**Consulting Report**

**Table of Contents**

1. **Introduction**

It’s always wonderful to see services customized to customer’s needs. Hence, business try to understand customer’s behaviour and adjust their offering so as to ensure “customer” feel attached to their services. Insaid Telecom, one of the leading telecom players, also understands that customizing offering is very important for its business to stay competitive. It is evident that age and gender play a significant role in the amount and type of phone and application usage. In this context, currently, they are 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.

1. **Project Description**

Insaid Telecom, one of the leading telecom players, understands that customizing offering is very important for its business to stay competitive. Currently, Insaid Telecom 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, Insaidians are expected to build a dashboard to understand 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 which are relevant to their users and catered to their preferences?

1. **Problem Statement**

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

**States to focus for Consulting**

Tamil Nadu

Manipur

Chandigarh

Tripura

Uttar Pradesh

Arunachal Pradesh

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

* Replaced Chinese words present in the fields phone\_brand and device\_model of phone\_brand\_device\_model dataset with equivalent English words given.
* Replaced missing values present in State column of events\_date with appropriate value. Use corresponding city value to populate the missing states.
* Replaced missing values present in longitude and latitude columns of events\_data with the help of device\_id.
* Found the data issue (inappropriate value) present in longitude and latitude fields of events\_data with the help of Plotly and Folium libraries. This co-ordinates issue of only 'Tamil Nadu', 'Manipur', 'Chandigarh', 'Tripura', 'Uttar Pradesh', 'Arunachal Pradesh' states have been analysed and resolved. We have only found outliers for Tamil Nadu and Uttar Pradesh. For Tamil Nadu, the incorrect data (longitude and latitude) present in 9 rows have been replaced with the correct longitudes and latitudes for the same device\_id. For Utter Pradesh, fix has been explained in below point.
* After analysing the data of Uttar Pradesh, we found that one of the cities named 'Kadi' has incorrect state. It is part of Gujarat state but as per data it is Uttar Pradesh. So, we updated the state of 947 rows having city=Kadi from Uttar Pradesh to Gujarat.
* Replaced the missing values present in device\_id column of events\_data with the help of corresponding longitude and latitude values.
* Converted timestamp column to Datetime.
* Filtered the states based upon problem statement i.e., only keep these states for our further EDA ('TamilNadu','Manipur','Chandigarh','Tripura','UttarPradesh','ArunachalPradesh'.
* Merged all Three Datasets into a single Data Frame for EDA

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

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.

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

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.

|  |  |
| --- | --- |
| 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' |

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

|  |  |
| --- | --- |
| **Observation** | **How do you want to resolve it (or) How did you resolve it?** |
| **Dataset 1: InsaidTelecom Events Data Frame** | |
| Events data has 7 columns and 3252950 rows | No Action Needed - Observation |
| event\_id is unique which means no duplicate events. | No Action needed |
| device\_id has only 60865 distinct values. It has 453 missing value | Replace device ID based on longitude, latitude values |
| timestamp is of object data type | To be converted to Datetime datatype |
| longitude and latitude fields has missing values. It has 423 missing values | This column may not be needed for EDA and can be dropped |
| city has a total 933 distinct values and no missing values | No Action needed |
| Delhi has maximum number of events per day | No Action needed |
| state has 32 distinct values and 377 missing values | Replace the missing values by looking into the City value |
|  |  |
| **Dataset 2: InsaidTelecom Demographic(gender\_age) Data Frame** | |
| Total 4 columns and 74645 rows | No Action needed - Observation |
| device\_id is unique which means no duplicate device. | No Action needed |
| More Male mobile users compared to Female | No Action needed - Observation |
| age of mobile users, minimum age: 1; maximum age: 96 | No Action needed - Observation |
| age is right skewed | No Action needed - Observation |
| Male 23-26 age group of people are maximum using the mobile | No Action needed - Observation |
|  |  |
| **Dataset 3: InsaidTelecom Phone Details(phone\_brand\_device\_model) Data Frame** | |
| Total 3 columns and 87726 rows | No Action needed - Observation |
| device\_id is unique which means no duplicate device. | No Action needed |
| phone\_brand and device\_model has no missing values | No Action needed |
| Total 116 phone brand out of which Xiaomi is preferred the most by the users. Almost 25% market is owned by Xiaomi followed by Samsung | No Action needed |
| Total 1467 different mobile models are used as per the data | No Action needed |
| Volume of Phone details is more than demographic details as one user may have multiple mobile devices. | No Action needed |
| Both brand and model columns have some Chinese words | Need to replace the Chinese words with equivalent English language |

1. **Proposed Solution for Customers (Describe your Analysis in Detail)**

* Scope to increase user base particularly in Manipur, Chandigarh, Tripura & Arunachal Pradesh through better network coverage and attractive plans particularly aimed at the age segment 20-40 years
* Xiaomi & Samsung being the most popular brands, we could tie up with them to provide InsaidTelecom connection along with the handset at an attractive price
* Special offers/ plans for females, launched on specials occasions like Women's Day, Mother’s Day etc could help increase Female user base
* Introduce special plans/offers for usage between 2:00 AM and 7:00 AM to incentivize usage at this time and avoid network congestion during peak hours.
* There is scope of improve Business in Uttar Pradesh and Tamil Nadu as well. These are densely populated area so we need to improve our customer base by providing more offers.

1. **Tools**

* DS Tools

০ Mysql for the purpose of a web database

০ Folium for visualizing geospatial data and for plotting interactive maps

০ Kaleido for generating static images for web based visualization

০ Numpy for working with arrays

০ pandas for data analysis

০ Seaborn for making statistical graphs in python and integrated with pandas

০ Matplot for data visualization ০ pandas profile for explorating data analysis within few minutes

০ Ploty.graph\_objs for represent parts of a figure

০ Jupyter notebook for project work Web UI Tools (You don’t have to explain this section)

* + PHP
  + JavaScript

1. **Conclusion**

* Maximum number of users are from Tamil Nadu (60.49%) followed by Uttar Pradesh (38.11%). Both these states accounts for 98.6% of the total users in these 6 states. This could be because they are the larger states and have more towers and better network coverage. Remaining 4 states (Tripura, Chandigarh, Arunachal Pradesh and Manipur) have only 133 users
* Xiaomi, Samsung, Huawei, Vivo, OPPO, Meizu, Coolpad, HTC, Lenovo, LeEco are the 10 most used phone brands
* Xiaomi (26.69%), Samsung (23.67%) and Huawei (16.67%) are the top 3 most used phone brands. These 3 companies comprise more than 67% of market share in these 6 states. Xiaomi and Samsung are market leaders when it comes to mobile as they launch models regularly as per market and customer demands. Also, they launch more budget friendly models with more service centre.
* Insaid Telecom has more Males users (62.65%) compared to Female users (37.35%)
* People in the range of 20 yrs. to 26 yrs. prefer this network the most compared to other age group. Most of the users (more than 80%) are of the age 20yrs to 40yrs.Young working class are the main users.
* Calcutta has the maximum number of users followed by Delhi, Bangalore, Mumbai, Chennai. This network has users mainly in the top 10 metropolitan cities showed in the graph
* Xiaomi is the most preferred mobile brand closely followed by Samsung for each age segments except 40+. Samsung has a slightly more preference over Xiaomi for Age segment 40+.
* Tamandu and Uttar Pradesh have the maximum number of users and it seems these people prefer Xiaomi phone followed by Samsung phones
* Both Male and Female prefer Xiaomi phones followed by Samsung. Huawei comes in 3rd number.
* In Tamil Nadu, Users of almost all age groups except 33yrs-40yrs and 40+ prefer Xiaomi phone brand closely followed by Samsung. However, users in group 33-40 and 40+ prefer Samsung.
* In Uttar Pradesh, Users of almost all age groups except 40+ prefer Xiaomi phone brand followed by Samsung. However, users in group 40+ prefer Samsung.
* More Male users compared to Female users in these 6 states
* All age segments have majority Male users. More male users compared to female for all age segments if we compare state wise (except Tripura)
* If we see the distribution of genders across top 10 phone brand, you would notice majority users for each brand are Male
* For all 6 states, majority of users belong to the age segment between 20yrs to 40yrs.
* Across top 10 phone brands, maximum users belong to the age group of 20yrs to 26yrs. Majority users come between 20yrs and 40yrs.
* Call Volume is Highest at night after 6:00 PM till 1:00 AM
* Call Volume is the lowest in the early hours between 2:00 AM and 7:00 AM after which it again starts rising