**Consulting Report**

**Table of Contents**

[Introduction](#_heading=h.d3xhek9tr4hh) 2

[Project Description](#_heading=h.hnngyi187774) 2

[Problem Statement](#_heading=h.w15njjkq4kv8) 2

[Problem Analysis (Strategy for the Problem Statement analysis)](#_heading=h.gkdww77mas48) 2

[Sources of Data (Explain about database connection, tables and their columns)](#_heading=h.pem96x9kvpt) 3

[Summary of Data Mining](#_heading=h.kz8a6520nezg) 4

[Proposed Solution for Customers (Describe your Analysis in Detail)](#_heading=h.69n1suwha5tr) 4

[Tools](#_heading=h.uey80fmzpy7e) 4

[Conclusion](#_heading=h.4ex7zagim4zz) 5

## Introduction

This project caters to the business needs of InsaidTelecom, who is one of the leading telecom players, by analyzing and showcasing its customer’s demographic characteristics so as to provide them a customizing offerings, if needed to stay competitive in business market

## Project Description

InsaidTelecom is seeking to leverage behavioral data from more than 60% of the 50 million mobile devices active daily in India. This analysis would help them to understand and interact their clients in better way.

## Problem Statement

* + Build a dashboard understand a user's demographic characteristics based on their mobile usage, geolocation, and mobile device properties help
  + Developers and brand advertisers would use these dashboards to pursue the marketing efforts which would be data-driven, relevant to individual customer and would focus on their preferences.

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

* + Exploratory Data Analytics (EDA) is done using Pandas, NumPy, Seaborn, Matplotlib, Pandas Profiling, MySQL connector
  + Univariate, Bi-Variate and Multivariate analysis is performed

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

* + **Events DataFrame** 
    - This database contains events details of the mobile users
    - When a user uses mobile on an INSAID Telecom network, the event gets logged in this database. This database is stored as a CSV file
    1. Table

       Description automatically generated

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 event\_id 3252950 non-null int64

1 device\_id 3252497 non-null float64

2 timestamp 3252950 non-null object

3 longitude 3252527 non-null float64

4 latitude 3252527 non-null float64

5 city 3252950 non-null object

6 state 3252573 non-null object

* Gender Age DataFrame
  + This database contains user details like their gender, age and the identification of their device

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 device\_id 74645 non-null int64

1 gender 74645 non-null object

2 age 74645 non-null int64

3 group 74645 non-null object

* **Phone Brand Device Model Data frame**
  + This database contains device details like device model, brand and the identification of device

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 device\_id 87726 non-null int64

1 phone\_brand 87726 non-null object

2 device\_model 87726 non-null object

## Summary of Data Mining

|  |  |
| --- | --- |
| **Observation** | **Resolution** |
| It is found that device\_id , state , latitude and longitude columns have missing values in event dataframe.  1. Device\_id - 453 rows missing values  2. State - 377 rows missing values 3.latitude and longitude - 423 rows missing values | 1.Replaced missing values present in device\_id columns of event\_data with the help of corresponding latitude value to populate the missing device\_id. 2. Populate State column with corresponding city |
| There were some discrepancies in latitude and longitude in event\_data for target states ('Tamil Nādu','Manipur','Chandigarh','Tripura','Uttar Pradesh', 'ArunachalPradesh'). Found this discrepancy with the help of plotly and folium libraries. | Fixed the outliers in lat and long of Tamil Nādu and Uttar Pradesh with help of same device\_id. |
| Tamil Nādu state found that 9 rows have incorrect lat & long data. | Have replaced with correct longitude and latitude for the same device\_id. |
| For Uttar Pradesh we found that one of the cities named 'Kadi' has incorrect state. | “Kadi” it is part of Gujarat state but in event data its mapped to Uttar Pradesh. This has been fixed with updating 947 rows of Kadi city to Gujarat state from Uttar Pradesh state. |
| Found that some brand/ device model names are in Chinese language in phone\_ brand and device\_model column. Need to replace/map with English name for better understanding. | Translated Chinese name present in phone\_brand and device\_model of phone\_model dataset with equivalent English name. |
| There were outliers in age column- 1,6,94 and 96 years are outliers. | Replaced these values by mapping the index of corresponding values in age column. Minimum age to 10 and maximum age to 89. |
| Found that dtypy of device\_id in gender\_age\_train and phone\_model dataset.  Need to change to float for merging that dataset. | Converted dtype to Float with .astype() function. |
| Found that timestamp dtype is object, it supposed to be datetime. | Converted the dtype of timestamp column to Datetime from object with .to\_datetime() function. |

## Proposed Solution for Customers (Describe your Analysis in Detail)

* Other than Tamil Nādu and UP, presence of users in others states are very less, hence INSAID telecom company needs to invest more with a strong marketing campaign, through better network coverage or some promotional offers in other states` like Manipur, Chandigarh, Tripura, Arunachal Pradesh
* InsaidTelecom can also introduce fiber connectivity for Tamil Nādu and Uttar Pradesh to increase their business, as these cities have more users.
* HTC, Lenovo. LeTV shows a very small % in terms of subscription of network from INSAID telecom, company needs to tie up with some offers in these brands
* People from 20-40 age group are having highest number of subscriptions to INSAID Telecom, their testimonial /feedback can be taken and can be published to attract more users
* senior citizens subscribe less, any issue if they are facing needed to be sought out by getting connected to such users
* Xeiomi, Samsung, Huawai brands are popular across all age groups, INSAID Telecom can attract these customers subscription by offering some lucrative offers specific to these brands
* % Of Women subscribers are less than Men, try to bring the marketing platform where women presence is more like Facebook, Instagram, Pinterest, can also keep marketing content related to women, company can also introduce some offers for female users on Women’s Day, Mother’s Day.
* More calls are happening in midnight, company has to come up with new talk time offers for these hours to attract more people.

## Tools

* **DS Tools**
  + Numpy
  + Panda
  + Matplotlib
  + Plotly
  + Folium
  + Translation tool

## Conclusion

Objective of this analysis is to help INSAID TELECOM to understand their customer’s behavior and a right way forward that would directly impact the company's offerings.

These are some of the quick insights from this exercise: (Analysis done on 6 states only, which are ‘Tamil Nādu’,’Manipur’,’Chandigarh’, ‘Tripura’, ‘Uttar Pradesh’,’ArunachalPradesh’.)

* Tamil Nādu (around 61%) and Uttar Pradesh (around 38%) attracts most of the customers, all other states have small size users. There are 9,405 users are from Tamil Nādu and Uttar Pradesh that’s around 98% of total users.
* Xiaomi, Samsung, Huawei, vivo, OPPO, Meizu, Cool, HTC, Lenovo and LeTV are Top 10 Phone Brands.
* Around 27% of users uses Xiaomi brand, 24% uses Samsung and 17% uses Huawei. These 3 companies comprise more than 68% of market share in these 6 states.
* Most of the people prefer Redmi Note followed by MI device model. Both device models are belonging to Xiaomi phone brand.
* Company have more Males users (62.7%) compare to female users (37.3%) across 6 states
* The minimum age of user is 10 and the maximum age of user is 89 years.
* More users are from 20 to 26 years age group who uses InsaidTelecom network more compare to other age groups.
* 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+.
* Among 6 sates Tamil Nādu and Uttar Pradesh have more mobile users and these people prefer Xiaomi brand followed by Samsung brand phones.
* InsaidTelecom have more Male users compare to Female Users. All age segments have majority of Male users in all 5 states except in Tripura female users are more.\
* For all states, more users belong to the age group of 20 to 26 years. Majority of users (more than 80% of all) falls under 20 to 40 years age group. Top states are Tamil Nādu and Uttar Pradesh.
* In Tamil Nādu, 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.
* Majority of users for each brand are Male and they prefer Xiaomi followed by Samsung phone.
* Most of the Events logged in May month only for the year 2016 and Male users have done more calls compare to Female users. Most of the events logged between 7AM to midnight 12AM.
* Age groups between 20 to 26 years and 29 to 32 years have done more calls between 8PM to 1 AM & 9 Am to 11 Am compare to other age groups.