**1.BUSINESS UNDERSTANDING.**

* **Business Overview**

MTN Côte d’Ivoire (MTNCI), is located in Abidjan, IvoryCoast and is part of the professional services sector industry. MTN COTE D’IVOIRE has 45 total employees across all of it’s locations and generated $5.44 million (usd) in sales. In this figure, employees and sales figures are modelled. There are 17 companies in the MTN Cote d’Ivoire corporate family. It is a subsidiary of South African-based MTN Group. It provides postpaid and prepaid GSM services including fixed line, mobile and internet services.

* **Business Objective**

The main objective of this report is to figure out how MTN Cote d’Ivoire should go about their infrastructure strategy upgrade in the given cities for its mobile users in the Ivory Coast.

* **Business Success Criteria**

To compile different data obtained from different cities and towns and find out the cities with the most usage of MTN Cote d’Ivoire services within the given period of time.

**Assessing the Situation.**

**Requirements, Assumptions and Constraints.**

* **Resources**
* Personal (Technical support, Data mining experts)
* Project Datasets used;

1. <https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing>
2. <https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing>
3. <https://drive.google.com/open?id=1cVoNXl25IO5-_yQk97ThdeqhE6yw8YTD>
4. <http://bit.ly/Telcom_dataset1>
5. <http://bit.ly/Telcom_dataset2>
6. <http://bit.ly/Telcom_dataset3>

* Computing resources
* Software ( Google Colaboratory, Google Drive, Github, Jira)

* **Assumptions**
* Data sampled will be an accurate representation of the whole Dataset.

**Data Mining Goals**

Our data mining goal is to figure out how MTN Cote d’Ivoire should go about their infrastructure strategy upgrade in the given cities.

Potential questions for research include:

* Which is the most used city for the three days?
* Which is the most used product in the given cities?
* What is the most popular hour of the given products?
* Which is the most used city during business and out of work hours?

**Project Plan**

Cross Industry Standard Process for Data Mining will be used as a guidance for conducting this research. Here is an overview of the plan.

| **Phase** | **Time** | **Resources** | **Risks** |
| --- | --- | --- | --- |
| Business Understanding | Less than an hour | Project Datasets  Data Analysts/Scientist |  |
| Data Understanding | 2 hours | Project Datasets  Data Analysts/Scientist |  |
| Data Preparation | 6 hours | Project Datasets  Data Analysts/Scientist |  |
| Data Modelling | Less than an hour | Project Datasets  Data Analysts/Scientist |  |
| Evaluation | 1 hour | Project Datasets  Data Analysts/Scientist |  |

**2.DATA UNDERSTANDING**

**Data understanding Overview**

The existing datasets consist of sample data collected from the station’s data about the usage of their products in Cote d’Ivoire for three days. We will not need any additional dataset apart from what we have to meet our objective. The datasets used were:

1. <https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing>
2. <https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing>
3. <https://drive.google.com/open?id=1cVoNXl25IO5-_yQk97ThdeqhE6yw8YTD>
4. <http://bit.ly/Telcom_dataset1>
5. <http://bit.ly/Telcom_dataset2>
6. <http://bit.ly/Telcom_dataset3>

**Collecting Initial Data**

The source of the data collected was the company’s database.

**Describing and exploring Data**

There are six datasets available. Three telcom datasets contain details of users of MTN Cote d’Ivoire . One geo dataset contains all of the geographical locations of MTN users. The rest contain descriptions of these datasets.

A further description of the provided datasets is as follows:

**Telcom dataset 1, 2 & 3 -** These datasets contain the details of the information of the users data. They have missing values, it has 10 columns and 5001 entries. All the values are represented in the data set and the data sets take some time to load. They open best in a new window..

**CDR Description -** this dataset has a description of the column data required for analysis.

**Geo dataset -** This dataset contains the geographical locations of the users.

**Cells geo description -** It has details about the fields in the geo dataset.

A data exploration report can be written for further investigation.

**Verifying Data Quality**

The dataset does contain missing values, null values and it has some errors. For example there are some unnecessary columns we had to drop and rename some fields.

**3. DATA PREPARATION**

**Selecting Data**

The following datasets were used for analysis in this project based on the relevance of our goals and data quality.

* Telcom 1,2 & 3 datasets
* CDR description
* Geo dataset
* Cells geo description

Used data frames to load data from a file, examine basic statistics of the data, change some values and finally get the output which is the results.

**Cleaning Data**

Data cleaning procedures performed during the analysis included:

* Dropped all unnecessary columns as per description for analysis so we could remain with the beneficial columns.
* Renamed the DATETIME to DATE\_TIME column then separated date and time.
* Renamed the miss-spelled columns as per the description.
* Dropped duplicate values and information from the dataset.
* Merged the 3 telcom datasets forming a new dataframe.
* After loading the geo dataset, dropping duplicates, null values and unnecessary columns, I merged the geo dataset and the 3 telcom datasets for full analysis.

**Constructing New Data**

* New data was created upon separating the date and time columns.
* New data was also created by creating a new data frame where I merged the 3 telcom datasets then merged them with the geo dataset.

**4. DATA ANALYSIS**

**During analysis, the following question was answered.**

1. Which was the most used city for the given period? Cocody
2. Which was the most used product? Voice

The above analysis was done using the python programming language. This analysis can be found in the following python notebook.

<https://colab.research.google.com/drive/1yIVh9_8OM16X14IZ2gLxPTLgMg8TeYpa?usp=sharing>

**5. RECOMMENDATION**

Following my analysis, the following recommendations are as provided.

* Voice should be given highest priority since it has the largest sum of users followed by data. According to the skewness and standard deviation values, data is really shooting up.

| **STANDARD DEVIATION** |  |
| --- | --- |
| PRODUCT |  |
| Voice | 236.315367 |
| data | 104.899256 |
| sms | 14.680076 |

| **SKEWNESS** |  |
| --- | --- |
| PRODUCT |  |
| Voice | 34.860563 |
| data | 34.559244 |
| sms | 2.469463 |

* The Abidjan North area has the most users, especially Concody. Here is the list of the top five. This should be the focus.

| COCODY | 354 |
| --- | --- |
| YOPOUGON | 289 |
| ABOBO | 244 |
| MARCORY | 141 |
| TREICHVILLE | 132 |

**6. EVALUATION**

From our business overview we have been able to determine the most used cities and the most used product.

#my github link