**DATA REPORT FOR MTN COTE D’IVOIRE**

MTN Cote d'Ivoire would like to upgrade its technology infrastructure for its mobile users in Ivory Coast, and therefore needs information concerning the use of their services throughout the various regions.

1. **Business Understanding**

The business objective is to determine which cities in Ivory Coast use most of the services provided by MTN.

The villes with the highest voice usage are Cocody, Yopougon and Abobo.

The most used city for the 3 days was Cocody for voice and Yopougon for sms

1. **Data Understanding**

MTN provided various files with the necessary data to undertake this project. The information was sample data collected over the course of 3 days from different cities in Ivory Coast.

A data file for each individual sample day was provided(highlighting the amount of time spent on voice, sms and data), a file containing each city and it’s information, and a Call Detail Record that details the customer information and their respective bills depending on their usage of MTN services in the 3 days.

The data collected had some missing values and some names of columns were misspelt.

1. **Data Preparation**

Using a colaboratory notebook, I uploaded all of the provided data files and loaded them into the notebook.

The columns in the ‘geo’ file were all joined so I had to separate them.

Some of the columns were incorrectly named so I corrected them.

After making the necessary corrections, I previewed the geo data, then proceeded to make concise summaries of the newly arranged data.

In the day 1, 2 and 3 files, we changed the datetime column from string to datetime format, and introduced an hour column in each dataset.

In the day 1 data file, a column ‘PRODUCT’ was misspelt as ‘PRODUTC’ so that was rectified.

In the day 2 data set, the columns were changed from ‘DATE\_TIME’, ‘DW\_A\_NUMBER’ and ‘DW\_B\_NUMBER’ to ‘DATETIME’, ‘DW\_A\_NUMBER\_INT’ and ‘DW\_B\_NUMBER\_INT’ respectively.

In the day 3 data set, the column "SIET\_ID" was changed to "SITE\_ID", column "DATE\_TIME" was changed to "DATETIME and "CELLID" was changed to "CELL\_ID".

The average billing price per day per product was determined, as well as the total billing price per day per product.

The dataframes for each day were joined so as to get the average billing price per product over the 3 days and the total billing price per product over the 3 days.

An inner join is then done by the site ID column to get rid of the null SITE ID values.

After joining the dataframes, I was able to then get the most used cities for the three days, the most used cities for each of the three days, and the product use for the three days.

1. **Analysis**

The top 3 most used cities for the total of the 3 days were: YOPOUGON with 39837, COCODY with 36023 and ABOBO with 23260.

The top 3 most used cities for each day were as follows:

Day 1:

* 2012-05-07 - YOPOUGON - 14518
* 2012-05-07 - COCODY - 10787
* 2012-05-07 - ABOBO - 8099

Day 2:

* 2012-05-08 - YOPOUGON - 11640
* 2012-05-08 - COCODY - 8758
* 2012-05-08 - ABOBO - 7609

Day 3:

* 2012-05-09 - COCODY - 16478
* 2012-05-09 - YOPOUGON - 13679
* 2012-05-09 - ABOBO - 7552

1. **Recommendation**

MTN should now have the necessary information to guide their infrastructure updates, since they should now have a clear view of where most of their customers are from within Ivory Coast, and what times people are mostly inactive so as to not disrupt customer experience.

The code used for this process can be accessed on git-hub, using the following link: https://github.com/Geoffrey-Chege/MORINGA-IP-W3-GEOFFREY-CHEGE.git