DATA REPORT

Data Scientist - > Ryan Mburu

Github Repository - > <https://github.com/mbururyan/WK3_IP.git>

Date - > 06/08/2021

# BUSINESS UNDERSTANDING

BUSINESS OVERVIEW

MTN is the leading telecommunications company in Cote D’Ivore and is in dire need of a technological and infrastructural upgrade. The telecommunications company needs to upgrade its infrastructure in the technology department, so as to provide better service for its mobile users.

BUSINESS OBJECTIVES

In order to improve the company’s infrastructure, certain objectives need to be outlined which will act as a guide:

* Upgrade the company’s technological infrastructure
* Improve the mobile subscribers’ experience when using MTN
* Give a guide that will assist the company to achieve the desired tasks
* Improve MTN’s financial profit
* Improve user satisfaction

BUSINESS SUCCESS CRITERIA

In order to achieve the desired results, data mining will be done on the reports and data sets given to us by MTN so as to derive results that will satisfy the objectives outlined above.

Since MTN is an established telecommunications company, the data reports are expected to be very extensive and data mining will come into play when seeking expected results, so as to meet the company’s demands and requirements.

SITUATION ASSESSMENT

MTN provided us with 6 data reports that we will work with when data mining, and will be covered in detail in the next section of the report.

As the Data Science team, we have our personnel and constraints that will be outlined so as to make the findings as genuine as possible

1. Personnel

* A Data Scientist (Ryan Mburu)
* Project Managers (Antonny Muiko & Nikita Njoroge)

1. Resources Inventory

* Laptop
* Github
* Python
* VS Code
* Google Collab
* Google Docs

1. Constraints

* Internet disconnection
* Numerous null values all over the tables

DATA MINING GOALS

In order to satisfy the company’s business goals, the data has to be mined and prepared. Here are the expected results when mining the provided data :

* The most used cities in three days, done via data mining
* The most used cities in business hours
* The most used cities in home hours
* Whether the majority of calls occurred in the capital of Abidjan or outside
* Gathering values of the calls/sms for financial reasons across three days
* Cleaning the data
* Gather information on which city had the most activity across the three days

DATA MINING SUCCESS CRITERIA

In order to achieve the data mining goals outlined above, certain methods have to be done :

* Methods to be used when mining the data include data cleaning, selection, sorting, and analysis of the data.

# DATA UNDERSTANDING

DATA UNDERSTANDING OVERVIEW

The data is understood to be in form of 6 tables, three of them giving communication information that occurred across the entire city in three separate days, two of them offer descriptions, and one highlighting the communications from a geographical standpoint.

The data sets provided include :

1. *Cells\_geo.csv*
2. *Cells\_geo\_description.xlsx*
3. *CDR\_description.xlsx*
4. *CDR 20120507*
5. *CDR 20120508*
6. *CDR 20120509*

DATA DESCRIPTION

We are to analyze the Company’s phone call details in a span of **three days**.

In the provided tables, I was able to analyze them and came up with a satisfactory analysis of the data and what the values represented.

In the (Cells\_geo.csv) file,

* VILLE indicate the city
* STATUS indicate if the cell is in service or not
* LOCALIZATION indicates if its in ABIDJAN or not
* DECOUPZONE indicates the geographical zone
* ZONE NAME indicates the name of the zone
* LONGITUDE indicates the zone’s longitude
* LATITUDE is the zones latitude
* REGION and AREA show the zone’s region and area respectively
* CELL\_ID represents the cell’s unique id
* SITE\_CODE represents the site

In the CDR tables: i.e (CDR 20120507, CDR 20120508, CDR 20120509) which show the details of the calls in three separate days, the columns represent :

* The nature of the communication (call or sms)
* The value of the communication,
* Date and Time of when exactly the communication occurred (yyyy, MM, dd hh: mm: ss)
* The caller’s phone number (anonymous)
* The receiver’s phone number (anonymous)
* Caller’s country
* Receiver’s country
* ID of the cell
* ID of the site

The other two tables i.e. (CDR\_description.xlsx and Cells\_geo\_description.xlsx) offer insightful descriptions of what each column in the other table means for easier mining, which I used to get the description outlined above.

# DATA PREPARATION

Here is where the data is worked on, via numerous mining techniques to be outlined.

1. I set up a GitHub repository to save the work and all necessary deliverables done
2. On the python notebook, Pandas and NumPy libraries are imported so as to provide access to manipulate the tables. Other imported libraries are DateTime and excel
3. The tables : (Cells\_geo.csv, Cells\_geo\_description.xlsx, CDR\_description.xlsx, CDR 20120507, CDR 20120508, and CDR 20120509) are promptly loaded onto the work environment
4. Data mining commences, where the first output gathered grouping the cities in the Cells\_geo table to avoid redundancy
5. Sorting occurs where the grouped cities are ordered from highest to lowest, so as to get the cities with the most activity across the three days.
6. Data merging is done to gather insight whether more activity occurred within Ivory Coast’s capital of Abidjan or outside the capital
7. Via data analysis, mathematical operations were done on the tables to get the total value of the calls and SMS.
8. The tables were merged so as to get matching values for better data mining
9. Data cleaning occurred when merging to clear NULL values
10. Data cleaning on misspelled Column names i.e CELL\_ID to CELLID and PRODUTC to PRODUCT
11. After merging, Grouping of the merged tables was done according to cities
12. Sorting occurred across the three merged and grouped tables, where they were sorted in Descending order
13. Cities that appeared most times across the three days was mined

# ANALYSIS

The overall results of the company’s desired data acquired via data mining is satisfactory, and will hopefully appeal to what the company is looking for.

The derived results are as follows:

1. The most used cities across the three days (in total) are :
2. **Cocody (1062)**
3. **Yopuogon (885)**
4. **Abobo (765)**
5. **Marcory (423)**
6. **Treichville (396)**
7. The majority of the communications using MTN was done **outside** Abidjan (2305) in the **Interiour** as opposed to within your country’s capital city Abidjan (1669)
8. The most used city in the three days is Cocody
9. The total amount of money earned from calls and sms is 383,307 XOF :

* Day 1 : 129,374 XOF
* Day 2 : 121,699 XOF
* Day 3 : 132,234 XOF

# RECOMMENDATION

As the mined data suggests, few changes need to be made so as to improve the infrastructure of MTN as a whole.

* Focus on the town Cocody as a company as it is where the majority of operations via MTN occur, and build MTN offices, etc so as to offer better services to the users, that way customers are satisfied and new users will be attracted.
* Plan your infrastructural plans majorly in your country’s capital city, Abidjan, as it is where the company will get the most revenue and many citizens who reside in the capital can afford higher rates and spend more when calling
* Focus on improving financial profits, either by higher rates or reaching more unregistered users, as the daily returns are really low compared to major telecommunications companies in Africa such as Safaricom, who rack in millions a day.

# EVALUATION / CONCLUSION

Besides a few hiccups when data mining such as failure to get activity rates during business hours and home hours, the majority of the desired objectives have been achieved by the Data Science team and presented accordingly.

The findings are deemed accurate or close to accurate as nothing is always perfect, and we are looking forward to working more closely with your telecommunication company, MTN, so as to take your company to new technological and financial heights.