**Consulting Report-Group 1026**

**Table of Contents**

1. Introduction
2. Project Description
3. Problem Statement
4. Sources of Data
5. Problem Analysis
6. Summary of Data Mining
7. Proposed Solution for Customers
8. Tools
9. Conclusion

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Introduction**

As part of our Capstone project, Group 1026 worked on the studying the demographics (gender and age) of users of InsaidTelecom based on their usage behaviours to provide actionable insights for impactful offerings by the company. Our analysis is restricted to users of the following states:

* MadhyaPradesh
* Chhattisgarh
* Uttaranchal
* Jammu and Kashmir
* Goa
* Nagaland

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**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.

In this consulting assignment, we are building 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.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Problem Statement**

Suggest actionable insights to help the company design its offering based on user behaviour and demographics.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Sources of Data**

Following three tables were extracted from Database using the “mysql.connector” package in Python.

|  |  |
| --- | --- |
| **Table Name** | **Description/Name of Columns** |
| gender\_age\_train | device ids  gender  age  age group |
| phone\_brand\_device\_model | device ids  brand  models phone\_brand |
| 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). |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Problem Analysis**

We divided the broad goal into following smaller goals:

1. Distribution of Users(device\_id) across States
2. Distribution of Users across Phone Brands
3. Distribution of Users across Gender
4. Distribution of Users across Age Segments
5. Distribution of Phone Brands for each Age Segment, State, Gender.
6. Distribution of Gender for each State, Age Segment and Phone Brand
7. Distribution of Age Segments for each State, Gender and Phone Brand

Note: We are considering 10 most used phone brands for analysis.

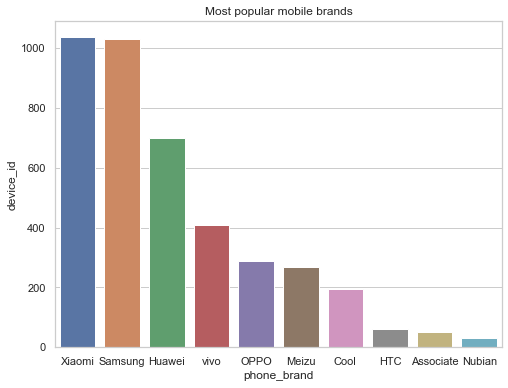
Collating the above points, we have built actionable insights for the company.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

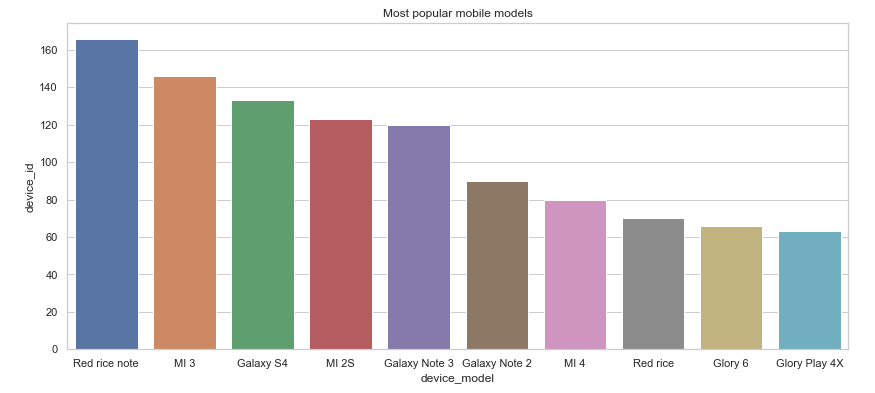
**Summary of Data Mining**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S no. | Observation | | | How we resolved it |
| 1 | In events\_data, following are the null values:  device\_id = 51 | | | It was observed that corresponding to 3 latitude and longitude values we have missing device\_id values (17 each). So, these null values were filled using the corresponding latitudes and longitudes values. |
|
|
| 2 | In events\_data, following are the null values:  longitude = 63  latitude = 63 | | | It was observed that the same device\_id values had missing longitude and latitude values. Corresponding to 3 device ids, we have latitude and longitude values (21 each). So, these null values were filled using the corresponding device ids. |
|
|
| 3 | In events\_data, Except Madhya Pradesh, No NULL values for other states | | | NA |
|
|
| 4 | In events\_data, Data is for duration April 30, 2016, to May 7, 2016, for all states, BUT for Jammu and Kashmir and Nagaland, it is for duration May 1 - May 7 | | | NA |
|
|
| 5 | In the given dataset table records, many values of device\_id are logged as negative. | | | The negative values are converted back to positive assuming they are logged as negative mistakenly and converted values are used for analysis. |
|
|
| 6 | In events\_data, Out of 261097, 232690 is from state MadhyaPradesh (89.12%) | | | NA |
|
|
| 7 | In gender\_age\_train, there are 12 groups, 6 for males and 6 for females. This groups are divided based on different age groups | | | NA |
|
|
| 8 | In gender\_age\_train, there are 2702(62.88%) males and 1595(37.11%) females | | | NA |
|
|
| 9 | In gender\_age\_train, Min age is 6 and Max age is 90 | | | NA |
|
|
| 10 | In gender\_age\_train, No Null value is present | | | NA |
|
|
| 11 | In phone\_brand\_model, No Null value is present | | | NA |
|
|
| 12 | In phone\_brand\_device\_model, Out of 60 phone brands, 52 are in Chinese.  Out of 676 models, 92 are in chinese | | | Translated using the Chinese to English mapping information present in the dataset description and Google translate package. |
|
|
| 13 | In phone\_brand\_device\_model, Many of them are a combination of chinese and English names like '红米Note2', '联想黄金斗士S8' which cannot be replaced using the method used for phone brand column.  These are new values like '联想黄金斗士S8' is 'Lenovo Gold Fighter S8' | | | Translated the values from English to Chinese using google translator python package |
|
|
| 15 | In phone\_brand\_device\_model  Corresponding to brand name : Samsung  Device name: 联想黄金斗士S8 (English google translation: 'Lenovo Gold Fighter S8' | | | Translated using google translator in python. Saved all the converted names in a list and then replaced in the dataframe. |
|
|
| 16 | In gender\_age\_train, Major data point lies in the age range of 23 to 43 | | | NA |
|
|
| 17 | In the three tables, device\_id is the common field to combine data from all the tables. | | | All three tables were merged on device id |

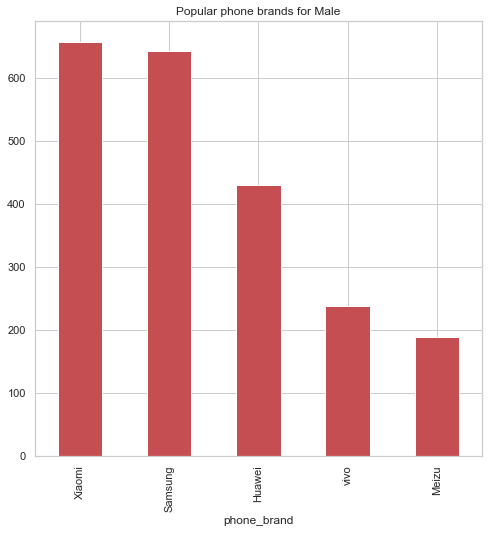
The most popular brands are shown below.



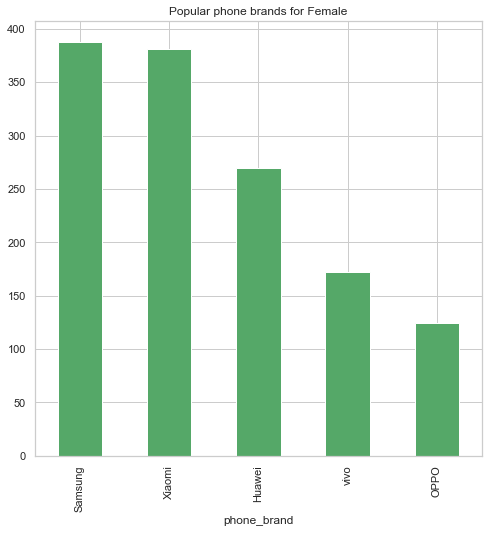
The most popular device models are shown below.

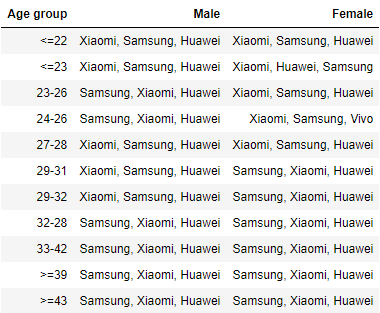
****

Mobile brands preferred by Male are shown below



Mobile brands preferred by Female are shown below-



The most popular brands for various age groups are shown below.****

**State-wise Analysis Report**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **State** | **City with most devices** | **Most recorded dates** | **Most recorded time** | **Brand preference** |
| Madhya Pradesh - 232690 events | Indore | May 03, 2016  May 04, 2016 | 8 PM - 10 PM | In Indore, Xiaomi and CoolPad  In other cities, Xiaomi is the best seller, followed by Samsung and Huawei.  Huawei is more popular with Female as compared to Male |
| Chattisgarh - 9754 events | Jabalpur | May 05 2016 and May 06 2016 | 3 AM | Customers with age groups till 25 prefer Xiaomi and Huawei.  Customer with age group 25-45, prefers Huawei,Xiaomi and Samsung in the order  Customers with age group >40 mostly prefers Samsung and Vivo |
| Uttrakhand - 7720 events | Kashipur | May 01 2016 has recorded maximum events followed by May 03 2016 | 6 AM to 9 AM | Samsung and Huawei are more populate with Male, with Xiaomi closely next  Female prefers Xiaomi followed by Samsung and Vivo  Customer with age group till 25 prefers Xiaomi and Huawei  Customer with age group 25-45, prefers Huawei,Xiaomi and Samsung in the order  Customers with age group >40 mostly prefers Samsung and Vivo |
| Jammu and Kashmir - 5385 events | Anantnag | 01 and 02 May 2016 | 4 AM - 4:30 AM | Samsung and Xiaomi are more popular with Male followed by Meizu and Oppo and Huawei  Female prefers Samsung then Xiaomi and Huawei  Xiaomi and Samsung are the most preferred ones across the age group |
| Goa - 3250 events | Mormugao | 02 May 2016 and 04 May 2016 | 12 AM | Female prefer Xiaomi and then Vivo  Male prefer Huawei, CoolPad and Samsung  Customers with age group till 25 prefers Vivo and Samsung  Other age group prefers Xiaomi and Meizu |
| Nagaland - 2298 events | Wokha | 07 May and 05 May 2016 | 6 PM - 9:30 PM | Male prefer Samsung and Xiaomi and Huawei  Female prefer Oppo followed by Samsung, Xiaomi and Huawei  Customer with age group till 25 prefers Samsung, Oppo and Xiaomi  Other age group prefers Samsung |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Proposed Solution for Customers**

|  |  |
| --- | --- |
| **Observation** | **Possible Actions** |
| 03 May 2016 has maximum events registered followed by 04 May and 06 May  03, 04, 06 and 05 May are weekdays while 07 May is weekend, so Weekdays have more events registered as compared to weekends | 1. Some coupons can be given on reaching a defined limit of phone usage duration on weekends. 2. Weekdays events could be business calls, so discounts on corporate connections can be considered. 3. To increase the user consumption on weekends, can introduce the complimentary talk time to be used on weekends for the users having high usage on weekdays. |
| In the night, around 8 PM, there are maximum events followed by 10 AM and 11 AM.  However, it varies from state-to-state. In Chattisgarh, the most preferred time is 3 AM. | 1. Give discounts and offers based on the preferred time, state and age group. 2. Company can target to maintain good signal strength to handle the traffic in peak hours. |
| Users with the age group of 23-30 have the majority of the mobile devices. | 1. Introduce more discounts and offers to promote sales in other age groups. |
| Xiaomi and Samsung have the majority of the customers - around 1000 each, while Huawei is next with around 650 customers.  For Male, overall order is - Xiaomi, Samsung and Huawei  For Female, overall order is - Samsung, Xiaomi and Huawei | 1. Introduce offer plans along with the top brands so that users can opt for the new mobiles along with the plan from the company. 2. Develop trust amongst customers by advertising using proper after sale service. 3. Have various marketing campaigns to acquire more customer base. |

**Tools**

**DS Tools** -

* + Jupyter notebooks
  + Pandas
  + Numpy
  + Matplotlib
  + Seaborn
  + Folium
  + Google translation APIs

**Conclusion**

* New plans can be introduced based on the different age groups,locations and their usage, like plans for data and plans for talktime based on the usage patterns.
* New plans can be based on the customer diversity, like for students, housewives and corporate employees have different requirements and usages, so can provide customized plans to them.
* Possibility to introduce combo plans consisting of attractive monthly or quarterly plans along with the top brands of mobiles, so that users can opt for the mobile devices of their choice with flexible payment options.
* Signal strength needs to be monitored and maintained in peak hours so that users should not face issues and promote the company to other future potential customers.
* Plan for more marketing campaigns in the areas with less customers to increase the customer base and can introduce some introductory offers for new customers in those areas.