A time in which I performed data analysis was this past semester in DSC495-011 to determine who the best QB was in the NFL during 2021. I outlined a process prior to starting, that factored in advanced analytics heavily without abandoning some of the more basic counting statistics. One of my intentions was to remove bias, so regardless of the results I stuck to the process no matter what. For the first graph, I created a scatterplot of EPA (Expected Points Added) vs CPOE (Completion % Above Expected) and subset the data frame to include only the QB’s above average in both. For the second graph, I repeated the same process but with passing yards and passing touchdowns. From the remaining subset of QB’s I applied a custom formula that was as follows: **Weighted QB Play**  = (.25 \* EPA Percentile) + (.25 \* CPOE Percentile) + (.2 \* WPA Percentile) + (.1 \* Passing Yards Percentile) + (.1 \* Passing TD’s Percentile) + (.1 \* Completion % Percentile). Percentiles were assigned to QB’s only from the final subset.

In the first graph (EPA vs CPOE), Graph 1, I was intrigued by the visuals. I was happy to get the team logos and labels working correctly while also being able to see exactly where each QB fell. Ultimately, I was disappointed that my process excluded Matthew Stafford and Justin Herbert from the final subset, but I was happy to see most of the QB’s that I expected to make the cut did. Additionally, I was not a fan of some QB’s that the fans consider game managers such as Teddy Bridgewater and Jimmy Garoppolo making the cut, so I created a supplemental graph, Graph 4, to see if swapping out CPOE for WPA (Win Probability Added) would change anything. This graph, although not used to avoid adding my bias, was able to include Stafford and Herbert but did exclude Josh Allen which I was surprised to see.

The second graph (Pass Yards vs Pass TDs), Graph 2, I opted for more basic, counting based statistics. While I had mentioned that I wanted to heavily use advanced analytics, I thought this graph could serve as a filter for what some people would call the eye test. It was also successful in removing some of the QBs that missed significant time during the season such as Tua Tagovailoa and Lamar Jackson. After this graph I was left with a subset of seven QBs and began assigning the percentiles and calculating Weighted QB Play. I created Graph 3 which showed how each QB ranked, with a horizontal bar graph, and bar colors corresponding to team colors. I was left with Aaron Rodgers as my QB1 for 2021 which makes sense given he is the reigning, back to back NFL MVP. My prediction prior to starting was that Tom Brady would be QB1, however he finished as QB4.

After reviewing my analysis, I determined that there were some ethical concerns to be addressed. This data could be used to make decisions based on how to pay a player or whether or not a team should replace their QB. This could have a significant impact on a real person’s life if that were to happen. This data was acquired legally and there are no concerns in that regard, but when analyzing someone’s performance in their profession, ethics should always be considered.

My final conclusions from this project are that I am excited to have developed a process that has some merit given that my QB1, matches popular opinion. I developed another supplemental graph, Graph 5 to determine if the subsets that I did were necessary, by applying the percentile and formula to all QB’s and came away noticing that while the top may remain the same, the formula allows for more average QB’s to sneak up the rankings, such as Garoppolo at QB5. For future analysis, which I do plan to do, I want to factor in a way for rushing statistics and interceptions to be included. I want to do this without punishing QBs who do not run too much, but award those that do since that is a big part of their game. I am very proud of this analysis, through this class and this project I have gained a lot of confidence in doing projects such as this one, on my own. I look forward to expanding upon this project, as well as creating plenty more, and sharing my analysis on Twitter and other platforms!

**Index**

Graph 1: Chart, scatter chart

Description automatically generated

Graph 2:

Chart, scatter chart

Description automatically generated

Graph 3:

Chart

Description automatically generated

Graph 4:

Chart, scatter chart

Description automatically generated

Graph 5:

Chart

Description automatically generated