2020

MAHARASHTRA TELECOM DATA CONSULTING REPORT

INSAID – CAPSTONE PROJECT 1

INSAID TELECOM CONSULTING REPORT – MAHARASHTRA (TEAM 1009)

Contents

1. INTRODUCTION

We INSAIDIANS, a group of Data Scientists from Team 1009 with following Team members have come together to analyze data for a leading Telecom player INSAID Telecom which is part of our Capstone Project-1, to understand users behavior and provide insights to customize their offerings and expand their business to stay competitive in the Telecom Market.

2. TEAM MEMBERS

Aditya	gpaditya11@gmail.com
Padmapriya M	padmapriyajain@gmail.com
Lingaraju C.H	director.lingu@gmail.com
Sulakshana	smdoiphode@gmail.com
Sayan Chatterjee	sayan22@yahoo.co.in
Hemanth Kumar G	chhera666@gmail.com
Chelladurai	chelladurai.w@gmail.com
Shyamkumar	Shyamkumar.chauhan@gmail.com
Karthik P	karthik.nttf07@gmail.com

3.   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 to 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, geo-location, and mobile device properties. This dashboard will help millions of developers and brand advertisers around the world pursue data-driven marketing efforts which are relevant to their users which cater to their preferences.

4.   PROBLEM STATEMENT

We need to understand and provide in depth user behavior analysis from the given telecom data and offer insights to help INSAID Telecom to identify new opportunities, maximize growth, customize existing offers/create new offers to satisfy their customers and also to avoid potential pitfalls.

5.   PROBLEM ANALYSIS/STRATEGY

To analyze user's behavior, we have considered the following dimensions of the users' data, there usage pattern and device specifications:

ENTITY	SL NO	DIMENSIONS
Users	1	Users Location
	2	Users Age
	3	Users Gender
	4	Users Age Group and Gender
	5	Users Phone Brand
	6	Users Device Model
	7	Users distribution by Location and Phone Brand
	8	Users distribution by Location and Device Model
	9	Users distribution by Location, Age and Gender
	10	Users Phone brand distribution by Location, Age and Gender
Usage Pattern	11	Usage by Timestamp
	12	Usage by Age
	13	Usage by Gender
	14	Usage by Group
	15	Usage by Location
	16	Peak Hours of Usage
	17	Usage by Location and Group
	18	Market share of Phone brands by Age
Usage Hours	19	Usage Peak Hours by Location
	20	Usage Peak Hours by Age
	21	Usage Peak Hours by Gender
	22	Usage Peak Hours by Age Group
	23	Usage Peak Hours by Weekday
	24	Usage Peak Hours by Location and Weekday
	25	Usage Peak Hours by Weekday and Gender
	26	Usage Peak Hours by Weekday and Age Group

6.   SOURCES OF DATA

The data is collected from mobile apps that use INSAID Telecom services. Below is the schema that was provided and the SQL database connection details for mapping data:

```
host="cpanel.insaid.co",
user="student",
passwd="student",
db="Capstone1"
```

TABLE NAME	SL NO	COLUMNS	DESCRIPTION
gender_age_train Source: Capstone1 SQL Database	1	device_id	Unique identification number of the user's device
	2	Gender	Gender of the user
	3	Age	Age of the user
	4	Group	Age Group of the user
phone_brand_device_model Source: Capstone1 SQL Database	1	device_id	Unique Identification Number of the users device
	2	phone_brand	Phone brand of the device used by the user
	3	device_model	Model of the device used by the user.
events_data Source: Project DataDescripton Link	1	event_id	Unique identification number for every activity done by the user using INSAID Telecom services.
	2	device_id	Unique identification number of the user's device
	3	timestamp	Timestamp of the event
	4	longitude	Longitude of the user's location.
	5	latitude	Latitude of the user's location
	6	City	City where the event has occurred.
	7	State	State where the event has occurred.

7. SUMMARY OF DATA MINING

We analyzed the subset of the whole dataset that was provided and considered the data only for the state of Maharashtra. Below are the challenges that we faced with their respective solutions:

SL NO	CHALLENGES	SOLUTION PROVIDED
1	Missing state names in the main events dataset	1. The cities for which states where missing in the event dataset were listed. 2. Pune was the only city in the list which belongs to Maharashtra 3. Hence, for all the events with Pune city was filled with state values as 'Maharashtra'.
2	The Maharashtra events dataset had 63 Latitude and 63 Longitude values missing.	1. The missing latitude and longitude values were identified to be only for Pune city. 2. Hence, the missing values were replaced with latitude and longitude of Pune city.
3	The Maharashtra events dataset had 72 missing Device Ids.	1. The missing device ids were identified to be only for Pune city and belong to only 3 specific locations (lat and long) in Pune. 2. Hence, the device ids were replaced with the mode value of device ids specific to those 3 specific latitudes and longitudes in Pune.
4	Latitude and Longitude outliers were found in the events dataset.	1. On plotting the events on map, we found that, there were 3 locations outside India. 2. These 3 outliers were identified to belong to Pune city. 3. Hence, the latitude and longitude of 3 outliers were replaced with the latitude and longitude of Pune.
5	In PBDM dataset, we found that there were brand names and device model names in Chinese.	1. A list of English language names for the corresponding Chinese brand names in the dataset were identified and replaced. 2. RE library was used to find the list of Chinese device model names and the corresponding English names were listed and replaced.
6	Device ids were in int64 data type in GAT and PBDM dataset and float in events dataset. We were not able to create the dashboard.	1. While reading the events main file, the device ids were read as string. 2. Device ids in GAT and PBDM dataset were also converted to 'str' data type.
7	Each age group values in the events dataset were non uniform with either very short or wide range for analysis.	1. A new column age group was created with shorter and more uniform age range for insightful analysis.

8.   **PROPOSED SOLUTION FOR CUSTOMERS**

I. In view of Target Marketing:

Mumbai and Pune cover 89% of user activity and 68% user base which is far greater compared to other cities in Maharashtra. Hence there can be certain offers for retaining existing customers and also attract new customers from low activity age group, region etc. Following is proposed solution based on various criteria.

A) Based on device model and phone brand

i) Relatively latest models are more popular among the men in Mumbai compared to Pune. Hence in future, tie up with such device models can be suggested for Mumbai region during the model launch.

ii) Tie up with top 5 popular phone brands is expected to benefit the business.

B) Based on peak hour activity

i) To encourage the activity of users, the customized offers can be given during morning hours in Mumbai and during late night hours in Pune.

C) Based on number of users in Pune and Mumbai

i) Usage of services by Pune users is almost thrice than that of Mumbai, so more offers (unlimited free calls between INSAID Telecom users, data, validity, attractive rental plan for post paid etc) can be launched in Mumbai.

D) Based on usage by men and women

i) Women have less phone activity compared to men according to the given data. So to increase women customers we can have promotions, targeting in particular the highly active 18-37 age group.

E) Based on weekday activity

i) For Mumbai, give offers from Sunday to Wednesday (currently low performing)

ii) In Pune, give offers from Friday to Monday (currently low performing)

iii) All across Maharashtra, offers on Sunday and Monday can be given as in general we find least number of events for these days.

F) Based on age group activity

i) Below the age of 18, the activity is very low. We recommend targeting students. Tie up with educational apps may encourage parents to buy INSAID Telecom services with focus on data pack.

ii) Most of the user age is between 18-37 years. We can categorize users in 18-25 age group as student class, 25-37 age group as working class because of difference in need for customized offers. Users in the Student Class are mostly prepaid users who prefer using mobile data more than calls. The company can offer good amount of data either free or at very low cost compared to other service providers to attract more youngsters. Group calls offer can also complement above internet package offer. The working class is mostly postpaid users and we can provide monthly offer for mobile data.

F) Based on general analysis

i) Mumbai city alone has 45% Migrant workers (Working class). They generally like free roaming calls, easy portability option & hassle-free new SIM activation. If we can address this issue, we can dominate this set of people.

II. In view of business expansion

i) For expansion of business in other cities like Navi Mumbai, Thane, Nagpur, Nashik etc, a feedback of current users about any network issue should be taken. Call offers for small scale businessmen can be offered to attract new users in cities like Ahmednagar, Nashik, Aurangabad, Kolhapur. These cities also have majority of farmer population which are mostly prepaid customers. Considering their utility, small top-up recharge plans and data plans can be launched with new sim cards.

ii) The customer care complaints regarding network, call drops etc. can be analyzed to understand why there is very less customer base from cities like Nashik, Aurangabad, Nagpur etc. as poor network may be the main reason people refraining from using services of a particular service provider.

iii) Improvement in the performance of customer care can be suggested giving quick response and resolution to the customer issues.

iv) We suggest tying up with large organizations in Mumbai as more users are of working class.

v) Advertise on services & offers in cities like Nashik, Aurangabad, Nagpur for expanding business in these cities.

9. TOOLS

1. DS TOOLS
- Jupyter Notebook
- Pandas
- Seaborn
- Matplotlib
- Folium
- Plotly

2. WEB UI TOOLS
- Google Colab
- Google Drive

10. CONCLUSION

We can conclude that from the company perspective and the data that was analyzed for Maharashtra, **Target marketing** and **further expanding business** are the areas that needs to be focused based on the below considerations achieved from our analysis.

Considering user behavior based on gender, age group, popular brands, device models and peak hours of activity, Mumbai and Pune customer bases are stronger compared to other cities in Maharashtra, hence we can focus on increasing the customer base in other cities.

The emphasis may also be given to retain the existing users in **Mumbai and Pune**. We suggest launching different offers which would be beneficial in increasing the user activity during non-peak hours.

We have also observed that men outnumber women in terms of customer base. To increase the women customer base, offers pertaining to age group **18-37** can be targeted. Also, the other cities in Maharashtra with larger population can be viewed as more probable locations for business expansion. Thus, attractive new offers can be launched with new connections.

New product recommendations can be analyzed if we can get hold of the related data.