

DATA SCIENCE PROJECT FOR MTN COTE-D'IVOIRE

Infrastructure Upgrade Strategy Analysis and Recommendation

BUSINESS UNDERSTANDING

General Information

MTN Cote d'Ivoire/Ivory Coast requires to upgrade their infrastructure for better user experience of it's subscribers. They have shared their dataset to myself as a data scientist for analysis and advice on strategy to undertake.

Data

Datasets below provided

- cells_geo_description.xlsx
[<https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQAQhrY9N/view?usp=sharing>]
- cells_geo.csv
[https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing]
- CDR_description.xlsx
[https://drive.google.com/open?id=1cVoNXl25lO5-_yQk97ThdeqhE6yw8YTD]
- CDR 20120507 [<http://bit.ly/TelecomDataset1>]
- CDR 20120508 [<http://bit.ly/TelecomDataset2>]
- CDR 20120509 [<http://bit.ly/TelecomDataset3>]

Requirement

- Study the provided given dataset
- Recommend the strategy for upgrading the infrastructure within the given cities

Data Description

We have 3 datasets available for this project. A detailed description of the datasets is provided as follows:

- Telecom Dataset1
- Telecom Dataset 2

- Telecom Dataset2

They have similar columns and rows. The columns are not well named so we have to name the columns well

Telecom Data Observations

Value column is of Integer data type. All other columns are strings

1. Site ID is the only column with null values across the three datasets
2. There are three unique values of product i.e Voice, SMS and Data
3. Duplicate records noted. More business understanding analysis to decide if to keep them or eliminate them needed

Actions

- a) Date_Time needs to be changed to date type
- b) Country A and Country B needs to be dropped as they are not required as per project description
- c) Since the 3 datasets have similar fields, they can be combined for ease of cleanup and analysis

Data Exploration

We have to explore the telecom so we can know which product is most used by clients for the first three days. From our data exploitation of the MTN data we have found out the following products are used in the following order respectively

- Voice
- Data
- SMS

Voice is the mostly used product from MTN

Recommendation

Results

1. Voice is the most popular product
2. 'ffa6759bb2' is the busiest cell. Expansion on this cell needed

GITHUB LINK

- https://github.com/kelvinnjunge/Moringa_Data_Science_Prep_W3_Independent_Project_2019_07_KELVIN_NJUNGE_DataReport_.ipynb.git