GITHUB REPO:

**1. BUSINESS UNDERSTANDING**

**Business overview**

Transmission of information by various types of technologies has its origin in the desire of humans for communication over a distance greater than that feasible with the human voice but with a similar scale of expediency; thus, slow systems for example postal mail are excluded from the field. MTN Côte d'Ivoire has therefore emerged to be the most known company in West Africa that has fulfilled the desire of humans in terms of distant communication channel and internet.

MTN Côte d'Ivoire is a telecommunications company which was founded on 1 July 2005. It is a subsidiary of South Africa-based MTN Group and an entity of MTN International of Loteny Telecom's mobile phone license (Telecel). Being an Ivorian telecommunications company, MTN Côte d'Ivoire has provided postpaid and prepaid GSM services including fixed line, mobile, and internet services to its customers enabling them to avail quality telephonic and internet access. with a mobile phone license (Telecel).

**Business objective**

i.  **General objective**

To identify strategies to be used in upgrading technological infrastructure.

ii.  **Specific objectives**

* To analyze the cities that were the most used in the three days
* To assess the cities that were most used during business hours and home hours
* To identify the city that was most used in the three days

**Business success strategy**

**To invest more on the cities that used the company’s services in general.**

**Assessing the situation**

**i. Resource inventory**

**Datasets :**

a. Cells\_geo\_describtion.xlx [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing)

b. Cells\_geo.csv[[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing)

c. CDR\_description.xlsx[[Link]](https://drive.google.com/open?id=1cVoNXl25IO5-_yQk97ThdeqhE6yw8YTD)

d. CDR20120507[[http://bit.ly/TelecomDataset1]](http://bit.ly/Telcom_dataset1)

e. CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2)

f. CDR 20120509 [[http://bit.ly/TelecomDataset3]](http://bit.ly/Telcom_dataset3)

**Software:** Pandas, Google Collaboratory and Git hub

ii.  **Assumptions.**

Data provided is relevant and correct

MTN Côte d'Ivoire has datasets that are up to date.

iii. **Constraints.**

There is no limitation in this analysis

**Goals for data mining**

Identifying strategies that the company can employ when upgrading technological infrastructure

Identify the cities that were the most used in the three days

Find out which cities that were most used during business hours and home hours

Realize identify the city that was most used in the three days

**2. DATA UNDERSTANDING**

**Data understanding overview**

In this project we are only going to use the datasets that have been provided by the company.

· Dataset1, Dataset2, and Dataset contains information on the services that have been uder by consumers in the days.

· Cells\_geo contains site information

**Data description**

Below is a detailed description of the date that has been provided by the company.

· Dataset1, Dataset2, and Dataset3 have the billing prices, voices, cell on site, cell id site id, date –time and product

· Cells\_geo has the city region, area, cells\_id, site code and status

**Verifying data quality**

There were no missing values in Dataset1, Dataset2, and Dataset3. Although there were duplicate files

For Cells\_geo there were missing values which could lead to some errors that might lead to estimation.

**3. DATA PREPARATION**

Below are the steps followed to load the data;

a**. Loading Data**

Pandas is used to load the data frames to our google colab working environment.

b.  **Cleaning Data**

While going through our datasets we realized that the naming of Dataset1, Dataset2, and Dataset3 did differ so we decided to set all column names for all the data frames to be the same to ease merging. The datasets also had duplicate records that we had to drop.

We also realized that the cell\_geo had missing values that we had to fill using various techniques.

c.  **Merging of the Datasets.**

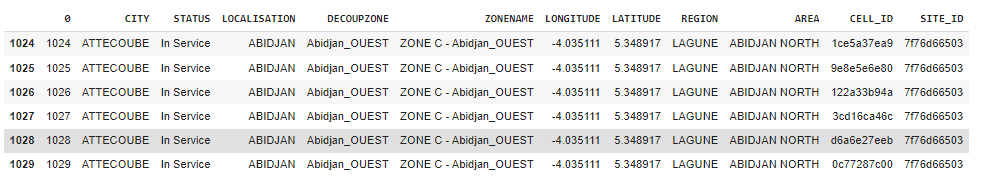
After cleaning the data, we concatenated the three datasets and merged them with the cell\_geo.

d. **Deriving New Attributes**

Finally, after merging the data frames, we came up with a data frame that will help in the analysis and achieve our goal.

**4. ANALYSIS**

During our analysis we were able to find the most used city as listed bellow.

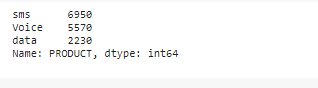


Attecoube was the most used city

It was also realised that there were no usage during business hours as shown below

which means that either people don't use the products during business hours or the data for business hours was not collected.

An for working hours, tye results show that customers use all product both after work (17.00hr to 23.59 hrs) and before work (00.00 hrs to 08.00hrs) which indicates that its either the data collected was for home hours only or customers only use the products while at home. The result is as shown below;



From the above results we can also note that the sms are mostly used

**5. RECOMMANDATIONS**

From our analysis we would like to recommend the following to the company.

* The company should prioritize investing in the top most used city. The reason behind this recommendation is, the cities are likely to generate more revenue as analysed from their usage.
* The company should also consider investing more on sms services since most customers use them most of the time.
* The company should also consider upgrading the infrastructure in the less most used cities i.e strong internet connection, and stable network as an incentive for customers to use more of their services.

6. **EVALUATION**

This analysis can be used and its methods, reports, and recommendations can be used and adapted to any situation related to this project. The company can use the findings to advise its users , improve human resources, and in the case of upgradding.

Evaluating the results on the most used product, the company should also consider finding reasons why the product is mostly used and try to implement the findings to other services.

**Bellow is the attached link to my github repo that can help in understanding the reports.**

Github Repository: