**Consulting Report**

Table of Contents

1. Introduction
2. Project Description
3. Problem Statement
4. Problem Analysis (Strategy for the Problem Statement analysis)
5. Sources of Data (Explain about database connection, tables and their columns)
6. Summary of Data Mining (What kind of challenges you faced with the Data and how you resolved them?, Summary of your Analysis)
7. Proposed Solution for Customers (Describe your Analysis in Detail)
8. Tools

* DS Tools
* Web UI Tools( You don’t have to explain this section)
  + PHP
  + JavaScript

1. Conclusion

**Introduction :**

This capstone project is developed by team group no: 1040 as part of our term assignment from INSAID. All the project work done, discussion, implementation , conclusion are done by three people – Purna Behrananda , Harjeet Kaur, Sunil Kumar Mano.

**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 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.

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

This analysis report 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 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. The analysis needs to be done for demographics of data and their telecom usage region wise. There are different 9 task assign to analyze the data into different aspect.

1. Distribution of Users(device\_id) across States.
2. Distribution of Users across Phone Brands(Consider only 10 Most used Phone Brands).
3. Distribution of Users across Gender.
4. Distribution of Users across Age Segments.
5. Distribution of Phone Brands(Consider only 10 Most used Phone Brands) for each Age Segment, State, Gender.
6. Distribution of Gender for each State, Age Segment and Phone Brand(Consider only 10 Most used Phone Brands).
7. Distribution of Age Segments for each State, Gender and Phone Brand(Consider only 10 Most used Phone Brands)
8. Hourly distribution of Phone Calls.
9. Plot the Users on the Map using any suitable package.

For our group the states of focus are

MadhyaPradesh, Chhattisgarh, Uttaranchal, JammuandKashmir, Goa, Nagaland

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

During initial analysis of data, we have found many observations by first looking at the data after importin to juypter notebooks. We have documented all our observation at google sheet [Observation\_1040](https://docs.google.com/spreadsheets/d/1gJQyEy_VfsK_erBUmXPTooct5BmngymuIYx28O5w7pA/edit#gid=0) . Some common observations were:

1. Negative device ID in all the tables
2. There are chinese phone brand names in phone\_brand\_device\_model table, that need to map corresponding English names.
3. Device information in gender\_age\_train and phone\_brand\_device\_model tables are unique.
4. we can join gender\_age\_train and phone\_brand\_device\_model tables to get the user age group information vs device details.
5. We have 12 different age groups found as follow: Male Age Group: M22- , M23-26 , M27-28, M29-31, M32-38, M39+ Female Age Group: F23- , F24-26, F27-28 , F29-32, F33-42, F43+
6. Different age group bucket for male and female, need to normalize it before use it.

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

The data is divided in to three tables mainly

**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 (latitude/Longitude), and the event corresponds to frequency of mobile usage,timestamp: when the user is using the mobile.

**For obtaining the data from these tables , we have created the sql connection using** "mysql.connector" package in Python. Also provided the credentials host,username , password and database name to fetch the data from these table to dataframes.

* host 'cpanel.insaid.co'
* user 'student'
* passwd 'student'
* database 'Capstone1'

We have created connection with sql databased using create\_engine python command and then read\_sql to read the data from table of that database. We converted the data into csv using to\_csv commands. Telecom\_EDA\_Data\_Collection.ipynb is attached in the zip .

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

During the Data cleaning part , we have worked upon observation that we gather in initial team discussion and captured at google sheet [Observation\_1040](https://docs.google.com/spreadsheets/d/1gJQyEy_VfsK_erBUmXPTooct5BmngymuIYx28O5w7pA/edit#gid=0) .

1. We have negative values of device\_ids in all the three tables (**gender\_age\_train** , **phone\_brand\_device\_model** ,events\_data) and removed them by using abs() function.
2. Phone\_brand\_device\_model has 87726 records with no null values and have information about device\_id,phone\_brand,device\_model.
3. We have some phone\_brands which has Chinese names and cannot be readable. So, we created a dictionary map to map the Chinese names to their English name. Based on the information provided on portal we created the following dictionary

phone\_brand\_map = {'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'} and also created function def getBrandValue(symbol) to map these Chinese brand name to their English name.

1. In events\_data table, we have filter the records based on our state of focus - MadhyaPradesh , JummuPradesh , Goa, Nagaland, Chhattisgarh, Uttaranchal using list and putting events record into another datarame which are present in the list using enumerate(states). There are total 261097 records in selectedstates dataframe.
2. In filtered states, there are 51 null values found for device\_id, so we have deleted those 51 records as they are too small number as compare to the total number of records.
3. There are 63 missing values of longitude and latitude of Indore city of Madhyapradesh . As they all belong to same city, so we replace them with the longitude and latitude value of Indore city.

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

Based on our analysis, we are proposing the following recommendation to the management.

1. Madhya Pradesh has the maximum number of INSAID Telecom users, so provide good customer services to keep the pace going.
2. Goa, Nagaland, Jammu & Kashmir and Uttaranchal have 100-300 customer only, so focus on the marketing strategies based on customer preferences and try to increase the customer base by providing low rate Telecom packages.
3. Xiaomi, Samsung, Huawei, vivo, OPPO are the top five popular brands among the userbase. Focus on providing attractive data packages to this customer segment.
4. The number of female customers are half that of males, Add some packages to attract female customers, add cookery app , watching online TV shows app, beautician & fitness app to attract female customers.
5. We have less customer base between the age segment of 27 to 28 yrs . Some job oriented, latest new, Netflix free subscription can help to add user in this segment.
6. OPPO, Meizu, Coolpad, vivo are some of the emerging phone brands among the users, we can add new packages around them.
7. Samsung, Xiaomi are pre-dominant phone brand of biggest customer base in Madhya Pradesh, Whereas HTC has very few customer in Nagaland and need to be worked upon.
8. Very low female customer base in Goa, Chhattisgarh and Jammu & Kashmir . We can add some offers to attract female customers.
9. Very few customer base in Goa, As it is the most popular tourist destination in India. We can provide some roaming call offers with attractive data packages to attract foreign tourists.
10. Samsung , Xioami, Huawei are the top three popular phone brand across all age segments and focus on these to hold the market.

**Tools**

* DS Tools
  + Mysql.connector , Panda, dataframe , folium ,matplotlib,seaborn
* Web UI Tools( You don’t have to explain this section)
  + Tableau

**Conclusion**

In Conclusion, INSAID Telecom should focus on increase the customer base in Goa , Nagaland, Uttaranchal , Chhatissgarh and focus attracting female customers. Xiaomi , Samsung , Huawei are the top phone brands and HTC, Coolpad, vivo, LeEEco are the emerging phone brands among users and need to focus more in that. Madhya Pradesh has the maximum customer base and focus to keep the momentum going.