This is the project of MTN Cote d'Ivoire Telecom.

In this project, we are required to analysis data with **CRISP-DM** process. The CRISP-DM process is below.

#### CRISP-DM (Cross-Industry Standard Process for Data Mining)

* Business Understanding
* Data Understanding
* Data Preparation
* Modeling
* Evaluation
* Deployment

So, in this kernel I analysis data with following this process.

## Business Understanding

Overview

The telecommunications industry is made up of cable companies, internet service providers, satellite companies, and telephone companies. Telecommunications is defined as communicating over a distance. The industry’s origin can be traced to postal courier services.

Postal courier services were used primarily to communicate with the armed forces. Over time, new communication methods, such as the telegraph and the telephone, came into being. Soon, important operating companies such as Bell Telephone Company and American Telegraph and Telephone Company (AT&T) were formed.

The entire telecommunications sector has evolved from just a few large players in the market to a much more decentralized market. In addition, the sector has so many functions involved, like service, hardware, and software, that have opened up many business opportunities.

Objective

Our main objective is to get methods of how to go about the upgrade of its infrastructure strategy within the given cities.

Fifth generation mobile networks (5G) are now poised for field testing and launch worldwide. The technology unlocks unprecedented potential to build seamless digital ecosystems, reshaping the way citizens live, work, and interact.

An equally transformative moment is coming with 5G, but with two important differences. First, the economic stakes are potentially much higher, where connected devices, applications and business models could dramatically stimulate economic productivity. Second, the United States is not as well prepared to take full advantage of the potential, lacking needed fiber infrastructure close to the end customers (deep fiber).

Business Success criteria

To discover how go about the upgrade of its infrastructure strategy within the given cities?

Assessing the Situation

1. **Resource Inventory**
   1. Datasets:
      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 20120507 [[http://bit.ly/TelecomDataset1]](http://bit.ly/Telcom_dataset1)
      4. CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2)
      5. CDR 20120509 [[http://bit.ly/TelecomDataset3]](http://bit.ly/Telcom_dataset3)
   2. Software( Github, Google Collaboratory, SQLite)
2. **Assumptions**
   1. The data provided is correct and up to date
3. **Constraints**
   1. There are no constraints

Data mining goals.

This project describes the concepts of data mining and their synergy with manufacturing environments. A generic process is introduced, which outlines data mining goals and techniques, supported by example scenarios. Various applications of manufacturing environments are shown in which data mining has been applied to successfully, and potential areas in which the outlined mechanisms are capable of being applied.

For this project, we are using the availed dataset by the company. These datasets are

* Cells\_geo\_description.- This dataset gives the glossary decription
* Cells\_geo.- This dataset show the population of company assets in regions.
* CDR 20120507,CDR 20120508 ,CDR 2012050 -These dataset describes the customer data for the company

Data Description/Understanding

We have three datasets available for this project representing the three business working days for the company. A detailed description of the datasets is provided as follows:

1. **Cells\_geo\_description-**his dataset gives the glossary decription. It consists of two columns; ***Column name*** ,Description and The data type***.***
2. CDR 20120507,CDR 20120508 ,CDR 2012050**-** This dataset, on the other hand, focuses on the company customer data. It contains the areas,regions,cities,lonituge,latitude where the company offers its services.
3. cells\_geo\_description.xlsx [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing)
4. cells\_geo.csv [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing)
5. CDR\_description.xlsx [[Link]](https://drive.google.com/open?id=1cVoNXl25IO5-_yQk97ThdeqhE6yw8YTD)
6. CDR 20120507 [[http://bit.ly/TelecomDataset1]](http://bit.ly/Telcom_dataset1)
7. CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2)
8. CDR 20120509 [[http://bit.ly/TelecomDataset3]](http://bit.ly/Telcom_dataset3)

Data in these files were stored in form of tables,columns and rows.

### **Verifying Data Quality**

Some of the data had missing values hence data cleaning.

## **Data Preparation**

These are the steps followed in preparing the data

#### **Loading Data**

Loaded the datasets from the CSV and then created an SQLite database from them.

#### **Cleaning Data**

Deleted a column that wasnt really necessary

Sorted the data by date

## **Recommendations**

From our analysis, We can see that some Cities lacked service,while other has low service coverage.These are the areas where the company should come in for development.