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? - Done
      * 1. Is the average skill count based on the higher pay, you are an expert at certain languages? Too busy managing? More consulting?- Done
        2. Jack of all trades vs knowledgeable- Done
        3. Create a line graph of the average- Done
   3. What are the top locations? Heat maps- Done
      * 1. Calculate the number of postings per state, percent\- Done
        2. All Jobs- Done
      1. What locations pay the most?- Done
         1. Percentage to job title count from each state- Done
         2. Highest paying job- Done
   4. Tally of top industries in highest pay category? (Keep the blanks) - No Blanks
      * 1. Do a count of each industry and calculate the percentage - Done
        2. Pick out the top 5 and put everything else into category, explaining the everything category. Display as pie chart for each salary category - Done
      1. Industry with the most listing? - Unnecessary
   5. What are the top skills by industries? - Done
      1. Sort by top earnings in each industry or top number of data sci jobs in each industry - Done
      2. Take top 3, and look at the top skills required by each of those - Done
2. Is there a difference between skill sets by job type? (Wednesday)- Done
   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

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>

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

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>

Colin

Paola

Kevin

Heran

Presentation:

Slide

1. Introduction
   1. Context
2. Summary Stats.Describe
   1. Income
      1. Median
      2. Mean
      3. Std
   2. No of Skills
      1. Median
      2. Mean
      3. Std
   3. # of skills tally - Table
   4. Table of States
3. Viz - 3 Bar Chart - Skills by Job Type
4. Viz - Bar Chart - Which skills are in high demand?
5. Viz - Line Graph - Higher # of Skills correlate to higher salary?
6. Viz - Heat Map - of job concentration by location
   1. Table
7. Viz - Heat Map - of job concentration by location for top paying jobs
   1. Table
8. Viz - Pie Chart - Concentration of Jobs by Industry - Top 5 + Others
9. Viz - Bar Chart -Tally of skills by income bracket
10. Viz - Bar Chart -Tally of skills by income industry
11. Takeaways