IST 652 – Scripting for Data Analysis
School of Information Studies
Syracuse University
Fall / 2021

Homework #5

Due date: October 13 / 2021

Part 1 (50 points)

Perform a search for publicly available datasets. Describe two datasets that you find interesting.
For each dataset provide the following information:
Note #1: You may not need to download the dataset to be able to describe it.
Note #2: DO NOT use datasets from **kaggle.com**
Note #3: Your selected dataset must have more than 300 records

- The URL of the dataset
- A high level description of the data contained in the dataset.
- A detailed description of the fields/columns in the dataset. (Note: some datasets have a "data dictionary" or similar document that provides this description).
- The number of entries (records, rows) in the dataset
- The size in Kbyte or Mbyte of the dataset if available.

Part 2 (50 points)

Find a publicly accessible API and build a notebook where you submit at least two different requests that result in different pieces of information (or objects) being returned from the API service. You should take each of the responses obtained from the requests and get them to display "nicely" the information contained in them by using pandas or other libraries to get the information to show in a tabular or graphical format.

Make sure you also include the following in your notebook

a. A description of the API you interacted mentioning its URL
b. The API requests that you submitted and how you submitted them

Include your notebook file (.ipynb) in your submission on Blackboard.

Note #1: Some good sources for APIs:
1. https://www.programmableweb.com/
2. https://api.data.gov/

Note #2: To prevent any loss of your work, create your notebook inside the "library" directory of your Jupyter environment