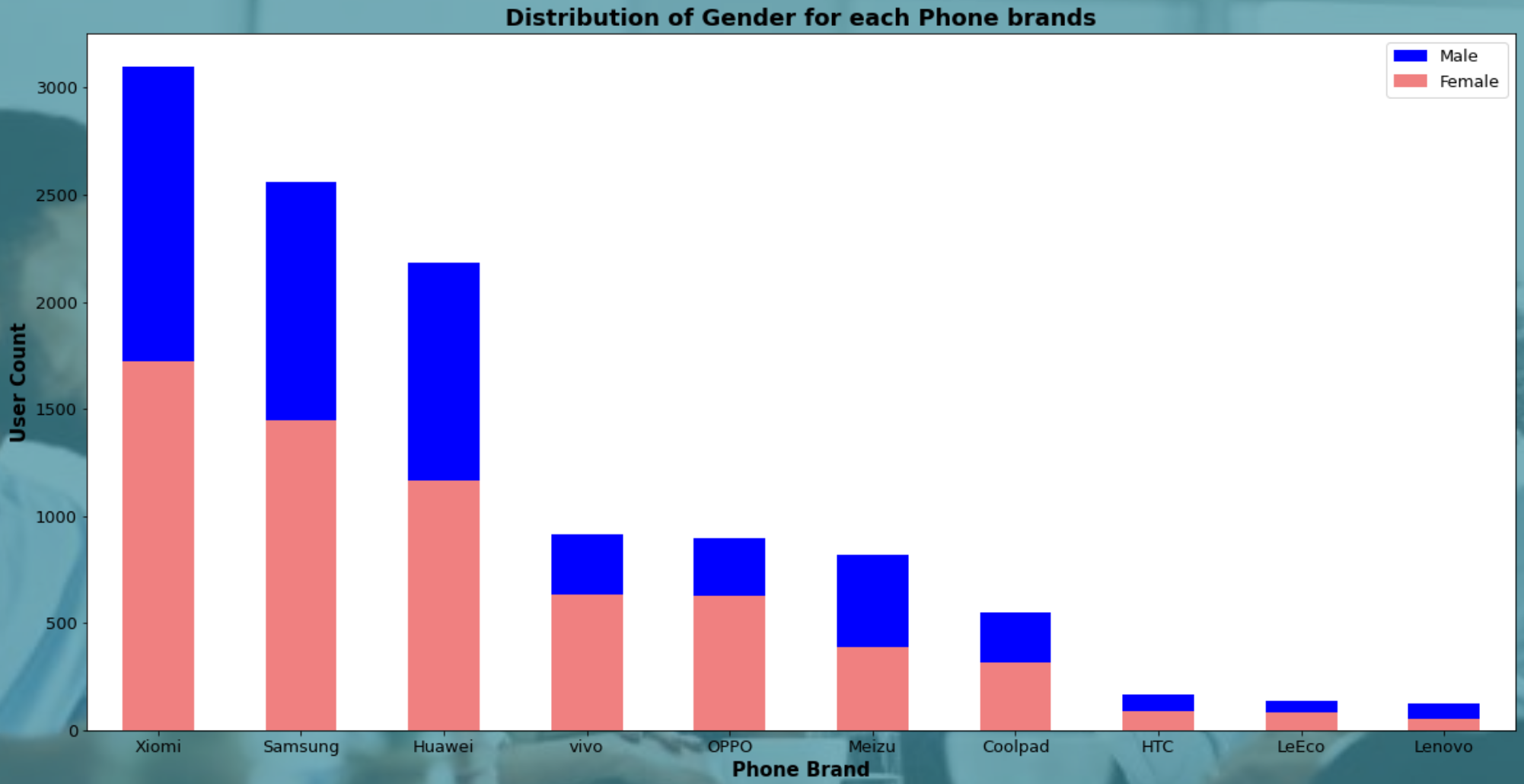
➢ How is the distribution of gender across each state?



Observation: In all the states male users are more than female users.

DATA ANALYSIS

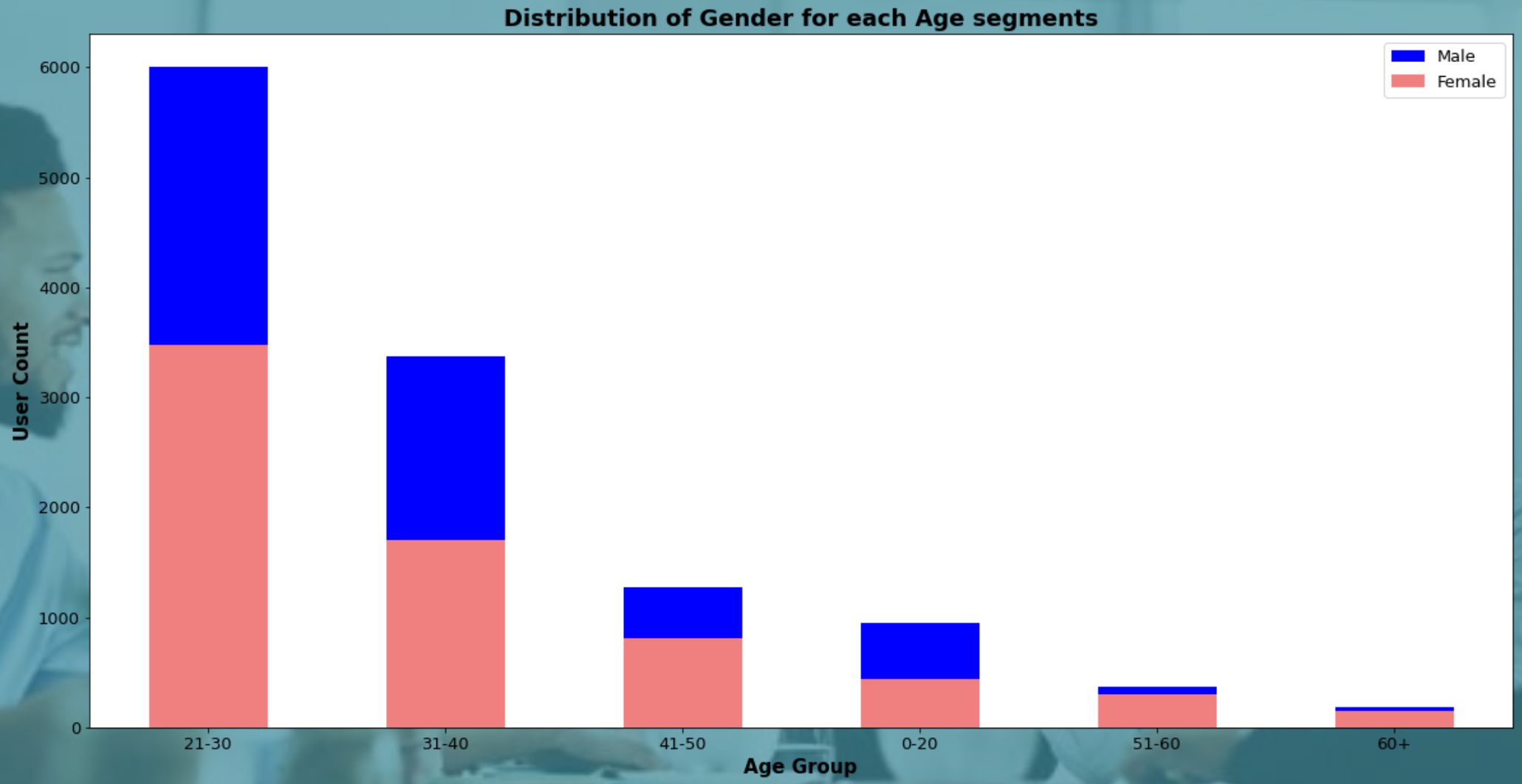
➢ How is the distribution of gender across each phone brands?



Observation: Across each Phone brands male users are more than female users.

DATA ANALYSIS

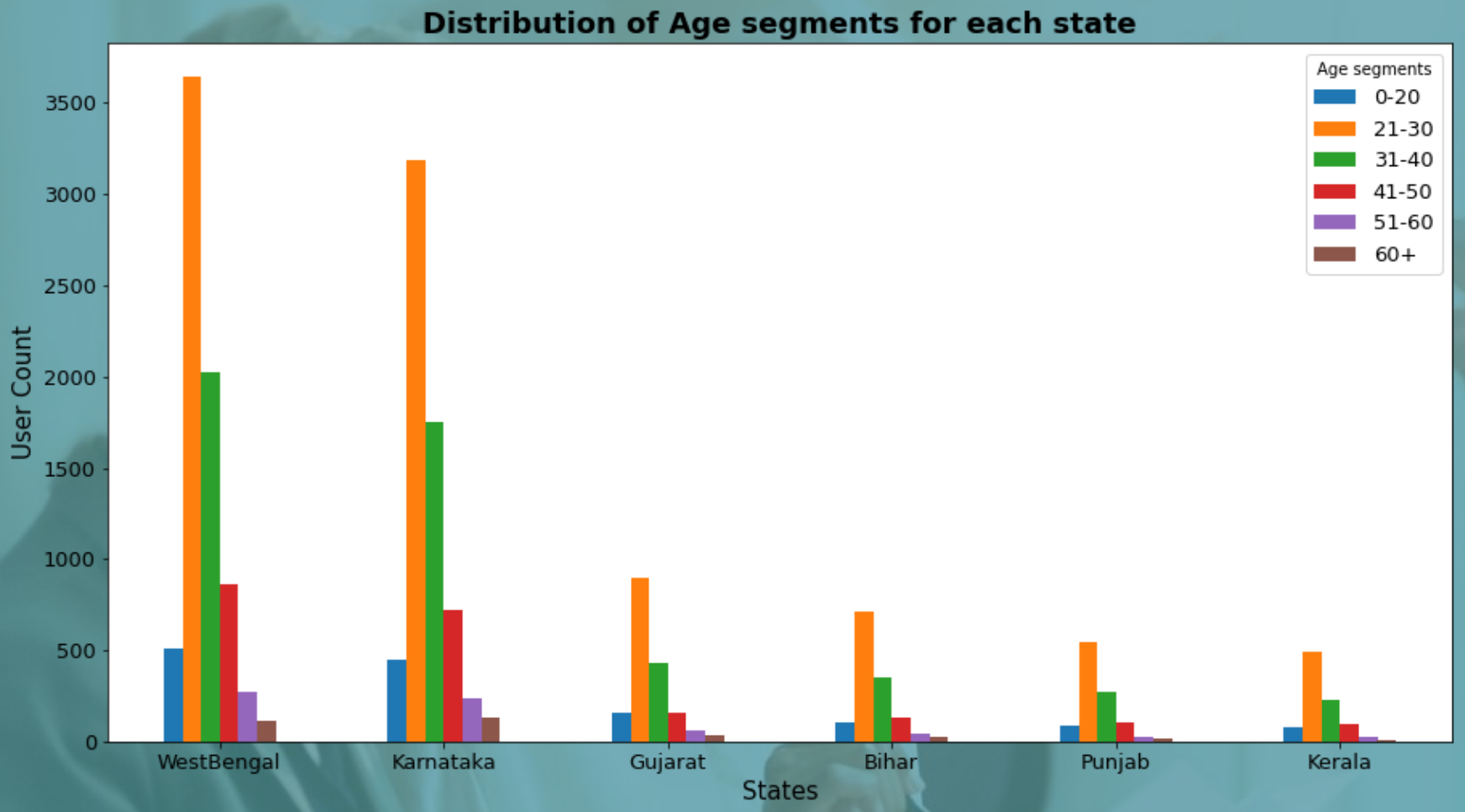
➢ How is the distribution of gender across each age segments?



Observation: Across each age segments male users are more than female users.

DATA ANALYSIS

➢ How is the distribution of Age segments for each state?

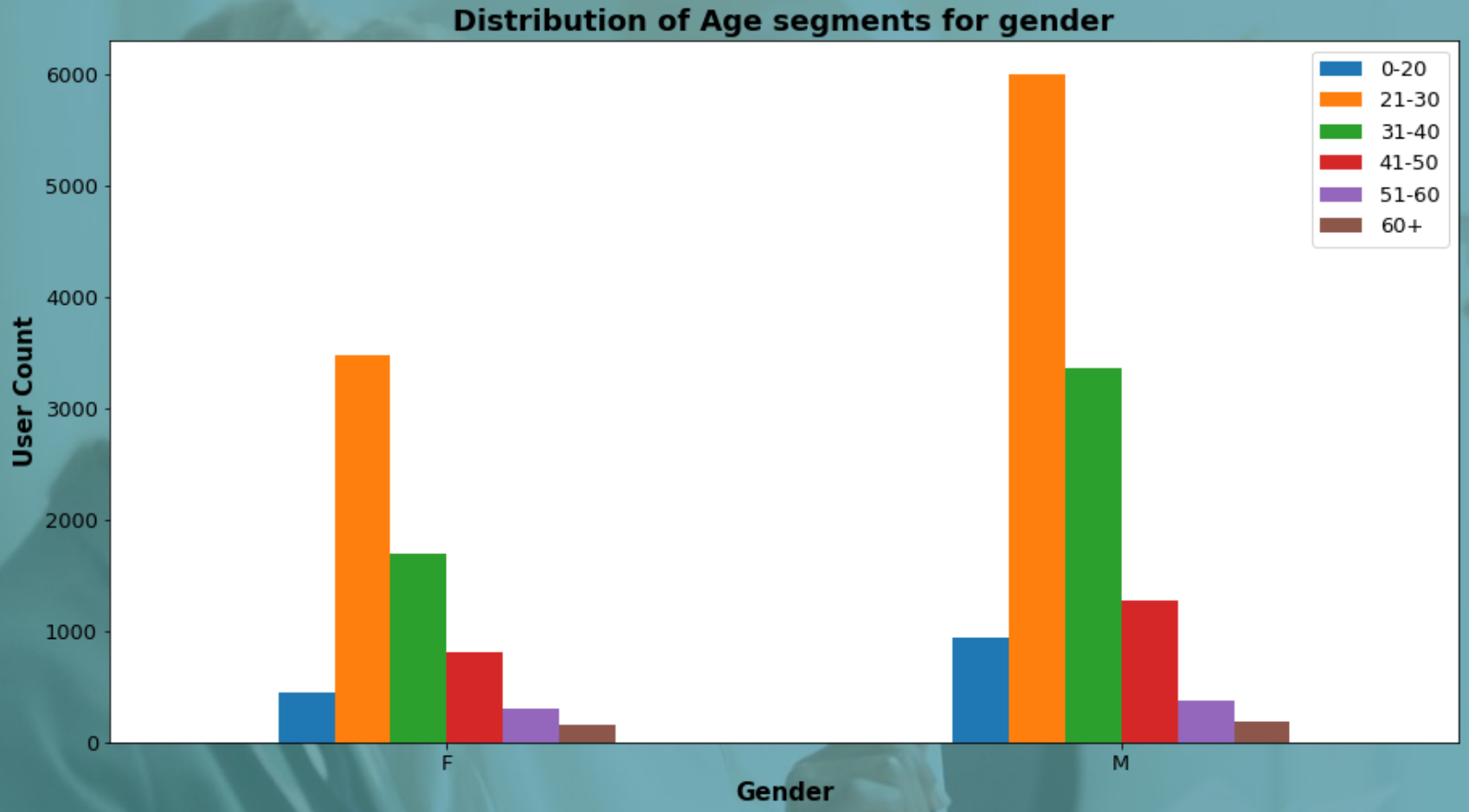


Observations:

- Across all the states, maximum users are of the age group 21-30yrs followed by 31-40 yrs.
- Young teenage population of 20- age group are more than the elderly population of 50+ years across all the states.

DATA ANALYSIS

➢ How is the distribution of Age Segments for each gender?

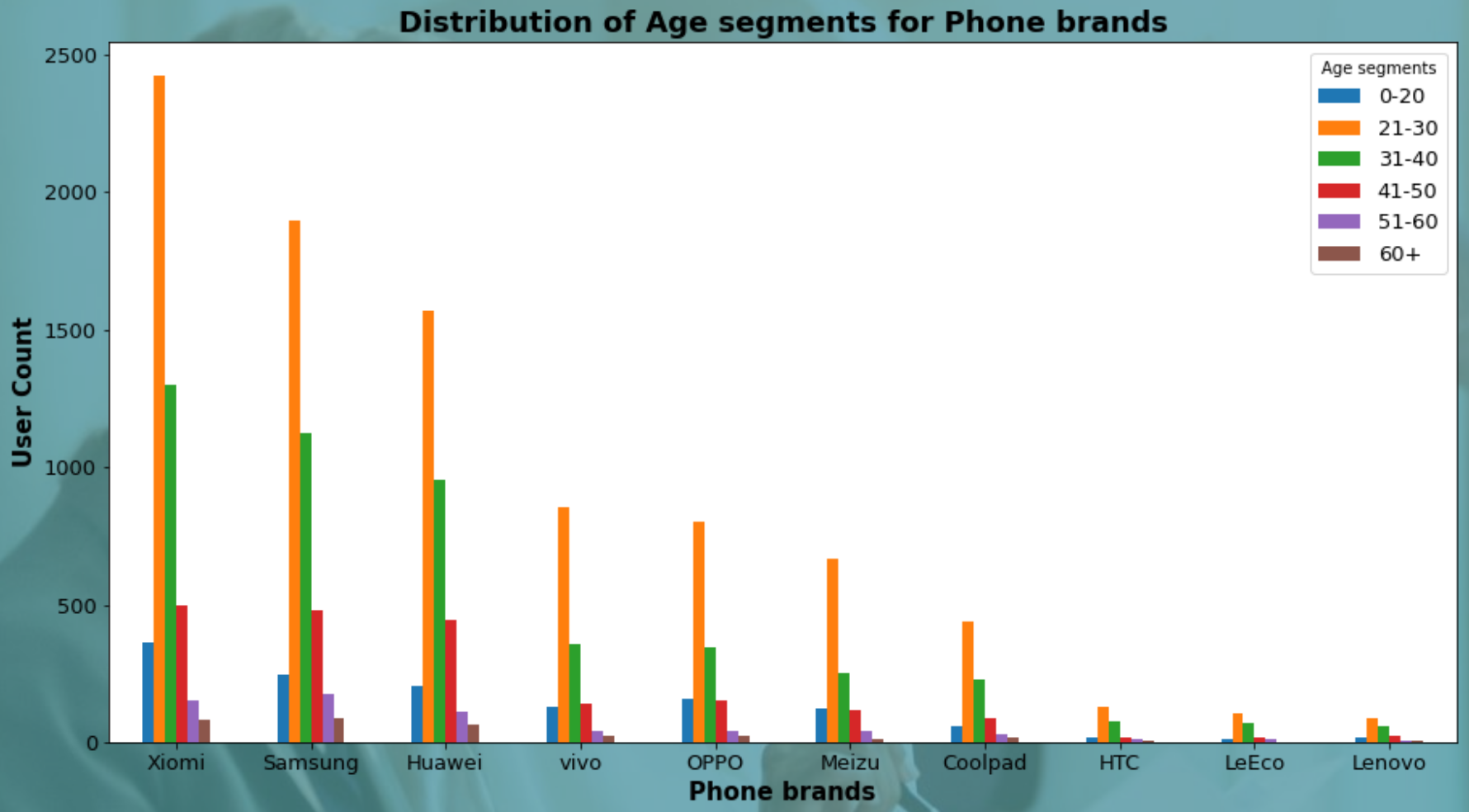


Observations:

- Across both the gender, maximum users are in the age group 21-30yrs followed by 31-40 and 41-50 years.

DATA ANALYSIS

➢ How is the distribution of Age Segments for each phone brands?

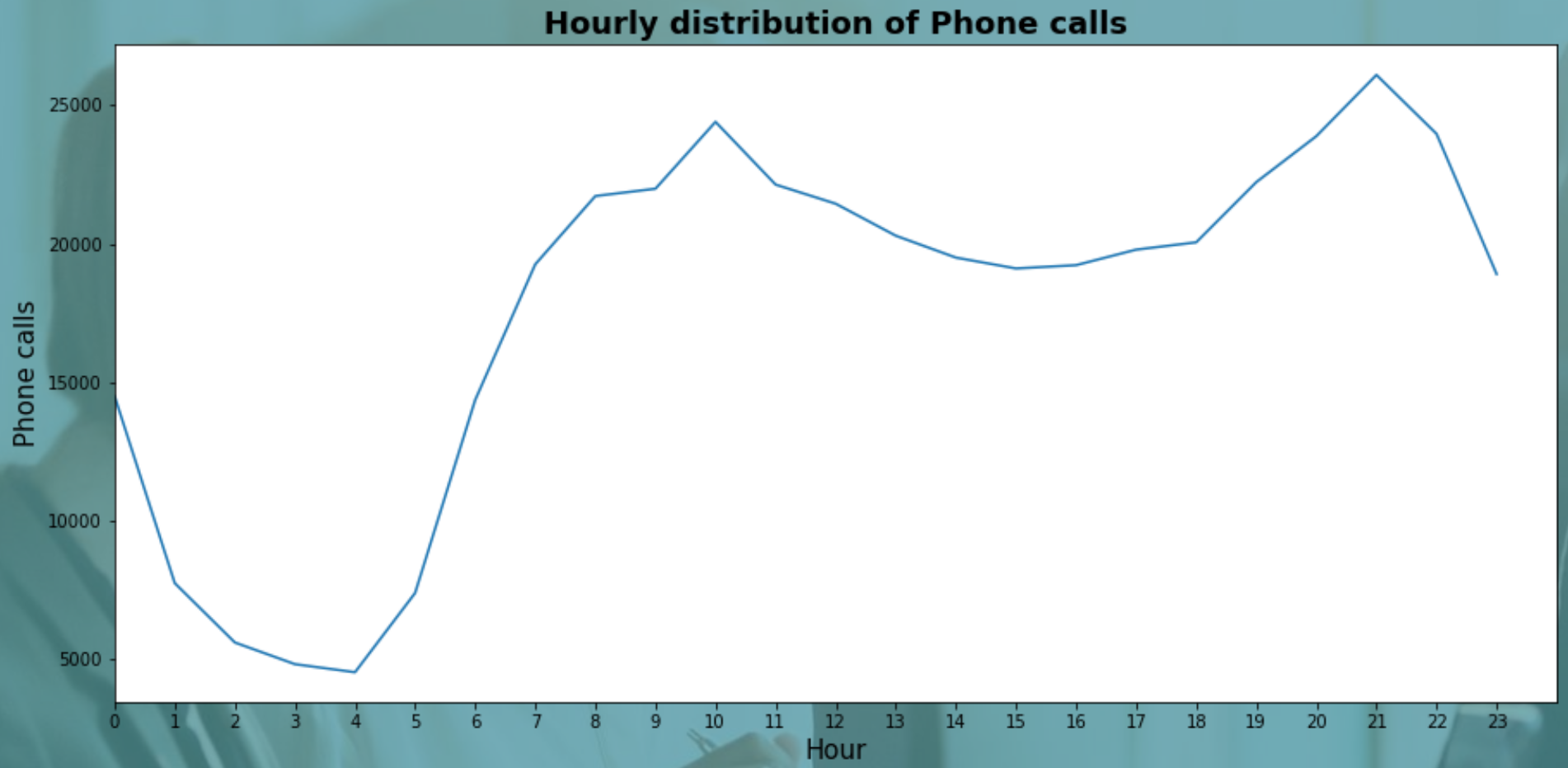


Observations:

- Among all the age segments, Xiaomi is the most popular brand followed by Samsung, Huawei.
- Vivo and Oppo are equally liked by the users.
- Meizu is preferred over Coolpad.
- HTC is preferred over LeEco and Lenovo among the age groups.
- LeEco and Lenovo are equally liked by the users.

DATA ANALYSIS

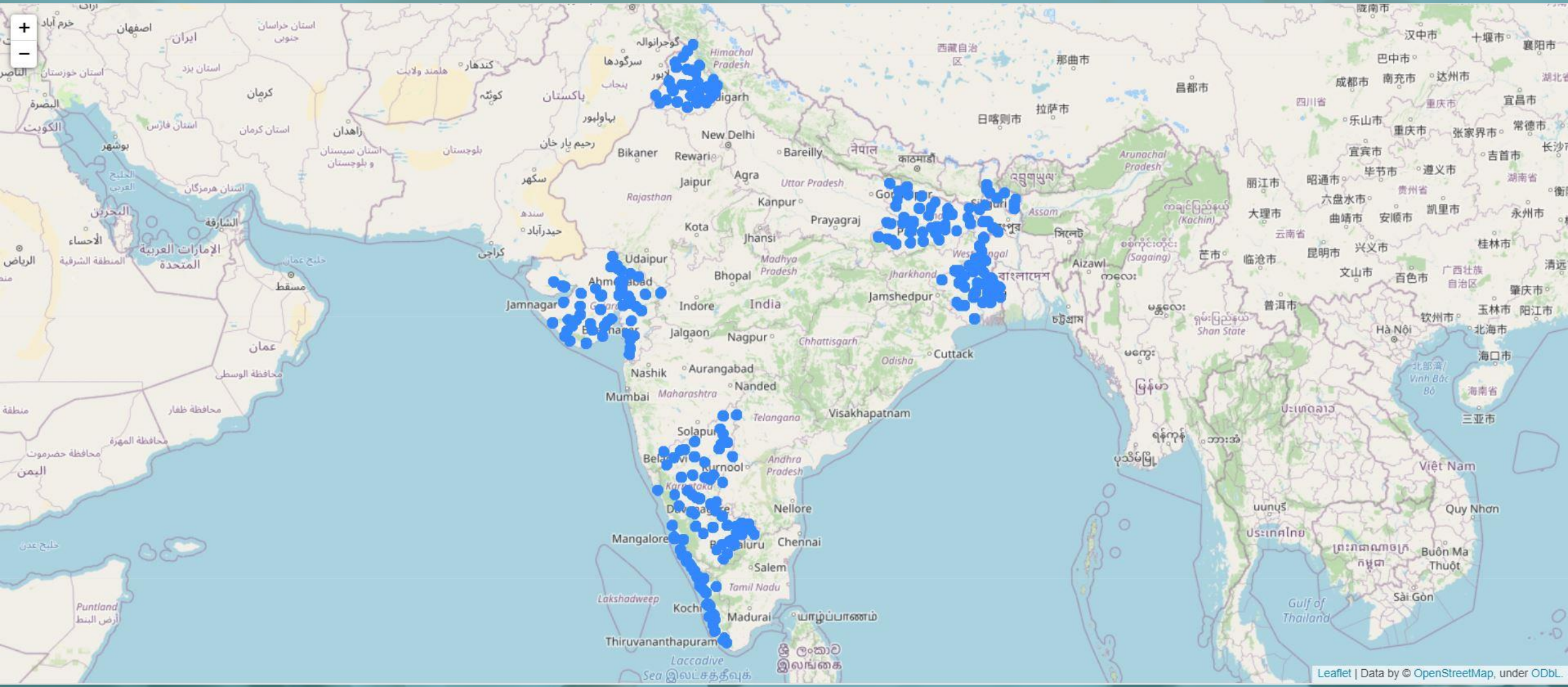
➢ How is hourly distribution of phone calls?



Observation:

- The phone calls starts rising from 4am till 9am.
- There is a steep rise between 9-10 am and drops down in next one hour till 3:00pm and then rise up sharply from 6 pm to 9pm .

DATA ANALYSIS ➢ Plotting users on the map



CONCLUSION

- West Bengal and Karnataka have the highest number of InsaidTelecom users. In other 4 states it has low presence may be due to poor network and services.
- Xiaomi, Samsung and Huawei are the leading phone brands in each state and across all age segments. They cover almost 63% of the market.
- Vivo and Oppo are equally preferred by the users followed by Meizu and Coolpad. Users using other brands are nearly 0-1%.
- Male users are more compared to female users across each state, age segments and phone brands.
- Around 50% of the users are young working population in the age group 21-30years and 75% users fall in the age group 21-40years across each state, gender and phone brands.
- The phone calls peeks between 9-10 am and drops down in next one hour till 3:00pm and then rise up sharply from 6 pm to 9pm.

INSIGHTS

- To improve their presence in the other 4 states (Gujarat, Punjab, Bihar and Kerala), InsiadTelecom can install more towers and offer some freebies to attract more customers.
- To retain their existing customers in West Bengal and Karnataka, they can engage them with some loyalty schemes.
- InsiadTelecom can tie up with Xiaomi, Samsung and Huawei for it's in-built apps or some attractive offers or deals with these brands to increase its user base.
- To increase the female users, InsiadTelecom can offer some freebies or discounted tariff plan for them!
- They can launch some family plan by allowing to add 2-3 family members in same tariff of existing users. Adding their spouse, child & parents in the same tariff will increase their user base.
- They can provide discounted subscription of some popular Music/Movie/Yoga/Medication/Health /Business/OTT apps to retain the customers.
- To attract students, housewives & senior citizens, company can tie up with some OTT channel which has content or program related to these customers.
- Company can improve the phone calls between 10 am to 3 pm by giving data offers and calls offers exclusively for this duration.

THANK YOU!

GROUP - 1007