Section 6 Homework – Capstone development

Due: Tuesday before class (30 May, 6:30pm CT)

1. **Create a definitive problem statement to be solved with data analytics.** 
   * Your statement needs to be written.
   * The statement should be 1-3 sentences.
   * Provide a brief overview of the topic if context is necessary to help understand the problem statement.
2. **Provide your datasets**.

* In .csv / sql format
* -or- a website that will be accessed with webscraping/API. If possible, include the Python code to be used for accessing this data.
* Provide an explanation of the datasets you submit.

1. **Describe your plan** on how to use this data to answer your problem statement.

* Create a brief write-up (typically 3-5 sentences) describing your plan of what you intend to do with that data to solve your question.

**Submit your homework to git hub, to include pushing your dataset files to your repo.**

*See example of completed homework on next page…*

1. Which of the following factors most impact EV adoption?

-Income level

-Climate (average temperature)

-Tax rebates and incentives

Context: With the increase in Electric vehicle sales in the United States, clearly some states are quicker to adopt this technology than others. Why is this? Do the above factors influence purchasing of such vehicles?

The problem statement above will be addressed by considering which U.S. states have more electric vehicles registered than other states, and if those listed factors influence these patterns.

1. Datasets:

**10962-ev-registration-counts-by-state\_6-30-22.xlsx**

* “Primary” dataset
* Shows number of electric vehicle registrations by state
* <https://afdc.energy.gov/data/search?q=ev>

**EVIncentivesbyState.xlsx**

* A manually created dataset based on various state tax incentive websites

**StatePopulations.xlsx**

* Simple table showing population by state

**StateTemperatureDataset.xlsx**

* Average temperature by state
* Will be used to address the “climate” question.
* <http://www.usa.com/rank/us--average-temperature--state-rank.htm>
* Webscraping option:

import pandas as pd

import requests

url = <http://www.usa.com/rank/us--average-temperature--state-rank.htm>

r = requests.get(url)

df\_list = pd.read\_html(r.text, header=0) # this parses all the tables in webpages to a list

print(df\_list)

**Data-gKZFY.csv**

* Table showing state income levels by income threshold
* Shows what income level equates to top 1% of earners, top 5%
* <https://smartasset.com/data-studies/what-it-takes-to-be-in-the-1-by-state-2023>

**h08.xlsx**

* Table showing all states median income levels.
* This table and the above “data-gKZFY.csv” will be used to address the “income level” factor.
* <https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-income-households/h08.xlsx>

**mv1.xlsx**

* Total state motor vehicle registrations
* Gives context to number of electric vehicle registrations per total registrations (rather than comparing to total population)
* <https://www.fhwa.dot.gov/policyinformation/statistics/2021/mv1.cfm>

1. **Plan:**

This is a three-part approach.

Step 1: I will use mv1.xlsx to compare the number of registered electric vehicles to the number of total vehicle registrations. I will do this using Excel.

Step 2: Joins. To be done with SQL and/or Python.

I will join the resulting table from Step 1 with StateTemperatures to see if there is a correlation between EV registrations and temperatures (Do warmer temperatures equate to more EV sales?)

I will do the same to join Step 1 results with EVincentivesByState.

I will do an additional join using Step 1 results with h08 table data to identify correlations between EV registrations and median income.

Time permitting, I will also compare EV registrations with Top 5% income earners by state to see if that shows a different result (EVs are, in general, expensive, and using top income levels instead of median income could give a different result).

After doing the above joins, data visualization in Tableau will help identify correlations between datasets.