**Accredian Telecom: A Perfect Choice for Next-Gen**

Project Team Description:

Group-1010

Ms. Akanksha Kaushik (Project Leader, Problem statement defining, concept designing, coding, report writing)

Mr. Harsh Saurav (concept designing, major contributor in coding, report writing)

Mr. Sai Brahamaiah (coding)

Mr. Naveen (coding)

Ms. Preeti Nassa (coding)

Ms. Anu Ramojjala (coding)

Contents

Contents

1. Introduction
2. Project Description
3. Problem Statement
4. Problem Analysis
5. Sources of Data
6. Summary of Data Mining
7. Proposed Solution for Customers
8. Tools Used
9. Summarized Conclusion

10. Actionable Insights

**Consulting Report**

# **Introduction**

Accredian Telecom is one of the pioneering Telecom industries which is focusing on how to elevate users’ telecom experience. With a vision to redefine the way users connect with their loved ones and the world around them, we at Accredian put our best efforts into making ourselves the ultimate destination for cutting-edge telecom networking services. As a frontrunner in the telecom industry, we are dedicated to enhancing your communication experience and taking it to unprecedented heights.

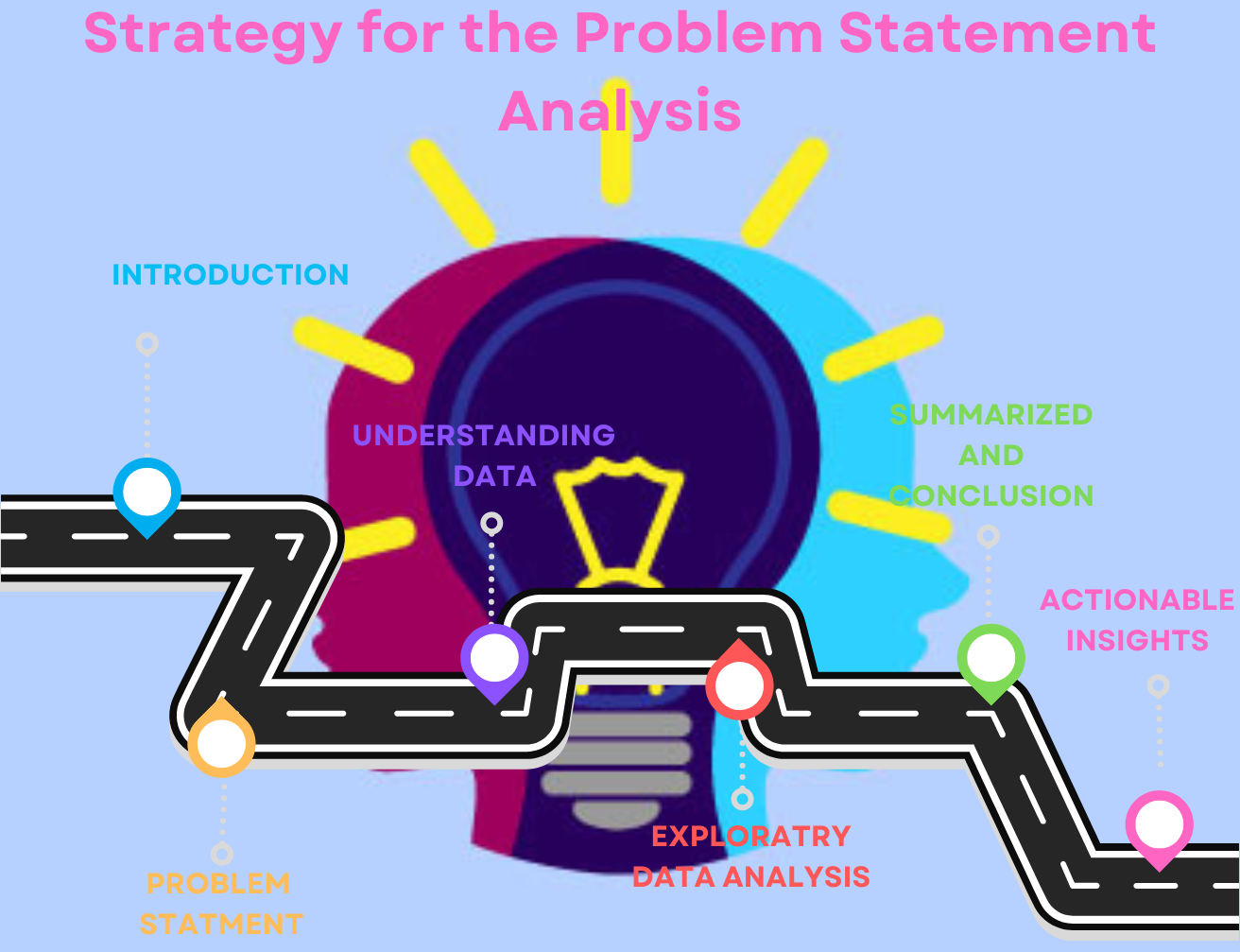
# **Project Description**

In this consulting assignment, we are expected to build a dashboard which will help the telecom industry to understand their 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 Statement**

Leaders at Accredian Telecom understand that customising offerings is very important for its business to stay competitive. Understanding users’ desires and behaviours on a new service or model is the key to running their business. Currently, Accredian Telecom is seeking to leverage behavioural data from more than 60% of the 50 million mobile devices active daily in India. Doing so will help their clients better understand and interact with their audiences.

# **Problem Analysis (Strategy for the Problem Statement analysis)**

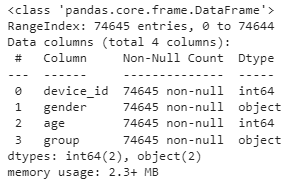


# **Sources of Data (Explain about database connection, tables and their columns)**

In this project, the demographics of a user (gender and age) based on their app download and usage behaviours, have been studied. The data source given was in two different formats/sources. One was the text file with .csv extension and the other was to be accessed remotely from Accredian Telecom’s SQL server. The CSV data was downloaded and saved on the local machine, while SQL data was accessed by connecting through the IDE (jupyter notebook).

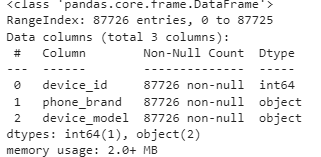
SQL Data: **1.** **gender\_age\_train**

**The attributes of this data are:**

****

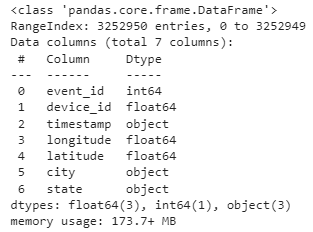
**2. phone\_brand\_device\_model**

**The attributes of this data are:**

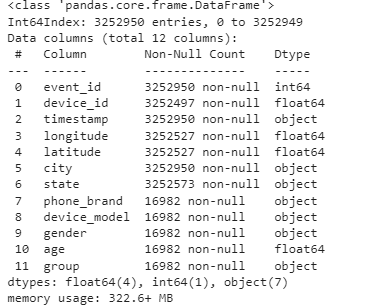
**.**

CSV Data: **events\_data.csv**

**The attributes of this data are:**

****

Merge Data Frame: **dataset\_final**

****

# **Summary of Data Mining (What kind of challenges you faced with the Data and how you resolved them?, Summary of your Analysis)**

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

1. **Age gender Train data**

a. Embark on a Grand Data Voyage: Behold the majestic expanse of our dataset, a sprawling landscape teeming with 74,645 data points, each a tiny universe of its own. Within this digital cosmos, we wield the power of 4 essential dimensions, meticulously curated to reveal the secrets of our digital realm. Brace yourself, for within these digital shores lies a unique

gem, a shimmering parameter known as 'device\_id', our true north, guiding our analytical odyssey as a distinguished primary key.

b. Unveiling Gender's Mysterious Veil: Amidst the sea of data, a tapestry of identity unfolds. At its heart, a dichotomy of letters, 'M' and 'F', like ethereal brushstrokes on a canvas. In this realm, they signify more than just characters; they are the keys that unlock the doors to the realm of 'Male' and 'Female', bringing forth a deeper understanding of human diversity. As we tread through the corridors of data, we unravel the profound story woven by these gender markers, revealing insights that enrich our narrative.

1. **Mobile Brand Data**

a. Dataset Dimension Enchantment: Embark on a journey through a dataset of remarkable proportions, with 87,726 entries cascading across three dimensions. The legendary 'device\_id' stands tall as our unique key, guiding us through this expansive realm.

b. Language Unison: Witness the symphony of languages as device names seamlessly blend Chinese and English. A linguistic duet that adds a touch of global allure to every entry.

c. Brands Beyond Borders: Mobile brands unite cultures, featuring both Chinese and English renditions. A testament to the worldwide tapestry of technology.

d. Translation Revelation: Our translation quest unveiled the power of "Simplified Chinese" in transcending language barriers. A tale of linguistic transformation, connecting worlds and illuminating meanings.

1. **Events Data**

a. Data Size Chronicles: Behold a colossal dataset, boasting a staggering 3,252,950 entries across seven dimensions. Upon initial inspection, behold the absence of duplicates, a testament to its pristine state.

b. Event Odyssey: Embark on a journey through 3,252,950 captivating events, each bearing its unique 'Event\_Id'. A symphony of device-driven endeavours.

c. Temporal Voyage: Our chronicle unfolds between April 30th, 2016, and May 8th, 2016, encapsulating a captivating week of data. Every day, an opportunity for insight, even within the limited observations on April 30th and May 8th.

d. Null Nebula: Amidst this rich tapestry, we encountered enigmas. Among them, is the enigmatic absence of data:

- 377 entries wander without a 'State'

- 453 entries shroud 'Device\_id' in mystery

- 423 entries remain cryptic in 'Latitude' and 'Longitude'

- 3235968 entries have null value in ‘Phone\_brand’,’device\_model’,’gender’,’age’,’group’

e. Missing Mosaic: Fear not, for patterns to emerge. Employing the power of latitude, longitude, and state, we piece together the missing links. The 'State,' 'Latitude,' and 'Longitude' nulls are illuminated.

f. Device Dilemma Dissolved: In the realm of 'Device\_id' nulls, a harmonious solution arises. Nearby brethren devices share their identity, resurrecting completeness.

g. Global Geocaching: Amidst our dataset, 42 entries dared to venture beyond India's borders. Swiftly, their latitude and longitude are realigned, a dance of data harmony.

h. For the Selected States in scope ‘TamilNadu’,’ Manipur’,’ Chandigarh’,’ Tripura’,’ Uttar Pradesh’,’ Arunachal Pradesh.

1. **Other Challenges and solutions**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # | Observation/Challenges faced | | | How we resolved problems identified in the dataset. |
| 1 | the missing value is present in gender, age, group, phone brand and device model | | | age and group can be filled with mean, gender and device model can be filled with mode |
|  |
|  |
| 2 | There are some Object variables available in the dataset | | | We can change it into a continuous variable by using dummification / one hot encoding |  |
|  |
|  |
| 3 | Convert all Chinese names into English | | | by using google translate and by using mapping method it will convert |  |
|  |
|  |
| 4 | A dataset is very skew | | | Log transformation is most likely the first thing you should do to remove skewness |  |
|  |
|  |
| 5 | There are some missing values in the state column | | | Mapping of the city with its corresponding state |  |
|  |
|  |
| 6 | The data type of a timestamp is an object, whereas it should be datetime | | | the datatype of the timestamp can be converted to DateTime using .todatetime() method |  |
| 7 | the device id has some negative/- value | | | replace the negative values with their equivalent positive value or we can simply remove '-' |  |
| 8 | Device id has a different character length | | | we can put a prefix in the device id to make all the device ids of the same length. |  |
| 9 |  | Age values lower than 22 are not assigned to any age groups |  | we can use apply method to resolve this |  |
| 10 | age cannot be of float datatype | | | The age column has been changed from Float to int |  |
| 11 | Device id cannot be null | | | After doing pre-processing as suggested, device ids are found to be not-null |  |

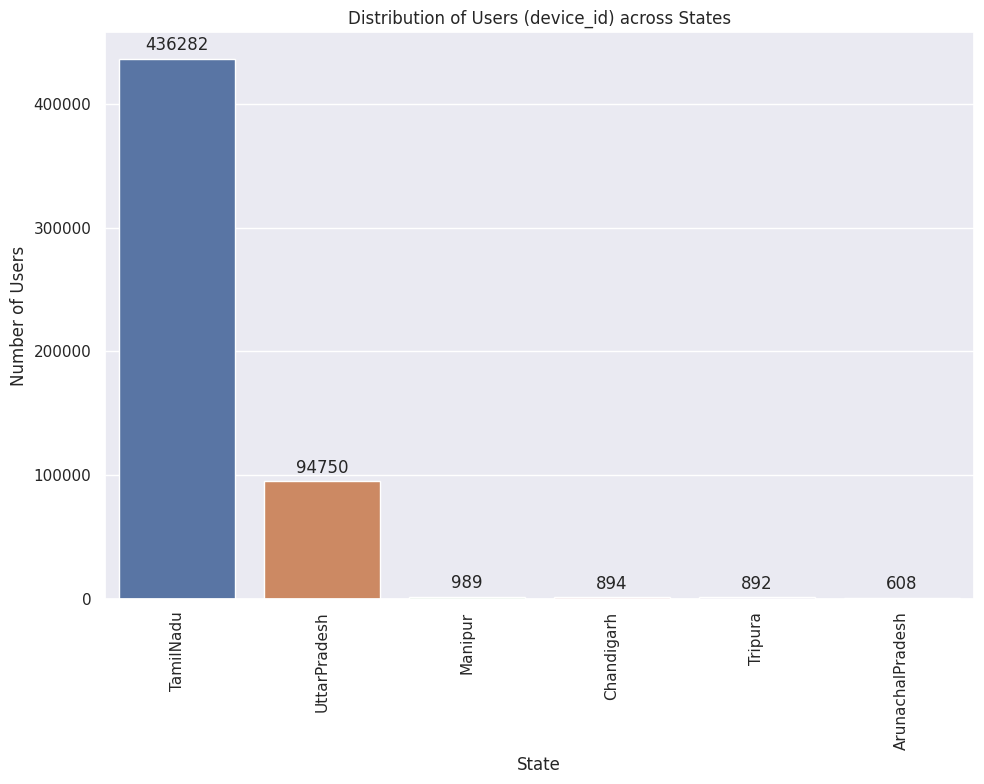
**Analysis and Insights**

1. **Number of Records analyzed:**

The team was presented with three datasets and was asked to analyze only for “TamilNadu”, “Manipur”,” Chandigarh”,” Tripura”,” UttarPradesh” and “ArunachalPradesh” states.

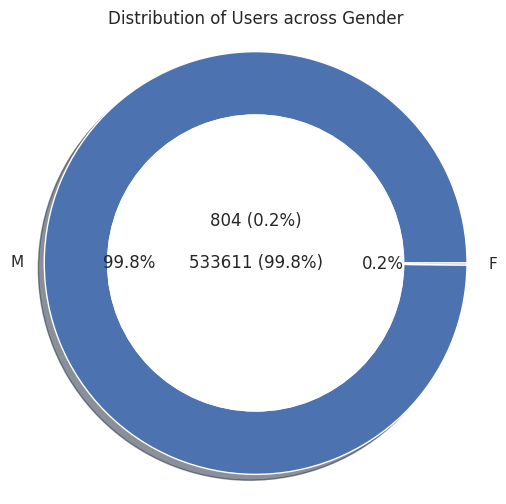
|  |  |
| --- | --- |
| event\_data | Analyzed 534415 rows of data out of 3252950 based on selected states |
| gender\_age\_train | Analyzed 74645 rows of data |
| phone\_brand\_device\_model | Analyzed 87726 rows of data |

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



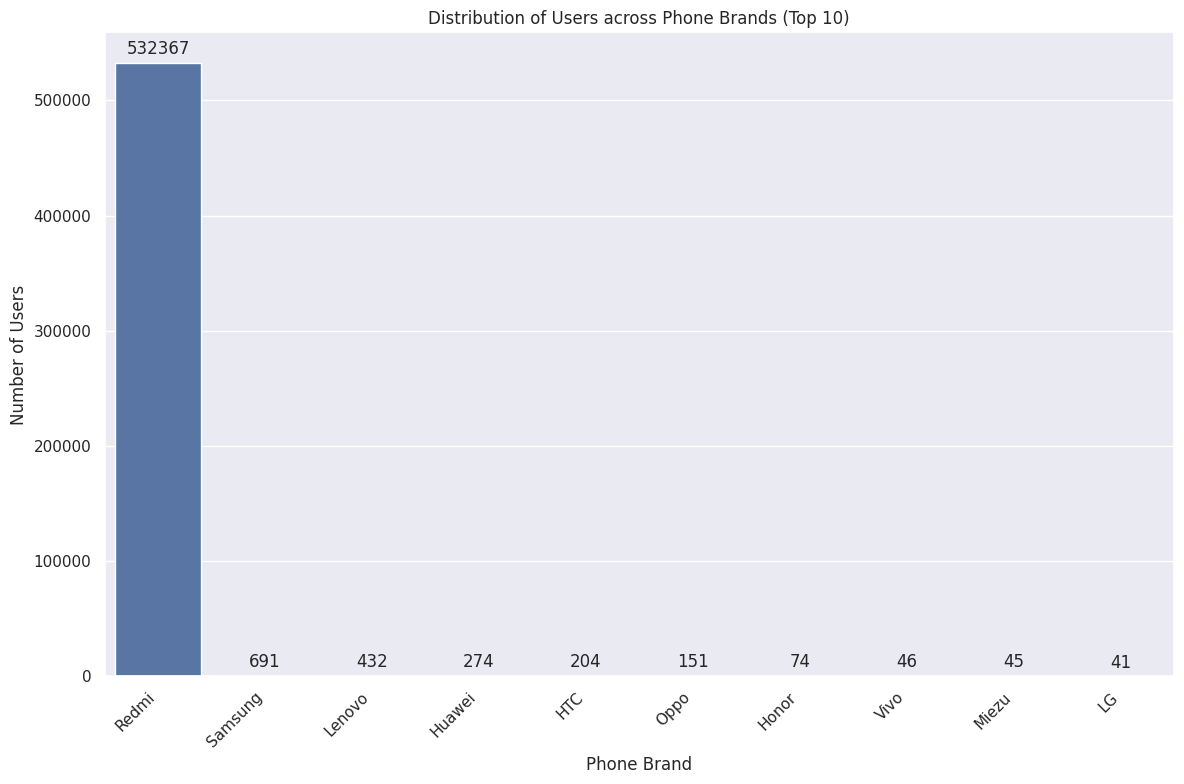
According to the above graph, it has been observed that **Tamil Nadu (Chennai)** has the **highest number** i.e **436282** of users across other states (UP, Manipur, Chandigarh, Tripura, Arunachal Pradesh)

1. **Distribution of users across gender:**

****

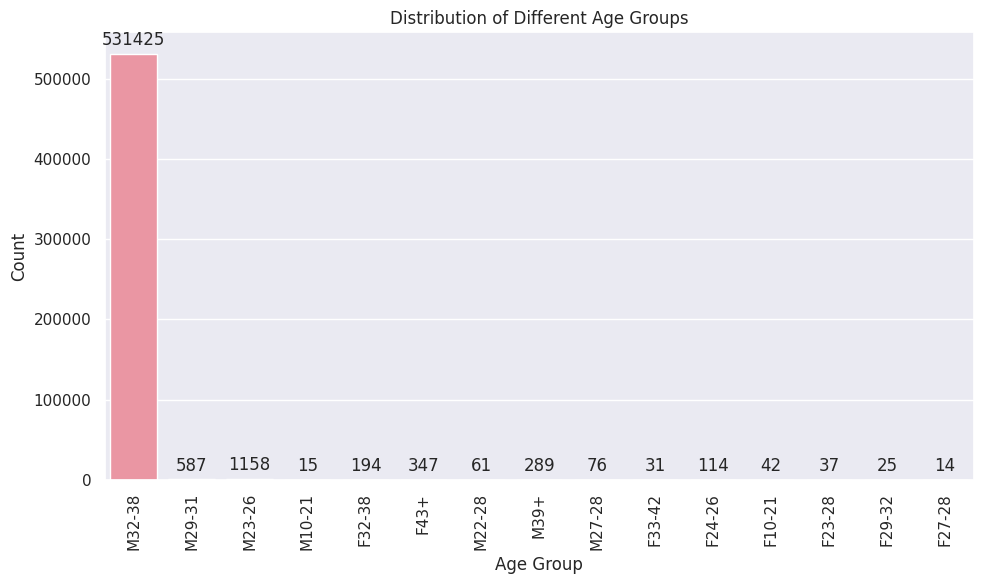
The above donut graph shows the distribution of users across gender. This graph reveals that the **maximum number of users are males** with a percentage of **99.8%** in comparison to **female users who are 0.2%.**

**4.Distribution of users across phone brands (Top 10 )**

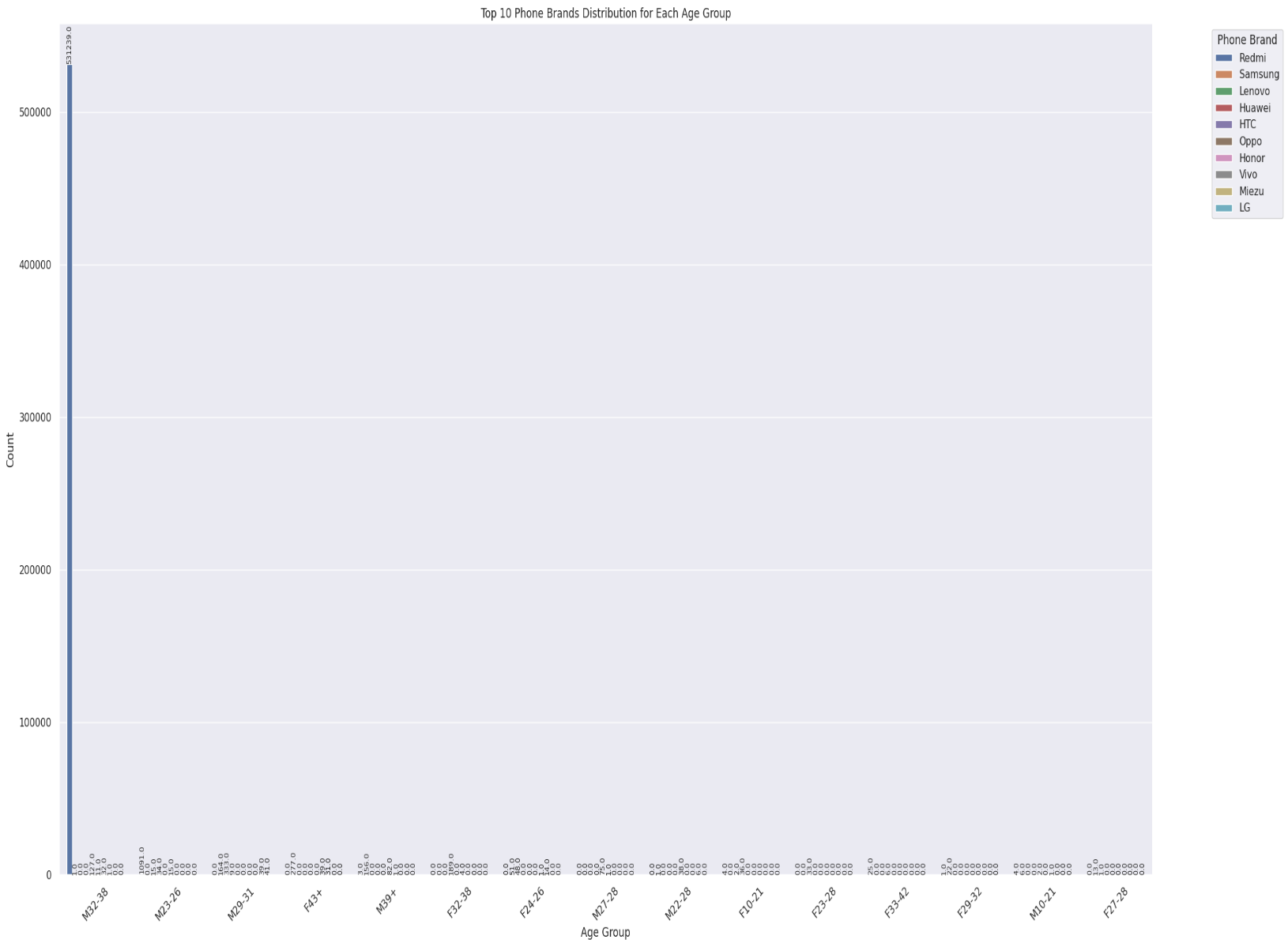
****

 It has been observed that amongst the **top 10 phone brands**, Redmi is leading the market, i.e.**Redmi** is dominating with above **5 million** users.

**5.Distribution of Different Age Groups**

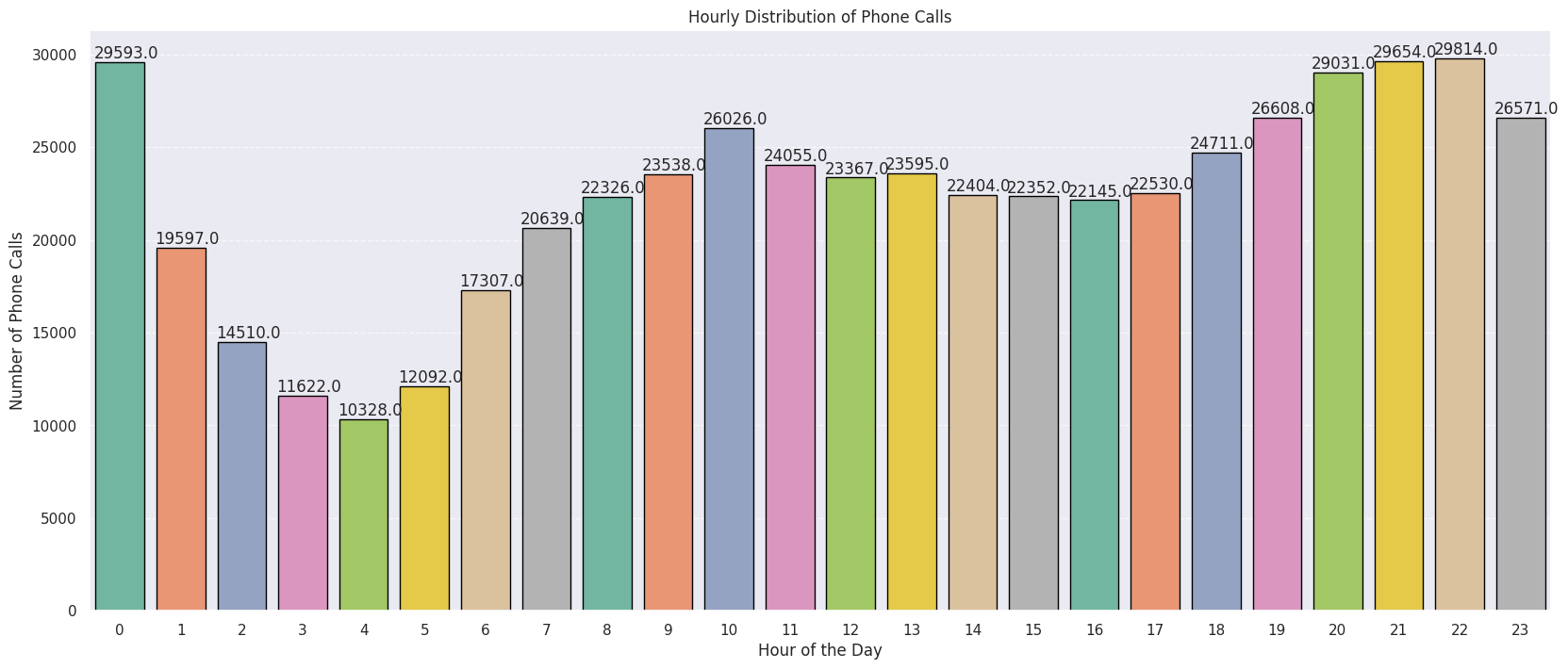
****

We have observed from the graph that the age group which is having a **maximum count** is **M 32-38**. I mean that **male users** within the age **group of 32-38 are the maximum in number**. There is a very less number of female users apart from that **most belong** to the **43+ age** group.

**6.** **Mobile Phone users based on Top 10 brands and Different Age Group  
**

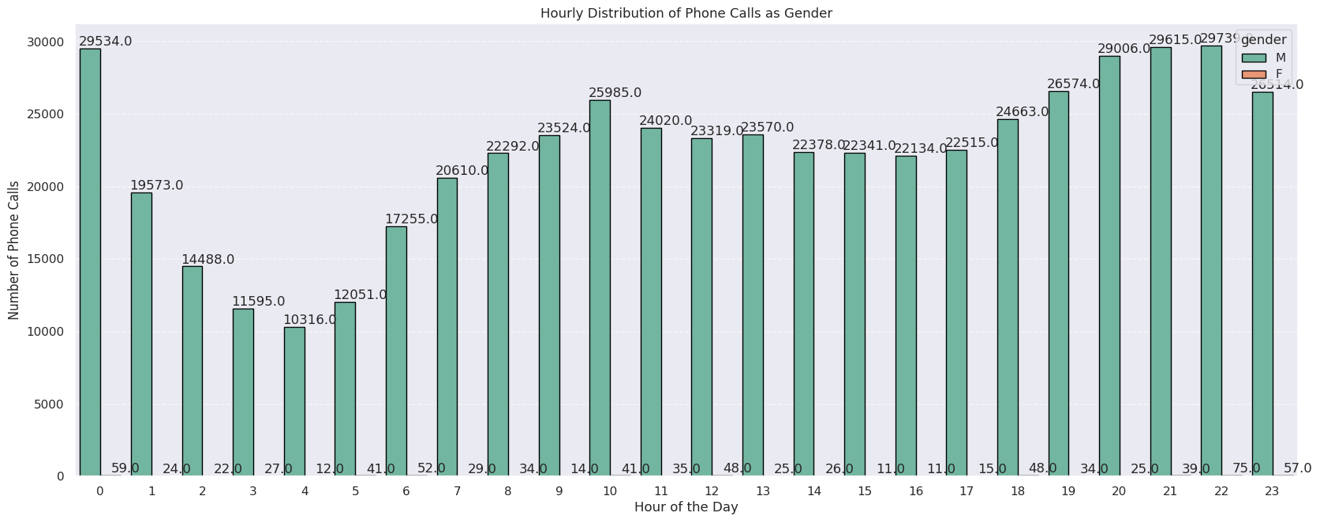
On looking at the bar, it can be easily predicted that the most **liked** brand is **Redmi** amongst the males of **age between 32 and 38**.

**7. Hourly Distribution of Phone calls**

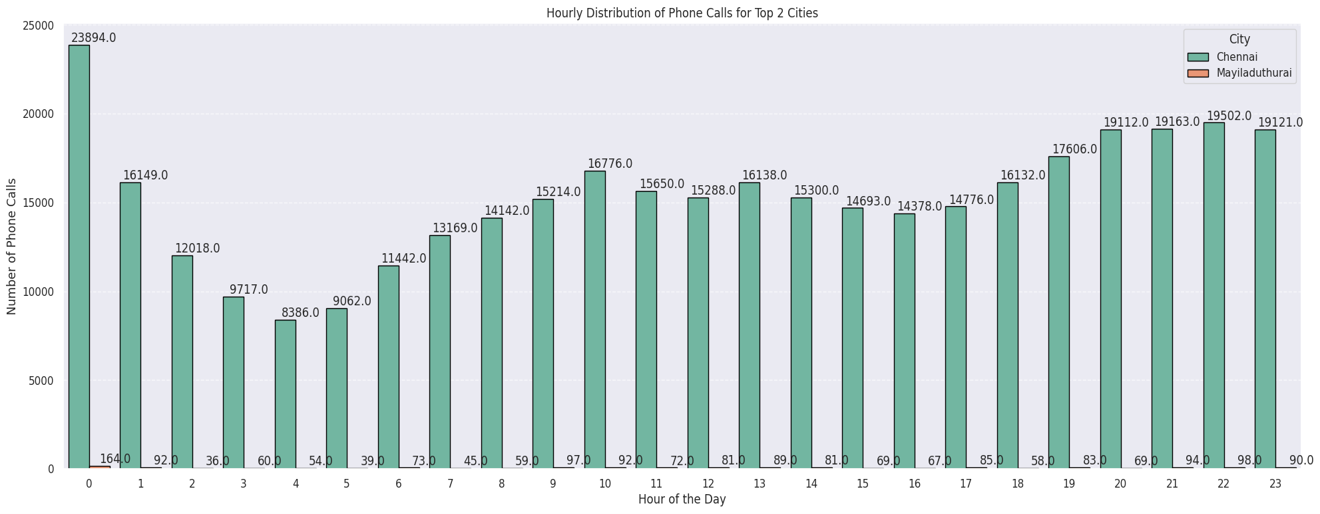
****

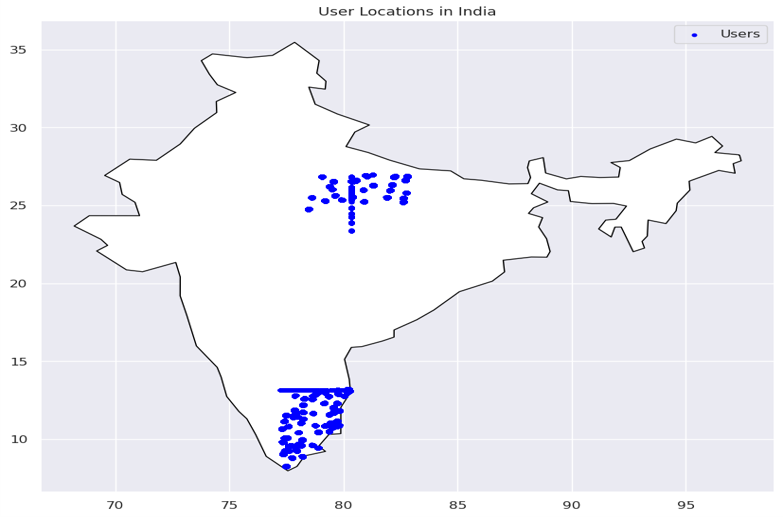
**The Hourly distribution of phone calls is shown in the above graph. It has certain observations:**

* The calls are done **mostly** at night, i.e. from **8 PM to 12 AM**
* The **least** number of calls are made during the **early morning time, i.e. 3 AM to 5 AM**
* There is a **little bump** in the number of calls made in the **morning, i.e. 10 AM.**

**8. Hourly Distribution of Phone calls as Gender  
**

The Hourly distribution of phone calls amongst the male and female users is shown in the above graph. This graph suggests that the maximum number of phone calls are made by users in the time duration of **8 PM to 12 AM** and the least number of calls are made in the early morning hours i.e. **3 AM to 5 AM**. Whereas the **least number of calls are made by female users.**

**9. Hourly Distribution of Phone calls for the Top 2 Cities  
**

The Hourly distribution of phone calls amongst the top two cities shows that **Chennai** has the **highest number** of phone calls made throughout the day, whereas the **2nd number** is achieved by **Mayiladuthurai**.  
  
**10. Distribution of Users**Most of the users are from **South and North India** and in Which two states are the most number of users i.e **Tamil Nadu and Uttar Pradesh.**

**Tools**

The following tools were utilized for the study

1. Jupyter / Google Colab
2. PyCharm

**Summary / Recommendations**

* Unravelling User Composition: The Gender Mosaic

- 99.8% Male Users, 0.2% Female Users - Unveiling the Majority!

- Dominance of Male Users - Fueling the Conversational Landscape.

* Geographical Gems: Tamil Nadu & Uttar Pradesh Shine

- Tamil Nadu & Uttar Pradesh Take the Spotlight - A Haven for User Engagement!

- The Heartbeat of India's Conversations - Witness the Vibrancy.

* Brand Royalty: Redmi, Samsung & Lenovo Lead the Way

- Redmi, Samsung, Lenovo - The Triumvirate of Trendsetting.

- Redmi's Meteoric Rise - A Journey to Fame and Favor!

* Age Chronicles: Pioneers of the 32-38 Era

- Ages 32 to 38 - Where the Magic Unfolds.

- Male Marvels in M32-38, Female Champions in F33-43+ - A Power-Packed Saga!

* Thriving Twilight Chats: The Call Hour Extravaganza

- The Night's Canvas - 8:00 PM to 12:00 AM - A Tapestry of Talk.

- Conversations Bloom as the Stars Twinkle - Embrace the Active Engagement!

* City Symphony: Chennai & Mayiladuthurai's Call Dance

-Chennai's Euphonic Calls - A Symphony of Communication.

- Mayiladuthurai's Silver Medal - Crafting Conversations, One Call at a Time.

**Actionable Insights.**

* The Digital Dynamo Phase: 32-38's Tech Symphony

- A Chapter of Tech Marvels - Ages 32 to 38 Take Center Stage.

- Embrace the Era of Digital Prowess and Technological Fluency!

* Spotlight on Relevance: Platform's Age-Defying Allure

- Our Platform's Magnetic Pull - A Magnet for the 32-38 Enthusiasts.

- Relevance Redefined - Where Tech Meets Taste!

* The Call Chronicles: Nights Ablaze from 8 to 12

- Lights, Camera, Conversation!

- Enter the Realm of 8 PM to Midnight - Where Calls Ignite and Connections Unite.

* Engagement Extravaganza: User Uprising in the Twilight Hours

- Engage, Enthrall, Energize - The Spectacle of User Involvement.

- Twilight Talks: More than Conversations - A Symphony of Interaction.

* Optimization Avenue: Paving the Path to User Delight

- Insights Carved in Time - Fueling the Future of Fulfillment.

- 8 PM to 12 AM - A Journey of Service Enhancement and Satisfaction.

* Unleash the Potential of the Digital Age and Experience Nightly Wonders!