**DATA REPORT**

The Collab notebook is both on the GitHub and Google Drive. On the GitHub it can be accessed via this link: <https://github.com/allanstar-byte/IP_WEEK_3.git>

**BUSINESS UNDERSTANDING**

* Business Objectives

MTN Cote d'Ivoire would like to upgrade its technology infrastructure for its mobile users in Ivory Coast. The insights from the data will then be reported to the management for decision making.

* Assessing the Situation

The data available is structured data(Tabulated Data) and unstructured for example Cells\_geo.csv dataset, which can be used to draw business insights.The data is having a risk of the biases of different cities, since we don't know the level of loyalty of the Individuals in the different regions.The risk can be mitigated by assigning an error value upon modeling for predictability.

* Determining Data Mining Goals

1.Identifying high-value cities based on recent data.

2.Building a model using available data to predict the convergence for each city

3.Assigning each city a rank based on both Population and region.

**Data Understanding**

* Describing data

The amount of data available is big enough to be modeled. since, large datasets can produce most accurate models.The datasets comprise both numeric, objects and string formats of data, of which our datasets contains

* Data exploration

The availability of for example, city, time, region, date seem promising for further analysis.

* Verification of data quality

The dataset contains null values and no duplicated values in our dataset.

There are columns which are wrongly labeled

**Data preparation**

* Data cleaning

Missing data - Exclude rows or characteristics, fill blanks with an estimated value.In our dataset I have dropped the null values for consistency

Data errors - I UseD logic to manually discover errors, replace, exclude characteristics and merge the datasets.

Coding inconsistencies - I decided upon a single coding scheme, of dropping all the missing values

* Integrating Data

1. Merging data - this involves merging the three datasets to form one big dataset.For example, cells\_geo.csv

2. Appending data - involves integrating two or more data sets with similar attributes but

different records. The data is integrated based upon similar fields.For example, Telcom\_dataset.csv and Telcom\_dataset2.csv

**Conclusion**

1. The most used cities in the three days are Kimbirilanord, Montcoupe,Naniefongo, Nambengue, Namassi
2. The most used cities during business and home hours are Plateau, Marcory, Bianouan, Biankouma
3. The most used city for the three days was Kimbirilanord.
4. The most used zone was E - SUD-OUEST
5. The most used region was ZANZAN
6. The most used product was sms
7. Sms and Data were mostly used in ZANZAN
8. Longitude is negatively skewed (-0.763337) while latitude is positively skewed(1.0973)

**Recommendations**

I recommend that:

-To put more efforts on advertising on the usage of sms and data which are showing great consumption and as well do a good product mix to improve on the consumption of voice calls

-To focus on the ZANZAN region which is showing a great consumption of both data and sms thus promising a great profitability margin.

-To focus on the top consuming cities such as Plateau, Marcory, Bianouan, Biankouma and do intense advertisement on the lagging cities.