Possible questions:

Current Questions:

Clean the data and remove the 0 skills

Colin

Paola

Kevin

Heran

Part A - Before Merge:

1. What are the top skills required for Data Scientist jobs
   1. What skills are correlated with highest pay?
      1. Bar chart for each program grouped in salary categories
   2. Is higher number of skills correlated with higher pay? How many skills are enough?
      * 1. Is the average skill count based on the higher pay, you are an expert at certain languages? Too busy managing? More consulting?
        2. Jack of all trades vs knowledgeable
        3. Create a line graph of the average
      1. Which skills explain the largest chunk of that?
   3. What are the top locations? Heat maps
      * 1. Calculate the number of postings per state, percent\
        2. All Jobs
      1. What locations pay the most?
         1. Percentage to job title count from each state
         2. Highest paying job
   4. What industries are correlated with highest pay? (Keep the blanks)
      * 1. Do a count of each industry and calculate the percentage
        2. Pick out the top 5 and put everything else into category, explaining the everything category. Display as pie chart for each salary category
      1. Industry with the most listing?
   5. What are the top skills by industries?
      1. Sort by top earnings in each industry or top number of data sci jobs in each industry
      2. Take top 3, and look at the top skills required by each of those
2. Is there a difference between skill sets by job type? (Wednesday)
   1. See viz 1

HOW TO DO CORRELATION BETWEEN NUMBER OF SKILLS vs JOB TYPE?

Part B -  After Merge:

1. Weight the salary figures by
   1. (Job sal - mean income) / mean income =skill premium %
2. Using the Average Cost of Living, after how many months will what I’m putting in my pocket will I be able to pay off the tuition and be able  to start saving = # of months until ROI is positive

Visualization ideas:

1. Top 10 skills by job type - bar chart
   1. See Q8
2. Heatmap with concentration of jobs by
   1. Job type
   2. Salary
   3. Industry
   4. Rating?
3. Industry and

Links:

Github Project Repository Link:

<https://github.com/cce24/Project-1---Group-2>

Good example: <https://www.kdnuggets.com/2020/08/employers-expecting-data-scientist-role-2020.html>

Possible API sources:

1. Indeed:

<https://opensource.indeedeng.io/api-documentation/>

<https://opensource.indeedeng.io/api-documentation/docs/job-search/>

1. Glassdoor: <https://www.glassdoor.com/developer/index.htm>
2. LinkedIn: <https://docs.microsoft.com/en-us/linkedin/shared/integrations/people/profile-api>
3. USAJOBS: <https://developer.usajobs.gov/General/Schemas>

Job Posting Datasets:

1. [Find Open Datasets and Machine Learning Projects](https://www.kaggle.com/datasets?search=employment)
2. <https://www.kaggle.com/madhab/jobposts/data>
3. <https://www.kaggle.com/andrewmvd/data-analyst-jobs>
4. <https://www.kaggle.com/datasets?search=job+posting>
5. <https://cloud.google.com/bigquery/docs/reference/rest/v2/jobs/get>
6. <https://data.world/datasets/jobs>
7. <https://data.world/dataremixed/linkedin-top-10-skills-by-year>
8. <https://www.glassdoor.com/research/app/uploads/sites/2/2019/04/Methodology-Glassdoor-Job-Market-Report-2-2.pdf>

 Prospective Datasets List:

1.      <https://www.kaggle.com/elroyggj/indeed-dataset-data-scientistanalystengineer?select=indeed_job_dataset.csv>

2.  <https://www.kaggle.com/sl6149/data-scientist-job-market-in-the-us>

3.  <https://www.kaggle.com/andrewmvd/data-scientist-jobs>

4. <https://www.kaggle.com/andrewmvd/data-analyst-jobs>

Bonus 1#:

Presentation/Analysis: <https://www.kdnuggets.com/2020/08/employers-expecting-data-scientist-role-2020.html>

Data: <https://www.kaggle.com/vikasbhadoria/data-scientist-role-in2020>

Bonus #2:

Canadian <https://www.kaggle.com/refaee/data-scientist-employment>

Appendix 1.

Old Questions:

1. Loop through skills and certain cities
   1. How many jobs have a combination of the skills
   2. Maybe 5 skills - Python, SQL, VBA, Excel, etc
   3. Sort or limit by date posted
2. How important is a college degree?
3. Experience or academic qualification - what matters more?
4. What skills are looked for as a data analyst? 5 tops
5. What skills is trilogy teaching and how does that correlate or what is missing?
   1. What skills?
   2. How relevant
   3. Are there skills that are not being taught that should be?
   4. Are there skills being taught that should not be?
6. Where are the data analytics jobs focused in, states, cities?
   1. Tech hubs - East Coast vs Best Coast
7. Different levels of data analyst? Government/Private/Public?
   1. Federal (DC Jobs) vs creative (West Coast jobs)
8. Salaries and skills correlation

To Do:

1. Email Steffanie Sweeney about Trilogy student hiring prospects data
   1. Keywords on employer search terms
   2. Salaries?
   3. Success rate after graduation?
   4. Data?