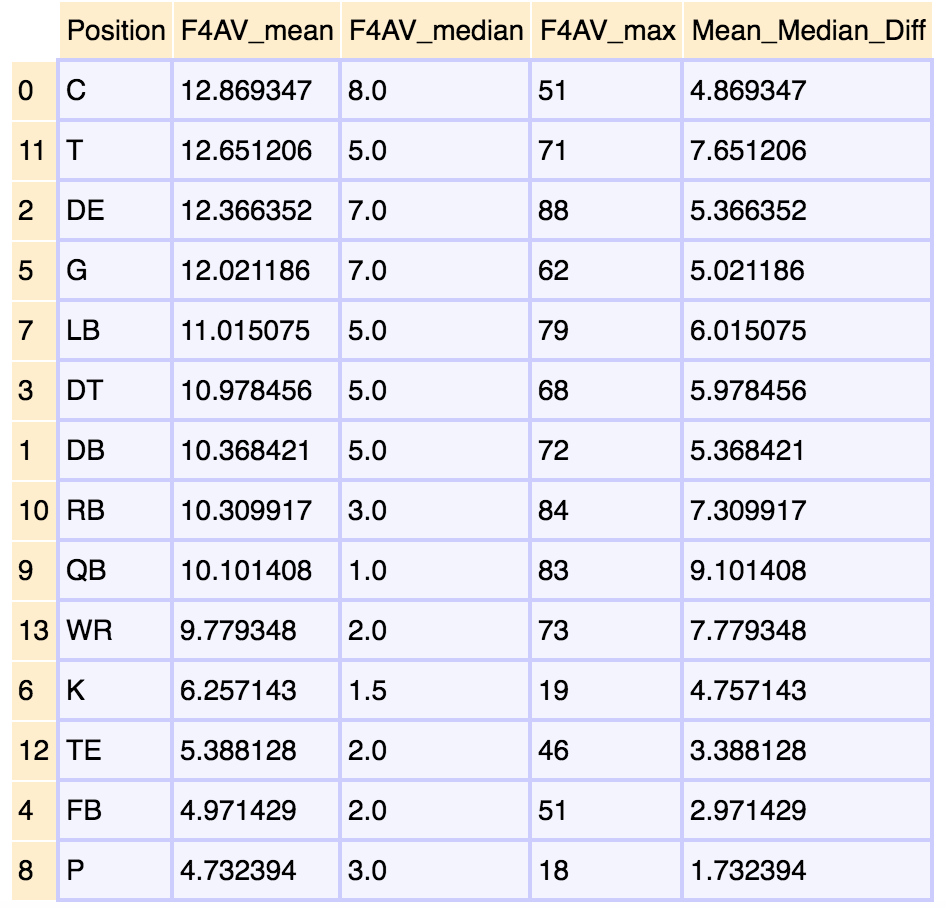
**SI 370 Final Project – NFL Data Analysis**

**Introduction**

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. Which I was unable to obtain and ended up changing the questions I wanted to answer from my proposal. This dataset is in one large csv with about 10,000 rows and 32 columns. I primarily used 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. After doing some analysis on the column First4AV, which is a number given to a player after four years in the league to measure their success. I found problems with this because I was missing data for some of the important quarterbacks drafted after 2011. Therefore, I created an additional scoring system similar to a fantasy football scoring system to rank quarterbacks. I decided to consider only quarterbacks based on conferences and draft round to see if I could find any trends to see what made for a successful quarterback in the NFL. The scoring system is calculated by the average points per game for a quarterback. In the system, I multiplied their pass completion percentage by 10, multiplied the pass yards per game by 2.5%, gave 6 points for each type of touchdown, subtracted 3 points for each interception, and then multiplied each rushing yard or receiving yard by 10%. I performed much of my analysis using joint grid plots, pivot tables, linear regressions, box plots, bar graphs, and some clustering visualizations.

**Questions**

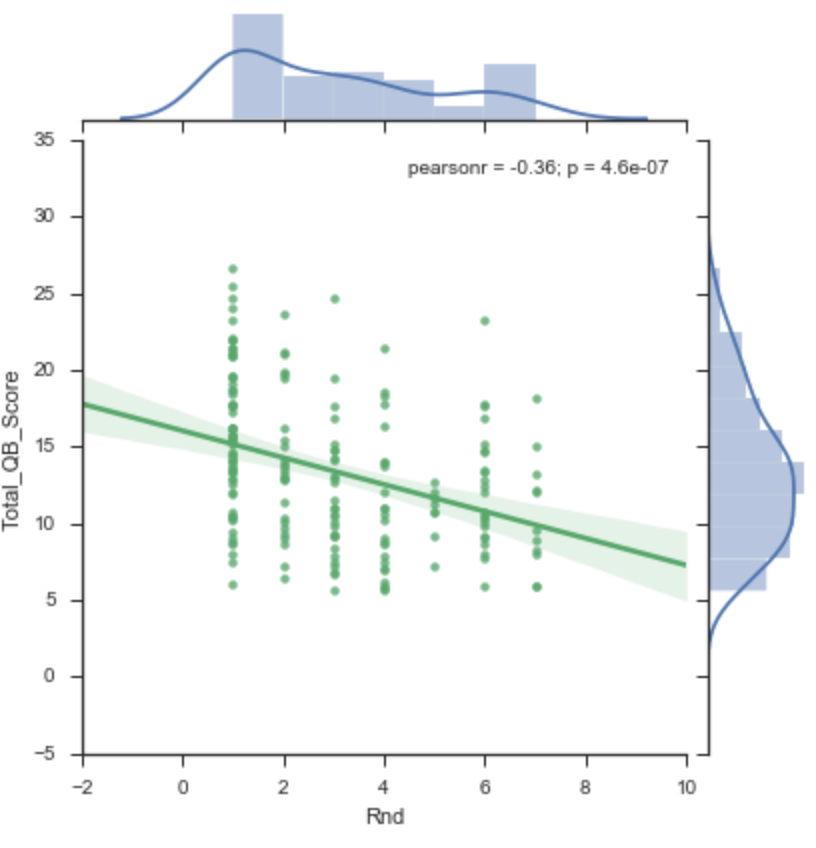
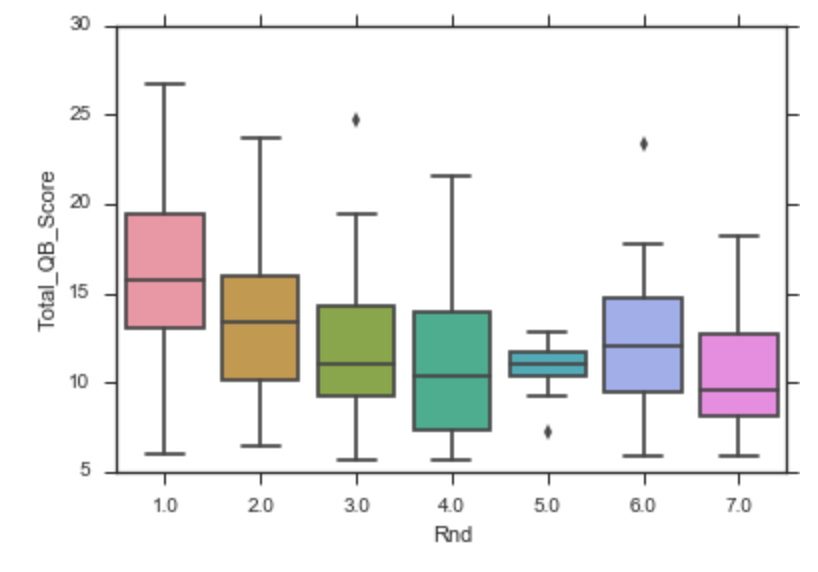
* Is there a correlation between the college the player attended and their success in the NFL?
* Does draft round important for a successful QB?
* Does the round a player is drafted contribute to how long they are going to survive in the NFL?

This was interesting to me because many people always think of WR, RB, and QB positions being the most valuable. But when considering this data, offensive and defensive lineman have a higher average value. The one thing that stuck out most to me is the difference in the QB Mean and Median being just over 9. This inspired me to do further analysis on the QB position and create an additional scoring system more like a fantasy football one to look at QB's further. I also do not like the First4AV stat as much because we are missing data on some of the more recent QB's.

**Questions**

**Question 1 🡪**

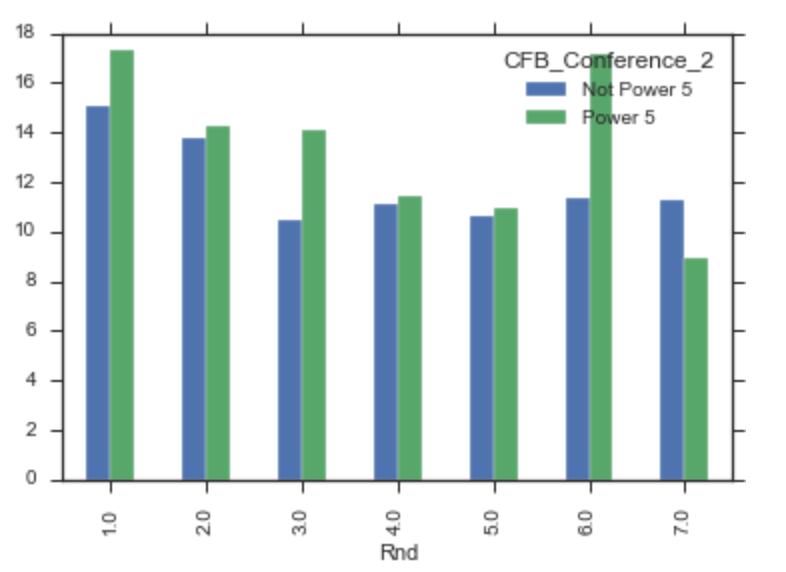
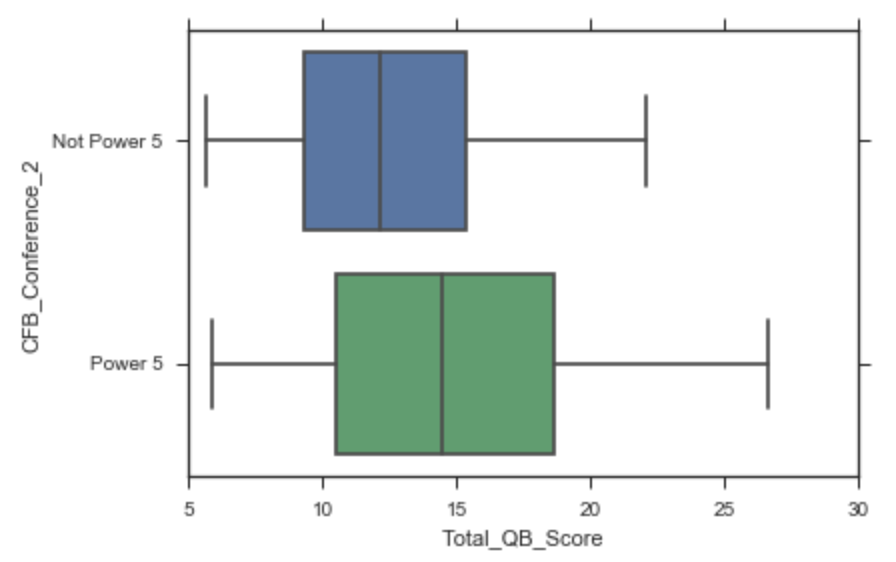
* Does draft round important for a successful QB?

****

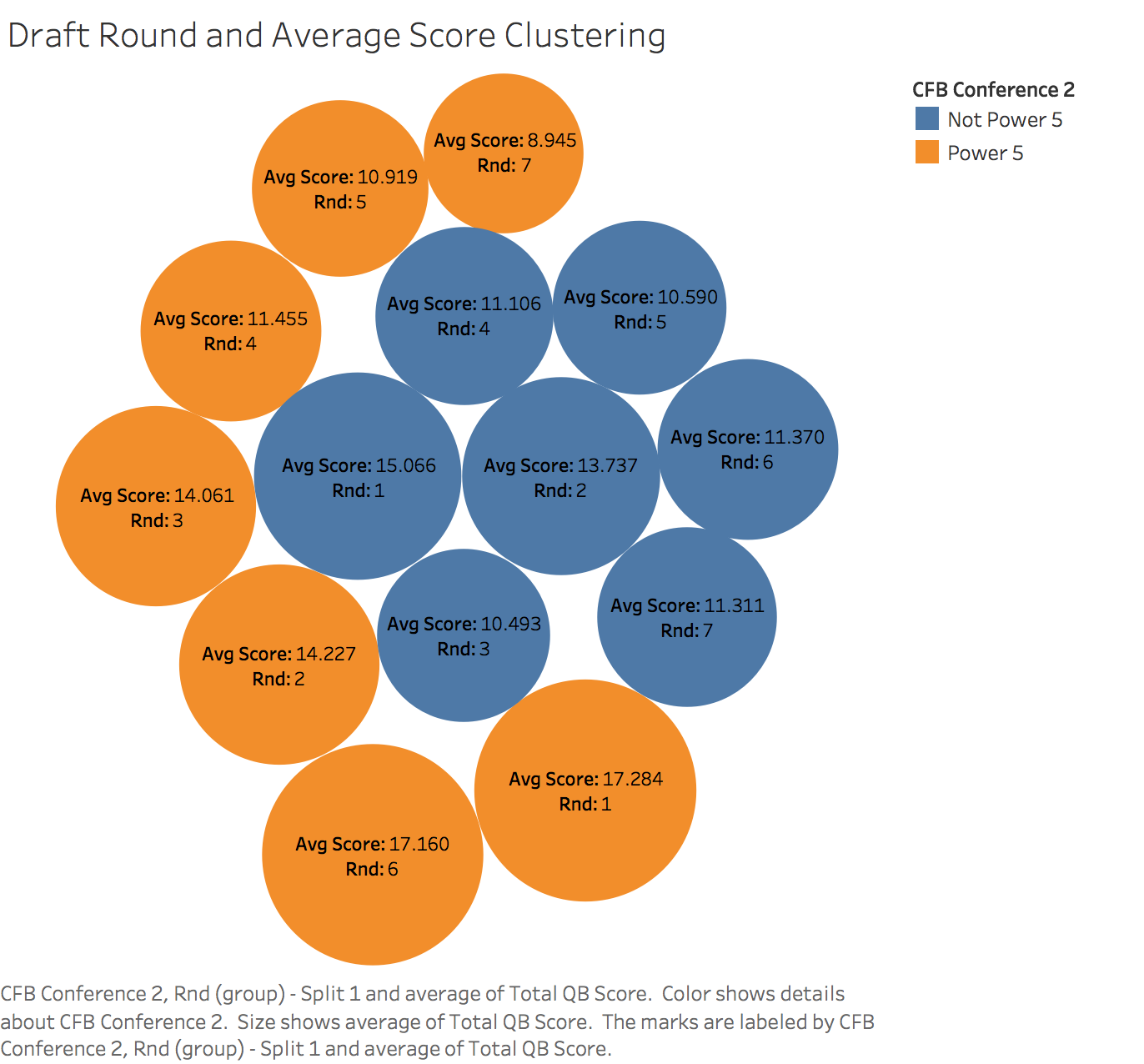
With a p-value of almost 0 I can conclude this does not tell us much because of the wide variety of values in each round. This is because there are many bust players in each round, but more successful players in the earlier rounds too. Also with a -0.36 Pearson r-value we can see that there is a negative correlation for Total\_QB\_Score and the Round that one is selected. These numbers are similar to First4AV, but show that typically being drafted in a higher round does not ensure that the quarterback will be successful. This is very true when you think about the NFL because there are a lot of bust quarterbacks taken in the earlier rounds that do not end up being successful in the NFL. The Box plot can more easily represent the data, when looking at the visualization one can see that there is a constant decrease from rounds 1 to 7 in mean Total\_QB\_Score. With the mean going from 15.99 in round 1 to 10.67 in round 7. With round 6 being an exception with a mean of 12.38 and when doing further analysis in the data I found that Tom Brady is an outlier in round 6 with his Total\_QB\_Score being 23.31.

**Question 2**

* Is there a correlation between the college the player attended and their success in the NFL?



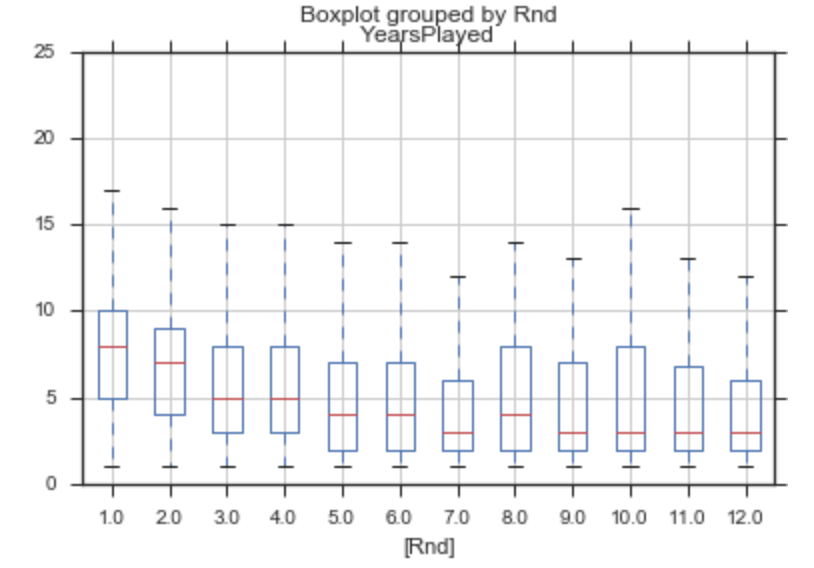
The data shows that the higher round you are drafter that you have a better chance of being successful and having a higher Total\_QB\_Score. On average when moving from round 1 to round 7 there is a constant decline when moving to the next round. When further analyzing the Power 5 conferences (Pac 12, SEC, ACC, Big 12, Big 10) vs. Non-Power 5 conferences I saw a correlation that if you attend a Power 5 university then you are more likely to be successful when comparing each round to the conferences. how much more likely you will be successful in the NFL. In addition to this many quarterback that played in a Power 5 conference are more likely to be successful when comparing each round to Power 5 and Not Power 5. When looking further into the Power 5 vs Non-Power 5 we can see that the mean for power 5 is about 15 where Non-Power 5 is 12. With Power 5 having a much larger maximum of 27 compared to 23 for Not Power-5. Generally speaking, this shows us that majority of the better quarterbacks are coming from power 5 conferences because of their competition at the college level and typically being better players.



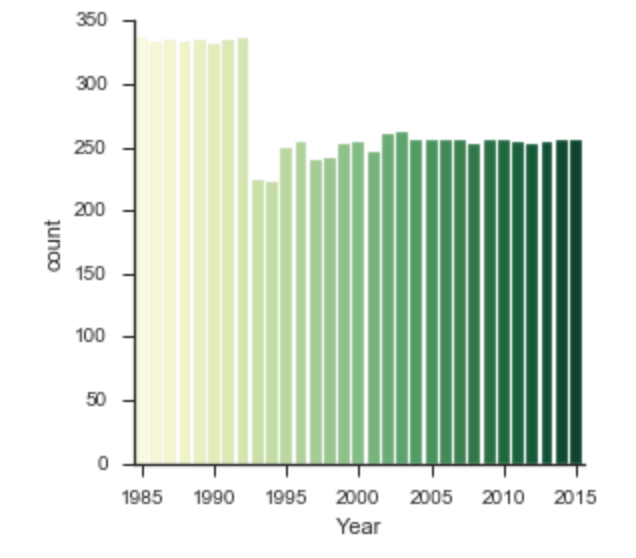
In this visualization, I used a clustering technique in tableau to show the correlation by the round and the average Total\_QB\_Score to the conferences. Again, this is showing us how the rounds and the conferences match up against each other. Additionally, proving how QB’s in the Power 5 conferences and drafted in a higher round are more likely to be successful. With round 6 for Power 5 conferences being an outlier because of Tom Brady.

**Question 3**

* Does the round a player is drafted contribute to how long they are going to survive in the NFL?



While looking at conference and draft round correlations to see a player’s success. I wanted to further consider if there was a correlation in what round a player was drafted to their life time in the NFL. After further analysis, it was evident that there was a negative correlation between round drafted and years in the NFL. With a constant decline from rounds 1 to 7 in the amount of years a player would last in the NFL. The round 8 to 12 were a little different and had some outliers, but after doing further analysis I found that there was very little data in these rounds because typically since we can remember the NFL Draft is only 7 rounds. That is why I ended up taking that data out of my data frame to do further analysis.



This graph shows the number of players drafted in each year. As you can see in 1993 is when they began to move to a 7-round draft style instead of 12. The NFL ended up changing to this style because they round it was more suitable for the quality and the number of players each team needed.

**Top 10 QB’s**



Based on my scoring system these are the quarterbacks that ranked in the top overall totals in scoring. With my above analysis, we can see that 8 out of 10 of these quarterbacks were drafted in the first two rounds and five of them in the top two picks. In addition to this 9 out of 10 of these quarterbacks were in a power 5 conference.

My overall findings have proven that there is a correlation to measure how to find a successful QB. If they are selected in one of the top rounds and in a Power 5 conference then they are more likely to be successful. Granted any quarterback in one of the top rounds is more likely to be successful, but there are also many busts like Johnny Manziel. There are also outliers as well, such as our very own Tom Brady who was drafted 199 overall in the 6th round. Although you never know how someone will perform in the league we can use this to predict how successful a QB will be in the league. In addition to this QB’s in the earlier rounds have a longer life span and make them more appealing for teams to invest in because they will typically last longer in the NFL.

**Moving Forward**

As mentioned in my proposal I wanted to consider more information regarding salaries. I was unable to obtain a data set with the salaries and could not get the API pull to work to incorporate this into my project. Moving forward I want to continue to work with the API to pull this salary data to do additional analysis on a QB’s success compared to their salary. Initially I was trying to do salaries for every player, but if I start small and then move on I will be able to accomplish this. I would also like to do additional analysis on RB’s and WR’s and see how those compare to the findings I was able to obtain on QB’s.