**Elizabeth Josephine**

**Data report for MTN Cote d'Ivoire tech infrastructural upgrade strategy**

In this data report will be a [**link** to my **git repository**](https://github.com/alfafimel/ELIZABETHJOSEPHINE-IPWK3.git) with the necessary python codes. These **python codes** will help in understanding how the analysis was done and prompt recommendations made. In a bid to **determine the best strategy for the telecom company to make the necessary upgrades**, I will use the CRISP-DM methodology to break down the processes and possible strategies. It is essential to note that we are dealing with big datasets which will make the analysis more accurate though very vital. Notably, the research question in this project is **“how mtn should go about the upgrade of its infrastructure strategy within the cities.”**

1. **Business understanding**

The first stage of every analysis requires an understanding of the objectives along with an assessment of the situation at hand. This phase ensures that all the accomplishments are understood and that too from the business perspective. For this project, it is essential that I note the significant factors that are able to influence the outcome of this analysis. For example the determination of the goals of data mining, the outputs that are desired and even the ability to develop an airtight plan that will work. This plan is the key to finding out the strategies.

Being a telecom company, MTN is expected to make profits from the voice calls and SMSs. However, without a good understanding of the datasets provided for the three days, then the data is a s good as nothing. From the research question, it is clear that the organization wants to provide their consumers with the best services.

1. **Data understanding**

This phase involves the description, exploration, verification along with the making of a quality report from the data provided. From the provided datasets, it is essential that the data quality is good to avoid discrepancies and errors in the analysis stage. This stage also helps in determining the attributes of the data while understanding what these attributes actually mean. An error in the analysis stage will make it hard to know the difference between accurate and faulty results. It is also crucial to understand the usage of the MTN network by its consumers.

For example, in the Geo description data: this data contains several areas, cell ids, geographical zone thereby referred toas the decoup zone, latitude and longitudinal coordinates, regions and different zone names. This data also has the localization of the data collected with specific references to Abidjan. This is to say that the localizations are either in Abidjan or they are not. Notably, there are several sites with their site codes in the data along with the villes or rather the cities from which the data was taken. The status was also important considering they had to be either in service or out of service. Notably, all the columns were in string format except for the longitudinal and latitudinal coordinates. These two had floats as their data types.

CDR description: this dataset had ten columns that is the product column which defined the product as either voice or sms. The value for all this was the billing price which was in a different column. There was also the date time format with its format coming from the year down to the seconds. While the cell on site and country daa was provided, it was not essential for the analysis. The ids in the data were for the cells and sites in play. Also there were two columns that had anonymous phone numbers though one had the contact info for the individual whose cell id and site are given while the other was the contact for the counterparty. Notably, all the columns had strings as their data types except for the billing price value and the cell on site data which is not essential for this analysis.

1. **Data preparation**

At this point in the CRISP-DM methodology, I am expected to select, clean, integrate and construct the data that is required for analysis. The datasets provided have several columns, most of which are useless. With the construction of required data, I am able to determine which sections of the provided dataset are needed for analysis and which ones are not. This stage involves removing duplicates, data scrubbing, aggregation levels, missing values and general data cleaning. It also involves data formatting. This stage will determine the quality of the data that I will be using for my analysis.

1. **Analysis**

This phase involves the selection of modelling techniques and identifying a model that will be used in determining the appropriate strategy. Everything that needs to be done properly can only be made possible by the existence of a working strategy. The only way that is possible is when all factors have been tabled and an understanding of the provided data achieved. For this stage, a model is built and assessed all the while generating the test designs.

1. **Evaluation**

This phase of the project helps in the determination of the accuracy and essentiality of the data. It helps determine if the results obtained from the data provided, after analysis, meet the objectives of the research question. The research question here as noted is that **“how mtn should go about the upgrade of its infrastructure strategy within the cities.”**

1. **Recommendation**

Normally and for any set of data, recommendations are made for the betterment of the project or similar future projects. For this project, i would recommend that:

1. The telecom company maximise its efforts on improving the quality of:
   1. Voice calls
   2. Internet service
   3. Network coverage.

This is to enable their strategies to work.