**DATA ANALYSIS REPORT THAT WILL ASSIST MTN COTE D'IVOIRE UPGRADE ITS INFRASTRUCTURE , MAY 2021**

**BY DENNIS KAMAU KIARIE**

1. **Business Understanding**

**Business Overview**

Ivory coast is a country in West Africa. It consists of a number of cities , zones and Regions. MTN Cote d'Ivoire is one of the companies offering Telecommunication services in Ivory coast.

Based on analysis carried out by the Worldometer elaboration of the latest United Nations data, Côte d'Ivoire current population is estimated to be 26,942,755,

population density of 83 per Km2 (215 people per mi2) , total land area is 318,000 Km2 (122,780 sq. miles) , 51.3 % of the population is urban (13,532,312 people in 2020) and the median age in Côte d'Ivoire is 18.9 years.

A high percentage of the population depends on MTN Cote d'Ivoire for mobile coverage.

**Business Objective**

The main objective of this report is to improve Telecommunication Services across mobile users in Ivory coast and drive more revenue to the MTN Cote d'Ivoire telecommunication company.

**Business Success Criteria**

To determine the upgrade strategy for MTN Cote d'Ivoire telecommunication equipment.

**Assessing the situation**

1. **Resource Inventory**

**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, python , numpy , pandas.

**3. Assumptions**

1. The data provided is correct and up to date

**4. Constraints**

There are no constraints

**Data Mining Goals**

Our data mining goal for this report is as follows:

To determine the following: Which city has the highest user traffic ? Which city has the lowest user traffic ? Which products are offered by MTN ? Which MTN product is mostly used per city ? Which MTN product is not mostly used per city ? Which time of the day there is more traffic coming from users? Which time of the day there is less traffic from the users ? Which Cell has high coverage ? Which Cell has low coverage ? Which Cell is receiving the most traffic ? Which areas support the Fibre optic ?

**Data Mining Success Criteria**

Our success criteria is to be measured with the following:

We target the Regions and Cities that have the highest frequency of calls, which products are mostly used, which Regions have the highest return per product and at which time there is more traffic from the user.

**2.Data Understanding**

**Overview**

For this project we are using six datasets which are provided by the company. These dataset are:Telecom dataset 1, Telecom dataset 2, Telecom dataset3, CDR description dataset ,

cell \_geo dataset and cell\_geo\_description dataset.

**Data Description**

The three Telecom Datasets contain the same column names which are : product, bill, datetime, call\_anonymized, recipient\_anonymized, cell\_id and site\_id.

Cell\_geo, cell\_geo\_description and CDR description datasets are joined together.

The two combined datasets are joined together using the cell\_id and site\_id which acts as a secondary key .

**Verifying data quality**

These are steps followed in preparing the data.

1. **Loading the Data:**

Load the data from CSV and EXCEL format.

1. **Cleaning the Data**

While doing the data exploration we noted that the combined datasets of Cell\_geo, cell\_geo\_description and CDR description had duplicate values, null values, missing values and unnamed columns. All these issues were addressed by dropping the missing column name, null values , duplicate values and renaming of columns.

1. **Merging the Datasets**

After cleaning the data we merged all the datasets based on secondary key that is cell\_id and site\_id

1. **Deriving new attributes**

Once merging was complete we created a list of cities and their corresponding Zone name based on the traffic frequency.

**4. Analysis**

During our analysis we were able to single out the Region with the highest traffic frequency. These regions are:

**LAGUNE 33261**

**LACS 2484**

**BAS SASSANDRA 2395**

**HAUT SASSANDRA 2304**

**VALLEE DU BANDAMA 1575**

**SUD BANDAMA 1533**

**SAVANE 997**

**FROMAGER 916**

**WORODOUGOU 795**

**18 MONTAGNES 783**

**MARAHOUE 621**

**AGNEBY 602**

**NZI COMOE 571**

**MOYEN CAVALLY 531**

**DENGUELE 398**

**ZANZAN 336**

**MOYEN COMOE 321**

**SUD COMOE 206**

**BAFING 102**

**Name: region, dtype: int64**