Analyzing NFL Quarterback 2020 Fantasy Performance

MA469 - Final Project

Michael Averill, Jed Martin, & Jack Michalowski

April 30, 2021

Table of Contents

Introduction & Web Scraping……..……..……..……...……3

Tableau Data Visualization…...……...…….4

Verifying Normal Forms & Relational Integrity…..……...……...7

MySQL + R + SQL Data Manipulation…….……..…..8

Conclusion……...…….…..……..…11

Introduction & Web Scraping

The data that will be used throughout this project will be fantasy football data courtesy of the NFL. More specifically, breaking down quarterback data from the 2020 NFL season. In this dataset, there are various statistics and fantasy point numbers from 25 different quarterbacks in the league. Through web-scraping using R, we obtained the names, total passing yards, total passing touchdowns, total interceptions thrown, and total fantasy points for the entire 2020 season. All of this was done by using different web-scraping techniques such as the use of html functions within R, and the rvest package to start off. After formatting all of this data, it was written into a text (.txt) file for Tableau functionality and later into a comma-separated value (.csv) file for SQL functionality. Tableau visualizations, databases, SQL queries, etc. will be put together as we analyze this data step by step.

Tableau Data Visualization

Quarterback Total Fantasy Points

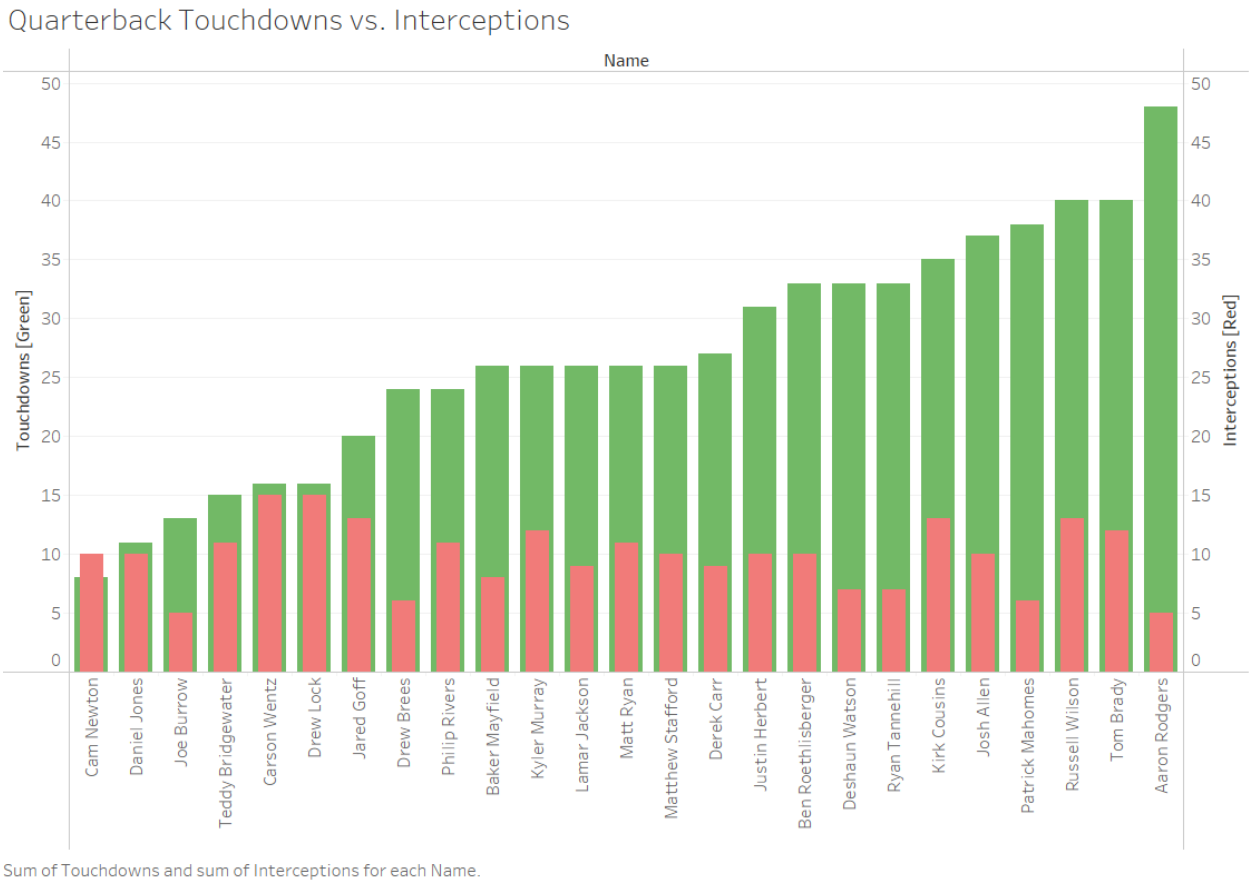
Name	Fantasy Points
Josh Allen	396.1
Aaron Rodgers	383.3
Kyler Murray	378.7
Patrick Mahomes	374.4
Deshaun Watson	369.3
Russell Wilson	359.8
Ryan Tannehill	344.4
Tom Brady	337.9
Justin Herbert	332.8
Lamar Jackson	332.8
Kirk Cousins	306.2
Matt Ryan	282.4
Derek Carr	272.1
Ben Roethlisberger	267.2
Cam Newton	261.0
Matthew Stafford	260.6
Baker Mayfield	248.6
Teddy Bridgewater	241.2
Jared Goff	240.0
Philip Rivers	240.0
Drew Brees	209.5
Carson Wentz	198.4
Drew Lock	181.3
Daniel Jones	180.0
Joe Burrow	173.7

Fantasy Points

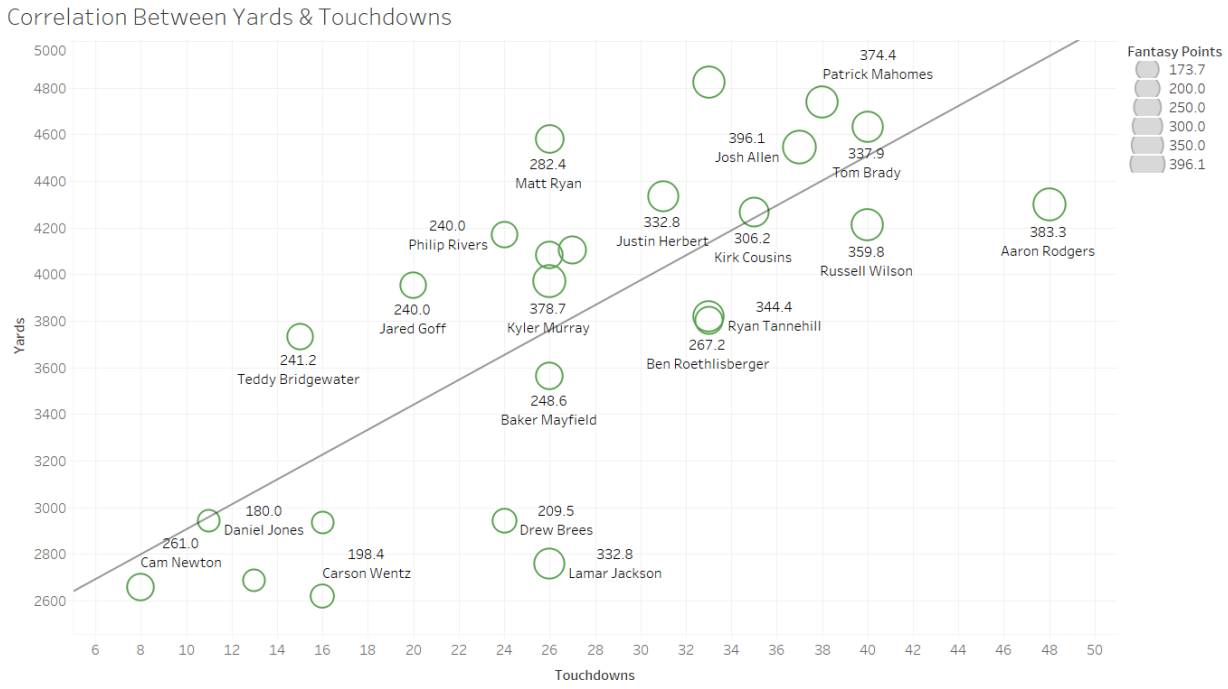
173.7 396.1

Sum of Fantasy Points broken down by Name. Color shows sum of Fantasy Points. The marks are labeled by sum of Fantasy Points.

This graph is a representation of quarterbacks and their corresponding total fantasy points. The darker blue corresponds to a better fantasy performance.



This bar graph displays the quarterback name along with their touchdown and interception numbers for the season. Touchdowns are the green bars and interceptions are the red bars, both of which have the same axis values to make it easier to interpret. It can be noted that as touchdowns increase, you do not see an increase in interceptions, that is likely because the quality of player increases as touchdowns thrown increases, so the quarterback is not subject to making more mistakes than the average player. It can also be seen that one quarterback, Cam Newton, has more interceptions than touchdowns.



Touchdowns vs. Yards. Size shows sum of Fantasy Points. The marks are labeled by sum of Fantasy Points and Name.

The dot plot represents the correlation between touchdowns (x-axis) and yards (y-axis). The legend shows that the larger the dot is, the more total fantasy points they had. Each dot displays the total fantasy points as well as the player's name. One interesting data point to note is Lamar Jackson. He has a much larger total fantasy point value than those around him. This is likely due to the fact that he runs the ball much more frequently than other quarterbacks, and rushing yards and rushing touchdowns are not accounted for in this plot.

Verifying Normal Forms & Relational Integrity

The original dataset that we webscraped was not in third normal form because the variable "Fantasy Points" has a partial dependency on the other three quantitative variables, "Yards", "Touchdowns", and "Interceptions", as those variables are in the calculation. However, many more statistics are used to calculate the Fantasy Points, which is a weighted average of all statistics measured for that player. To combat this issue and transform this dataset into third normal form, the data will be split into two tables. The first table will contain the "Name", "Player ID" (which will be a variable that is added for each unique player), and "Fantasy Points" variables. The second table will have the "Name", "Player ID", "Yards", "Touchdowns", and "Interceptions". Once the data is split into these tables, there will not be any functional dependencies on any two variables.

Table 1

ID	playerName	FantasyPoints
001	Josh Allen	396.06
002	Aaron Rodgers	383.26
003	Kyler Murray	378.74
004	Patrick Mahomes	374.40
005	Deshaun Watson	369.32
006	Russell Wilson	359.78
007	Ryan Tannehill	344.36
008	Tom Brady	337.92
009	Justin Herbert	332.84
010	Lamar Jackson	332.78

Table 2

ID	playerName	Yards	Touchdowns	Interceptions
001	Josh Allen	4544	37	10
002	Aaron Rodgers	4299	48	5
003	Kyler Murray	3971	26	12
004	Patrick Mahomes	4740	38	6
005	Deshaun Watson	4823	33	7
006	Russell Wilson	4212	40	13
007	Ryan Tannehill	3819	33	7
008	Tom Brady	4633	40	12
009	Justin Herbert	4336	31	10
010	Lamar Jackson	2757	26	9

The tables above display the new data tables. The new primary key of the dataset is "Player ID" since the "Player Name" variable is too long and ID numbers are easier for users to quickly identify. Since "Fantasy Points" is not in the same table as all other quantitative variables anymore, there are no more functional dependencies in the data, no non-primary key attribute is transitively dependent on the primary key, and the data has been officially converted into third normal form.

MySQL + R + SQL Data Manipulation

The first queries that were written show the average number of touchdowns, yards, and interceptions thrown per quarterback last season. Those queries are shown below.

avg(Touchdowns)
26.8800

avg(Yards)
3806.8000

avg(Interceptions)
9.9200

The next query shows took the count of the player's names that had thrown more than 30 touchdowns last season. The result below shows that 10 quarterbacks threw for 30 or more touchdowns last season.

NumberOfPeople
10

The query below shows the names of the quarterbacks that threw more interceptions than they did touchdowns during the season. The result shows that Cam Newton was the only quarterback to throw for more interceptions than touchdowns.

playerName
Cam Newton

The next two queries show the minimum and maximum number of touchdowns and interceptions thrown by all of the quarterbacks last season. The query on the left shows the most and the least number of touchdowns and the query on the right shows the most and least number of interceptions thrown. These queries show the range of the variables and describe how touchdowns are typically more common than interceptions, but there is also a wider range of number of touchdowns thrown in a season.

MostTouchdowns	LeastTouchdowns
48	8

MostInterceptions	LeastInterceptions
15	5

The next query uses the first table, which is the fantasy points table. This query is taking the average number of fantasy points from all of the quarterbacks last season.

avg(FantasyPoints)
286.868000

This query below also uses the fantasy points table and builds off the query that was built above. It shows a list of the quarterbacks and the number of fantasy points they had who had more points than the average number of fantasy points, which was calculated in the last query. The result shows eleven out of the twenty-five quarterbacks had more fantasy points than the average.

playerName	FantasyPoints
Josh Allen	396.06
Aaron Rodgers	383.26
Kyler Murray	378.74
Patrick Mahomes	374.40
Deshaun Watson	369.32
Russell Wilson	359.78
Ryan Tannehill	344.36
Tom Brady	337.92
Justin Herbert	332.84
Lamar Jackson	332.78
Kirk Cousins	306.20

This next query joins the two tables together and displays all of the variables in one table. The quarterbacks that are listed in this query have a total number of fantasy points that are above the average number of fantasy points *or* they have a total number of yards over 3,500 yards.

ID	playerName	Yards	Touchdowns	Interceptions	FantasyPoints
001	Josh Allen	4544	37	10	396.06
002	Aaron Rodgers	4299	48	5	383.26
003	Kyler Murray	3971	26	12	378.74
004	Patrick Mahomes	4740	38	6	374.40
005	Deshaun Watson	4823	33	7	369.32
006	Russell Wilson	4212	40	13	359.78
007	Ryan Tannehill	3819	33	7	344.36
008	Tom Brady	4633	40	12	337.92
009	Justin Herbert	4336	31	10	332.84
010	Lamar Jackson	2757	26	9	332.78
011	Kirk Cousins	4265	35	13	306.20
012	Matt Ryan	4581	26	11	282.44
013	Derek Carr	4103	27	9	272.12
014	Ben Roethlisberger	3803	33	10	267.22
016	Matthew Stafford	4084	26	10	260.56
017	Baker Mayfield	3563	26	8	248.62
018	Teddy Bridgewater	3733	15	11	241.22
019	Jared Goff	3952	20	13	239.98
020	Philip Rivers	4169	24	11	239.96

Similar to the query above, this next one joins the two tables together and displays all of the variables in one table. The quarterbacks listed in this query have a total number of fantasy points that are above the average number of fantasy points *and* they have a total number of yards over 3,500 yards. It was important to show both queries because

ID	playerName	Yards	Touchdowns	Interceptions	FantasyPoints
001	Josh Allen	4544	37	10	396.06
002	Aaron Rodgers	4299	48	5	383.26
003	Kyler Murray	3971	26	12	378.74
004	Patrick Mahomes	4740	38	6	374.40
005	Deshaun Watson	4823	33	7	369.32
006	Russell Wilson	4212	40	13	359.78
007	Ryan Tannehill	3819	33	7	344.36
008	Tom Brady	4633	40	12	337.92
009	Justin Herbert	4336	31	10	332.84
011	Kirk Cousins	4265	35	13	306.20

This last query also joins both tables and displays the quarterbacks who threw for more than 3,500 yards in a season, but is sorted by the total number of touchdowns they threw in that same season. This query shows how the two variables are not always perfectly correlated and some quarterbacks tend to throw more touchdowns in a season and some throw more yards in a season.

playerName	Yards	Touchdowns
Aaron Rodgers	4299	48
Russell Wilson	4212	40
Tom Brady	4633	40
Patrick Mahomes	4740	38
Josh Allen	4544	37
Kirk Cousins	4265	35
Deshaun Watson	4823	33
Ryan Tannehill	3819	33
Ben Roethlisberger	3803	33
Justin Herbert	4336	31
Derek Carr	4103	27
Kyler Murray	3971	26
Matt Ryan	4581	26
Matthew Stafford	4084	26
Baker Mayfield	3563	26
Philip Rivers	4169	24
Jared Goff	3952	20
Teddy Bridgewater	3733	15

Conclusion

We began this project by scraping data from fantasy.nfl.com using R. We pulled quarterback statistics like total fantasy points, touchdowns, yards, and interception for 25 starting quarterbacks. We then formatted this data and put it into a .txt file for use in Tableau. Using Tableau, we were able to generate three different visualizations that did a great job of exposing different things in the dataset. We then verified the normality of the data, and came to the conclusion that our data was not in the third normal form, so we separated the data into two relations and created a second table. Our data was now in third normal form. After that, we manipulated the data in SQL to answer some questions about the data we were curious about. One example of a question we had that we were able to answer incredibly quickly and efficiently was if there was a quarterback(s) who threw more interceptions than touchdowns and to show their name(s). Upon finishing the visualizations and queries, we walked away finding some information we didn't expect to find before beginning the project.