## **Business Understanding**

### **Business Overview**

MTN Cote d’Ivoire is a telecommunications company operating in Ivory Coast It offers a variety of telecommunication services over call, internet wire, radio, optical electromagnetic systems.

.

### **Business Objective**

The main objective of this report is to find out a strategy that can roll out an effective upgrade of the technological infrastructure for mobile users in Cote d’ Voire by analyzing our data in different cities in the country.

**Business Success Criteria**

To give insight on effective roll out of the infrastructure in the country by applying resources effectively to the cities that have high usage of the mobile infrastructure

**Assessing the Situation**

1. **Resource Inventory**
   1. Datasets:

cells\_geo\_description.xlsx [[Link](https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing)

cells\_geo.csv [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing)

CDR\_description.xlsx [[Link]](https://drive.google.com/open?id=1cVoNXl25IO5-_yQk97ThdeqhE6yw8YTD)

CDR 20120507 [[http://bit.ly/TelecomDataset1]](http://bit.ly/Telcom_dataset1)

CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2)

CDR 20120509 [[http://bit.ly/TelecomDataset](http://bit.ly/Telcom_dataset3)

* 1. Software( Github, Google Collaboratory)

1. **Assumptions**
   1. The data provided is correct and up to date
2. **Constraints**
   1. There are no constraints

### **Data Mining Goals**

Our data mining goals for this project are as follows:

* Which ones were the most used city for the three days?
* Which cities were the most used during business and home hours?
* Most used city for the three days?
* Which regions have the highest frequency of calls sms data
* Which city brings most revenue

**Data Mining Success Criteria**

Our success criteria will be measured by the following criteria;

* We target the cities that generate the most revenue and highest usage of sms and calls

## **Data Understanding**

### **Data Understanding Overview**

For this project, we are using the availed dataset by the company. These datasets are

* Cells\_geo\_description.xlsx - description of our cell geographic data
* Cells\_geo.csv - data of our cell geographic data
* CDR\_description.xlsx - description of our call detail record
* CDR 20120507 - Day one of our call detail record
* CDR 20120508 - Day two of our call detail record
* CDR 20120509 - Day three of our call detail record

### **Data Description**

We have two datasets available for this project. A detailed description of the datasets is provided as follows:

* Cells\_geo\_description - description of columns in cell geographic data
* Cells\_geo.csv - detailed description of mappings of our cell geographic data from latitude longitude region cell id site id
* CDR\_description.xlsx - description of columns of the call detail record
* CDR 20120507- [Day one of the data collected of usage of telecommunication services eg sms and calls](http://bit.ly/Telcom_dataset1)
* CDR 20120508 - [Day two of the data collected of usage of telecommunication services eg sms and calls](http://bit.ly/Telcom_dataset1)
* CDR 20120509 -[Day three of the data collected of usage of telecommunication services eg sms and calls](http://bit.ly/Telcom_dataset1)

### **Verifying Data Quality**

All of the four datasets had any missing values. There were data errors in the datasets like null values and duplicates

## **Data Preparation**

These are the steps followed in preparing the data

#### **Loading Data**

Loaded the datasets from the CSV using pandas

#### **Cleaning Data**

While doing data exploration, we noticed column names were not the same along the call detail record datasets and the cell geographic data so we had to rename them so as to be persistent in all the analysis. We had to clean the null values remaining with the first in the set of duplicates. We had to get rid of null values since they were skewing our datasets a lot. We restrained on merging our datasets early so we individually cleaned all of our datasets.

#### **Merging of the Datasets**

After cleaning the data, it was time to merge the four datasets.We used pd.concat to concatenate the three CDR datasets and later merged our cell geography data

#### **Deriving New Attributes**

We merged on site id since its a unique connector of our datasets and its where our cells relay their information from.

## **Analysis**

During our analysis, we were able to single out the following cities to be targeted first

1. CITY
2. COCODY
3. YOPOUGON
4. ABOBO
5. TREICHVILLE
6. MARCORY
7. SAN-PEDRO
8. ADJAME
9. KOUMASSI
10. YAMOUSSOUKRO
11. DAOUKRO
12. GAGNOA
13. BOUAKE
14. DALOA
15. PORT-BOUET
16. GRAND-BASSAM
17. DIVO
18. ATTECOUBE
19. ABENGOUROU
20. SIPILOU

## **Recommendations**

From our analysis, we would recommend that the company would prioritize the above-listed cities. Our main reason behind this recommendation would be that the states listed above have the highest revenue spend and ratio between the population and the revenue is high.