1. **BUSINESS UNDERSTANDING**

Business Objectives

MTN, a leading telecommunication company in Cote D'Ivoire, would like to improve its technology infrastructure for its mobile users in Ivory Coast. In this project we would like to look into which cities have the most mobile traffic so as to know where to invest most in the technology infrastructure. We want to answer questions such as which cities have had the most traffic in the last three days, which cities had the most traffic during business hours and home hours, which method of communication was most used in the three days such as sms, calls e.t.c.

Desired Outputs

We want to get relevant information on the mobile phone traffic in the cities in Ivory Coast and in discovering the areas with the most traffic. With this information we will be able to tell MTN which cities it should invest more in terms of telecommunications infrastructure. With improved infrastructure the cities with the highest traffic will be able to experience better service from their mobile provider and would allow for even more people to start using MTN as their service provider for their mobiles.

Assessing the current situation

Our resources are data files on the geographical descriptions and data files on three days of mobile phone usage in the areas. Assumptions at this point are that the data given to us are correct and relevant enough to give us the information needed for our desired outputs. The constraint at the moment is time because it is a project that has a lot of information to look into but not enough time to do it. Another constraint is that it seems there might be a problem with the dataset because the hours in which we find activity are only 2300hrs and 0000hrs.

Data mining goals

We want to obtain data that shows relevant information on mobile traffic in the cities of Ivory Coast. We should obtain information on which hours there was the most mobile traffic and which means the subscribers are using most, be it sms or voice calls over the three days provided.

Data mining success criteria

Our success will be based on whether we will gather the relevant information on mobile traffic in the cities of Ivory coast. Over the three days we should know the values of how many sms’ and voice calls took place in the cities and which had the highest values.

1. **DATA UNDERSTANDING**

Overview

We had six datasets in which one was the cells and their geographical locations with an accompanying dataset which gave a description of what the titles meant. The other important datasets were the Telcom datasets for each of the three days in question and an accompanying dataset to give the description of the titles in the Telcom datasets.

Data Description

The datasets were mainly two important ones which were the cells\_geo which gave the cell data with the accompanying geographical locations and the Telcom datasets for the three days. We were given two other datasets called cells\_geo\_description and CDR\_description which gave the descriptions needed to understand the titles of the two important datasets.

Verifying Data Quality

First off there was a correction to be made with the cells\_geo data set when it came to the spacing so that it would be able to give relevant information. There were also errors with the names of some columns and the data such as city names in the “Decoupzone” which had to be corrected. There were also some missing values in the column of areas which were replaced with null values. Longitude and latitude were also given as objects instead of float values.

1. **DATA PREPARATION**

Selecting the Data

The cell\_geo dataset and the Telcom datasets for each day were loaded together with the xlsx descriptions of the datasets.

Cleaning Data

First we began by correcting the spacing on the cell\_geo dataset so that the data would be able to make sense and relate to the titles of the columns. Next we changed the names of the columns that were misspelled and afterwards the values inside the columns that had incorrect naming such as '"""Abidjan\_EST"' to the correct "Abidjan\_EST". We then looked for unique values in the dataset in which we found that in the AREAS column there were some missing values just given as ‘ ’ which we changed to null values. Afterwards we changed the datetime column from string to datetime format. After Merging the datasets we also dropped the duplicates in the data and reset the index so as to get cleaner data.

Constructing Required Data

We introduced the hour column so as to be able to tell which hours had the most traffic.

Merging the Datasets

We merged the Telcom datasets for the three days into one dataset called all\_days and we concatenated them so that the three days follow each other from day 1 to day 3. We then merged the data sets for all the days with the geo dataset.

1. **ANALYSIS**

We came to find that the top 5 cities which had the most traffic in the three days were;

1. Yopougon.
2. Cocody.
3. Abobo
4. Koumassi
5. Adjame

On the first day the highest traffic was in Yopougon, On the second day the highest traffic was in Yopougon as well but on the last day the most traffic was in Cocody.

For each day it was voice calls that was the most used product followed by sms and last was data. When it comes to voice calls the top three cities with the most traffic were;

1. Cocody.
2. Yopougon.
3. Abobo.

**5. RECOMMENDATIONS**

MTN should improve their infrastructure in the cities with the highest traffic such as Yopougon, Cocody and Abobo where they can get a higher return on investment. With better technology infrastructures in these areas the users will be getting better services and hence will start using their mobiles more which will lead to an increase in revenue. With such high usage in these areas we can assume that these are the areas with high populations and hence a larger market to obtain, with better technologies they will be able to cater to even more users in these areas. Since the sms and data services are not being used as much as voice calls, we could recommend that they start giving promotions or discounts on these services so that the users will start using them more and they can increase their revenue. In cities such as Zaroko and Zaibo where the sms usage is very low, marketing can be targeted more towards these areas so as to create more traffic in these areas.

**6. EVALUATION**

We have been able to achieve the business objectives we set out to find such as the cities which have the most traffic during the three days and we also got which of the products provided by MTN were in most use. We were unable to get what times between home hours or business hours had the most traffic because it seems there was some trouble with the dataset or that users were only using the services around midnight which seems unlikely.