Project 2

Extract, Transform, & Load

Requirements

- Gather Data from at least 2 sources (screen scrape, API save, downloaded csv files)
  - One large source is acceptable if the transformation process is larger.
- Store Data in csv, JSON, or MongoDB
- Extract the data from the raw data and read with python / pandas
- Clean the data and transform the data to SQL database
- Create ERD for the Database (At least 2 normalizations / deduplications / joins / etc.)
- Load the transformed data into the database

Deliverables:

- 1 - Python / Jupyter Notebook to collect the data
- 1 – Python / Jupyter Notebook to transform the data and load into SQL database
  - This can be two files, one to transform data and one to load the data.
- 1 – ERD Diagram for the SQL Database
- 1 – Script to create database schema
- 1 – Documentation on how to run in the projects README.md file
- 1 – Link to the project GitHub
  - Do not include the data, only the instructions for getting the data