**Indian Mobile Users Demographics**

****

**Data Analysis Report**

****

**by INSAID Group - 1054**

**Document Information**

|  |  |
| --- | --- |
| **Document Name** | Indian Mobile Demographics – data analysis report |
| **Released Version** | 1.0 |
| **Released Date** | 16-03-2020 |

**Document Revision History**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Version** | **Author** | **Date** | **Changes incorporated** | **Reviewed By** | **Review Date** |
| 1.0 | Group -1054 | 16-03-2020 | Initial Document | Group – 1054 | 16-03-2020 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**Disclaimer:** This document contains proprietary confidential information of INSAID.

**Table of Contents**

[1 Introduction 4](#_Toc35966428)

[2 Project Description 5](#_Toc35966429)

[3 Problem Statement 6](#_Toc35966430)

[4 Problem Analysis & Data Processing 7](#_Toc35966431)

[5 Sources of Data 18](#_Toc35966432)

[6 Summary of Data Analysis 20](#_Toc35966433)

[6.1 Overall data analysis 20](#_Toc35966434)

[6.2 Data Analysis focused on 6 states 23](#_Toc35966435)

[6.3 Dashboard 26](#_Toc35966436)

[7 Observations & Conclusions 31](#_Toc35966437)

[8 Tools used 32](#_Toc35966438)

# Introduction

**INSAID Telecom** is 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.

We Group -1054, working on shared data to provide insight to INSAID Telecom.

# Project Description

In this consulting assignment, Group- 1054 are expected to build a dashboard to understand user's demographic characteristics based on their mobile usage, geolocation, and mobile device properties.

Aim of the project is to 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.

Aim to study the demographics of a user (gender and age) based on their app download and usage behaviours.

The Data is collected from mobile apps that use INSAID Telecom services.

# Problem Statement

To help the INSAID Telecom, we being approached to have depth of clarity in the underlying data.

For analysis we being asked to focus mainly on following 6 states

1. Madhya Pradesh
2. Chhattisgarh
3. Uttaranchal
4. Jammu and Kashmir
5. Goa
6. Nagaland

While conducting the analysis on complete dataset, following issues observed in data

1. 453 records found with device Ids blank
2. 423 records with Latitude and Longitude NULL values
3. 377 records found with States having NULL values
4. Many records found with wrong Latitude and Longitude values
5. Many devices details not available but event details are present
6. Detected Outliers in data

# Problem Analysis & Data Processing

We being asked to focus on 6 States for analysis but we decided to perform data Processing of problems found on complete data set.

While analysis we first checked the data for missing values and following result obtained

|  |
| --- |
| Total Percent  device\_id 453 0.000139  latitude 423 0.000130  longitude 423 0.000130  state 377 0.000116  city 0 0.000000  timestamp 0 0.000000  event\_id 0 0.000000 |

1. **Device ID NULL -**

453 records analysed based on other data values.

1. Listed the state and city for which Device Ids missing from record

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| State | City | event\_id | device\_id | timestamp | longitude | latitude |
| AndhraPradesh | **Visakhapatnam** | 69 | 0 | 69 | 69 | 69 |
| Delhi | **Delhi** | 69 | 0 | 69 | 69 | 69 |
| Gujarat | **Bardoli** | 16 | 0 | 16 | 16 | 16 |
| **Jetpur** | 16 | 0 | 16 | 16 | 16 |
| MadhyaPradesh | **Indore** | 51 | 0 | 51 | 51 | 51 |
| Maharashtra | **Pune** | 72 | 0 | 72 | 72 | 72 |
| Punjab | **Hoshiarpur** | 16 | 0 | 16 | 16 | 16 |
| Rajasthan | **Jaipur** | 81 | 0 | 81 | 81 | 81 |
| TamilNadu | **Chennai** | 63 | 0 | 63 | 63 | 63 |

1. Listed the records where similar Device IDs found with proper Latitude and Longitude with state and city
2. Only 3 devices found with proper data
3. So we decided to drop 453 records where device ID not found
4. **Latitude and Longitude NULL -**

423 records analysed based on device ID, city and state values

1. Listed the records with combination of device ID, city and state values having proper Latitude and Longitude values.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| State | City | event\_id | device\_id | timestamp | longitude | latitude |
| AndhraPradesh | **Visakhapatnam** | 63 | 63 | 63 | 0 | 0 |
| Bihar | **Araria** | 14 | 14 | 14 | 0 | 0 |
| **Bagaha** | 14 | 14 | 14 | 0 | 0 |
| Delhi | **Delhi** | 63 | 63 | 63 | 0 | 0 |
| MadhyaPradesh | **Indore** | 63 | 63 | 63 | 0 | 0 |
| Maharashtra | **Pune** | 63 | 63 | 63 | 0 | 0 |
| Punjab | **Moga** | 14 | 14 | 14 | 0 | 0 |
| Rajasthan | **Jaipur** | 66 | 66 | 66 | 0 | 0 |
| TamilNadu | **Chennai** | 63 | 63 | 63 | 0 | 0 |

1. Updated all records based on the records found where combination found matching.
2. Multiple device IDs found in cities where lat and longs are NULL

i. Visakhapatnam - multiple devices

ii. Araria - one Device

iii. Bagaha - one device

iv. Delhi - multiple devices

e. Indore - multiple devices

f. Pune - multiple devices

g. Moga - one device

h. Jaipur - Multiple devices

i. Chennai - Multiple Devices

Updated all devices with correct Latitude and Longitude data

1. **State value NULL**
2. 377 states found records with state value NULL. Listed the data with city wise

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| City | event\_id | device\_id | timestamp | longitude | latitude | state |
| Arambagh | 16 | 16 | 16 | 16 | 16 | 0 |
| Channapatna | 14 | 14 | 14 | 14 | 14 | 0 |
| Chennai | 47 | 47 | 47 | 47 | 47 | 0 |
| Delhi | 64 | 64 | 64 | 64 | 64 | 0 |
| Gangarampur | 26 | 26 | 26 | 26 | 26 | 0 |
| Indore | 51 | 51 | 51 | 51 | 51 | 0 |
| Jaipur | 49 | 49 | 49 | 49 | 49 | 0 |
| Pune | 63 | 63 | 63 | 63 | 63 | 0 |
| Visakhapatnam | 47 | 47 | 47 | 47 | 47 | 0 |

1. For these 377 records, city and other data found present
2. Based on these combination state value updated
3. On the basis of city we can update states as follows

1. Arambaugh  - WestBengal

2. Channapatna - Karnataka

3. Chennai - Tamilnadu

4. Delhi - Delhi

5. Gangarampur - WestBengal

6. Indore - Madhyapradesh

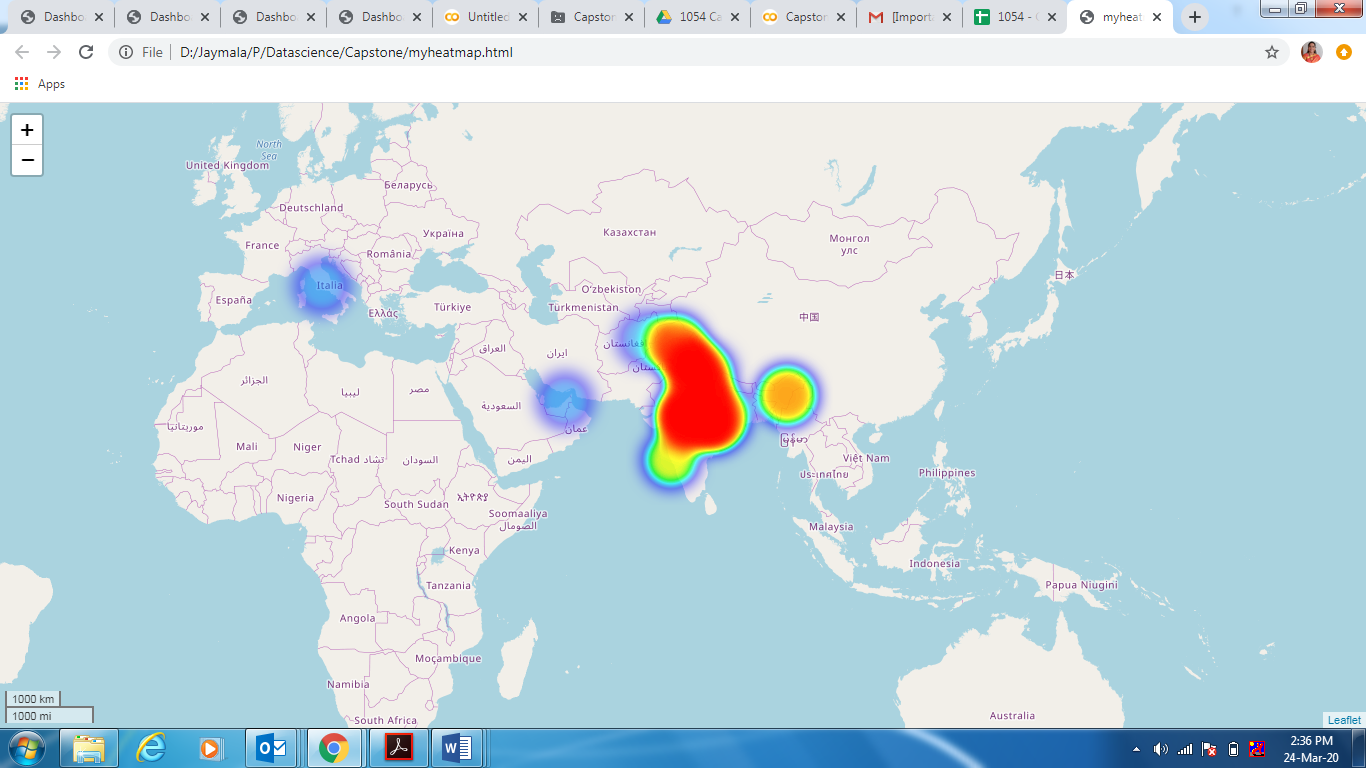
7. Jaipur - Rajasthan

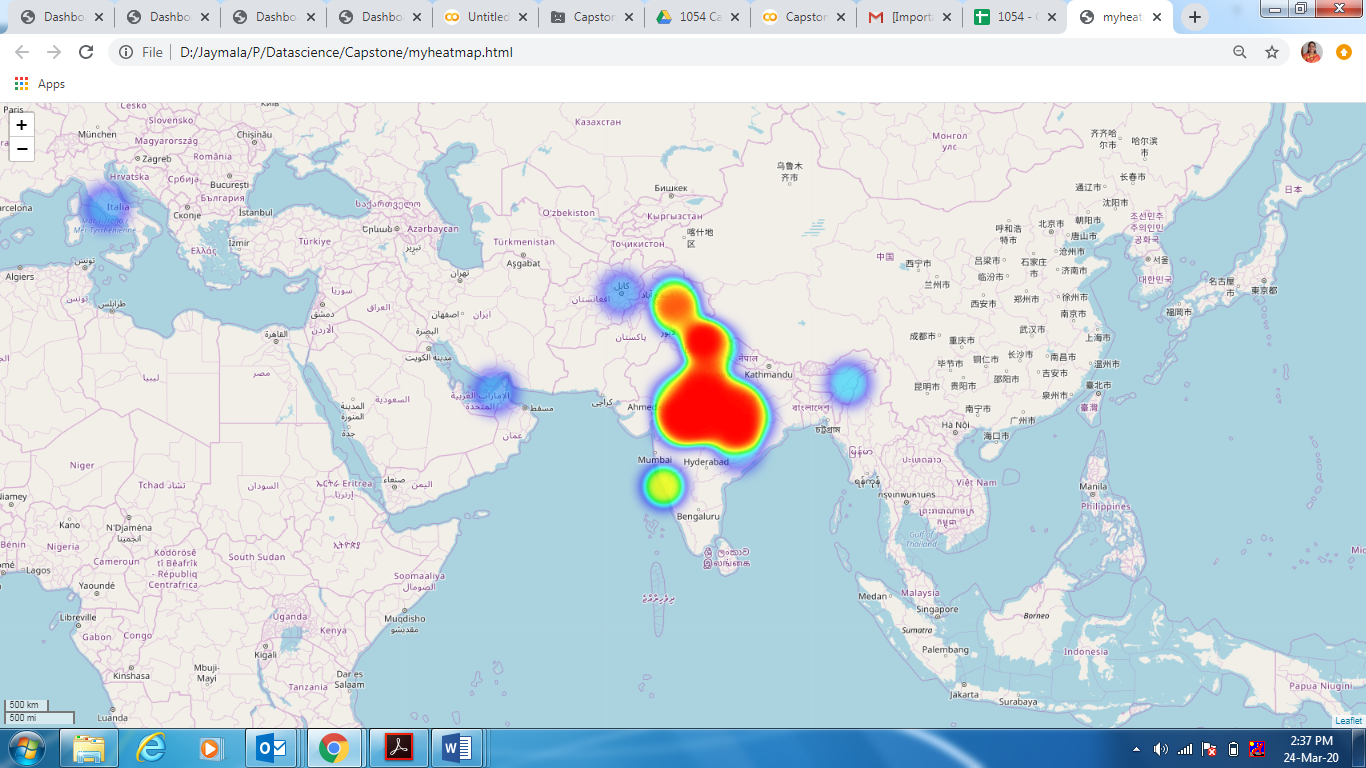
8. Pune - Maharashtra

9. Visakhapatnam - Andhrapradesh

1. **Wrong Latitude and Longitude**

While analysing data we plotted the values on Map and observed that few entries plotted out of India.





* 1. We found that there are wrong entries of Longitude. Two wrong entries of Longitude

| **12.567000** | **12.567400** |
| --- | --- |

* 1. Listed the data for above 2 wrong values

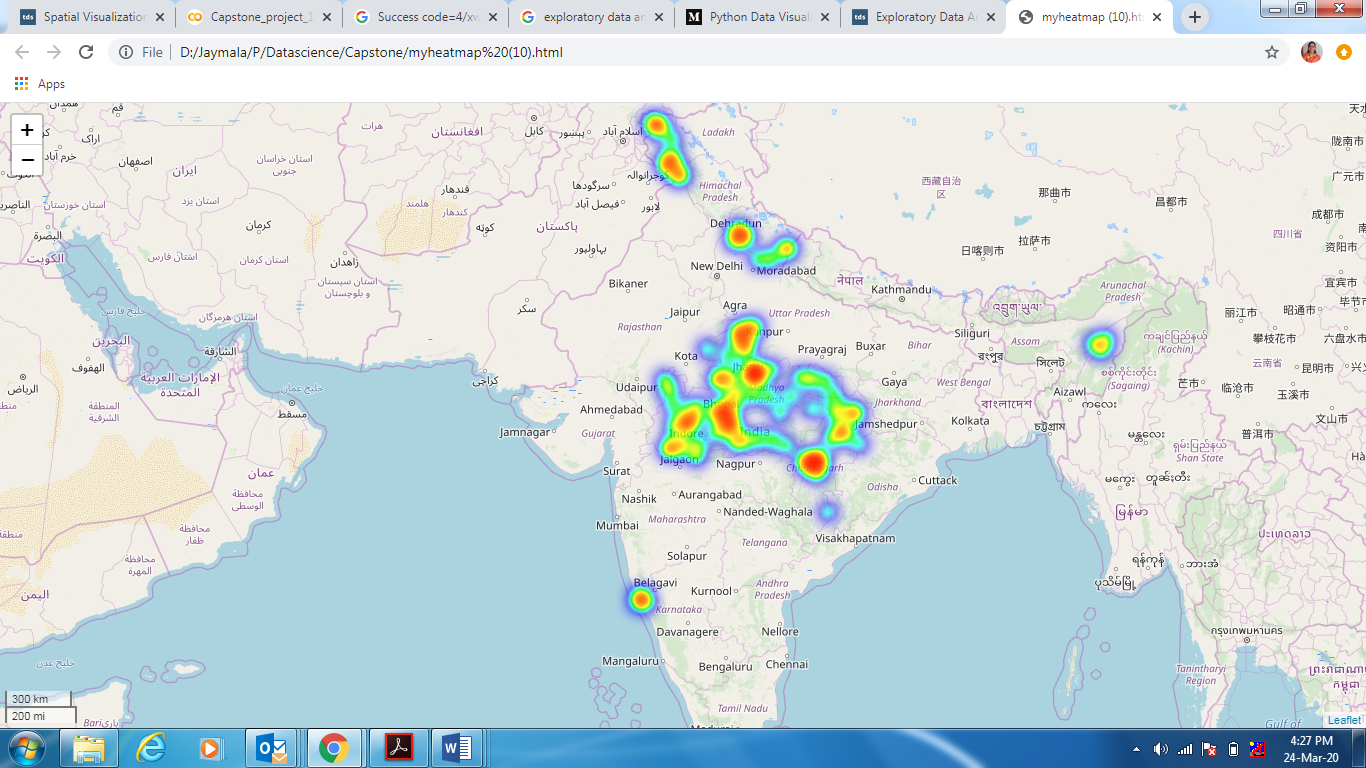
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Event ID | Device ID | TimeStamp | Longitude | Latitude | City | State |
| 282400.000000 | -1448078833416777984 | 2016-05-05 17:33:57 | 12.567400 | 41.871900 | Delhi | Delhi |
| 419288.000000 | -1448078833416777984 | 2016-05-04 09:06:44 | 12.567400 | 41.871900 | Delhi | Delhi |
| 452839.000000 | -1448078833416777984 | 2016-05-01 16:44:00 | 12.567400 | 41.871900 | Delhi | Delhi |
| 311979.000000 | -3300282771940051456 | 2016-05-04 18:12:47 | 12.567000 | 41.871900 | Pune | Maharashtra |
| 451415.000000 | -3300282771940051456 | 2016-05-06 14:37:28 | 12.567000 | 41.871900 | Pune | Maharashtra |
| 483087.000000 | -3300282771940051456 | 2016-05-01 10:57:32 | 12.567000 | 41.871900 | Pune | Maharashtra |
| 351833.000000 | -170254198178616704 | 2016-05-04 16:26:38 | 12.567400 | 41.871900 | Chennai | TamilNadu |
| 563738.000000 | -170254198178616704 | 2016-05-02 01:05:02 | 12.567400 | 41.871900 | Chennai | TamilNadu |
| 594503.000000 | -170254198178616704 | 2016-05-06 06:17:27 | 12.567400 | 41.871900 | Chennai | TamilNadu |
| 226298.000000 | 9078282559555458048 | 2016-05-05 15:21:49 | 12.567400 | 41.871900 | Visakhapatnam | AndhraPradesh |
| 2976723.000000 | 9078282559555458048 | 2016-05-06 17:00:08 | 12.567400 | 41.871900 | Visakhapatnam | AndhraPradesh |
| 3020905.000000 | 9078282559555458048 | 2016-05-05 11:57:30 | 12.567400 | 41.871900 | Visakhapatnam | AndhraPradesh |
| 387968.000000 | -2796107298017636352 | 2016-05-07 17:53:20 | 12.567400 | 41.871900 | Indore | MadhyaPradesh |
| 571466.000000 | -2796107298017636352 | 2016-05-05 10:07:28 | 12.567400 | 41.871900 | Indore | MadhyaPradesh |
| 596824.000000 | -2796107298017636352 | 2016-05-06 09:44:36 | 12.567400 | 41.871900 | Indore | MadhyaPradesh |
| 170433.000000 | 4718752803626400768 | 2016-05-07 01:39:40 | 12.567400 | 41.871900 | Jaipur | Rajasthan |
| 3188661.000000 | 4718752803626400768 | 2016-05-07 16:08:20 | 12.567400 | 41.871900 | Jaipur | Rajasthan |
| 3228359.000000 | 4718752803626400768 | 2016-05-02 11:50:26 | 12.567400 | 41.871900 | Jaipur | Rajasthan |
| 286380.000000 | 6171733141545328640 | 2016-05-02 21:12:53 | 12.567400 | 41.871900 | Purnia | Bihar |
| 515181.000000 | 6171733141545328640 | 2016-05-02 21:23:51 | 12.567400 | 41.871900 | Purnia | Bihar |
| 560457.000000 | 6171733141545328640 | 2016-05-04 22:35:39 | 12.567400 | 41.871900 | Purnia | Bihar |

* 1. Updated above data with correct Latitude and Longitude by finding correct values with device ID, state & City
  2. Also we observed in data that same combination of Latitude and Longitude applied for multiple cities and state with different Device IDs

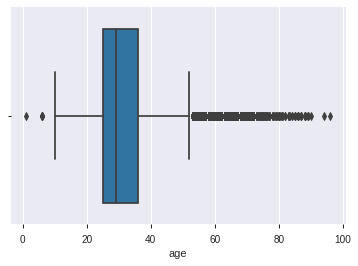
|  |  |  |  |
| --- | --- | --- | --- |
| latitude | longitude | state | city |
| 8.190110 | **77.511124** | TamilNaduTamilNaduTamilNadu | NagercoilNagercoilNagercoil |
| 8.193192 | **77.466580** | TamilNaduTamilNaduTamilNaduTamilNaduTamilNaduT... | NagercoilNagercoilNagercoilNagercoilNagercoilN... |
| 8.199913 | **77.441830** | TamilNaduTamilNaduTamilNaduTamilNaduTamilNadu | NagercoilNagercoilNagercoilNagercoilNagercoil |
| 8.200381 | **77.521803** | TamilNadu | Nagercoil |
| 8.200689 | **77.487619** | TamilNaduTamilNadu | NagercoilNagercoil |
| 34.371760 | **74.491055** | JammuandKashmirJammuandKashmirJammuandKashmirJ... | SopurSopurSopurSopurSopurSopurSopurSopurSopurS... |
| 34.379609 | **74.525620** | JammuandKashmirJammuandKashmirJammuandKashmirJ... | SopurSopurSopurSopurSopurSopur |
| 34.381385 | **74.512342** | JammuandKashmirJammuandKashmirJammuandKashmirJ... | SopurSopurSopurSopurSopurSopurSopurSopurSopurS... |
| 34.391257 | **74.563104** | JammuandKashmirJammuandKashmirJammuandKashmirJ... | SopurSopurSopurSopurSopur |
| 34.555300 | **69.207500** | DelhiDelhiDelhiMaharashtraMaharashtraMaharasht... | DelhiDelhiDelhiPunePunePuneChennaiChennaiChenn... |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| event\_id | device\_id | timestamp | longitude | latitude | city | state |
| 251785.000000 | 1057289835566390784 | 2016-05-07 14:23:56 | 69.207500 | 34.555300 | Delhi | Delhi |
| 378539.000000 | 1057289835566390784 | 2016-05-01 17:30:29 | 69.207500 | 34.555300 | Delhi | Delhi |
| 409004.000000 | 1057289835566390784 | 2016-05-06 12:28:27 | 69.207500 | 34.555300 | Delhi | Delhi |
| 360331.000000 | 4995976326034236416 | 2016-05-07 21:25:54 | 69.207500 | 34.555300 | Pune | Maharashtra |
| 526715.000000 | 4995976326034236416 | 2016-05-01 20:58:46 | 69.207500 | 34.555300 | Pune | Maharashtra |
| 560421.000000 | 4995976326034236416 | 2016-05-04 22:45:37 | 69.207500 | 34.555300 | Pune | Maharashtra |
| 2896853.000000 | 7758074839281444864 | 2016-05-02 05:48:31 | 69.207500 | 34.555300 | Chennai | TamilNadu |
| 3175643.000000 | 7758074839281444864 | 2016-05-01 05:13:10 | 69.207500 | 34.555300 | Chennai | TamilNadu |
| 3203806.000000 | 7758074839281444864 | 2016-05-01 21:16:33 | 69.207500 | 34.555300 | Chennai | TamilNadu |
| 172407.000000 | -601883305299256832 | 2016-05-03 12:37:36 | 69.207500 | 34.555300 | Visakhapatnam | AndhraPradesh |
| 2949383.000000 | -601883305299256832 | 2016-05-06 18:29:57 | 69.207500 | 34.555300 | Visakhapatnam | AndhraPradesh |
| 2997772.000000 | -601883305299256832 | 2016-05-04 14:45:21 | 69.207500 | 34.555300 | Visakhapatnam | AndhraPradesh |
| 358230.000000 | 8997995259821809664 | 2016-05-07 20:16:27 | 69.207500 | 34.555300 | Indore | MadhyaPradesh |
| 561149.000000 | 8997995259821809664 | 2016-05-01 23:29:12 | 69.207500 | 34.555300 | Indore | MadhyaPradesh |
| 604811.000000 | 8997995259821809664 | 2016-05-03 15:22:47 | 69.207500 | 34.555300 | Indore | MadhyaPradesh |
| 252730.000000 | -3693510914308259840 | 2016-05-04 15:13:00 | 69.207500 | 34.555300 | Jaipur | Rajasthan |
| 3195812.000000 | -3693510914308259840 | 2016-05-07 12:15:05 | 69.207500 | 34.555300 | Jaipur | Rajasthan |
| 3223664.000000 | -3693510914308259840 | 2016-05-01 07:59:07 | 69.207500 | 34.555300 | Jaipur | Rajasthan |
| 263848.000000 | -7165635156926407680 | 2016-05-07 23:26:20 | 69.207500 | 34.555300 | Masaurhi | Bihar |
| 398252.000000 | -7165635156926407680 | 2016-05-06 00:17:53 | 69.207500 | 34.555300 | Masaurhi | Bihar |
| 414695.000000 | -7165635156926407680 | 2016-05-03 05:24:01 | 69.207500 | 34.555300 | Masaurhi | Bihar |

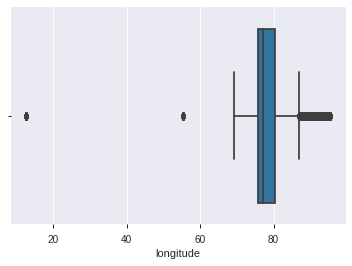
* After Updation of wrong Longitude and Latitudes



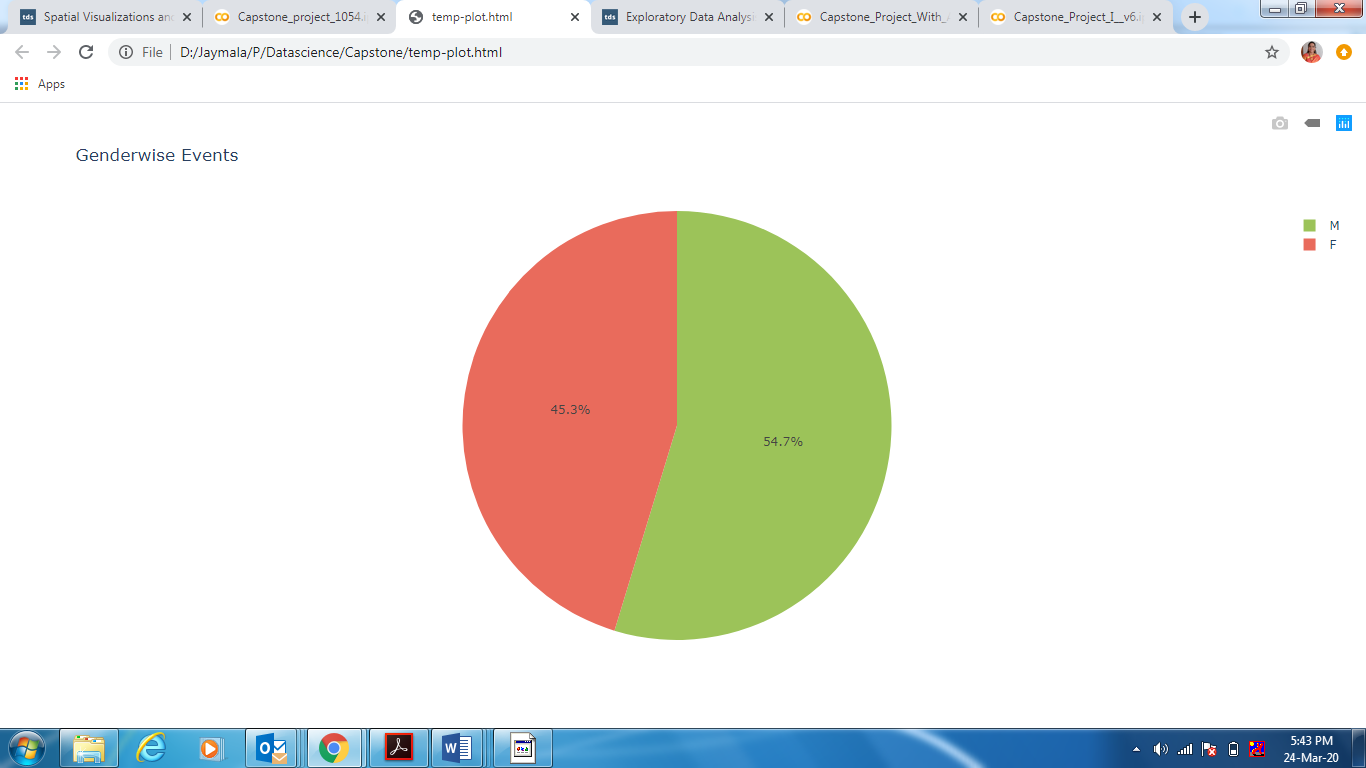
1. **Outliers in Data**
   1. **Age :**

****

* 1. **Longitude :**



1. **Gender wise**



Male - 54.7 %

Female – 45.3%

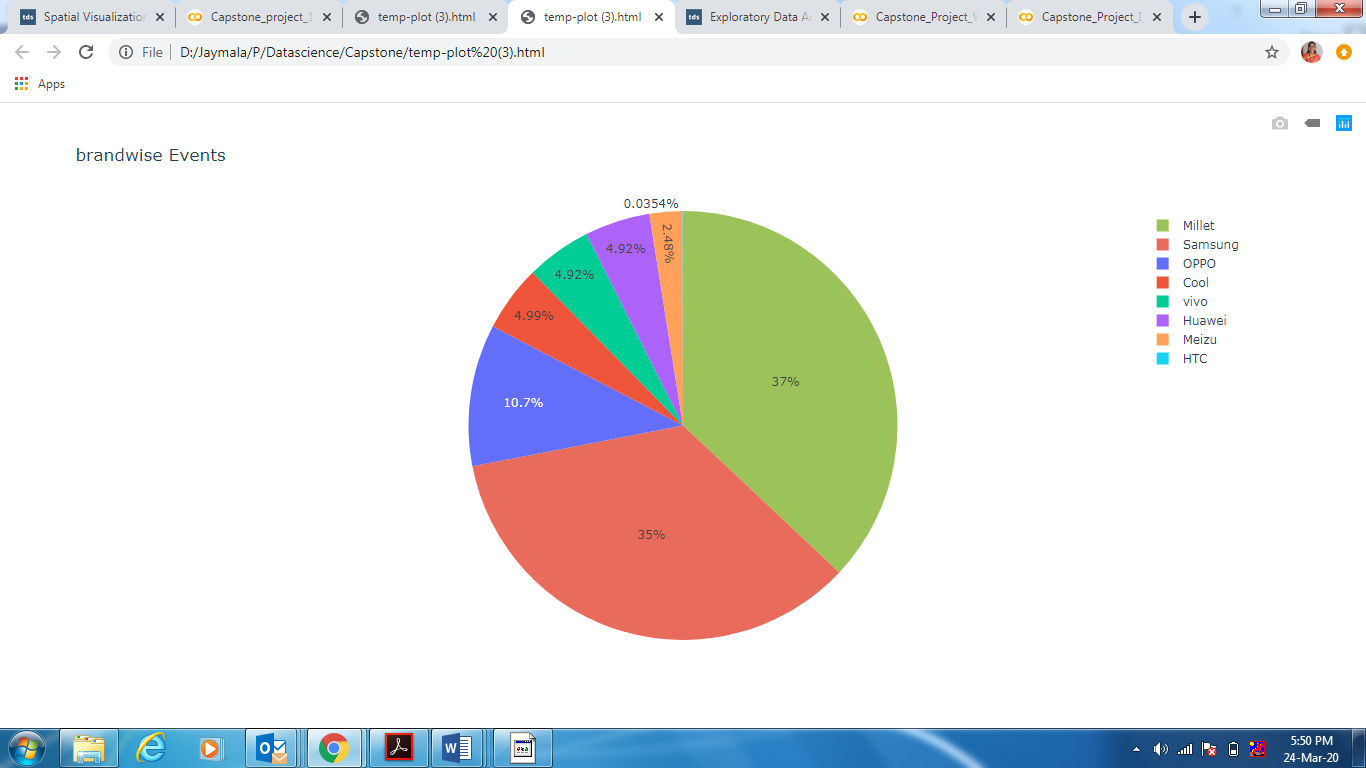
1. **Brand found with language other than English**
   1. Listed the Unique brands in data. 116 Unique Brands found in data

**Unique Brands :**

|  |
| --- |
| distinct brand: {'大Q', '摩乐', '飞利浦', '宝捷讯', 'OPPO', '梦米', 'E人E本', '神舟', '本为', '丰米', '酷派', '海信', '米奇', '努比亚', '小米', '白米', '贝尔丰', '虾米', '联想 ', '小杨树', '糯米', '普耐尔', 'vivo', '酷比魔方', '昂达', '优购', '优语', 'LOGO', '赛博宇华', '广信', '朵唯', '海尔', '大显', '青葱', '易派', '长虹', '谷歌', 'E派', '瑞高', 'MIL', '青橙', '尼比鲁', '欧沃', '华硕', '糖葫芦', '飞秒', '瑞米', '天语', 'SUGAR', '斐讯', '大可乐', '米歌', '聆韵', '邦华', '鲜米', '德赛', 'PPTV', '亿通', '欧新', '酷珀', '诺亚信', '乐视', '恒宇丰', '魅族', '欧博信', '维图', 'Lovme', 'LG', '至尊宝', '惠普', '奇酷', 'ZUK', '百加', '果米', '诺基亚', '爱派尔', '凯利通', '帷幄', '欧奇', 'HTC', '艾优尼', '语信', '欧比', '夏新', '三星', '纽曼', '西米', '锤子', '基伍', '西门子', '一加', '华为', '摩托罗拉', '先锋', '台电', '智镁', '优米', '中国移动', '美图', '唯比', '富可视', '酷比', '百立丰', '沃普丰', '波导', '蓝魔', '黑米', '金星数码', '欧乐迪', '康佳', '乡米', '世纪星', '奥克斯', 'TCL', '唯米', '首云'} |

Used GoogleTrans to convert the data to English

1. Brandwise Events



# Sources of Data

* 1. **gender\_age\_train** - Devices and their respective user gender, age and age\_group

|  |
| --- |
| RangeIndex: 74645 entries, 0 to 74644  Data columns (total 4 columns):  device\_id 74645 non-null int64  gender 74645 non-null object  age 74645 non-null int64  age\_group 74645 non-null object  dtypes: int64(2), object(2) |

Total Records = 74,645

No Null Records Found

* 1. **phone\_brand\_device\_model** - device ids, brand, and models phone\_brand

|  |
| --- |
| RangeIndex: 87726 entries, 0 to 87725  Data columns (total 3 columns):  device\_id 87726 non-null int64  brand 87726 non-null object  model 87726 non-null object  dtypes: int64(1), object(2) |

Total Records = 87, 726

No Null Records Found

* 1. **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.

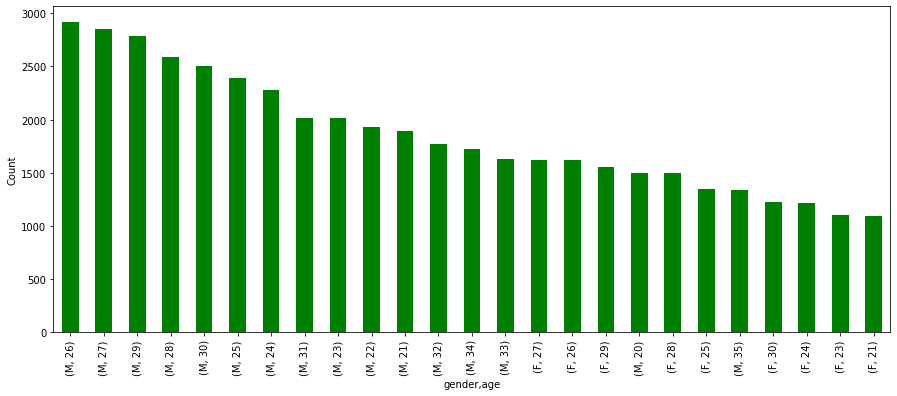
|  |
| --- |
| 1. RangeIndex: 3252950 entries, 0 to 3252949 2. Data columns (total 7 columns): 3. event\_id float64 4. device\_id float64 5. timestamp object 6. longitude float64 7. latitude float64 8. city object 9. state object 10. dtypes: float64(4), object(3) |

Total Records = 32,52,950

# Summary of Data Analysis

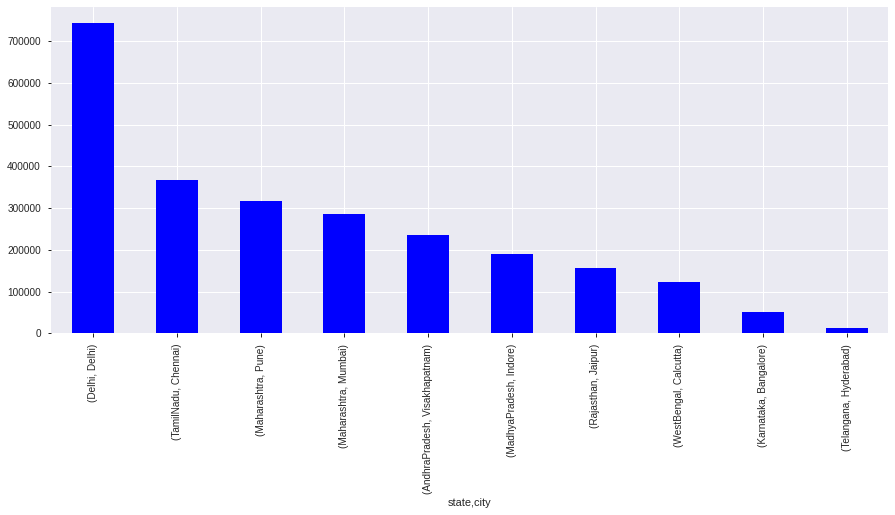
## Overall data analysis

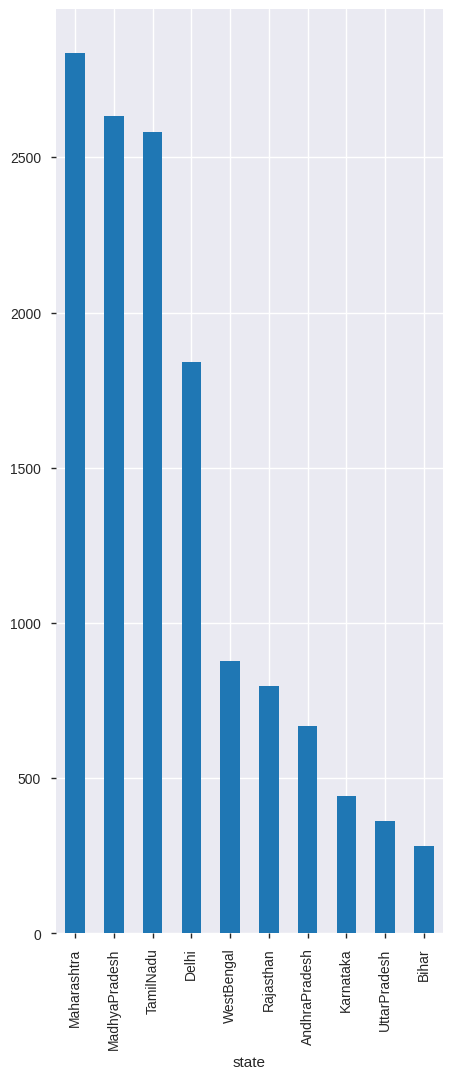
1. Gender and Age wise Count of Events Occurrence



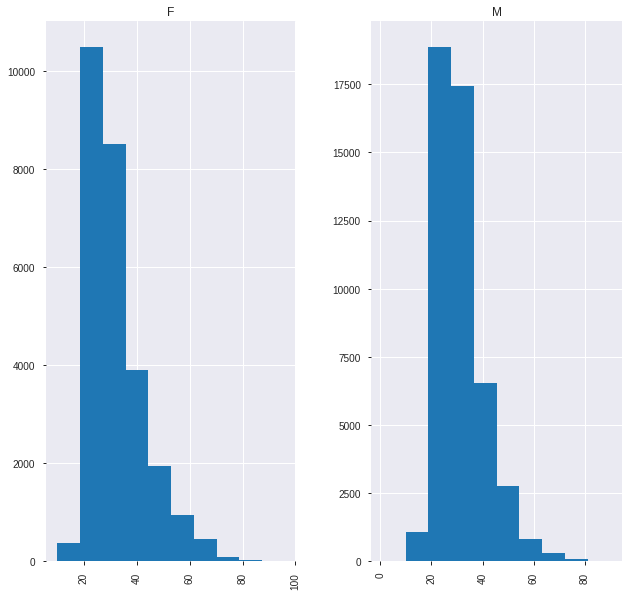
Top 5 users are of Male category and in the age group 26 to 30 years

1. State wise highest Events occurrence

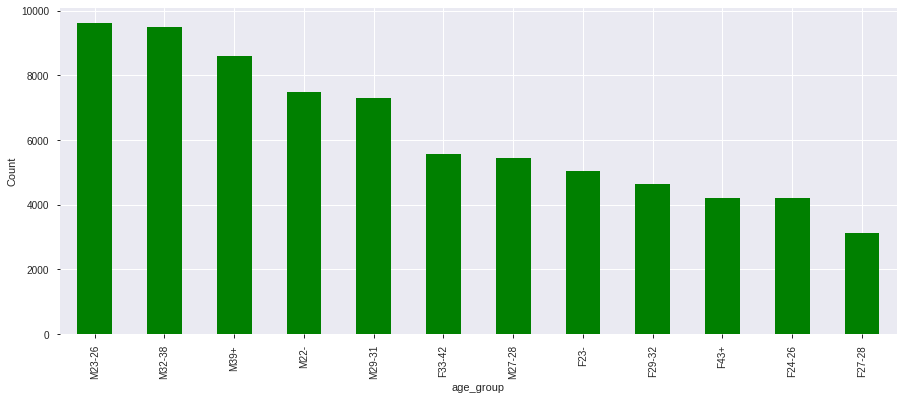




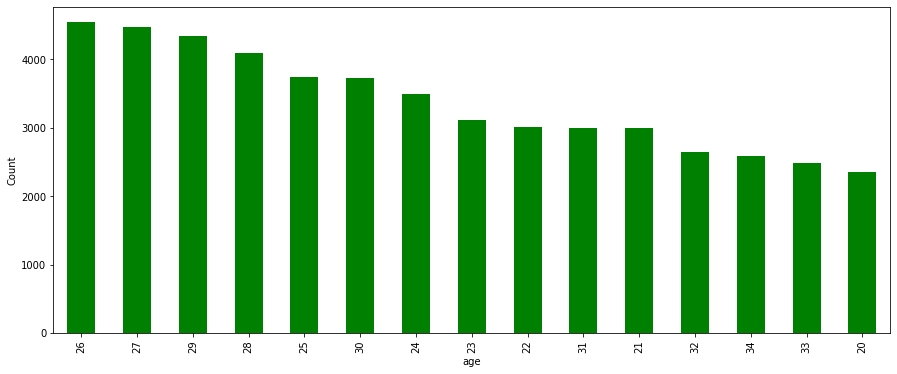
1. Gender and Age wise Analysis



1. Age Group wise event count

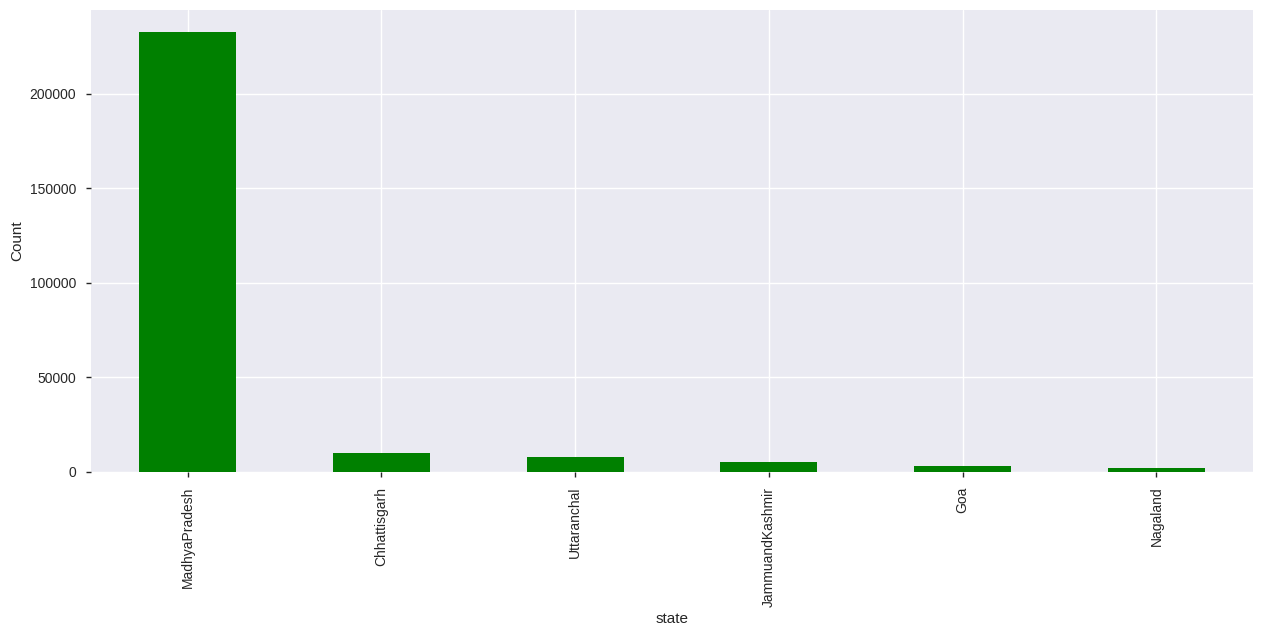


1. Age wise event Count

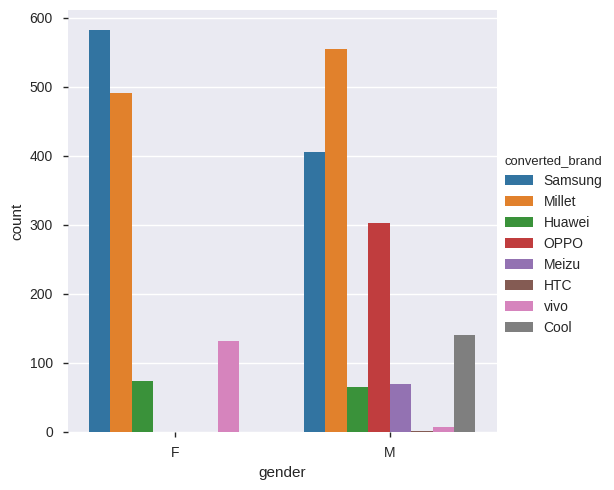


## Data Analysis focused on 6 states

1. State wise highest Events occurrence



1. Gender and brand wise

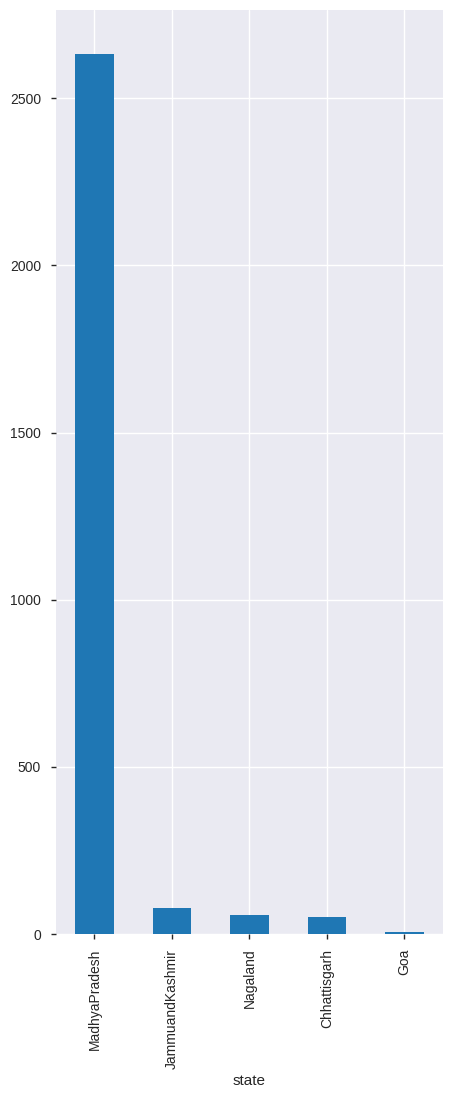


Devices details recorded but their Events not found or vice versa

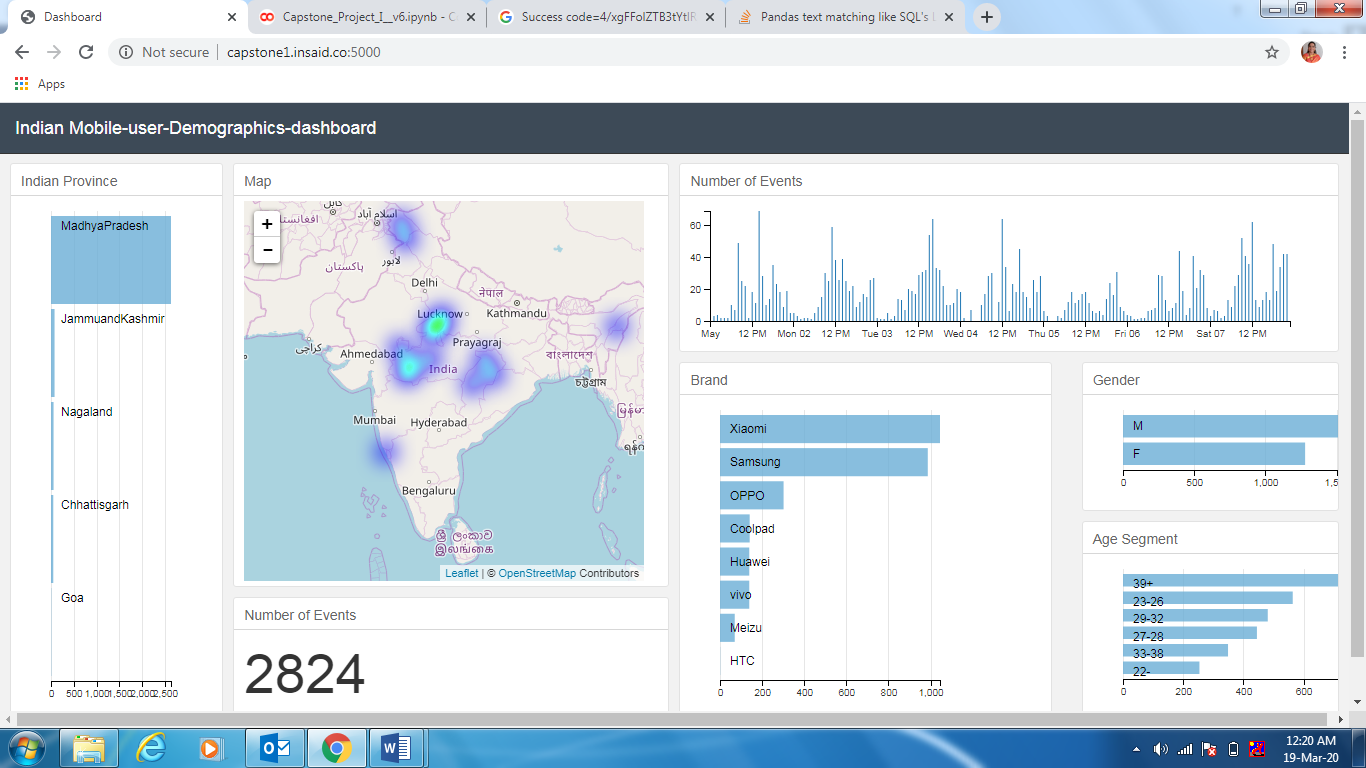
|  |
| --- |
| device\_id 77434 non-null int64  gender 77434 non-null object  age 77434 non-null int64  age\_group 77434 non-null object  event\_id 2824 non-null float64  timestamp 2824 non-null object  longitude 2824 non-null float64  latitude 2824 non-null float64  city 2824 non-null object  state 2824 non-null object |

Based on Matched records state wise as follows. Total records **2824. Only 5 state records found matched with other details. Uttaranchal doesn’t have single record matching**

|  |
| --- |
| MadhyaPradesh 2633  JammuandKashmir 78  Nagaland 56  Chhattisgarh 50  Goa 7 |



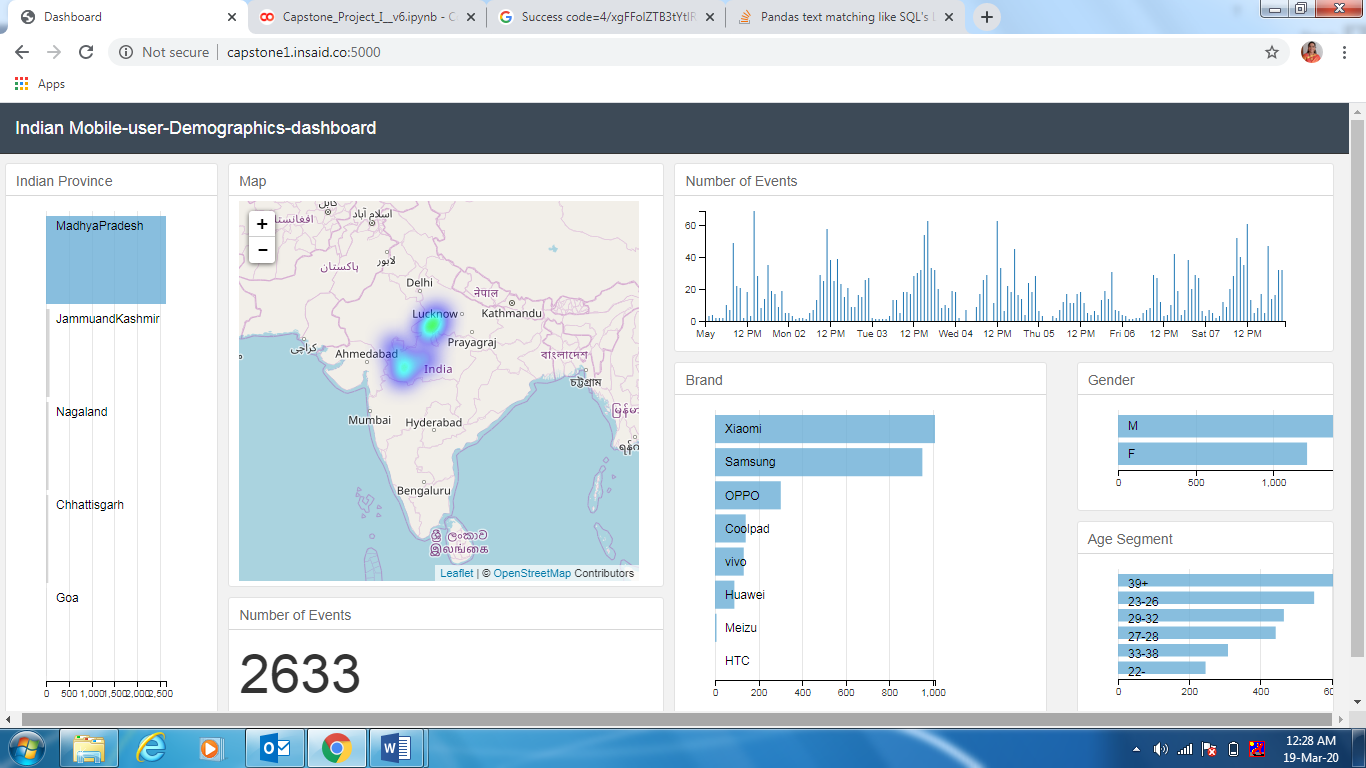
## Dashboard



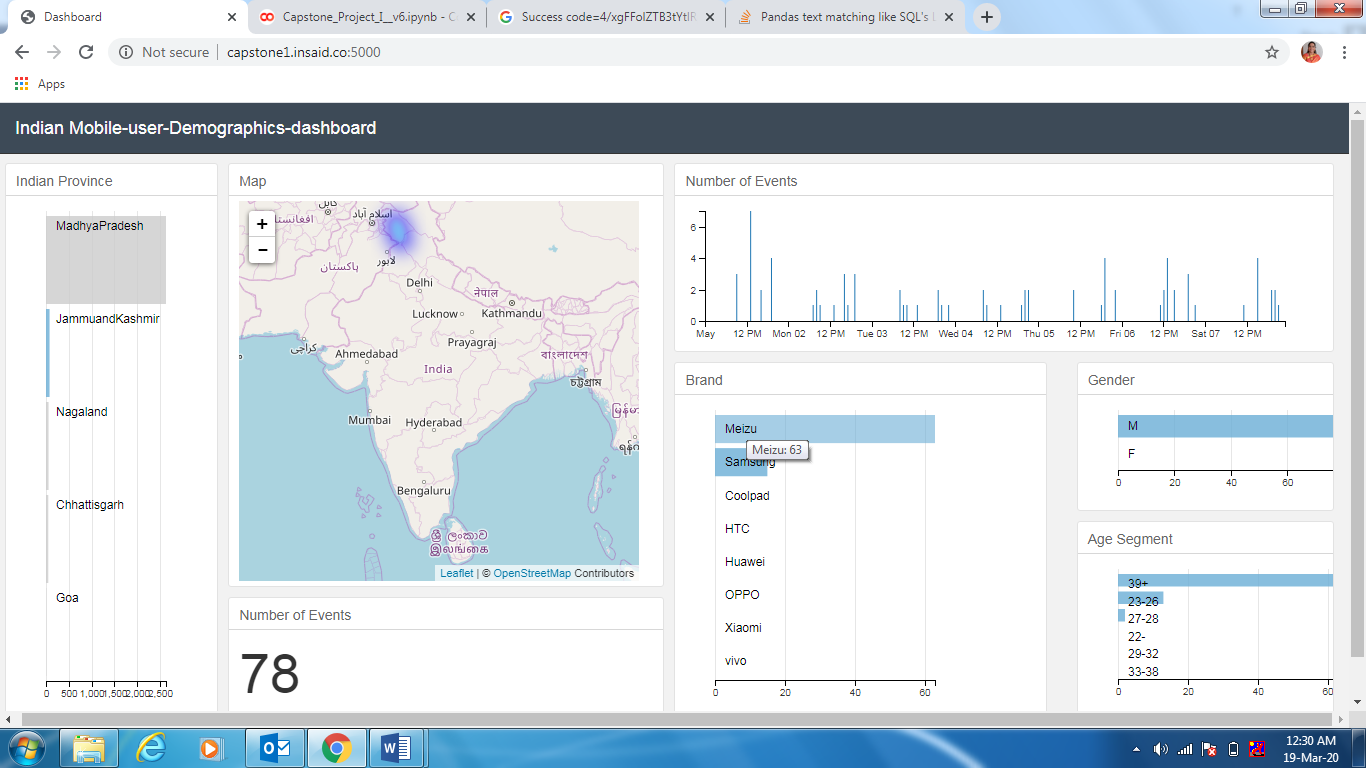
1. Madhapradesh has maximum records of events
2. Maximum users are of age more than 39 years
3. Top 2 brands of mobiles used are Xiomi and Samsung
4. Male users are more than Female Users

Now let’s focus on Individual state.

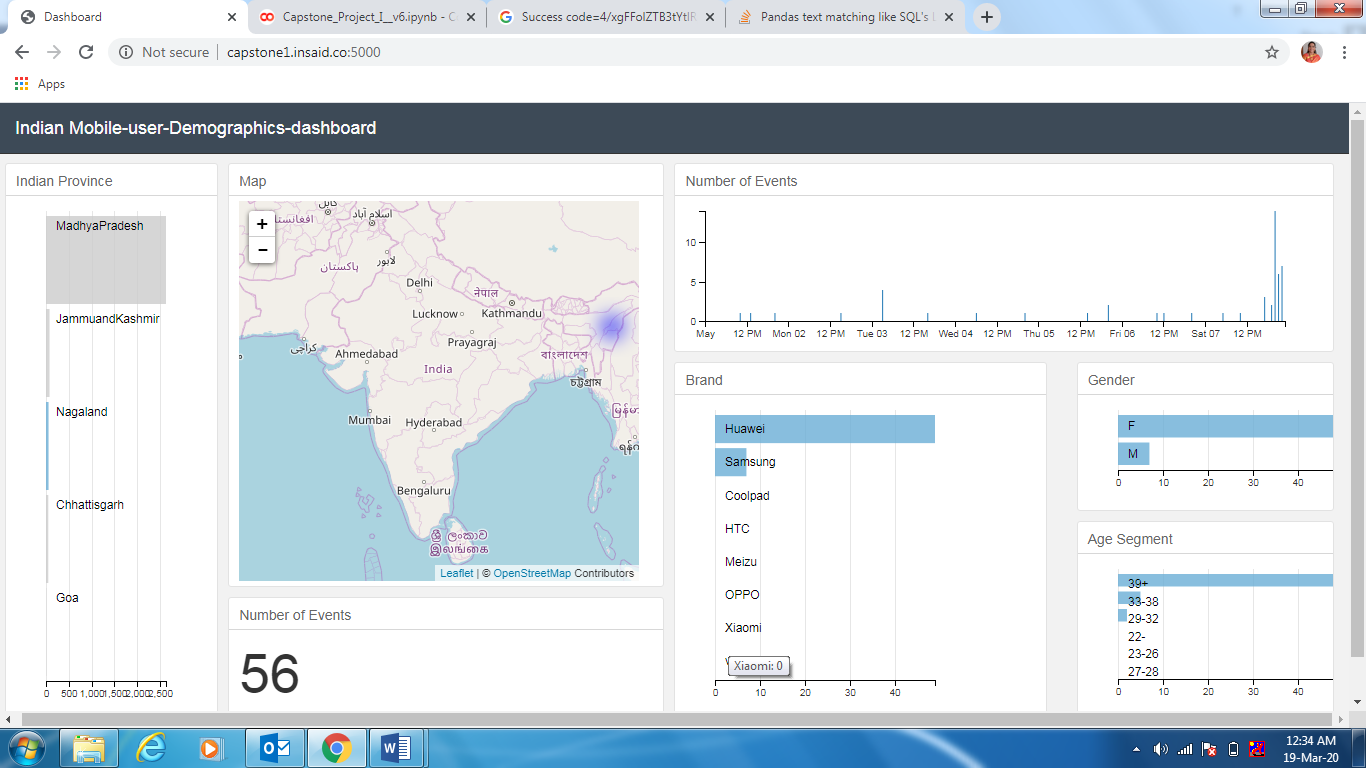
1. Madhyapradesh - Overall result is driven by Madhyapradesh data



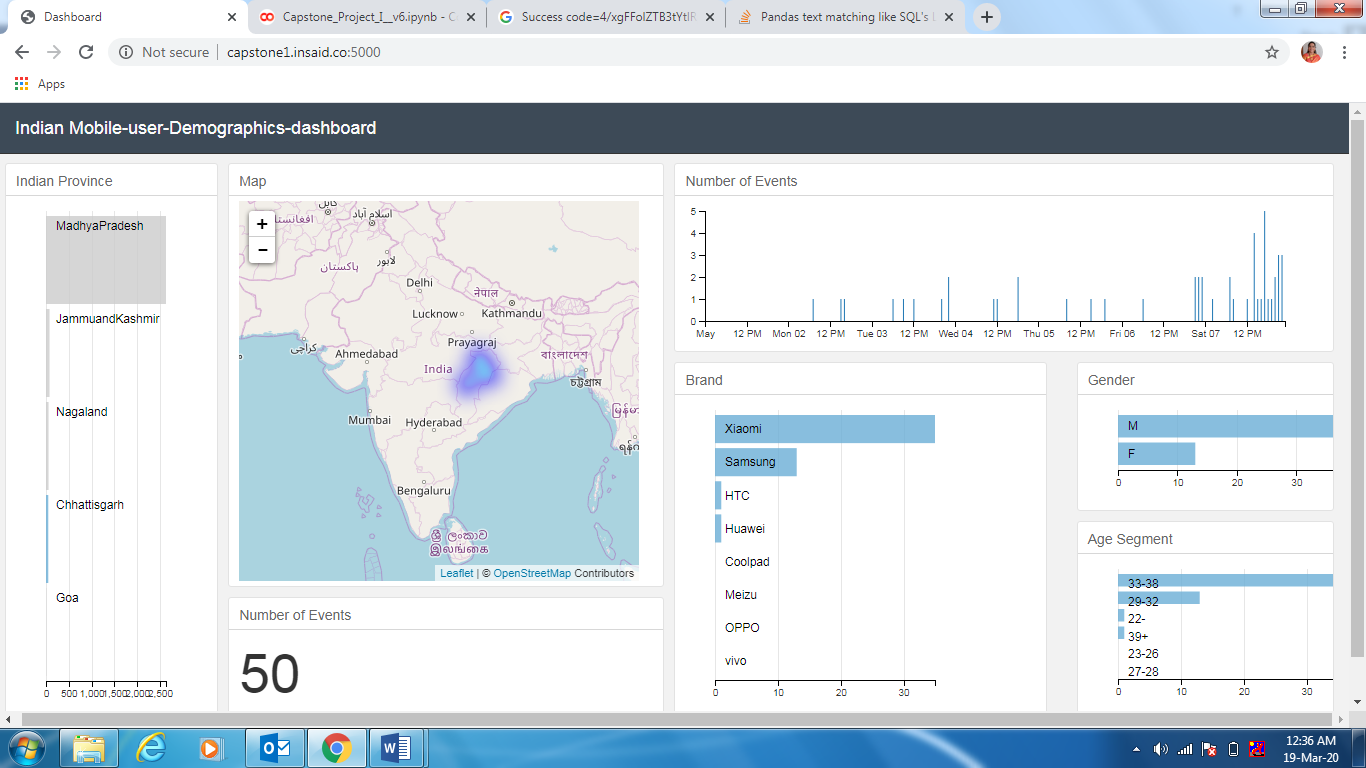
1. Total 2633 records of events
2. Male users are more than Female users
3. Majority users are in age group of 39+ years
4. Top Phone brand used is Xiomi
5. Jammu and Kashmir



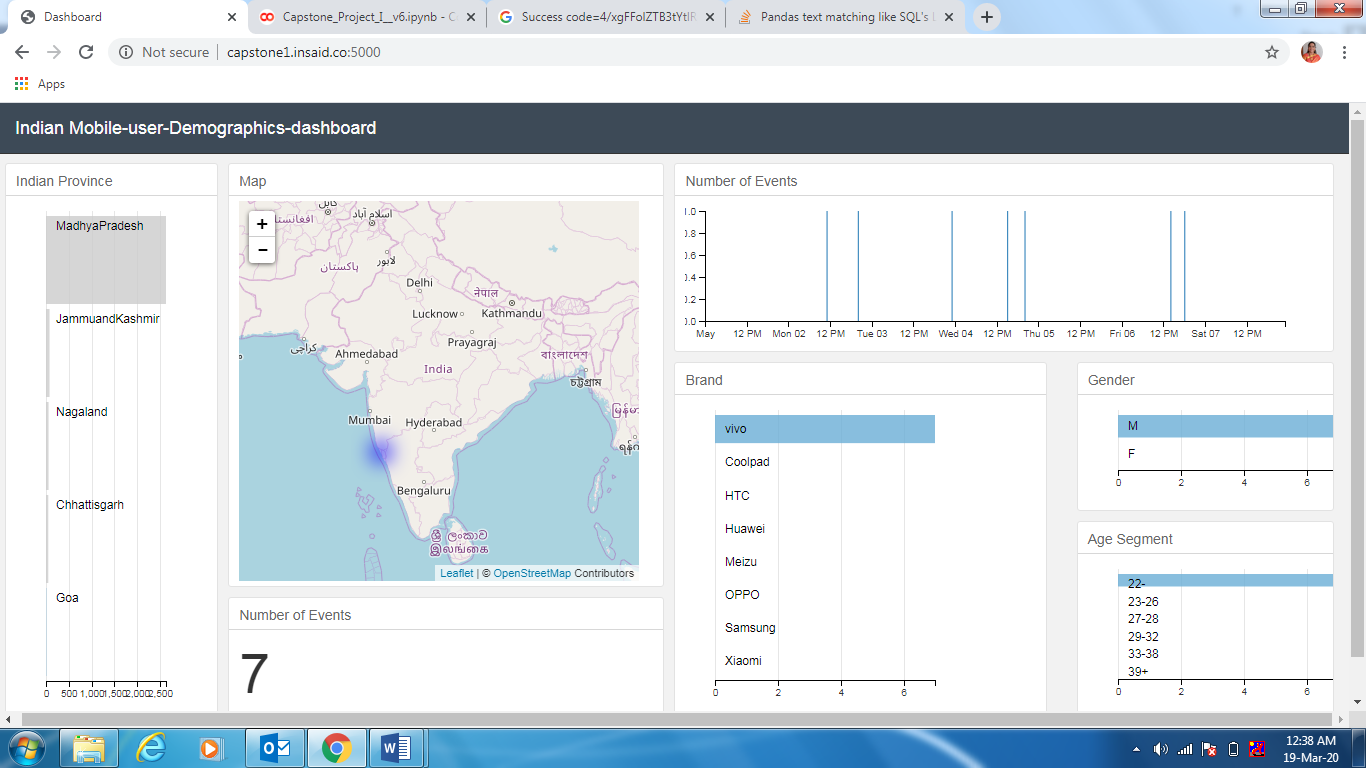
1. Total 78 records of events
2. All are male users
3. Majority users are in age group of 39+ years
4. Top Phone brand used is Meizu
5. Nagalad



1. Total 56 records of events
2. All are female users are more than male
3. Majority users are in age group of 39+ years
4. Top Phone brand used is Huawai
5. Chattisgarh :



1. Total 50 records of events
2. Male users are more than Female users
3. Majority users are in age group of 33 to 38 years ( less than 39 years)
4. Top Phone brand used is Xiomi
5. Goa



1. Total 7 records of events
2. All are male users
3. Majority users are having age less than 22 years
4. Top Phone brand used is Vivo

# Observations & Conclusions

* Device, age and Gender details need to capture with re-survey. As in majority cases this details are missing even though events details are available
* It seems SAMSUNG is not focusing on smaller States and Cities. Some Smaller brands like Meizue and Huwai are popular in states like Jammu-Kashmir and Nagaland.
* Considering events are more in the afternoon and higher in female users, Operators should focus varied data plans based on time slots
* For the digital content providers; they should focus on gender based content
* Advertisement should focus female product ads for afternoon zones and they can also increase the CMP rates
* Users with age more than 39+ using Xiomi brand more. Need to check the various factors why particular age group using same brand. Maybe Screen size, Screen resolution, presence of native vernacular application may be impacting the brand preference.
* Xiomi used more by male users than female users
* Some of the Latitude-Longitude were from outside India, which proves that some events are getting generated via proxy servers located outside India

# Tools used

|  |  |  |
| --- | --- | --- |
| Sr. No | Tools | Purpose |
| 1. | Colab | To write / execute python code |
| 2. | Jupiter | To write / execute python code |
| 3. | Python, Pandas, Numpy | For scripting to perform analysis on data |
| 5. | GoogleTrans | For Conversion from other language to English |
| 6. | MySql Connector | To fetch data from MySQL |
| 7. | Folium | To plot Latitude and Longitude on plot |
| 8. | Matplotlib | For plotting data for analysis |
| 9. | Seaborn | For plotting data for analysis |