**GIT-HUB LINK-** <https://github.com/RaymondKiplangat/WEEK-3-IP.git>

**DATA REPORT**

**1.BUSINESS UNDERSTANDING**

**Introduction.**

MTN Côte d’Ivoire (MTNCI), a subsidiary of South Africa-based MTN Group, is an Ivorian telecommunications company that provides postpaid and prepaid GSM services including fixed line, mobile, and internet services. (Invested in 2007.)

**Objectives.**

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?

**Requirements, Assumptions and Constraints**

**Requirements-**To complete these project approximated time is 3 weeks .(my case)

**Constraints-**No constraints at all.

**Success criteria**

Best of my knowledge the analysis i did is the best.and can be used to implement the project.

**Project plan.**

In regard to this project in followed the CRISP DM methodology to do my analysis and research concerning this project that is:

1. Business understanding
2. Data understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Deployment

**2.DATA UNDERSTANDING**

I used different types of data to do my project,links for my data are listed below.

Cells\_geo\_description.xlsx:<https://bit.ly/2RXVE0R>

cells\_geo.csv :<https://bit.ly/33NjIGt>

CDR\_description.xlsx :<https://bit.ly/3bsTptn>

CDR 20120507 :<http://bit.ly/Telcom_dataset1>

CDR 20120508 :<http://bit.ly/Telcom_dataset2>

CDR 20120509:<http://bit.ly/Telcom_dataset3>

**Data quality.**

In terms of completeness I can say the data was not complete(approximately 96% complete),for sure some of the values were missing and others were repeated(duplicates).

but all the data were correct for this project(mend for this project)

To achieve a better quality in future projects I will recommend the collection of data for areas that were missing and also duplicated areas.

**3 .DATA PREPARATIONS**

I can say this is the stage that matters the success of any project. For my case there are many techniques I used to clean my data.

**3.1 loading my data.**

I first decided to load all my dataset so that I can easily identify the errors.in the dataset and indeed there many errors in data examples are:

1.some of the columns were not written correctly eg. PRODUTC instead of PRODUCT.

2.some of the columns were not matching the columns of the other dataset.eg

In the other dataset the column was DW\_NUMBER and in the other was DW\_NUMBER\_INT so in this case I am required to correct so that the tables can merge during merging.

3. Another column was written SIET\_ID instead of SITE\_ID and the other was SITEID

**3.2 cleaning my data.**

I did all the data cleaning process listed above, and also I went ahead to remove the NaN values .

Some of the columns in my dataset were not required e.g. the COUNTRY\_A and COUNTRY \_B columns . so i drop them

I had to check the duplicated rows and columns and drop them as well.

**3.3 integrating my data.**

After data cleaning I felt that my data was good and without errors .So I went to merge them all.

So I decided to merge my dataset in two steps .

1.merging all the similar data sets together (that's Telkom 1,Telkom 2 and Telkom 3)

2. The second merging was Telkom1,2,3\_merged with the csv\_geo.

The merging is very clear in my python notebook.

After integrating the data was good for analysis.

**4.DATA ANALYSIS**

In data analysis stage i break down in four steps:

1.product count

2.relationship between product and value.

3.relationship between product value and cities

4.relationship between time and product count.

**1.product count**

To come up with the number of people using the product I decided to run a product count because there was a column for DW\_NUMBER. This simply means if i count the product it's just equivalent to the number of people using that product.

And it is clear that ;

1. Sms
2. Voice
3. Data

This simply means that many people use sms ,voice and data in that order.

**2.relationship between product and value.**

I also did analysis on the relationship between product and value .and the results i came up with is as show below:

1. Voice
2. Data
3. sms

This means that in terms of value the voice is high followed by the data then sms.

**3.relationship between product value and cities**

In order to come up with the cities that are using the product and also to check how they are doing in terms of value i run a group by check and come up with the product value with each city.

And some of the cities that i listed in the recommendation section require some upgrade

These cities are:

1.ZUENOULA

2.ZOUKOUGBEU

3.ZON-HOUNIEN

For the company to do well in all the products I will recommend checking this analysis of cities,product and value .It gives a good idea on what is supposed to be upgraded in each city.

**4.relationship between time and product count.**

In this section of analysis I come up with the time and product count and realise that most of the people use these products during the night.

The analysis on this is very clear on the last part of the notebook in the link above.

**5.RECOMMENDATION**

Something small I coil recommend is that upgrades should be done in the cities with the low usage so that they can also enjoy the service. Some of this cities with low usage are:

1.ZUENOULA

2.ZOUKOUGBEU

3.ZON-HOUNIEN

I also recommend that the MTN company upgrade the SMS and VOICE to enable them to return more value because many people are using SMS but return low value compared to the voice.

**6.CONCLUSION.**

Have done a brief comparison between the product , product and value,cities and product and values .

In my conclusion and opinion is that many people use SMS but in terms of value voice returns a lot of money so to increase the profit to the MTN company they should upgrade the SMS and VOICE to enable them to return more value .

To bring out and see the relationship between those three things .Kindly check the relationship of all the three in python notebook,link to my github repository above .

**GIT-HUB LINK-** <https://github.com/RaymondKiplangat/WEEK-3-IP.git>

.