**BUSINESS UNDERSTANDING.**

BUSINESS OVERVIEW

MTN Côte d'Ivoire, a leading telecom would like to upgrade its technology infrastructure for its mobile users in Ivory Coast. So given a number of datasets I as a data scientist will go through them and find the best way for MTN Cote d'ivoire go about the upgrade of its infrastructure strategy within the given cities.

BUSINESS OBJECTIVES.

The main objective of this project is to upgrade the technological infrastructure within the given cities in MTN Cote d'ivoire.

Project plan

These are plans that will help the Data Scientist achieve the business objectives.

1 Finding the most used city for the 3 days

2 Finding the most used cities during business and home hours.

3 Most used cities for the three days.

ASSESSING THE CURRENT SITUATION

RESOURCES USED

So we have a number of resources to use during this project and this includes data ,computer resources and softwares.

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)

Softwares

The softwares i will use on this project are mostly Pandas inPython and Git ,which we will then upload on our Git repository on github.

DATA MINING GOALS.

Our data mining goals for this project is as follows;

* Merging the Telecom tables for the three days into one table
* Merge the one big table with the with the Geo table
* Data cleaning
* Analysing the data

2 **DATA UNDERSTANDING**

**Overview**

We have mentioned a number of data resources that we are going to be using on this particular project.

1. Telcom Data sets ;These are datasets that

CDR 20120507 [[http://bit.ly/TelecomDataset1](http://bit.ly/Telcom_dataset1)

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

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

This are telecom data sets that were obtained for the three days.We also have a set of data which gives a more in depth explanation of the column variables in the Telecom datasets.

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

1. We also have a set of data called the cell geo,and as set of describing the variables in depth and below are the links

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

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

3 .GITHUB REPOSITORY LINK ; https://github.com/NorahEkuttan/hello-.git

Explore data

Verify data

The data had some missing values hence we deleted all the rows with a missing value to get a more clean data for analysis

Data quality report

**3 DATA PREPARATION**

**Loading and importing data into panda**

So the first thing I did was import the datasets as panda.I imported the telecom datasets separately.

**Merging the datasets**

The next thing was to merge the three datasets vertically into one big data

**Cleaning**

I then cleaned the data by noting the missing values,and deleting rows with the missing values.

This applies to both Geo\_csv table and the merged table as well

4. ANALYSIS

In the Geo csv i analysed my data by using the described function just to find out the most frequent city to use the telephone service that is both sms and voice

count 3880

unique 501

top COCODY

freq 354

Name: VILLES, dtype: object

And here i found out the city with the most frequent mobile survive is COCODY

Telcom csv datasets

On the Telcom csv data sets here were the results

count 13006.000000

mean 27.304398

std 111.025088

min 0.000000

25% 0.000000

50% 0.000000

75% 25.000000

max 6750.000000

Name: VALUE, dtype: float64

The count after deleting rows with missing values was 13006

The mean for values that is the voice and sms altogether was 27.30

Min number of voice or sms was 0 while the maximum was 6750

Thee standard deviation was 111.01

5 RECOMMENDATION

Since the goal is to improve technological infrastructure I would advise telecom to focus its resources in cities like COCODY because most people use the service there.

GIT/GITHUB REPOSITORY.

Here is the link for my github repository containing my python work

<https://github.com/NorahEkuttan/hello-.git>

<https://github.com/NorahEkuttan/hello-.git>