**BUSINESS UNDERSTANDING**

**Business Overview**

In MTN Cote d’Ivore, a leading Telcom company provides services in voice and sms in Cote d’Ivore. There are six regions where the company operates. This happens in respect to the traffic of the different regions. The regions in our analysis in the Abidjan North, Abidjan South, Centre North, East, North West and South West. Finding out which region has the highest number of users will help in making an intuitive decision on the right amount of resources to use in each region.

**Business Objective**

The regions that we should give a priority to know the amount of resources,we should allocate in improving infrastructure.

**Business Success Criteria**

To do descriptive statistics on the resource inventories.

Clean the data; that is drop duplicates in the merged datasets and drop missing values.

**Assessing the Situation**

1. **Resource Inventory**

**Datasets:**

**i) Cells geo description**

**ii)Cells\_geo.csv**

**iii)CDR\_description**

**iv)CDR 20120507**

**v)CDR 20120508**

**vi)CDR 20120509**

**b)**Software( Github, Python, Google Colab Notebook)

1. **Assumption**

The data provided was correct.

1. **Constraints**

Identification of column names in the cells geo dataset

**Data Mining Goals**

The projects data mining goals are:-

1. Find the most used region for the three days
2. Find the product that is mostly used.
3. Find out which region uses the product more.
4. Find out the usage of different cells on site or the number in the region.

**Data Mining Success Criteria**

Our success shall be measured by the following criteria:

i) Get the mean of product usage in the different regions

ii)Get to know the regions that have the highest traffic

**2. DATA UNDERSTANDING**

**Data Understanding Overview**

In the analysis, we are going to use the datasets provided by MTN and the descriptions for the three days the data was collected, the description of the data collected and the cells geo-graphical locations which contain the transactions done.

**Data Description**

There are four datasets that were available for the project. Below is the detailed decription of the datasets:

1. Telcom Datasets: There were three datasets that contained data collected for three days. The dataset contained the value described as billing price, the anonymized phone numbers for both parties, the countries they used to call to, the identity of the cell phone and the identification of the site described as the SITE ID.
2. The Cells geographical location. This has had the cell’s geographical location, the cell phone codes, the site identification, the cities and the regions.

**Verification of Data Quality**

The merged dataset had missing values and duplicated values which I dropped.

**3.DATA PREPARATION**

**Loading Data**

I loaded the data in the Python Colab Notebook from the csv files

**Cleaning the data**

As I explored the data, there were errors on the column names which had to be corrected such as the “PRODUCT” which was written as “PRODUTC” and the SITE\_Id which was written as SIET ID. There were a number of unnecessary columns such as the country A and country B codes which I dropped.

I also did a ground check on the concatenated dataset of nun-values which were not present.

I created new datasets with values of interest which helped me in the final analysis.

**4.MERGING THE DATASETS**

After cleaning each dataset, I merged the datasets.

**5.ANALYSIS**

During the analysis, this is the order of priority of resources which should be directed to the different regions.

1. ABIDJAN NORTH: 258
2. CENTRE NORTH: 155
3. NORTH WEST: 133
4. SOUTH WEST: 125
5. ABIDJAN SOUTH: 123
6. EAST: 70

The above analysis was done in [Python](https://github.com/BrianChegeGichau/Moringa_Data_Science_W3).

**6.RECOMMENDATION**

It was also found the Voice product was mostly used in Abidjan North recording a high of 258.

It surpassed the Text Product in all the regions.