Predicting Fantasy Football

One of the hardest things to predict is the outcome of sports games. Even harder so is trying to predict how a player will perform in any given game. So many different variables can affect a game at any moment. How is their team performing, how well does the defense matchup, is the player dealing with any injuries, just to name-a-few. For this experiment, we are trying to predict high fantasy production from quarterbacks in the National Football League (NFL) by looking at the stats they accumulate throughout the season. Essentially, we are trying to know which quarterback statistics are highly correlated with producing high fantasy point seasons. Fantasy sports have been very popular over the last few decades, and now, with the implementation of gambling, they are at an all-time high in popularity. People were always trying to get an edge on their opponents by using statistical models, but now that high-stakes money is involved, there have been many different experiments run to try and determine fantasy production. I chose quarterbacks as my unit of analysis because of their ability to "make or break" an NFL offense. They are the engine that makes the car run, so if your quarterback is underperforming, odds are your offense is as well. I will be using data from the Fantasy Footballers website (https://fantasyfootballers.org/qb-quarterback-nfl-stats/). They record many different quarterbacking statistics. They have some well-known variables like passing yards per game, passing completion percentage, and passing touchdowns. However, they have important fantasy statistics that we will be using for our analysis, like fantasy points per game and fantasy points per passing attempt. One of the best ways to compare variables is to visualize with a regression model and find the correlation coefficient between the variables. We're going to experiment with about thirty-five quarterbacks from each year, over the past six years, and try to make a prediction about which stats are important to a quarterback's fantasy output. My goal is

to find statistics, outside of the "obvious" ones, that are most correlated with producing high fantasy point totals. With these findings, and multiple regression, we could try to make an equation that accurately predicts a quarterback's stats for any given season.

Before we begin the experiment, there are some things that need to be addressed before looking into our findings. We want to know how the statistics we use are scored. When we refer to "fantasy points", we are talking about DraftKings fantasy scoring. Similarly, the variable DK/Gm stands for DraftKings fantasy points per game, meaning we are taking DraftKings scoring as our broad term for fantasy points. Here is how the DraftKings scoring works for quarterbacks: every passing touchdown = 4 points, every 25 passing yards = 1 point, every interception = -1 point, every fumble = -1 point, every rushing touchdown = 6 points, every 10 rushing yards = 1 point, every 300+ yard passing game receives a 3-point bonus, and every 100-yard rushing game receives a 3-point bonus. To interpret our findings for the correlation coefficient (Pearson's r), we will be using Cohen's scale. Cohen's scale will be interpreted as such:

Scale of correlation coefficient	Value
$0 < r \leq 0.19$	Very Low Correlation
$0.2 \leq r \leq 0.39$	Low Correlation
$0.4 \leq r \leq 0.59$	Moderate Correlation
$0.6 \leq r \leq 0.79$	High Correlation
$0.8 \leq r \leq 1.0$	Very High Correlation

We will be using correlation coefficients for the relationships of each variable as showing the linear regression model will clutter the article. Also, it is important to note that this experiment deals with fantasy production only. Elite fantasy production doesn't always mean you will be a great quarterback (i.e. Jalen Hurts or Kirk Cousins).

The way to achieve the correlation coefficients through Python is by importing the libraries pandas, numpy, and sklearn. With these libraries we are able to import our datasets into Python, clean up the data for exploratory data analysis (EDA), and output our wanted information. In this case for the correlation coefficients and the multiple regression analysis. We start by importing the datasets into Python using pandas. Fortunately, pandas has its own form of correlation coefficient, thus we are able to use pandas to find the relationships between each of the variables. If we wanted to view the graph of a particular relationship between variables, we would create a scatterplot with a linear regression line using sklearn. Although it is quite simple, we are able to reveal a lot of usable information from these datasets and the two libraries in Python. With all that being said, these were the findings:

Table 1. Fantasy Points per Game vs. Test Variables

Y	X	2021	2020	2019	2018	2017	2016
DK/Gm	Pass Comp/Gm	0.57	0.56	0.28	0.54	0.58	0.61
	Pass Comp %	0.28	0.51	0.24	0.47	0.45	0.55
	Yds/Att	0.55	0.72	0.6	0.69	0.69	N/A
	FP/Att	0.83	0.81	0.77	0.76	0.89	N/A

Yds/Comp	0.51	0.62	0.55	0.49	0.56	0.41
Pass Yds/Gm	0.78	0.75	0.52	0.76	0.73	0.79
Pass Att/Gm	0.54	0.42	0.17	0.41	0.51	0.48
Pass TD/Gm	0.83	0.77	0.84	0.84	0.85	0.86
INTs/Gm	0.09	-0.47	-0.08	-0.04	-0.32	-0.37
Sacks/Gm	-0.22	-0.21	-0.08	-0.38	-0.06	-0.2
Rush Att/Gm	0.48	0.4	0.6	0.23	0.35	0.21
Rush Yds/Gm	0.44	0.41	0.63	0.13	0.27	0.18
Rush TDs/Gm	0.33	0.49	0.38	0.33	-0.08	0.11

By Table 1, we can see that yards per attempt (Yds/Att), fantasy points per attempt (FP/Att), passing yards per game (Pass Yds/Gm), and passing touchdowns per game (Pass TD/Gm) are the most positively correlated with fantasy points per game (DK/Gm). (The Fantasy Footballers website didn't record Yds/Att and FP/Att until 2017).

Table 2. Yards per Attempt vs. Test Variables

Y	X	2021	2020	2019	2018	2017
Yds/Att	Pass Comp/Gm	0.20	0.33	0.16	0.24	0.6
	Pass Comp %	0.58	0.61	0.6	0.54	0.6
	FP/Att	0.61	0.66	0.55	0.67	0.57
	Yds/Comp	0.87	0.9	0.82	0.77	0.72
	Pass Yds/Gm	0.62	0.68	0.57	0.63	0.83
	Pass Att/Gm	-0.01	0.12	-0.08	0.03	0.42
	Pass TD/Gm	0.62	0.65	0.63	0.7	0.56
	INTs/Gm	0.10	-0.44	-0.02	0.13	-0.46
	Sacks/Gm	-0.08	-0.32	-0.12	-0.17	-0.22
	Rush Att/Gm	0.03	0.1	0.1	0.02	-0.05
	Rush Yds/Gm	-0.06	0.15	0.11	-0.12	-0.18
	Rush TDs/Gm	-0.01	0.13	-0.01	0.03	-0.25

By Table 2, we can see that there is a strong, positive correlation between yards per completion (Yds/Comp) and yards per attempt (Yds/Att). Fantasy points per attempt (FP/Att), passing yards per game (Pass Yds/Gm), and passing touchdowns per game (Pass TD/Gm) also all have high correlations. Sacks per game (Sacks/Gm) and interceptions per game (INTs/Gm) have an overall low negative correlation with yards per attempt (Yds/Att), on average.

Table 3. Fantasy Points per Attempt vs. Test Variables

Y	X	2021	2020	2019	2018	2017
FP/Att	Pass Comp/Gm	0.07	0	-0.34	-0.07	0.25
	Pass Comp %	0.18	0.46	0.38	0.34	0.38
	Yds/Comp	0.67	0.6	0.46	0.59	0.43
	Pass Yds/Gm	0.40	0.28	-0.05	0.22	0.42
	Pass Att/Gm	0.01	-0.18	-0.48	-0.26	0.1
	Pass TD/Gm	0.62	0.6	0.66	0.59	0.67
	INTs/Gm	-0.12	-0.55	-0.39	-0.09	-0.45
	Sacks/Gm	-0.01	-0.27	-0.22	-0.09	0

	Rush Att/Gm	0.59	0.65	0.76	0.56	0.6
	Rush Yds/Gm	0.57	0.67	0.77	0.51	0.52
	Rush TDs/Gm	0.54	0.57	0.59	0.52	0.2

By Table 3, we can see that passing touchdowns per game (Pass TD/Gm) have a high, positive correlation and interceptions per game (INTs/Gm) have a moderate/high, negative correlation with fantasy points per attempt (FP/Att). Sacks per game (Sacks/Gm), on average, are becoming negatively correlated with fantasy points per attempt.

Table 4. Rushing Statistics vs. Fantasy Points per Game

Rush Att/Gm	0.48	0.4	0.6	0.23	0.35	0.21
Rush Yds/Gm	0.44	0.41	0.63	0.13	0.27	0.18
Rush TDs/Gm	0.33	0.49	0.38	0.33	-0.08	0.11

Table 5. Rushing Statistics vs. Yards per Attempt

Rush Att/Gm	0.03	0.1	0.1	0.02	-0.05
Rush Yds/Gm	-0.06	0.15	0.11	-0.12	-0.18
Rush TDs/Gm	-0.01	0.13	-0.01	0.03	-0.25

Table 6. Rushing Statistics vs. Fantasy Points per Attempt

Rush Att/Gm	0.59	0.65	0.76	0.56	0.6
Rush Yds/Gm	0.57	0.67	0.77	0.51	0.52
Rush TDs/Gm	0.54	0.57	0.59	0.52	0.2

Table 7. Rushing Statistics vs. Passing Yards per Game

Rush Att/Gm	0.00	-0.16	-0.24	-0.3	-0.26	-0.32
Rush Yds/Gm	-0.08	-0.15	-0.21	-0.44	-0.36	-0.38
Rush TDs/Gm	-0.17	0.07	-0.29	-0.11	-0.45	-0.28

Table 8. Rushing Statistics vs. Passing Touchdowns per Game

Rush Att/Gm	-0.01	-0.05	0.29	-0.12	-0.01	-0.15
Rush Yds/Gm	-0.03	0	0.33	-0.23	-0.09	-0.16
Rush TDs/Gm	-0.06	-0.05	0.03	-0.14	-0.47	-0.29

By Table 4-8, we can see that rushing statistics, from 2016 to 2021, are becoming more positively correlated with fantasy points per game (DK/Gm) and the other highly, positively correlated variables with fantasy points per game.

After testing our hypothesis, we can see that the variables that have the highest positive correlation with fantasy points per game are yards per attempt, fantasy points per attempt, passing yards per game, and passing touchdowns per game. Based on these findings, we can predict that a quarterback with higher yards per attempt, fantasy points per attempt, passing yards per game, and passing touchdowns per game will have more fantasy points per game. Passing yards per game and passing touchdowns per game are two of the most used statistics when predicting a quarterback's fantasy season, so we expect them to be highly correlated with fantasy points per game. However, yards per attempt and fantasy points per attempt are lesser used statistics, so they could give us a better insight into how to predict a quarterback's fantasy season.

We can test the independent variables we used against fantasy points per game against yards per attempt. We see that there is a high/very high correlation between yards per attempt and yards per completion. If we follow the trends of the NFL, this makes sense as offensive aggressiveness is higher now than in years past. Meaning teams are attempting more deep throws and connecting on them. Fantasy points per attempt, passing yards per game, and passing touchdowns per game also all have high, positive correlations. We can conclude that quarterbacks who are an equal amount aggressive and smart tend to do better than those who are too aggressive or too passive. We see that sacks per game and interceptions per game have an overall low negative correlation with yards per attempt. Sacks are correlated with offensive line play, having a better offensive line could lead to higher yards per attempt. This could also lead to fewer interceptions. Therefore, this could be interpreted as having a better offensive line allows for better quarterback play, or a quarterback who is more elusive and is sacked less tends to have higher yards per attempt.

We can test the independent variables we used for the previous two dependent variables again, now against fantasy points per attempt. We see that passing touchdowns per game has a high, positive correlation, and interceptions per game have a moderate/high, negative correlation with fantasy points per attempt, which should be expected. Sacks, on average, are becoming negatively correlated with fantasy points per attempt. The negative correlation makes sense since sacks could lead to fumbles, which lose fantasy points. However, sacks require the offense to travel further for a first down, positively impacting fantasy points, should they succeed. Something of note, rushing stats are becoming more positively correlated with fantasy points per attempt, which is positively correlated with high fantasy points per game production.

We can predict that a quarterback's rushing ability is becoming more important for becoming a more successful fantasy quarterback. NFL franchises are beginning to draft more mobile quarterbacks out of college in hopes of being able to create more plays with their legs as opposed to traditional pocket passers. Quality starting mobile quarterbacks were less frequent in the mid-2010s than now as NFL coaches have adapted and implemented college schemes into their gameplans to allow for mobile quarterbacks to succeed. With more college schemes coming to the NFL, we can predict that a quarterback's rushing ability will allow for more fantasy points than a quarterback who is less mobile.

We've seen, through correlations, that we can find relationships between fantasy points per game and obscure, lesser-known statistics. Although variables like passing touchdowns per game and passing yards per game are more obviously positively correlated with fantasy points per game, we found that yards per attempt and fantasy points per attempt are two lesser-known variables with statistically significant positive correlations to fantasy points per game. We can predict that quarterbacks who are more aggressive, yet smart with the football, while being

efficient runners, tend to be better fantasy producers. This makes intuitive sense when you think about big arm quarterbacks such as Justin Herbert and Patrick Mahomes, who frequently score many fantasy points. A great example of someone who does not throw far downfield is Mac Jones of the New England Patriots. His average depth of target is much lower, and therefore, he scores fewer fantasy points than the average quarterback, even though he is very efficient with those passes.

We can make an equation to predict the fantasy output of each quarterback depending on their projected stats for the next year. We will make a multiple regression equation in Python with the numpy and sklearn libraries. Using multiple regression with the stats from up above, a formula for fantasy points would look like, Formula 1: Fantasy Points per Game = -4.60 + 0.78*Yds/Att + 0.04*Pass Yds/Gm + 3.234*Pass TDs/Gm + 0.10*Rush Yds/Gm + 3.85*Rush TDs/Gm. Looking at the R squared, approximately 78.1% of the variance in fantasy points scored can be explained by yards per attempt, passing yards per game, passing touchdowns per game, rushing yards per game, and rushing touchdowns per game. I removed fantasy points per attempt since that variable would make most of the other variable irrelevant in the equation. With that variable implemented into the multiple regression, the equation would become, Formula 2: Fantasy Points per Game = -3.47 – 1.23*Yds/Att + 28.57*FP/Att + 0.07*Pass Yds/Gm. Looking at the R squared, approximately 91.6% of the variance in fantasy points scored can be explained by yards per attempt, fantasy points per attempt, and passing yards per game, more than the R squared without fantasy points per attempt.

Using these formulas' outputs, and comparing them to the 2021 season, we can see if they are accurate for predicting a quarterback's fantasy production. We will pick three quarterbacks at random. The three quarterbacks will be: Tom Brady, Jalen Hurts, and Daniel

Jones. Their actual fantasy points per game scoring for 2021 is 24.51, 21.94, and 16.41 for Brady, Hurts, and Jones, respectively. Using Formula 1, we obtain a predicted per game average of 22.80, 20.77, and 15.94, respectively. Using Formula 2, we obtain a predicted per game average of 24.16, 22.72, and 16.79, respectively. We can see that both equations are accurate for predicting fantasy points, but Formula 2 is a more accurate.

Overall, it is important to remember that correlation does not imply causation, and fantasy production can vary from game to game due to the high number of variables that can impact it. With that being said, a very accurate model for predicting fantasy output is out there, we just need to create it.

Works Cited

Fantasy Footballer.net. 2022. *Quarterback QB Stats Cheat Sheet NFL Current Season Statistics*.

https://fantasyfootballer.net/qb-quarterback-nfl-stats/.