**Consulting Report – Team 1034**

**Table of Contents**

1. **Introduction**

In this Capstone assignment, we are going to study the demographics of a user (gender and age) based on their app download and usage behaviors.

1. **Project Description**

**InsaidTelecom**, one of the leading telecom players, understands that customizing offering  
is very important for its business to stay competitive. Currently, InsaidTelecom 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. The Data is collected from mobile apps that use InsaidTelecom services. Full recognition and consent from individual user of those apps have been obtained, and appropriate anonymization have been performed to protect privacy. 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.

1. **Problem Statement**

In this project, we need to build a dashboard to understand user's demographic characteristics based on their mobile usage, geolocation, and mobile device properties.

1. **Problem Analysis (Strategy for the Problem Statement analysis)**

We will perform EDA on the data sets and explore the various relationships between different data provided. Plotting charts and graphs will provide insights on the data that will help InsaidTelecom to improve their services.

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

We have three datasets in this project. They are

* **Events data**

This data contains all the information related to the events generated by the device. The data contains following fields

* + event\_id: ID column of the dataset. Unique id for each events captured
  + device\_id: the ID of the device from which the data is captured
  + timestamp: timestamp of the event
  + longitude & latitude: map value of the event
  + city & state: city and state values of the event occurred
* **Gender data**

This data contains user specific information like the gender, age and age group of the user. The fields in the datasets are

* + device\_id: The ID of the device used by the user
  + gender & age: gender and age of the user
  + group: the age group in which the user falls
* **Phone brand data**

This data contains device specific information like the brand and model. Fields in the datasets are

* + device\_id: the ID of the device used
  + phone\_brand: brand name of the phone
  + device\_model: model name of the phone

The above mentioned data are fetched from two different sources.

* MySQL database (Gender and phone brand data)
  + Host name: ‘cpanel.insaid.co’
  + User: ‘student’
  + Database: ‘Capstone1’

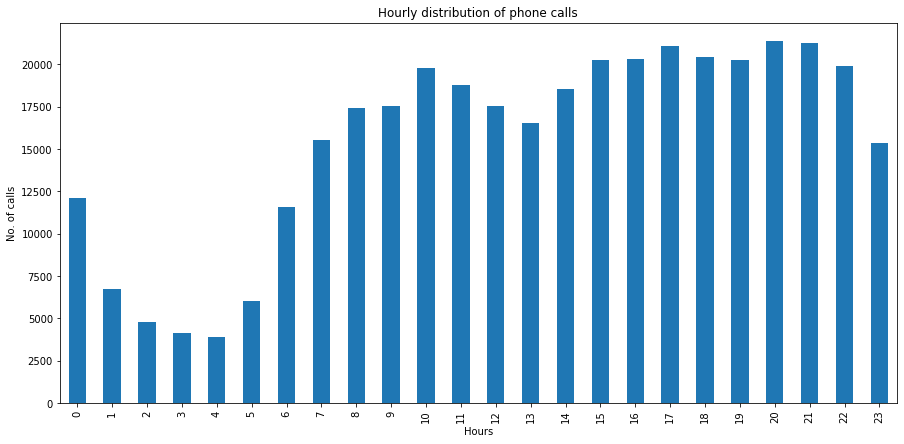
Note: Due to security reasons, password for the above database is not shared in the document

* Comma-separated value (CSV) file (Events data)

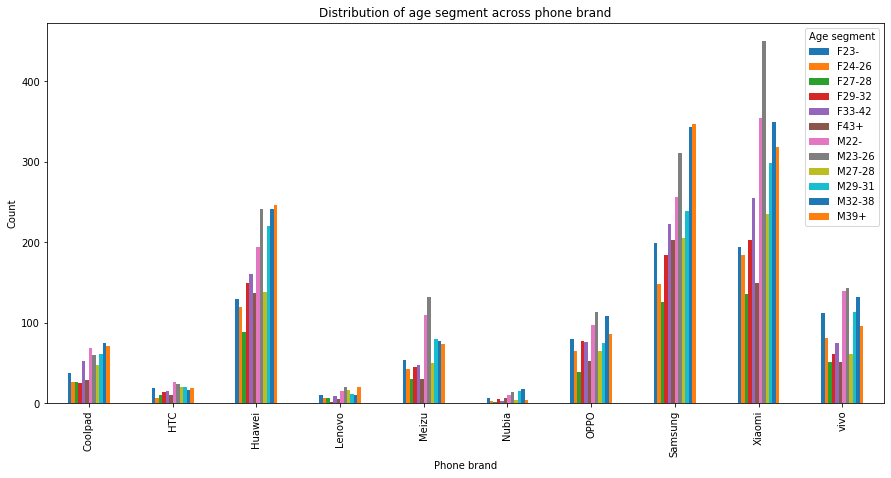
1. **Summary of Data Mining (What kind of challenges you faced with the Data and how you resolved them?, Summary of your Analysis)**

* We found some of the State values are missing in Events data. We filled up the state values by comparing with existing City and state values
* We found missing values in Latitude and longitude columns. We filled up the data based on mean of latitude and longitude of corresponding city values
* We dropped missing device\_id records
* We found some of the brand names in foreign language. In order to make it easier we have replaced all the foreign names into English.
* We also found that some of the latitude and longitude values are incorrect. Plotted the data in map and updated it with the correct value.

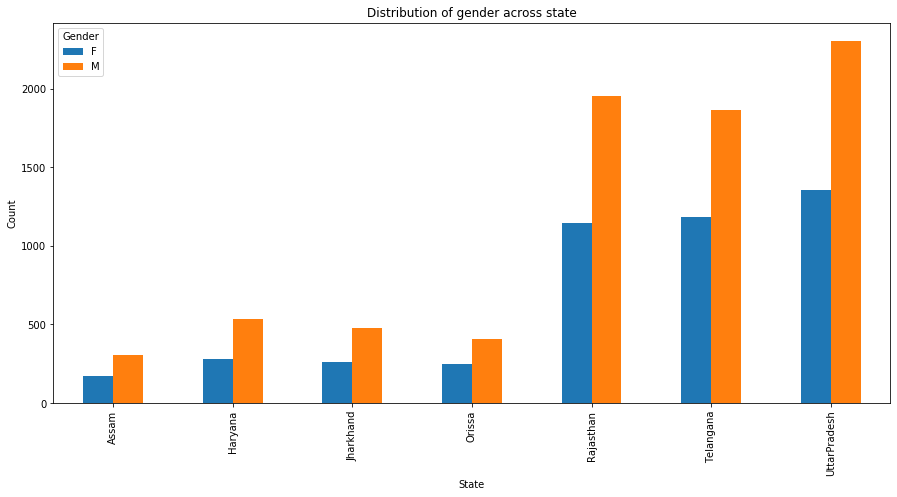
1. **Proposed Solution for Customers (Describe your Analysis in Detail)**
2. Based on the analysis of below ‘Hourly distribution of phone calls’ diagram, the calls distributions are lower in the range from 12:00 AM to 6:00AM. So we could offer free calls or call rate cutter plans to the customers during this hour range to increase the activities



1. Based on below diagram ‘Phone brand across state’, the brand Xiaomi has more count of users across state, so InsaidTelecom could do the tie-up with Xiaomi to boost up the market. In other hand Samsung has second highest count of users in all states. We can consider that too.



1. Based on the gender distribution data, InsaidTelecom can come up with plans to increase the female customer base.



1. **Tools**

The tools that are used in the capstone project are listed below

* Anaconda
* Jupyter
* MySQL Workbench

Python packages used

* Pandas
* Matplotlib
* Seaborn
* Folium
* Folium plugins
* MySQL Connector

1. **Conclusion**

Based on the EDA performed, we have analyzed the data for insights and came up with some proposals that might help InsaidTelecom to build a larger customer base and also generates more profit.