**1. BUSINESS UNDERSTANDING**

**Business Overview**

MTN Cote d'Ivoire is a leading telecom company In Ivory Coast. It would like to improve

Its infrastructure for its mobile users in Ivory Coast, through this their income rises.

**Business Objective**

The main objective of this report is to identify:

1. Which were the most used cities for the past three days from the data given.

2. Which were cities were the most using business and home hours.

**3**. which cities are not in service.

**Business Success Criteria**

1. To compile a list of cities which are not in service so as to improve their technological

infrastructure.

2. To compile a list of the most used cities from the past 3 days, to know which cities

need improvement in technology.

**Assesing the situation**

**1. Resourse Inventory**

1. cells\_geo\_description.xlsx
2. cells\_geo.csv
3. CDR\_description.xlsx
4. CDR 20120507(Dataset1)
5. CDR 20120508(Dataset2)
6. CDR 20120508(Dataset3)

**2. Assumptions**

The data provided above is correct and up to date.

**3. Constraints**

There were no constraints.

**Data Mining Goals**

The data mining goals are as follows:

* Find cities with maximum and minimum value.
* Find the hours when the value are high by sorting them both in a descending order and ascending.

**Data Mining Success Criteria**

Success will be achieved when:

* The target cities which require an improvement in their technological infrastructure have been attained.

**2. DATA UNDERSTANDING**

**Data Understanding Overview**

For this project I’m going to use the availed dataset by the company. These are:

* cells\_geo.csv
* CDR 20120507(Dataset1)
* CDR 20120508(Dataset2)
* CDR 20120508(Dataset3)

**Data Description**

* cells\_geo.csv **-** contains: cities, their status(if they are in service or not), their localization, their decoupzone, longitude, latitude, region and their site code.
* Dataset1 – contains the details for day one that is: product, value, datetime, cell on site, DW\_A\_NUMBER\_INT(Anonymized phone number of the person for which the CELL\_ID and SITE\_ID are given), DW\_B\_NUMBER\_INT(), cell ID and Site ID.
* Dataset2 – contains the details for day two that is: product, value, datetime, cell on site, DW\_A\_NUMBER\_INT(Anonymized phone number of the person for which the CELL\_ID and SITE\_ID are given), DW\_B\_NUMBER\_INT(), cell ID and Site ID.
* Dataset3 – contains the details for day three that is: product, value, datetime, cell on site, DW\_A\_NUMBER\_INT(Anonymized phone number of the person for which the CELL\_ID and SITE\_ID are given), DW\_B\_NUMBER\_INT(), cell ID and Site ID.

**Verifying Data Quality**

None of these datasets have any missing valuess.

**3. DATA PREPARATION**

1. **Loading the data**Loaded the datasets from the csv into a data frame.

**4. ANALYSIS**

From the analysis:

* 67 cities were not in service.
* SITE\_ID 2e828e607c had the highest value in day 2.
* SITE\_ID 2e828e607c had the highest value in day 3.
* /TJNe+Mmtu, /+cKZKIp41, /lrUXVGVcN, +38u2u/rfx, 09UjmmY1Ds etc have zero values in day one.
* +854AcBQT2,+laSrk7g6q, /MmplBlWjh, /0FxP3az1x etc have zero values in day two.
* +,+lafgSrk7g6q, *plBsdflWjh, udsifg*0/9 etc have zero values in day two.

**5. RECONSIDERATIONS**

* The cities with not service the technological infrastructure can be upgraded here or even be built.
* Sites with zero values technological infrastructure should be built here.