**Follow this link to access my GitHub repository:**

[**https://github.com/CindyMG/week3IP**](https://github.com/CindyMG/week3IP)

1. **BUSINESS UNDERSTANDING**

**Business overview**

MTN Cote d’Ivoire is one of Africa’s leading telecommunications companies. For years this company has been facilitating mobile transmissions, radio, and television broadcasts across Cote d’Ivoire and has gained a huge market where 80% of mobile users and those who own radios and televisions utilize their telecom services. In addition, their profits have been steadily increasing through the years as a result of their wide clientele. However, as of last year, the reputable company has been facing competition from new but successful telecom companies which have implemented modern and efficient telecommunications resources and technologies for mobile users such as 5G networks, which are taking the world by storm.

As a data scientist here at MTN Cote d’Ivoire, my team has been tasked with collecting, cleaning, and analyzing all possible datasets concerning changing technology trends as well as customer needs and preferences to ensure that this company maintains its well-established reputation and prove to the world that it deserves the title of ‘leading telecommunication company in Africa.

**Business objective.**

* Use datasets to come up with an effective plan to upgrade MTN Cote d’Ivoire’s technology infrastructure for mobile users so as to meet modern tech standards as well as widen our customer base across the major cities.

**Business success criteria.**

* Use data to map out the major cities with the highest percentage of mobile users across the country and come up with an effective and upgraded technology infrastructure strategy for mobile users hence widening the company’s market and boosting profits.

**Assessing the situation.**

* Resources - The following datasets are available for analysis and interpretation by the company’s data scientists to accomplish the main objective stated above:

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)

* Assumptions:
* The datasets across the three days are accurate.
* Constraints:
* There are no constraints.

**Data Mining goals.**

* Identify the top-most used cities for the three days worth of datasets.
* Compute the cities that were most frequently used during business and home hours.
* Identify the most used city for the three days.

1. **DATA UNDERSTANDING**

**Dataset overview and description.**

- cells\_geo\_description.xlsx [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing) (These datasets contain column names on various data such as cities, latitude, longitude as well as a description of the data outlined and the format of the data; for example, string or float)

- cells\_geo.csv [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing) (These datasets contain various data such as cities, latitude, longitude, and area names)

- CDR\_description.xlsx [[Link]](https://drive.google.com/open?id=1cVoNXl25IO5-_yQk97ThdeqhE6yw8YTD) (These datasets contain column names on various data such as data\_time, cell\_ID, site\_ID as well as a description of the data outlined and the format of the data; for example, string or integer)

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

- CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2)

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

**Verifying data quality.**

* Some of the datasets had missing or inaccurate data values.

1. **DATA PREPARATION**

**Loading data.**

Before starting to work on the program, I had to upload the outlined CSV and excel files from the internet into the Python notebook and create an SQLite database from them.

**Cleaning (splitting, merging and dropping columns)**

I noticed that the table ‘cells\_geo’ had merged values hence I had to split the various data categories into different columns for easier viewing and interpretation.

1. ANALYSIS.
2. **RECOMMENDATIONS.**

* MTN Cote d’Ivoire as a telecommunication company should emphasize on the densely populated cities and strive to set up more network boosters in the said ideas so as to improve its services to the public.

1. **EVALUATION.**

* The highly populated cities indeed consumed most of the telecommunication services provided by MTN Cote d’Ivoire hence should be of priority when it later comes to setting up more network boosters.