Proposal Guidelines:

1. The data set that I chose was from Kaggle and it is a dataset of the past 30 years of the NFL draft. The data set primarily summarizes each draft pick from the past 30 years, where they attended school, and then their statistics and success with in the NFL for their career. It also shows their team, number of games played, and other factors that I will be able to consider when determining what their success will be. The one thing it is missing is starting salary, which is something I am going to further consider if time permits later in the project. This dataset is in one large csv with about 10,000 rows and 32 columns. I will primarily use tables, numpy, and pandas to manipulate the data and variables to further consider them to try and find trends that lead to a successful player. From the beginning, I feel that the important aspects to consider will be round drafted, by what team, salary, position, and all of the statistics for each player. The player statistics are what will be used to measure their success by looking at those numbers based on other players in a similar position and compare from there. I will be using mean, media, mode, standard deviation, and regressions to generate my statistical analysis. Once these have been computed I will use seaborn for these statistical measures and ideally use tableau for visualizations. The types of variables are either integers or strings.

https://www.kaggle.com/ronaldjgrafjr/nfl-draft-outcomes

2.

- How important is rounded drafted/salary important for the success of a player?
- Is the team that a player was drafted to important for a players success and how the team has done overall?
- Is there a correlation between the college the player attended and their success in the NFL?
- Is there a relationship between the position (QB,TE,RB, ect..) a player was drafted for and the round they were chosen on the success for a player?

3.

- I will use the integer column for round drafted and overall pick number and find the average performance for a player in a certain position to then see how successful players rank against each other. I will use pandas and numpy to create these separate datasets and then use the statistical analysis functions to see how each play ranks based on success to draft round. I will initially use the statistical visualizations we have used in class and then use tableau or Bokeh to make a better representation once I better understand the data.
- I will use the team that a player was drafted to and find additional information on that team's overall record to see the relationship between these variables. Once I have extracted the data in tables that I feel fit I will run regressions to determine how these factors affect each other. Again, I will use tableau and Bokeh.
- Similar to my first approach I will use a similar approach except instead of using round drafted, I will use college attended and compare to the success of the players.

- In this approach I will further break down the players by position and compare that to how successful a player is. In this I want to look more into if there are more successful positions in certain years and will use a similar approach as above. Many of these will build off of each other and I will continue to further build off of them once I begin to analyze the data.