**1. BUSINESS UNDERSTANDING**

**Business Overview**

MTN Cote d'Ivoire is a leading telecom company, which would like to upgrade its technology and infrastructure for its mobile users in Ivory Coast.

**Business Objective**

Improve the network area coverage of the MTN Cote d'Ivoire throughout the Ivory Coast.

**Assessing the Situation**

1. Resource Inventory.
   * cells\_geo\_description.xlsx [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing)
   * cells\_geo.csv [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing)
   * CDR\_description.xlsx [[Link]](https://drive.google.com/open?id=1cVoNXl25IO5-_yQk97ThdeqhE6yw8YTD)
   * 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)
2. Software:
   * Python
   * Using libraries: pandas and NumPy
3. Assumptions
   * The data provided is accurate.
   * The data provided is up to date.

**Procedure Plan**

The procedure plan is as follows:

| **Procedure** | **Duration** |
| --- | --- |
| Importing the libraries to be used | 11:31 AM |
| Loading the datasets to be used | 11:35 AM |
| Verifying the data integrity | 11:40 AM-12:00 PM |
| Data Cleaning | 12:00 PM - 2:00 PM |
| Data analysis | 2:00 PM - 4:00 PM |

**Business Success Criteria**

Compiling a list of cities, where the technology infrastructure needs to be upgraded the most.

**2. DATA UNDERSTANDING**

**Describing the resource inventory**

1. cells\_geo\_description.xlsx [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing)
   * This dataset describes the fields in the cells\_geo.csv [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing) dataset.
2. cells\_geo.csv [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing)
   * The description of this dataset is as shown below:

| **Column name** | **Description** | **Format** |
| --- | --- | --- |
| VILLES | City | String |
| STATUS | In Service or not | String |
| LOCALISATION | in ABIDJAN or not | String |
| DECOUPZONE | Geographical Zone | String |
| ZONENAME | Name of Zone | String |
| LONGITUDE | Longitude | Float |
| LATITUDE | Latitude | Float |
| REGION | Region | String |
| AREA | Area | String |
| CELL\_ID | ID of the cell | String |
| SITE\_ID | ID of the SITE | String |

1. CDR\_description.xlsx [[Link]](https://drive.google.com/open?id=1cVoNXl25IO5-_yQk97ThdeqhE6yw8YTD)
   * This dataset describes the fields in the: CDR 20120507 [[http://bit.ly/TelecomDataset1]](http://bit.ly/Telcom_dataset1), CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2) and CDR 20120509 [[http://bit.ly/TelecomDataset3]](http://bit.ly/Telcom_dataset3) dataset.
2. CDR 20120507 [[http://bit.ly/TelecomDataset1]](http://bit.ly/Telcom_dataset1), CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2) and CDR 20120509 [[http://bit.ly/TelecomDataset3]](http://bit.ly/Telcom_dataset3) dataset.
   * These are the datasets collected in the course of three days.
   * The description of these datasets are given below

| Column name | Description | Format |
| --- | --- | --- |
| PRODUCT | Voice or SMS | String |
| VALUE | Billing price | Integer |
| DATE\_TIME | Time in format yyyy-MM-dd hh:mm:ss.0 | String |
| CELL\_ON\_SITE | Which cell in the site was used (not needed here) | Integer |
| DW\_A\_NUMBER\_INT | Anonymized phone number of the person for which the CELL\_ID and SITE\_ID are given | String |
| DW\_B\_NUMBER\_INT | Anonymized phone number of the counterparty | String |
| COUNTRY\_A | Country of party A (useless here) | String |
| COUNTRY\_B | Country of party B (useless here) | String |
| CELL\_ID | ID of the cell | String |
| SITE\_ID | ID of the SITE | String |

**Data exploration**

* The column name ‘Product’ has been misplaced in the telcom\_dataset1.
* The column name ‘SITE\_ID’ has been misplaced in the telcom\_dataset3
* The Datetime column is a 24 hrs manner.

**Verifying data integrity**

1. Cells\_geo dataset
   1. According to analysis, the following columns had null values with:
      1. STATUS with 67 null values
      2. ZONENAME with 6 null values
      3. AREA with 23 null values
   2. None of the records had duplicate values.
2. Telcom\_dataset1
   1. According to analysis, the following column had null values in it:
      1. SITE\_ID with 716 null values.
   2. The number of duplicated records is 82.
3. Telcom\_dataset2
   1. According to analysis, the following column had null values in it:
      1. SITE\_ID with 676 null values.
   2. The total number of duplicates is 78.
4. Telcom\_dataset3
   1. According to analysis, the following column had null values in it:
      1. SITE\_ID with 605 null values.
   2. The total number of duplicates is 93.

**3. DATA PREPARATION**

The following datasets are to be used for this project:

1. cells\_geo.csv [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing)
2. CDR 20120507 [[http://bit.ly/TelecomDataset1]](http://bit.ly/Telcom_dataset1)
3. CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2)
4. CDR 20120509 [[http://bit.ly/TelecomDataset3]](http://bit.ly/Telcom_dataset3)

**Data Cleaning**

1. Cleaning the cells\_geo dataset
   * Since it has only records that contain null values, I’ll drop all of them since they will hinder the analysis process.
   * Filling null values in the status column with ‘Not in service’
   * Then came up with a new dataset known as cleaned\_cells\_geo
2. Cleaning the telcom\_dataset1
   * I’ll drop all the records containing null values and duplicated records, to make the analysis process easier.
   * For telcom\_dataset1, I changed the column name from ‘PRODUCTC’ to ‘PRODUCT’ to make the merging process easier.
   * For telcom\_dataset3, I changed the column name from SIET\_ID to 'SITE\_ID’ to make the merging process easier.
   * Dropped the column ‘CELL\_ON\_SITE’ on each dataset since it isn’t required.
   * Then came up with new datasets, which are:
     1. cleaned\_telcom\_dataset1
     2. cleaned\_telcom\_dataset2
     3. cleaned\_telcom\_dataset3

**Data Integrating**

* I merged the datasets: cleaned\_telcom\_dataset1, cleaned\_telcom\_dataset2 and cleaned\_telcom\_dataset3 into one dataset known as combined\_telcom\_dataset.
* After that, I merged the combined\_telcom\_dataset with the cleaned\_geo\_dataset.

**4. DATA ANALYSIS**

1. I’m going to work with the region where the localization is Abdijan, since it’s located in Côte d'Ivoire.
2. Determining the most used city/cities for the given three days. This is determining the cities with the maximum billing value in the past three days.
   * The city is Cocody with a mean billing price of 6750.
3. Which cities were mostly used during business hours?
   * **NB:** Business hours are between 9 am and 5 pm.
   * According to the analysis, there were no records containing business hours.
4. Which cities were mostly used during home hours?
   * **NB:** Home hours are between 5:01 pm and 8:59 am.
   * The most used city is Cocody with a mean of 6750.
5. Cities that need an upgrade are located in an excel file, here is [Cities that need upgrades](https://docs.google.com/spreadsheets/d/1ef5J3s9lQx_XnDpbASzoiZPAa0xGkEkW/edit?usp=sharing&ouid=117307705693519969794&rtpof=true&sd=true)

**5. RECOMMENDATION**

The cities in this spreadsheet file ([Cities that need upgrades](https://docs.google.com/spreadsheets/d/1ef5J3s9lQx_XnDpbASzoiZPAa0xGkEkW/edit?usp=sharing&ouid=117307705693519969794&rtpof=true&sd=true)) need to be considered for technological advancement.

**6. EVALUATION**

The business success criteria have been met since, I was able to compile a list of cities, where the technology infrastructure needs to be upgraded the most.

**Appendix**

* **Villes** - cities