**CRISP-DM METHODOLOGY ON THE MTN IVORY COAST MTN TELECOM LEADING COMPANY**

Using the (CRISP-DM), Cross-Industry Standard Process for Data Mining, as an industry proven way to guide data mining efforts. I have used this methodology to analyse the given question in this week's independent project which is to solve a research question for the MTN ivory coast Telecom company.

**BUSINESS UNDERSTANDING**

**1. Business overview:**

MTN Cote d'Ivoire, is a leading telecom company. They would like to upgrade their technology infrastructure for its mobile users in Ivory Coast and how to upgrade its infrastructure strategy for mobile users within given cities.

**2. Determining business objectives**

The main objectives of this report are;

1. To identify how MTN Cote d'Ivoire can go about upgrading its mobile user infrastructure within the given cities that they feel should be prioritized.
2. Improving the technology infrastructure to prevent users from leaving the network for another telecommunication line.

**3. 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\_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)
2. **Assumptions**

The data provided is correct and up to date

1. **Constraints**

There are no constraints

**4. Determining data mining goals**

Using the available data , I can give an account for;

1. Which ones were the most used city for the three days?
2. Which cities were the most used during business and home hours?
3. Most used city for the three days?

**DATA UNDERSTANDING**

**1. Data understanding overview**

For this project, we are using the availed a set of 6 datasets that are required for the analysis of the data. These datasets are;

* + 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)

**2. Collecting initial data**

The data required for the required task was given in 6 datasets.

**3. Describing data**

The Data contained wasn’t 100% understandable upon site but the dataset contained valuable information.

**4. Exploring data**

The data given contained cities, time zones, latitudes, longitudes,Data types, Voice or

Sms products and villies.

**5. Verifying Data quality**

The data wasn’t completely readable upon site, some tables had clumped values. Other tables were not completely understandable columns and values.

**DATA PREPARATION**

**1. Selecting data**

I loaded the datasets from the CSV files provided.

**2. Cleaning data**

I didn’t clean the data , but I identified various errors located in the data , especially in the column names

**3. Constructing new data**

By merging 2 tables in calculating the number of business hours and home hours in the dataset, I was able to construct new data used in analysis.

**4. Integrating data**

I was able to manipulate the data to my advantage.

**ANALYSIS**

I was able to tabulate three key things that will assist in solving the problem stated.

I was able to :

* Identify the cities most used in three days.
* Identify the city where the network is mostly used during business and home hours.
* Identify the most used city for the three days.

**RECOMMENDATIONS**

I had a tough time analysing the data put it in order to enable me to manipulate the data

*Recommendation*

The data should be collected in an orderly format by introducing appropriate columns and rows for the data and ensuring that the data is precise, easy to read and understand.