## **Business Understanding**

### **Overview of the Business**

MTN Cote D’Ivoire is a telecommunications company registered with south African-based MTN Group. It provides postpaid and prepaid GSM services on fixed-line, mobile, and internet services with an estimate of about twelve million subscribers on mobile and the internet. It is the number two telecommunications operator in the country.

### **Business Objective**

1. To identify the most used city for the three days
2. To identify the cities used during business and home hours

### **Assessing the Situation**

The datasets to be used are provided for the process as shown below:

1. cells\_geo\_description.xlsx [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing)
2. cells\_geo.csv [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing)
3. CDR\_description.xlsx [[Link]](https://drive.google.com/open?id=1cVoNXl25IO5-_yQk97ThdeqhE6yw8YTD)
4. CDR 20120507 [[http://bit.ly/TelecomDataset1]](http://bit.ly/Telcom_dataset1)
5. CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2)
6. CDR 20120509 [[http://bit.ly/TelecomDataset3]](http://bit.ly/Telcom_dataset3)

The data has some missing values and needs some data cleaning and will use the following software for the process:

* Python
* Google collaboratory
* GitHub

### **Data Mining Goals**

Our main goal is to determine how much the telecommunication company will go about the upgrade of its infrastructure strategy within the given cities.

## **Data Understanding**

### **Data Understanding Overview**

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

* Geographical description - This dataset gives the description per geographical area
* CDR description - This dataset shows the description per the data used

### **Data Description**

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

* **Geographical data set -** This dataset contains the description of the data as per the geographical area.
* **CDR data set -** This dataset contains the description of data as per the different products , billing price etc.

### **Verifying Data Quality**

The data contains missing values hence data cleansing should be done using python to fill in the missing values.

## **Data Preparation**

These are the steps followed in preparing the data

#### **Loading Data**

Loading the data set in csv format through github

**Cleaning Data**

Thereafter we clean the data and fill in the missing values

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

After cleaning the data, it was time to merge the datasets.

## **Analysis**

Analyzing the data using python to clean the data