Friday: Independent Project - Week 3

Agenda

- **Instructions**
  - Learning Outcomes
  - Deliverables
- **Assessment**
- **Evaluation**

Instructions

During this week's independent project, you will get to test the skills that you learnt during this week's sessions. Specifically, you will get the test your understanding of the following the learning outcomes of this week.

Learning Outcomes

By the end of this week, you should be able to;

1. Recall the basics of Python programming for data science.

2. Write code and document your workflow in a programming environment.
3. Understand mechanisms for missing data, outliers and analytic implications.

Deliverables

The deliverables for this week's Independent project include:

1. Github Repository
2. Python Notebook
3. Data Report (Google Docs Document)

Assessment

Overview

In this week's independent project, you will be working as Data Scientist for MTN Cote d'Ivoire, a leading telecom company and you will be solving for the following research question.

- Currently MTN Cote d'Ivoire would like to upgrade its technology infrastructure for its mobile users in Ivory Coast. Studying the given dataset, how does MTN Cote d'Ivoire go about the upgrade of its infrastructure strategy within the given cities?

Your final deliverable will be a Data Report which will comprise the following sections;

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Analysis
5. Recommendation
6. Evaluation

You can use the CRISP-DM methodology to guide you while working on the Data Report. Your Data Report will also need to have an objective account, with insights majorly coming from the dataset. However, you can refer to external information for supporting information.

Below are some questions that can get you started;

1. Which ones were the most used city for the three days?
2. Which cities were the most used during business and home hours?
3. Most used city for the three days?
4. etc.

The telecom data provided for this project is only a sample (i.e. for only three days). The data files that you will need for this project will be as follows:

1. cells_geo_description.xlsx [Link] (https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing)
2. cells_geo.csv [Link] (https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing)
3. CDR_description.xlsx [Link] (https://drive.google.com/open?id=1cVoNXl25lO5-_yQk97ThdeqhE6yw8YTD)
4. CDR 20120507 [http://bit.ly/TelecomDataset1] (http://bit.ly/Telcom_dataset1)
5. CDR 20120508 [http://bit.ly/TelecomDataset2] (http://bit.ly/Telcom_dataset2)
6. CDR 20120509 [http://bit.ly/TelecomDataset3] (http://bit.ly/Telcom_dataset3)

You will use Python for your analysis.

Submission & Evaluation

- The submission to this week's Independent Project should be made here [Link] (https://moringaschool.instructure.com/courses/91/assignments/424) .

- Links to your Github repository should be given in your report. Your Github repository will contain your Python Notebook that will contain your data preparation and data analysis work.
- This submission will be a link to the Google Drive folder that will contain your Data Report.
  - Use the following naming convention for the above Submission:
    - Moringa_Data_Science_Prep_W3_Independent_Project_2019_07_FirstName_LastName_DataReport.
- You should also ensure that you set the sharing permissions of the folder to "Anyone with the link can view".
- Late submissions or without full viewing access and assessments that have been edited after the deadline will not be assessed.
- Do not seek to copy someone else's work while working on this Independent project. You deny yourself an opportunity to learn whenever you resolve to plagiarism.
- The deadline for this assessment is 6.00 pm on the day of the assessment.