Project 2 Proposal

Mathew Joseph Jacob Willard Churmin Park

Our project is 4-fold, as there are many questions that you can ask when thinking about college. Is college worth it? Is grad school worth it? What's the gender gap actually like? With the datasets provided, we can tackle these questions.

Question 1: What is the relationship between a college major (in UC Berkeley -- the L&S Majors vs Engineering) and unemployment rate? What is the relationship between STEM and unemployment rate?

Datasets used: https://github.com/fivethirtyeight/data/blob/master/college-majors/majors-list.csv
https://github.com/fivethirtyeight/data/blob/master/college-majors/grad-students.csv
https://github.com/fivethirtyeight/data/blob/master/college-majors/recent-grads.csv

Answer: In order to attempt to answer this question, we must look at the majors listed in recent-grads.csv, and compare the STEM majors vs the rest. This will give us an idea of how many of us who are taking this class will do after college, and it is useful information in general to have.

Question 2: Is grad school worth it?

Datasets used:
https://github.com/fivethirtyeight/data/blob/master/college-majors/grad-students.csv
https://github.com/fivethirtyeight/data/blob/master/college-majors/recent-grads.csv

Answer: We will compare grad-students.csv's majors to recent grads majors. These two are fundamentally different, as getting a PhD/Masters will give many recent grads an opportunity to get a high paying job. So how much does it really help to go to grad school? We will also look for data for the cost of grad school and compare it to the median pay of grad school students. This will give us an idea if grad school is really worth it for a major, information that many people are wondering (including ourselves!)

Here are some screenshots of the data tables:

In [8]:
```
import pandas as pd

recent_grads = pd.read_csv("recent-grads.csv")
graduates = pd.read_csv("grad-students.csv")

recent_grads
```

Out[8]:

nployed	...	Part_time	Full_time_year_round	Unemployed	Unemployment_rate	Median	P25th	P75th	College_jobs	Non_college_jobs	Low_wag
976	...	270	1207	37	0.018381	110000	95000	125000	1534	364	193
40	...	170	388	85	0.117241	75000	55000	90000	350	257	50
48	...	133	340	16	0.024096	73000	50000	105000	456	176	0
58	...	150	692	40	0.050125	70000	43000	80000	529	102	0
5694	...	5180	16697	1672	0.061098	65000	50000	75000	18314	4440	972

In [9]:
```
graduates
```

Out[9]:

sample_size	Grad_employed	Grad_full_time_year_round	Grad_unemployed	Grad_unemployment_rate	Grad_median	...	Nongrad_total	Nongrad_
	7098	6511	681	0.087543	75000	...	86062	73607
	40492	29553	2482	0.057756	60000	...	461977	347166
	18368	14784	1465	0.073867	65000	...	179335	145597
	3590	2701	316	0.080901	47000	...	37575	29738
	7512	5622	466	0.058411	57000	...	53819	43163

In [ ]:

As you can see, it'll be very easy for us to use pandas to compare unemployment_rate to grad_unemployment_rate, and the median salaries. There are many useful columns here: we can take a look at all of the columns shown above.

Question 3: Is college worth it?

Dataset used:
https://github.com/fivethirtyeight/data/blob/master/college-majors/recent-grads.csv

Answer: This one is going to be a very cool one. There is a hot debate going on right now, spearheaded by Peter Thiel, asking if college is really worth it. With recent-grads.csv, we can check the numbers (College_jobs column vs Non_college_jobs column) in relation to the average pay that jobs with many more college_jobs have to the ones that non_college_jobs are high in.

Question 4: What is the gender gap like?

Dataset used:
https://github.com/fivethirtyeight/data/blob/master/college-majors/women-stem.csv

Answer: We will compare the ShareWomen data with the major category, and see what the gender gap in many professions is actually like. Looking at it right now at a glance, women are definitely underrepresented in STEM.