Peer-Graded Assignment: Milestone 1: Project Proposal and Data Selection/Preparation

Step 1: Preparing for Your Proposal

Client/Data Set

I decided to use a data set that was not provided by Coursera for the project. After some initial research, I decided to use the American Trends Panel Wave 83 data set from the Pew Research Center conducted from February 16, 2021, to February 21, 2021, which covers various topics related to coronavirus vaccines and restrictions. I thought the responses would be interesting to analyze.

Steps to Import and Clean the Data

I downloaded the data set from the Pew Research Center website and converted the file to a .csv file. I then imported the data into my notebook to view dataframes using the pandas library. The data was overall very clean, so I focused more on my initial exploration.

Initial Exploration of Data

Below are some screenshots of my initial exploration of the data.

```
[4]: df_atp.describe()
```

[4]:

	Variable	QKEY	INTERVIEW_START_W83	INTERVIEW_END_W83	DEVICE_TYPE_W83	LANG_W83	FORM_W83	XKNOWPAT_W83	FLAG_W83	SC1_W83	...	F_IDEO
count	1	10122	10122	10122	10122	10122	10122	10122	10122	10122	...	10122
unique	1	10122	9059	9037	8	5	5	5	7	9	...	13
top	Label	Unique ID	17-Feb-2021 15:20:48	17-Feb-2021 14:30:38	2	1.00	1	2	3	1	...	3.00
freq	1	1	9	11	4705	8048	4115	4492	8048	6114	...	2857

4 rows × 102 columns

```
df_user = df_atp[['F_METRO', 'F_CREGION', 'F_CDIVISION', 'F_AGECAT', 'F_GENDER', 'F_EDUCCAT', 'F_EDUCCAT2',
                  'F_HISP', 'F_HISP_ORIGIN', 'F_YEARSINUS', 'F_RACECMB', 'F_RACETHNMOD', 'F_CITIZEN',
                  'F_BIRTHPLACE2', 'F_MARITAL', 'F_RELIG', 'F_BORN', 'F_RELIGCAT1', 'F_ATTEND', 'F_PARTY_FINAL',
                  'F_PARTYLN_FINAL', 'F_PARTYSUM_FINAL', 'F_PARTYSUMIDEO', 'F_INC_SDT1', 'F_REG', 'F_IDEO', 'F_INT
                  'F_VOLSUM', 'F_INC_TIER2']]

df_user
```

[7]:

	F_METRO	F_CREGION	F_CDIVISION	F_AGECAT	F_GENDER	F_EDUCCAT	F_EDUCCAT2	F_HISP	F_HISP_ORIGIN	F_YEARSINUS	...	F_PARTY_FINAL
0	Metropolitan area indicator	Census region	Census division	Age category	Gender	Education level category	Education level category 2	Hispanic identification	Hispanic origin	Years lived in U.S. (excluding Puerto Rico or ...	...	Party
1	1	2	4	4	2.00	2	3	2.00	NaN	1.00	...	2
2	1	4	9	4	1.00	1	5	2.00	NaN	1.00	...	1
3	1	1	2	2	2.00	1	6	2.00	NaN	1.00	...	2
4	1	1	1	3	2.00	1	6	2.00	NaN	1.00	...	2
...	...	...	...	...	...	...	...	...	...	...	...	...
10117	1	2	3	2	2.0	2	4	2.0	NaN	1.0	...	4
10118	1	1	2	4	1.0	1	5	2.0	NaN	1.0	...	1
10119	1	1	2	2	1.0	1	6	2.0	NaN	1.0	...	2
10120	1	3	5	3	2.0	1	5	2.0	NaN	1.0	...	2

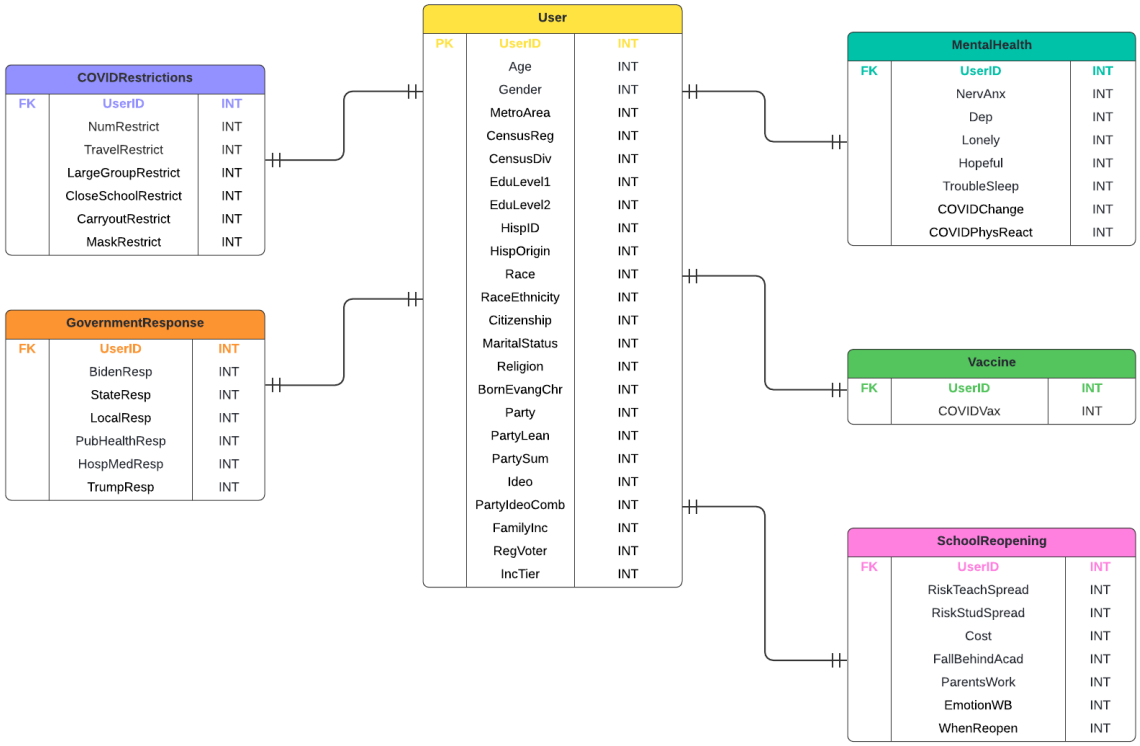
[37]:
```
df_trump_count = df_trump.groupby(['COVIDEGFPDT_W83'])['COVIDEGFPDT_W83'].count().reset_index(name="count")

df_trump_count
```

[37]:

	COVIDEGFPDT_W83	count
0	1	1509
1	2	1807
2	3	1273
3	4	5499
4	99	33

Proposed Entity–Relationship Diagram

I drew a proposed ERD using Lucidchart. Please see below.



Step 2: Develop Project Proposal

Description

I used the American Trends Panel Wave 83 data set from the Pew Research Center, which covers various topics related to coronavirus vaccines and restrictions. The data set includes questions related to schools reopening, vaccination status, COVID-19 restrictions, and respondents' mental health. I'm interested to see if the respondent demographics have any effect on their responses. This data could be important to politicians (to understand how the public feels about the COVID-19 response from the government) or other government officials. Journalists could also use this data to report findings to the public.

Questions

- Based on the results of the survey, how do Americans generally feel about how Donald Trump responded to the coronavirus (COVID-19) outbreak?
- Based on the results of the survey, how do Americans feel about schools reopening?
- Based on the results of the survey, how has COVID-19 affected Americans' mental health?

- Based on the results of the survey, what percentage of Americans are vaccinated against COVID-19?
- How do the demographics for this survey affect Americans' responses?

Hypotheses

- The majority of Americans are vaccinated against COVID-19.
- The majority of Americans are not happy with the way Donald Trump handled the response to COVID-19 during his term as president.
- The majority of Americans want schools to reopen (and in general want a return to normalcy in society).
- Americans with a lower level of education and a lower level of household income likely did not receive the vaccine for the coronavirus.

Approach

First I want to look at the respondent demographics (age, gender, education level, religious views, etc.). Then I will start to examine more specific questions related to schools reopening, the government response to the COVID-19 pandemic, and COVID-19 restrictions. I'm interested to see if respondents' religious beliefs, education levels, and levels of household income have any relation to whether they are vaccinated and how they feel government officials have responded to the pandemic. I plan to use the count aggregate function to group responses and view an overall summary of the data for each survey question. Then I would like to examine how respondent demographics may affect responses to the survey questions by filtering the data by certain criteria.