### Overview:

### We have learned how to collect, clean, and visualize data using Python. This assignment is designed to test your capability of what we have covered. In order to finish this assignment, you have to understand some core functions widely used in data science including basic knowledge about Python programming language, how RESTful API works, and Python libraries such as Pandas, Altair, requests, and .json.

### What to Do:

1. **A code (PA3\_1.py/ipynb)** that collects data from your API-of-interest (presumably APIs related to your project) to collect data. Execution of this py code should result in creating a new dataset or additional data columns through RESTful API. In collecting the data, make sure you cast **\*at least\* 10 RESTful API calls** with **more than 150 rows to be saved** (check [this article](https://medium.com/greedygame-engineering/an-elegant-way-to-run-periodic-tasks-in-python-61b7c477b679) for planning how to make periodic, multiple API calls). The 10 RESTful API calls should use one API endpoint. If your code requires some interval between API calls (e.g., a day), you can also submit what you have collected. Related files you should submit (5%):
   1. Your code
   2. Input file (in case you are adding a new data column from existing CSV)
   3. output file (CSV)
2. **A code (PA3\_2.ipynb)** that:
   1. **Builds DataFrame** based on a dataset you made for your project or built for the PA3\_1.py assignment. Note that the DataFrame **rows should be larger than 150** and the **columns should be more than 10**. Your code
   2. Has to **ADD/MODIFY at least 5 new columns** (e.g., converting an existing quantitative column to a nominal or ordinal column, or combining two quantitative columns using weighted sum--something similar to make a new combined MPG using City MPG and Hwy MPG in our exercise). Finally, your code must
   3. **Create at least 5 different visualization layouts using Altair**. Note that this assignment is an individual assignment and even if you use the same dataset with your teammates, your visualization should be different. The layouts should be different from each other. For example, creating two histograms will not be considered to be different layouts. Submit this with your input file in the input folder. Related files you should submit (5%):
   4. Your code
   5. Input file (CSV)
   6. Output file(s) (CSV)
3. **A write-up (PA3\_report.pdf):** you should also submit a document that explains your two code files. Explain (5%):
   1. For the PA3\_1 (2%),
      1. API/CSV data you used
      2. How did you clean up and merged 10+ different API calls
   2. For the PA3\_2 (3%),
      1. Explanation of how you added/modified columns based on which reason
      2. Briefly explain the visualization results you made (for PA3\_2)
      3. Conduct at least three statistical testing methods and report.
4. **Submit the files specified in this assignment** using zip (**PA3\_yoursurname.zip**) through MyMasonPortal until next Wednesday (2022-11-13) at 11:59 pm.
5. **Strictly follow the naming convention** when submitting your work.
6. Needless to say, your code should work on GTA's computer.

Please submit your write-up to the class Blackboard Assignment section in zip format no later than the deadline. Failing at any specifications in this document will result in the deduction of your grade. We will strictly follow our late submission policy. If you have any concerns/questions about this assignment, feel free to reach out to GTA (cc Ray).