Data Wrangling with Python

Project Brief

Project Deliverable

- Your deliverable will be a python notebook that will contain your solution.
- You will need to submit the shareable link to your notebook

Problem Statement

As a Data Science Consultant, you need to provide recommendations on reducing the cases of breakdown of buses in the city of New York.

The dataset you have provided comes from the Bus Breakdown and Delay system, which collects information from school bus vendors operating in the field in real-time. Bus staff that encounter delays during the route are instructed to radio the dispatcher at the bus vendor's central office. The bus vendor staff log into the Bus Breakdown and Delay system to record the event and notify OPT. OPT customer service agents use this system to inform parents who call with questions regarding bus service. The Bus Breakdown and Delay system is publicly accessible and contains real-time updates. All information in the system is by school bus vendor staff.

You start this project by reading, exploring, cleaning, and later, performing an analysis of your dataset.

You need to answer the following guiding questions as part of your analysis. Feel free to come up with many more other questions.

1. Which bus companies had the highest breakdowns?
2. What were the top 3 reasons for bus delays?
3. How many students were on the buses when they broke down?
4. Which were the most frequent reasons for bus breakdowns?
5. What were the most frequent reasons for the bus running late?
6. What was the average delay time of each reason type?

You will be required to use the following template [https://bit.ly/DataWranglingProjectTemplate] as a guide. You can delete the guiding notes.

Note that you will be using CRISP-DM as the approach for this project. We will cover CRISP-DM in the coming weeks.

Dataset

- Dataset Link: https://bit.ly/BusBreakdownDataset
- Dataset Source:
  https://data.cityofnewyork.us/Transportation/Bus-Breakdown-and-Delays/ez4e-fazm