**1.Business Understanding**

**Business Overview**

**Côte D'Ivoire MTN Mobile Money is an MTN service that allows subscribers to receive and send money, day or night, from your mobile, anywhere in Côte d’Ivoire and without even moving.**

**Business Objective**

The main objective of this report is to upgrade the technology and infrastructure of Cote D’Ivoire for its mobile users in Ivory Coast.

**Business Success Criteria**

To compile a list of telecom datasets that will help to upgrade its technology infrastructure for its mobile users in Ivory Coast.

**Assessing the situation**

**1. Resource Inventory**

**a. 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)

b.Software

The software used is (Github, Google collabs, Google docs, SQLite)

**2.Assumptions**

a. The data provided is correct and up to date.

**3.Constraints**

There are no constraints.

**Data Mining Goals**

Our data mining goals for this project are as follows:

* Calculate the average usage of voice bundle in the three datasets.
* Getting the total number of sms from the three datasets
* Getting the total number of voice calls from the three datasets
* Select certain columns to use in some of the datasets since all are not similar
* Merging of the datasets
* Determine the cells that generate highest revenue
* Sort the cells to find the ones that might generate the highest profit

**Data Mining Success Criteria**

Our success criteria will be measured by the following criteria:

* Determining the cells that generate the highest revenue

**2.Data Understanding**

**Data Understanding Overview**

We have three data sets available for this project. A detailed description of the data sets is

Provided as follows :

* Cells\_geo\_description - it contains the column names and their description
* Cells\_geo - it contains the location of the villes
* CDR description - it contains column names
* CDR 20120507 - it contains a description of the services with the logs for the month of June.
* CDR 20120508 -it contains a description of the services with the logs for the month of July.
* CDR 20120509 - it contains a description of services with the logs for the month of August.

**Data Description**

All three datasets contain :

Product , date\_time, cell\_on\_site, dw\_a\_number, dw\_a\_number, country\_a, country\_b, cell\_id, site\_id

**Verifying Data Quality**

None of the provided datasets had any missing values.

**3.Data Preparation**

These are steps followed in preparing the data:

**1.Loading Data**

Loaded the data sets from the CSV and then created an SQLite database from them.

**2.Cleaning Data**

While doing data exploration we noticed the cells\_geo dataset the column was written as SITE\_CODE and in the other datasets the corresponding columns were SITE\_ID. Therefore we decided to convert the SITE\_CODE to SITE\_ID so that it corresponds with the values in the other datasets. As a result it would make the merging of the datasets easier.

**3.Merging of the datasets**

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

**4.Analysis**

While analysing we started by doing calculations the calculations involved:

Calculating the average usage of voice calls in the three datasets noting that dataset three had the highest usage of voice calls hence seemed to be more promising.

Calculating the average usage of sms bundles in the three datasets noting that dataset three again has the highest mean of sms bundles therefore seemed to be more promising again.

Then moved on to working through columns. We used the Product, Value, Date\_Time and the CELL\_ID in the datasets.

Then got the total number of sms and voice calls from the different datasets under the Product column.

Then got the most used infrastructure from the datasets.

ffcc507579

ffcc507579

ffb86d66d2

ffcc507579

fed537f3b2

Ff99709d62

We obtained the above listed as the most used infrastructure.

Then determined the cells that generated the highest revenue

.

And finalised by having the following cells:

ffa6759bb2

2def69183e

3f5ac97214

MpOhP3OKNN

5a7f6cf0fa

**RECOMMENDATION**

We would recommend that MTN Ivory Coast should concentrate on the above infrastructure that will generate higher profit.

**GITHUB LINK**

[**https://github.com/Alakay-cloud/Moringa\_Data\_Science\_Prep\_W3\_Independent\_Project\_2020\_03\_Alex\_Mburu**](https://github.com/Alakay-cloud/Moringa_Data_Science_Prep_W3_Independent_Project_2020_03_Alex_Mburu)

**5.Recomendations**