**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

* Python Jupyter Notebook
* Microsoft Word & Excel

1. Conclusion
2. **Introduction:**

India's telecommunication network is the second largest in the world by number of telephone users (both fixed and mobile phone). The Total subscriber base, wireless subscription as well as wired broadband subscription have grown consistently Tele-density stood at **84.56%** as of December 2022. Total broadband subscriptions grew to 832.2 million until Dec'2022 and the total subscriber base stood at 1170.38 million in Dec'2022.

India has the **world's second-largest Internet user-base** with 747.41 million broadband internet subscribers in the country. Highest data usage per smartphone is at an **average of 9.8 GB per Month**.

With 70% of population staying in rural areas & with rural subscribers forming **43.69%** of the total telephone subscribers of Jan 2020. India's rural market will be the key to growth in the coming years. Government of India unreeled the **National Broadland Mission** with an aim to provide Broadland access to all villages by 2022.

Gross Revenue of the telecom sector stood at Rs. **76.408 Crore** (US$ 9.3 Billion) in the first quarter of FY23. Total wireless data usage in India grew at a rate of 6.65% from 37,626 PB in Sep'2021 to 40,126 PB in Sep'2022. Contribution of 2G, 3G & 4G data usage to the total volume of wireless data usage was at 0.16%, 1.02% and 98.81%.

Revenue of telecommunications services industry in India from 2020 to 2022 with forecasts until 2025(in trillion Indian rupees)



India added over 500 Mn new smartphone users over the last decade. We are expected to have **850 Mn smartphone users by 2026**, representing ~55% of the total population.

1. **Project Description:**

* **Accredian Telecom**, one of the leading telecom players, understands that customizing offerings is very important for its business to stay competitive.
* Currently, Accredian Telecom is seeking to leverage behavioral data from more than 60% of the **50 million mobile devices** active daily in India
* They are doing this to help their clients better understand and interact with their audiences.
* In this assignment, Accredianians are expected to build a dashboard which will help us to understand a 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**. These efforts are relevant to their users and cater to their preferences.

1. **Problem Statement:**

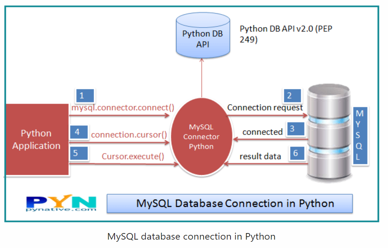
* To help the Company understand what is the **right way forward** and suggest actionable insights from Marketing and Product terms.

1. **Problem Analysis**

* Firstly, we plan to study the User behavioural data which will help us to **understand the current users** across state, gender, Age group which directly impact the company offerings.
* Based on the **data-driven** study, Plan to share the Actionable insights for customizing offers to expand their business to stay competitive.

1. **Sources of Data**

* **5.1. Data base connection:**
  + Data Acquisition was done using Connect() method of a mysql connection object to create a cursor object to perform various sql operations query i.e. Show Tables in Capstone1
  + This Connection.cursor() allows python code to execute postgrSQL command in a database session
  + Cursor.execute() methods run the SQL query & return the results & Cursor.fetchall() is used to read the query results.



* **5.2.Tables:**

3 Tables are available in the provided cases;

1. **Gender\_age\_train Table:**

* No Null Values
* All Data Types are OK
* Age alone needs to analysis based on IQR method to remove the Outliers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item** | **Device ID** | **Gender** | **Age** | **Group** |
| Data Count | 74645 | 74645 | 74645 | 74645 |
| Unique | 74645 | 2 | 85 | 12 |

1. **Phone\_brand\_device\_model Table :**

* No Null values
* All Data Types are OK
* Phone Brand is around 116 but given is 10 Phone Brand for Chinese to English Conversion

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Device ID** | **Phone Brand** | **Device Model** |
| Data Count | 87726 | 87726 | 87726 |
| Unique | 87726 | 116 | 1467 |

1. **Event\_Data Table :**

* Missing values are there in Device ID, Longitude, Latitude & State.
* Duplication of Device ID also exist but which differentiated by event id.
* Data types of Device ID & Timestamp need to update

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Item** | **Event ID** | **Device ID** | **Timestamp** | **Long** | **Lat** | **City** | **State** |
| Data Count | 3252950 | 3252497 | 3252950 | 3252527 | 3252527 | 3252950 | 3252573 |
| Unique | 3252950 | 60865 | 588126 | 60095 | 60177 | 933 | 32 |
| Missing Value | 0 | 453 | 0 | 423 | 423 | 0 | 377 |

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

* **6.1. Handling of Missing values in Event Data:**
  + Faced difficulty to fill the Missing values in Device ID, Longitude, Latitude & State.
  + However, we overcome by making the **temporary database by dropping the Nan Values** & then by **defining the function to fill the Nan Value Rows.**
* **6.2. Plot the User across map by using Longitude & Latitude**
  + Never used **Folium or GeoPandas** beforehand in training or other forum which is one of the difficult area.
  + However, we removed the outlier using .drop() with exact latitude & Longitude
* **6.3. Merging of Data Frame:**
  + One of the Most difficult & tedious task across this Project term is merging of Data Frame (With Event Data)
  + Because we used Merge with Outer/Inner, but we couldn’t get the result as expected.
  + After several iteration, we merged the data successfully using **Concat() function.**

1. **Proposed Solution for Customers**

* Filled the missing values in event data table
* As a next step, removed the Outliers using Longitude & Latitude
* Furthermore, we merged all the Data frame & confirmed that there is no missing values.
* As a Final step, Analysed the data & suggest the following actionable insights as a **data driven solution**.
* Maharastra has a Highest User. Suggest to focus on the Top 10 Cities by cutomizing the Offers to expand the Market
* For Regions with lowest user, Suggest to provide Customized offer considering ROI
* Suggest to make a tie up with the popular Phone Brands to have a default Accredian Network SIM locked Mobile
* Suggests to provide "Best Offers" for the New Users of these Phone Brands
* Roll Out more Customized offers for Females
* Tie up with Corporate - CUG Offers
* Customizing the Offers for working & Business Professionals
* Offers with Customized Hour Package can be rolled out

1. **Tools**
   * Python Jupyter Notebook
   * Microsoft Word & Excel
2. **Conclusion**

* This Project helped us to understand the User Behaviour across Age Segment, Gender & State.
* Total of 3252887 event id with 60865 Device id were provided which were analysed.
* Following are the Observations which we used to suggest action items;
  + - Top 10 States in Mobile usage are Maharastra, West Bengal, Karnataka, Tamilnadu, Andra Pradesh, Delhi, Uttra Pradesh, Madhya Pradesh, Rajasthan & Telegana
    - Males are the Most users
    - Top 3 Popular brand across Users are Xiaomi, Samsung & Hauwei
    - Peak Hour Usage is 10 o'clock in the morning followed by 8~10 in the evening
    - Lowest Usage is on night 1~5 AM
* Following the suggested actionable insight, the market can be expanded by Accredian.