**A DATA REPORT ON HOW MTN COTE D'IVOIRE CAN UPGRADE ITS TECHNOLOGY INFRASTRUCTURE FOR ITS MOBILE USERS IN IVORY COAST.**

1. **Business Understanding**

Our client MTN Cote D’Ivoire wants to upgrade its technology infrastructure in Ivory Coast so that they can be in a better position to deliver quality telecom services to its clients and probably add additional services based on clients’ needs.

One noticeable challenge facing MTN Cote D’Ivoire is that the data they have obtained from their clients’ is not formatted correctly and this inadvertently hinders them from deriving any meaningful information for making business related decisions on the upgrade.

MTN Cote D’Ivoire needs their datasets cleaned, analysed and recommendations/conclusions drawn from it. The objectives and project plan are as follows;

1. Breaking down the whole problem into smaller problems that are easily rectified i.e.

* Identifying incorrectly formatted datasets and standardising the format.
* Creating DataFrames for each dataset and merging all DataFrames into one.

1. Answering questions that lead us to uncover how we can analyse the new DataFrame to

derive meaningful insights to upgrade MTN’s Technology Infrastructure. i.e.

* Which ones were the most used city for the three days?
* Which cities were the most used during business and home hours?
* Most used city for the three days?
* What are the descriptive statistics of MTN’S datasets?

1. **Data Understanding**

The following are the resources/datasets provided by MTN Côte D’Ivoire. To understand the data, we first have to describe what it looks like and the shape/form it has taken so that we can identify how we want to format it for easier analysis. The table below describes the shape and gives links to preview the datasets as they are before any preparation takes place. The shapes of the datasets were obtained using Python Pandas.

|  |  |  |
| --- | --- | --- |
| **Dataset** | **Shape/Size** | **Preview** |
| cells\_geo.csv | 3974 rows × 1 columns  (The dataset has multiple columns in one column) | [cells\_geo](https://drive.google.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view) |
| CDR 20120507 (TelecomDataset, Day 1) | 5001 rows × 10 columns | [TelecomDataset](http://bit.ly/TelecomDataset1) |
| CDR 20120508 (TelecomDataset2, Day 2) | 5001 rows × 10 columns | [TelecomDataset2](http://bit.ly/Telcom_dataset2) |
| CDR 20120509 (TelecomDataset3, Day 3) | 5001 rows × 10 columns | [TelecomDataset3](http://bit.ly/TelecomDataset3) |

The remaining two datasets provided namely, “cells\_geo\_description.xlsx” and “CDR\_description.xlsx” give a description of the column names in cells\_geo and TelecomDatasets respectively. Therefore, we will have no data collection since we have been provided with the datasets.

1. **Data Preparation**

After describing the data, we cleaned, explored, and verified its integrity/quality for analysis. We started by;

* Splitting the column in cells\_geo dataset to the respective columns.
* Identifying and standardising misspelt column names in TelecomDataset, TelecomDataset2 and TelecomDataset2.
* Merging TelecomDataset, TelecomDataset2 and TelecomDataset3 into one DataFrame named df\_threedays.
* Merging the cells\_geo dataset to df\_threedays DataFrame and creating a new DataFrame with all datasets provided for use in analysis

The main tool we used for data cleaning and analysis was Python Pandas using a Google Collaboratory Notebook. See Section 1.0 Data Preparation in the Google Collaboratory Notebook.

1. **Analysis**

From the datasets provided by MTN, we analysed the dataset using descriptive statistics using Pandas to derive information from the data. Additionally, we answered the questions posed in the business understanding section in order to identify the gaps that MTN needs to fill in order to successfully upgrade its infrastructure.

Here is the link to the copy of a google collaboratory notebook on Github that contains the data preparation and analysis ;

<https://github.com/davidmuna/Moringa_Data_Science_Prep_W3_Independent_Project_2020_07/blob/master/Moringa_Data_Science_Prep_W3_Independent_Project_2020_07_David_Muna_Python_Notebook_.ipynb>

1. **Recommendation**

The results obtained from the analysis are in line with my expectations. The cities which were the most used in the three days given were Cocody, Yopuogon, Abobo and Marcory in that order. This means that Cocody was the most used city and this is the city that MTN should focus on primarily. Additional telecom products by MTN should be first introduced to Cocody followed by the next cities in terms of the most usage statistics.

This is because, uptake of new technological products, services and improved infrastructure will first be absorbed/adapted by the users who require it the most.

Additionally, cities with lower use cases need to be targeted as well because improving infrastructure in those areas will lead to acquisition of new clients who inadvertently lead to more sales and profits for MTN.

Finally, MTN Cote D’Ivoire must focus on upgrading its data infrastructure as seen from the results in the link to the analysis above under section 5.0 showing the different parameters analysed for voice, data and sms, indicating that MTN’S data has a small client base and needs an upgrade.

1. **Evaluation**

In this stage, we took a look at the generality of the analysis and the accuracy and/or degree to which it met the expectations of the business success criteria. From the results obtained from the analysis as seen from the link above, we can conclude that the model is approximately 70% accurate and this brings the need for MTN Cote D’Ivoire to continuously improve their pipelines for data collection and storage.

Improving the pipelines for data collection reduced the time taken to clean the data and verify its integrity. This gives more time for analysis and generation of better analysis models/tools thereby increasing the accuracy of the newly formulated model and business success criteria for approval.