# 1.Business Understanding

**Business overview**

MTN Africa is a company invested in the telecommunications Industry.

It was first pioneered in South Africa in 1994. Since then, it has ensured growth through connectivity, communication and collaboration solutions over world-class ICT infrastructure, as a committed partner to SMEs, corporates and the public sector. MTN has branched off further across the continent to several different countries.

Our focus will magnify MTN’s Cote d'ivoire branch. MTN Cote d’ivoire branches its services further through various cities. Each of these cities has installed technology infrastructure to drive their mobile platform.

**Business objective**

Our main objective will be to identify the most efficient way MTN Cote d'Ivoire will go about the upgrade of its infrastructure strategy, within the given cities.

**Business success criteria**

Identify the top cities that have the highest rates of MTN mobile subscription.

**Situation Assessment**

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 :
2. Github
3. Ubuntu Windows and Linux Subsystem
4. Google Collaboratory
5. **Assumptions**
   1. The sample data provided is correct and up to date
6. **Constraints**
   1. There are no constraints

**Data Mining Goals**

Our data mining goals for this project are as follows:

* Which product provides the most value and in which locations does it do so.
* Which regions have a high MTN subscription and does a high subscription relate to high value customers
* Determine which were the most used cities for the three days the sample data was collected.
* Investigate which was the city with the highest number of cellular usage for the three days

**Data Mining Success Criteria**

Our success criteria will be measured by the number of cities with the highest number of MTN subscriptions.

# 2. Data Understanding

**Data Understanding Overview**

We shall be focusing on Four of the Six datasets containing collected sample data by MTN

Cote’d’Ivoire.

* cells\_geo.csv [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing) -This dataset contains all necessary geographical information
* CDR 20120507 [[http://bit.ly/TelecomDataset1]](http://bit.ly/Telcom_dataset1) -This dataset contains necessary customer and cell site information
* CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2)-This dataset contains necessary customer and cell site information
* CDR 20120509 [[http://bit.ly/TelecomDataset3]](http://bit.ly/Telcom_dataset3)-This dataset contains necessary customer and cell site information

**Data Description**

We have two datasets available for this project. A detailed description of the

datasets is provided as follows:

* **Telecom dataset -** The dataset mainly constitutes of customers information indexed by Cell\_Id as its primary key.It also contains relevant information on the cell sites which are necessary for reference of data in the Cell Geography dataset
* **Cell Geography dataset -** This dataset focuses on the Geographical locations of different cities and zones

**Quality of Data**

The two datasets had a number missing values. There were also a number

of known data errors in the datasets.

# 3.Data Preparation

These are the steps followed in preparing the data :

**Importing Libraries**

Import pandas and Numpy libraries into the Google Collaboratory environment and run them

#### **Loading Data**

Loaded the datasets from the CSV and Excel files provided

#### **Cleaning Data**

Telecom dataset

While conducting our data exploration, we noticed that in the Telecom datasets, we had to rename a column name was misspelled, PRODUTC hence we had to change it to PRODUCT.We also noted that the VALUE column contained Null values which would not influence the outcome of the business objectives hence they were removed.The CELL\_ID column contained characters which interfered with the correctness of the data hence they were removed.

Cell Geography dataset

While conducting our data exploration in Cell\_Geo datasheet we noticed that there was a number of columns that we would drop i.e LATITUDE, LONGITUDE,DECOUPZONE

All of which was as a result of ineffectiveness to making a change in the

overall business

objectives.We also noted that STATUS column contains Null values which

we promptly deleted

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

After cleaning the data, we went ahead to merge the two datasets.

# 4.Analysis

The above analysis was done using Python, pandas library. The full

analysis can be found in the following notebook in the Git repository.

<https://github.com/SydneyTsuma/Autolib_Project.git>

# 5.Recommendations

From our analysis MTN Cote’d’ivoire should continue to invest in supplying the population with MTN airtime as it registers the highest use by the population.

We also recommend that more cells on site should be developed because they registered the lowest number of billing.