This assignment is a **Group Assignment**. One member of each project group should submit the group's work as a Word document (.doc, .docx) or a PDF (.pdf).

For this assignment, I would like you to describe your initial plan for the final report and presentation. It can be early stages, and you are not locked in to what you provide in this assignment, but I would like you to think through and comment on the following questions:

1. What type of data analysis problem would you like to tackle with the project, and why? If there is a particular sport, team, city, etc. that you are interested in, include it in your answer as well.

We want to analyze and compare NFL teams/players based on offensive and defensive metrics from the past year. We want to look at the league as a whole for comparison instead of focusing on an individual team. It will be interesting to see what teams are good on offensive and defensive or just good on one side of the ball. We then can see if one strength or weakness leads to a higher winning percentage.

We can even go further once we see what teams are good/bad. We can look into individual player statistics of these teams to see where their production or lack of production is coming from.

2. What possible data sources could you use for your task?

Football reference should be able to be used to get team data year by year, player data year by year, and team/player salaries. This will be the first place we can look to gather our data.

If football reference doesn’t have the info we need then we can look on NFL, ESPN, CBS or other sports websites to scrap the data we need.

3. From what we have covered in class so far, how can you apply Python to work towards generating interesting information about the problem you defined in 1?

From our class the first skill we will use is extracting the data from either the website we choose or football reference. We can use simple extracting/exporting a csv or use more advanced web scraping tools. We will use the panda function to read in our data and save it in a way that can be analyzed.

Next we will clean up the data in a way that makes it more efficient and easier to analyze the data. Indexing and removing unnecessary columns are the first steps in this process.

We will use the functions from the numpy package to extract information on the best offensive and defensive etc using max/avg passing yards, rushing yards, touchdowns etc and the amount given up for defensive metrics. The numpy package will allow us to “rank” teams and compare metrics to each other.

We will use the packages matplotlib and plotly in order to visualize the results of our analysis.