**Table of Contents**

**Consulting Report**

*Capstone Project on Data Analytics*

*Group: 1004*

### Introduction

### Project Description

### Problem Statement

### Sources of Data

### Loading Data

### Data Observations

### Quick look at the sample data

### Summary of Data Mining

### Missing Values

### Feature extraction

### Chinese phone brand names translation

### Outliers check in dfevents (identifying incorrect longitude and latitude)

### Analysis on user behaviour

### Analysis on who the users are & when do they use the services

### Analysis on how users are accessing the INSAID Telecom services

### Analysis on INSAID Telecom services used across markets

### Recommendations

### Increase user base

### [Increase user experience](#_Increase_user_experience)

**Introduction:**

It's always wonderful to see services customized to our needs. Businesses try to understand our behaviour and adjust their offerings so as to ensure we feel attached to their services.

Communication Service Providers handle and store a lot of data! They have to deal with millions of users’ call detail records, personal details and transaction data, usage patterns, technical fault data, and location data, to mention some.

To gain insights into the avalanche of data that they have at their disposal, Communication Service Providers are increasingly starting to take helps from the consulting agencies to turn their data into valuable business insights. Following are some of the benefits of consultations:

* + Targeted marketing and Personalization
  + Churn Prediction and Prevention
  + Network Planning, Optimization and Expansion
  + Prevention of Fraud and Revenue Loss
  + Data Monitization

**Project Description:**

**InsaidTelecom**, one of the leading telecom players, understands that customizing offering is very important for its business to stay competitive. **InsaidTelecom** is seeking to leverage behavioural 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.

In this **consulting assignment,** Group 1004 is expected to build a dashboard to understand user's demographic characteristics based on their mobile usage, geographic location and mobile device properties. Doing so 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.

**Problem Statement:**

For this business problem we had to study the demographics of a user based on their app download and usage behaviours. To help the customer and have an impact on their offerings, we are expected to find user behaviour and suggest the right way forward on the actionable insights with respect to marketing and product terms.

**Sources of Data:**

In this assignment, we are going to study the demographics of a user (gender and age) based on their app download and usage behaviors. 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. Due to confidentiality, we won't provide details on how the gender and age data was obtained.  
Please treat them as accurate ground truth for prediction. The data schema can be represented in the following table:

* **gender\_age\_train** - Devices and their respective user gender, age and age\_group
* **phone\_brand\_device\_model** - device ids, brand, and models phone\_brand: note that few brands are in Chinese

|  |  |
| --- | --- |
| Brand Name | Brand English Mapping |
| '华为' | 'Huawei' |
| '小米' | 'Xiaomi' |
| '三星' | 'Samsung' |
| 'vivo' | 'vivo' |
| 'OPPO' | 'OPPO' |
| '魅族' | 'Meizu' |
| '酷派' | 'Coolpad' |
| '乐视' | 'LeEco' |
| '联想 ' | 'Lenovo' |
| 'HTC' | 'HTC' |

* **events\_data** - when a user uses mobile on **INSAID Telecom network**, the event gets logged in this data. Each event has an **event id, location (lat/long)**, and the event corresponds to frequency of mobile usage. timestamp: when the user is using the mobile.

**Loading Data:**

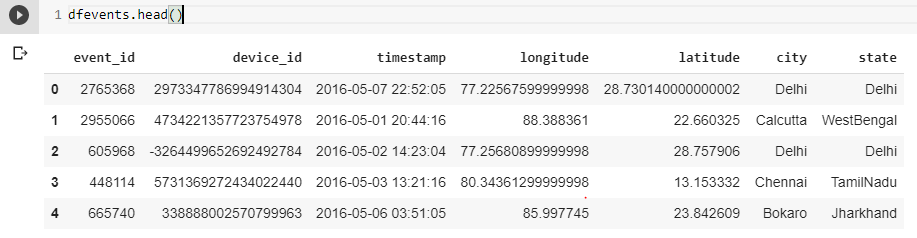
* Data for events is available from **events\_data.csv** file placed in **google drive.**  
  We first mounted google drive to access the file and then used pandas read\_csv to load data in **dfevents** data frame.
* Data for users and phone brands is available in **gender\_age\_train** and **phone\_brand\_device\_model** tables on **mysql server**, we used python mysql connector library to load data from tables in to data frames.

**Data Observations:**

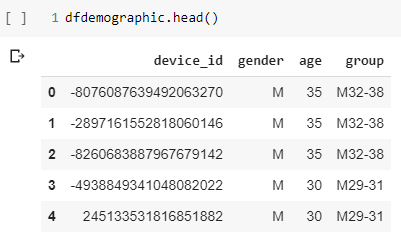
* device\_id field is a common field in all 3 tables.
* all coulmns in **dfevents** are imported forcefully as object(string), by default device\_id column was of float type which was resulting in loss of huge number of rows post merging.
* device\_id field is **int** in dfdemographic and dfphone data.
* We convert device\_id from **dfdemographic** and **dfphone** to **string(object)**

**Quick look at the sample data:**

**dfevents**



**dfdemographic**



**dfphone**



**Summary of Data Mining:**

**Missing Values:**

* Device\_id has 453 missing values.  
  Missing device ID’s are imputed by looking up for **same longitude and latitudes**.
* longitude and latitude have 423 missing values each  
  Missing longitudes and latitudes are imputed by looking up , longitude and latitude for same **device\_id**
* State has 423 missing values.  
  Missing states are imputed by looking up at corresponding cities.
* There are no missing values for **dfdemographic** and **dfphone**

**Feature extraction:**

* year, month, hour, day  fields are extracted from Timestamp which will be used for hourly usage analysis.

**df\_events** is filtered to have data specific to states in scope (AndhraPradesh, Pondicherry, Mizoram, AndamanandNicobarIslands, Meghalaya,   
HimachalPradesh, Karnataka)

There are no major issues seen with **dfdemographic** and **dfphone**

**Chinese phone brand names translation:**

* Using GoogleTrans API, we translated the brands name. However, we cannot assure the authenticity of the translation whether it is correct.
* A quick look at a small sample suggests that the translation went per the expectation.
* As there are many phone brands available globally, it is difficult to cross-check whether such a brand indeed exist. We will rely on GoogleTrans' o/p & proceed further.
* Also, we skipped the translation of 1467 unique device models. We do not see it is important for our analysis.

**Outliers check in dfevents (identifying incorrect longitude and latitude)**

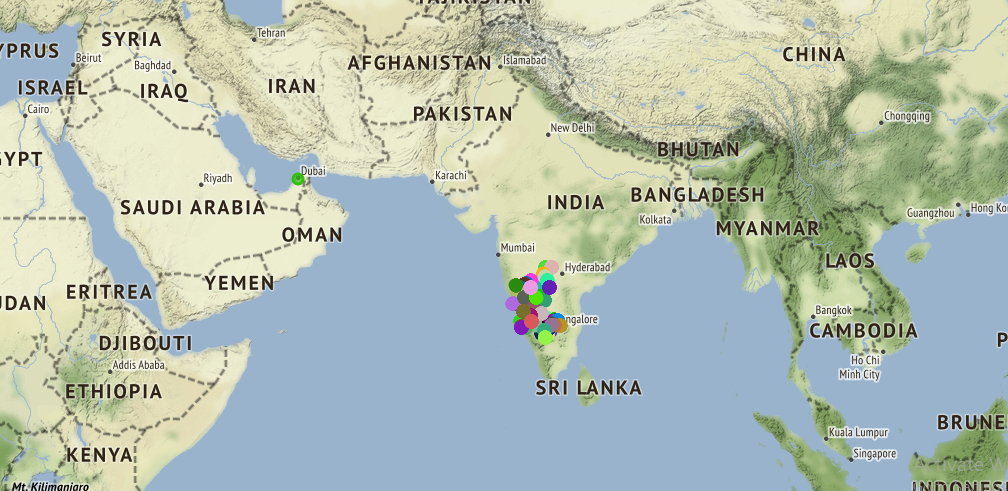
For this we plotted coordinates on map for each state to look for issue.

**Map for Andhra Pradesh**



9 records in AndhraPradesh have incorrect coordinates  
Below bad coordinates have 3 records each which need to be replaced based on their corresponding city-state values Latitude:34.5553, Longitude:69.2075 -- Kabul  
Latitude:25.2048, Longitude:55.2708 -- Dubai  
Latitude:41.8719,Longitude:12.5674 – Rome

**Map for Karnataka**



**3 records in Karnataka have incorrect coordinates**  
There are 3 records having below bad coordinates which need to be replaced based on their corresponding city-state values Latitude:25.2048,Longitude:55.270 – Dubai

**Map for Pondicherry**



No issues found for Pondicherry

**Map for Mizoram**



No issues found for Mizoram

**Map for Andaman and Nicobar Island**



No issues found for Andaman and Nicobar Island

**Map for** **Meghalaya**



No issues found for Meghalaya

**Map for Himachal Pradesh**



No issues found for Himachal Pradesh

* Using Folium, bad coordinates have been identified for the state of AndhraPradesh & Karnataka which need to be replaced with the right ones.
* There are 63 records having null latitude-longitude. All these records are in the state of AndhraPradesh. These need to be replaced with the right ones.

Treating bad coordinates:

* extract the city & state information of missing record
* get the most frequent coordinates for city & state pair
* replace the bad coordinates with the right ones

**Analysis on user behaviour**

**Analysis on who the users are & when do they use the services**

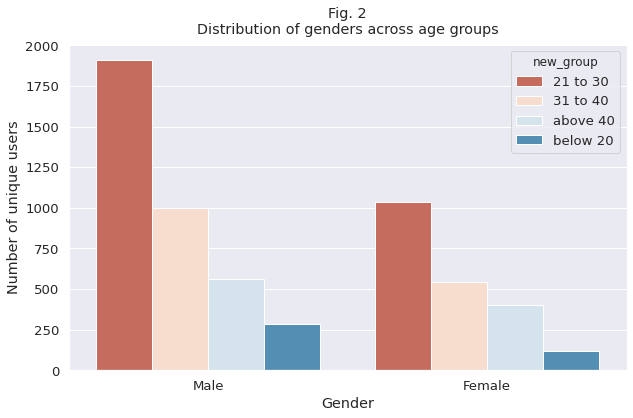
[Figure 1] shows the hourly usage of the INSAID Telecom call service to understand when during the day the users use the INSAID Telecom call services the most and the least.

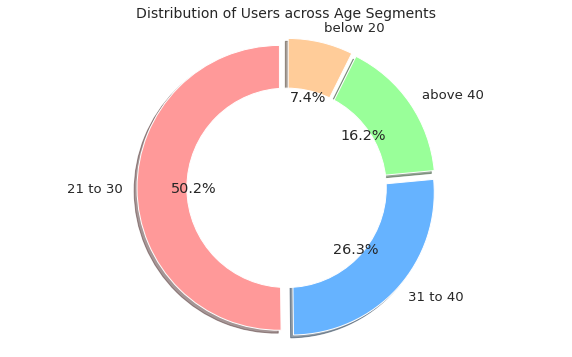


**Key findings:**

* The usage picks up at 6am and follows the trend till 10am when the usage is at its **peak**. The trend is following the normal day activity as people start off their day with commuting and communicating.
* The pattern shows that post 10am the usage starts declining and remains **constant** for most part of the day till 6pm. The trend is in-line with people generally settling down in their daily schedule and carrying on with predicted daily activities.
* The usage starts **declining** at 9pm till 4am when the usage is at its **low**, again in-line with normal daily routine when people generally mark their day off and rest for another day.

[Figure 2] shows slicing of the user type based on gender across age-groups to understand the different segments of the INSAID Telecom users.



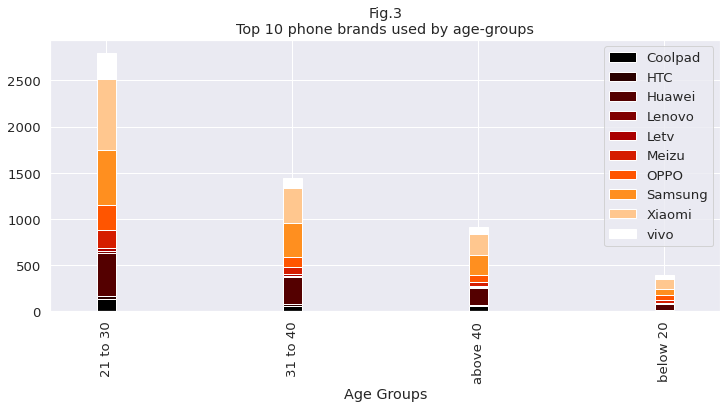


**Key findings:**

* INSAID Telecom has more male subscribers than the female subscribers across *all* the age groups
* **50%** of the users fall between 21 to 30 years of age. The 31 to 40 age-group has the second highest user base with **26%** of the share.
* Age-groups Below 20 and Above 40 seem to have low preference for INSAID Telecom services

**Analysis on how users are accessing the INSAID Telecom services**

[Figure 3] shows which top 10 mobile phone brands are widely used by the users of different age-groups & genders.

****

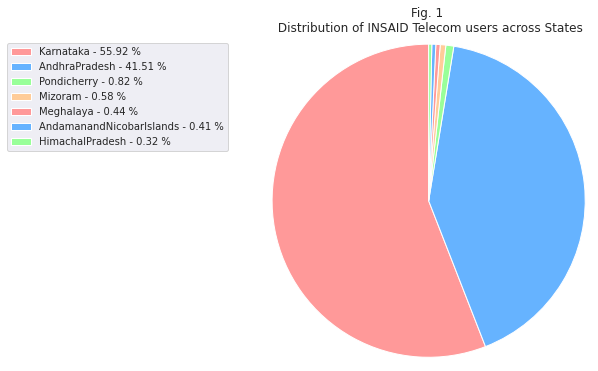
****

**Key findings:**

* In the top 10 phone brands, Chinese brands are having the lion's share. There are **8** Chinese brands in the top 10 list!
* The top 10 brands account about **50%** of the users
* The Chinese manufacturing giant Xiomi leads the favourite phone devices among both males and females.
* Samsung brand, a South Korean giant, is next most preferred phone brand after Xiomi among both males and females
* Chinese brands are preferred by all the age-groups. For instance, Xiomi mobile phones are the most preferred handsets across all age-groups.

**Analysis on INSAID Telecom services used across markets**

[Figure 1] shows the distribution of the users across 7 states/UT.

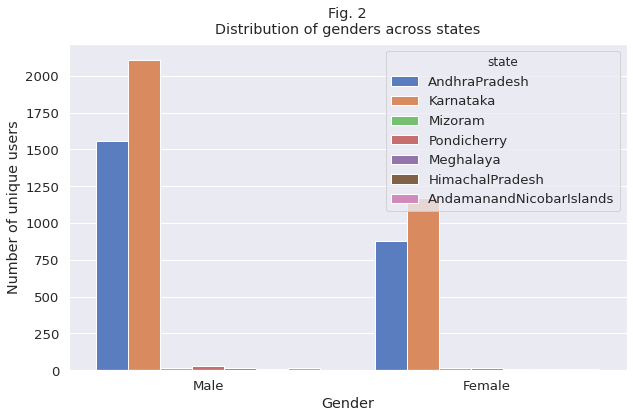


**Key findings:**

* Among 7 regions, INSAID Telecom has most of the users in the states of Karnataka & AndhraPradesh, **53%** & **42%** respectively.
* Karnataka & AndhraPradesh account for **95%** of the user base which is highly concentrated. Other 5 states have negligible user base.
* Though the states of Karnataka and AndhraPradesh have high population (as per the consensus data) which might be one of the driving factors for more user base, INSAID Telecom has *not* penetrated in other potential markets

[Figure 2] shows the distribution of the genders across 7 states/UT.

By analyzing the distribution of the genders (a high-level view on a type of the user) across states, we uncovered the *ratio* among the genders.



**Key findings:**

* Across all the states, INSAID Telecom has more male users than the female users
* States of Karnataka and AndhraPradesh which account for the major user base, have **38%** & **28%** respectively more male users than the female users.

**Recommendations:**

India is home to the second largest and youngest nations in the globe. Over the past few years, the telecom sector has seen a huge growth due to innovation in technology (smartphones, network spectrum) & change in user aspirations & needs. Still INSAID Telecom has low user base. Refer below census data as of 2011 [[source](https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_population)]. INSAID Telecom users are just a fraction of the census. Based on this and above key findings, we recommend some strategies to increase the user base and user experience.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| State | Population | Sex-Ratio | Literacy %age | Per capita income(NSDP) | INSAID users | %age |
| [Andhra Pradesh](https://www.census2011.co.in/census/state/andhra+pradesh.html) | 4,95,77,103 | 993 | 67.02 | ₹ 1,51,173 | 4930 | 0.0099 |
| [Karnataka](https://www.census2011.co.in/census/state/karnataka.html) | 6,10,95,297 | 973 | 75.36 | ₹2,31,246 | 6481 | 0.011 |
| [Mizoram](https://www.census2011.co.in/census/state/mizoram.html) | 10,97,206 | 976 | 91.33 | ₹ 1,58,621 | 70 | 0.0063 |
| [Andaman and Nicobar Islands](https://www.census2011.co.in/census/state/andaman+and+nicobar+islands.html) | 3,80,581 | 876 | 86.63 | NA | 41 | 0.010 |
| [Puducherry](https://www.census2011.co.in/census/state/puducherry.html) | 12,47,953 | 1037 | 85.85 | ₹ 2,37,279 | 96 | 0.007 |
| [Himachal Pradesh](https://www.census2011.co.in/census/state/himachal+pradesh.html) | 68,64,602 | 972 | 82.8 | ₹ 1,95,255 | 31 | 0.00045 |
| [Meghalaya](https://www.census2011.co.in/census/state/meghalaya.html) | 29,66,889 | 989 | 74.43 | ₹ 98,151 | 56 | 0.0019 |

**Increase user base:**

* INSAID Telecom should revise its marketing strategy in the under-penetrated markets according to the needs of the local markets.
* Identify market segments (males, females, professionals, students, businesses, etc.), and relevant product mix (Calls, SMS, Data, etc.) to match with the local markets
* Different offerings for different markets. For instance, small markets (in terms of population) might have different needs (for instance, low call usage and high data usage)
* Targeted offerings for age-groups. For instance, tie up with online educational apps ex. Byju’s learning app for students to attract students (age group < 20 years).
* Offer attractive data plans ranging from a variety of speeds and data for as short as 1-day to multi-year validity plans.

**Increase user experience:**

* Conduct a survey to understand the user experiences across all markets. This might uncover the quality of the INSAID Telecom services in different regions
* Focusing increasingly on data and value-added services apart from cutting costs through innovative models
* 24x7 customer support services. AI Chatbot options for customer support, extensive self-help library and community
* Monitor visitor actions to predict likelihood of purchase, and then proactively engage these visitors with chat or targeted content. Probe for their needs and interests, and answer their questions in real time, dramatically increasing online conversions
* Providing assistance with account registration, online bill-pay, and other processes up-front reduces contacts needed later, improves collections, and increases customer satisfaction and retention
* Conduct an audit on Telecom infrastructure & internal systems to analyze the quality of the services across all regions and major mobile brands.