## **Business Understanding**

### **Business Overview**

MTN Cote d’Ivoire is a leading telecom operation company (OpCo) in Ivory Coast and is part of the global MTN operations in Middle East and Africa.

MTN Cote d’Ivoire would like to upgrade its technology infrastructure for its mobile users in Ivory Coast.

### **Business Objective**

The main objective of this report is to identify the best strategy for MTN Cote d’Ivoire to upgrade infrastructure within its cities.

### **Business Success Criteria**

To compile a list of cities that will increase the return on investment of the upgrade resulting in collecting more revenue from the areas with growth portential.

### **Assessing the Situation**

1. **Resource Inventory**
   1. Datasets:
      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)
   2. Software ( Github, Google Collaboratory)
2. **Assumptions**
   1. The data provided is correct and up to date. There will be some cleaning up but there will be no fundamental change of the data.
   2. There is no failures in neighboring areas that can change traffic distribution pattern on a particular site.
3. **Constraints**
   1. The available is only for 3 days. These don’t represent a complete picture
      1. Each day is not a full 24-hour cycle that will reflect true picture of subscriber bahaviour.
      2. The 3 days cannot represent a whole month usage pattern to show true normal subscriber pattern. For example, the usage on weekends, night, weekday, day (business hours) cannot be the same.

### **Data Mining Goals**

My data mining goal for this project are: -

* 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?

**Data Mining Success Criteria**

Our success criteria will be measured by: -

* I target the cities that have most usage states that have the usage value sum that we will use as a priority list for upgrading the MTN Cote D’Ivoire infrastructure.

## **Data Understanding**

### **Data Understanding Overview**

For this project, we are using the availed dataset by the MTN Cote d’Ivoire. These datasets are: -

* The Sites/Cells distribution per city - This dataset gives the number of cells grouped into sites per city.
* Call data records (CDR) – This data for 3 days between 23:00hrs and 01:00hrs gives the actual network usage for three services; voice, sms and data.

### **Data Description**

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

* **Cells, Site infrastructure: - by State dataset -** There are several cells per site, several sites per City, several Cities per Zone, etc. I will use the VILLES (City), CELL\_ID and SITE\_IDs for this project.
* **Three day Call Data Records -** This dataset for 3 days, will the sample used to analyse the network usage by substructures. The toal value usage will be calculated on the 3 services of SMS, Voice and Data.

### **Verifying Data Quality**

The two datasets have missing values that will not have an impact on the final outcome of the data mining. There was only a mispelt SITE\_ID (as SIET\_ID). Other than, that rest of data was good quality for the purpose.

## **Data Preparation**

These are the steps followed in preparing the data

#### **Loading Data**

Loaded the datasets from the CSV and created pandas datasets.

#### **Cleaning Data**

The data was loaded and expolored. There were missing data. However, the missing data is not used in the current project, so it will not have a material effect on the outcome. A little misplet error for sites (SIET\_ID) was corrected to SITE\_ID. The material data was clean enough to be used without further cleaning.

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

After cleaning the data, the two data sets of Cells and CDRs were merged to get the priority list for infrastructure upgrade.

## **Analysis**

During our analysis of the three-day CDR data set, I was able to work out the top 3 cities in total value as follows: -

* 1. 2012-05-07 CDR data set

|  |  |  |
| --- | --- | --- |
| **#** | **VILLES** | **VALUE** |
| **1** | **YOPOUGON** | 86408 |
| **2** | **COCODY** | 58739 |
| **3** | **ABOBO** | 35746 |

* 1. 2012-05-08 CDR data set

|  |  |  |
| --- | --- | --- |
| **#** | **VILLES** | **VALUE** |
| **1** | **YOPOUGON** | 86408 |
| **2** | **COCODY** | 58739 |
| **3** | **ABOBO** | 35746 |

* 1. 2012-05-09 CDR data set

|  |  |  |
| --- | --- | --- |
| **#** | **VILLES** | **VALUE** |
| **1** | **COCODY** | 86923 |
| **2** | **YOPOUGON** | 71022 |
| **3** | **ABOBO** | 44912 |

* From the analysis: -
* **Yopougon** was the most used city for the three days. This was followed by Cocody and Abobo in that order.
* From the dataset, one cannot determine the most used city during the busioness hours since the dataset was only between the 23:00hrs and 01:00hrs. This is only within home hours and when traffic is low.
* Within the 23:00hrs and 01:00hrs, **Yopougon** was the most used city for the three days.

The above analysis was done using pandas on a Jupyter notebook. The full analysis can be found in the following github[[Moringa\_Data\_Science\_Prep\_W3\_Independent\_Project\_2021\_09\_Lawrence\_Ondieki\_Python\_Notebook.ipynb](https://github.com/ogwora/IP_Week3/blob/eae2cfd3f6ef826cbb73e0c34d1b59a1d83a657c/Moringa_Data_Science_Prep_W3_Independent_Project_2021_09_Lawrence_Ondieki_Python_Notebook.ipynb)].

## **Recommendations**

From our analysis, I would recommend that MTN Cote d’Ivoire upgrade infrasture in top three cities in terms of usage of Yopougon, Cocody and Abobo. These three cities have the most usage value from the data collected for the three days.This will give MTN a good return on investiment.