Friday: Independent Project - Week 4

Agenda

- **Instructions**
    1. Learning Outcomes
    2. Deliverables
- **Assessment**
- **Submission**

Instructions

During this week's independent project, you will get to test the skills that you learned 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. Obtain and manipulate data from various types of databases using the SQL language.
3. Write code and document your workflow in a programming environment.
4. Understand mechanisms for missing data, outliers and analytic implications.
5. Evaluate the integrity of data by making decisions on data quality issues.
6. Perform the extraction, querying, and aggregation of data for analysis in multiple projects through common techniques and tools.

Deliverables

The deliverables for this week's Independent project include:

1. Python Notebook
2. Data Report (Google Docs Document)
3. Github Repository
4. JIRA Kanban Board

Assessment

Overview

In this week's independent project, you will be working as a data scientist working for an electric car-sharing service

company. You have been tasked to process stations data to understand electric car usage over time by solving for the following research question;

Research Question

- Identify the most popular hour of the day for picking up a shared electric car (Bluecar) in the city of Paris over the month of April 2018.

Bonus Questions (Optional)

- What is the most popular hour for returning cars?
- What station is the most popular?
  - Overall?
  - At the most popular picking hour?
- What postal code is the most popular for picking up Blue cars? Does the most popular station belong to that postal code?
  - Overall?
  - At the most popular picking hour?
- Do the results change if you consider Utilib and Utilib 1.4 instead of Blue cars?

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 get the data and the dataset description for this Independent project here [**http://bit.ly/autolib_dataset (http://bit.ly/autolib_dataset)**] and here respectively **[Link]**

(https://drive.google.com/a/moringaschool.com/file/d/13DXF 2CFWQLeYxxHFekng8HJnH_jtbfpN/view?usp=sharing) . The dataset contains data collected for a period of 9 days. The dataset may take a bit of some time to load [~ 10 minutes].

Hint:

> To compute usage, we will need to understand that we have to join successive (in time) measures/counters for a given station, as the difference will tell whether a car was picked up, returned, or nothing happened.

The CRISP-DM methodology will 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.

You can use either SQL/Python for this project.

Submission

- The submission to this week's Independent Project should be made here [Link] (https://moringaschool.instructure.com/courses/91/assignments/427) .

- This submission will be a link to the Google Drive folder that will contain your Data report and the respective Colaboratory Notebook.

- You should also ensure that you set the sharing permissions of the folder to "Anyone with the link can view".

- Links to your Github repository should be given in your report. The Github repository will contain your Python notebook.

- Share your JIRA Kanban board with your Technical Mentor by adding them to the board.

- 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 6.00pm on the day of the assessment.