

# NFL 2021 QB Data Projections

Jayson Rhea

09/19/2023

## Dataset

Retrieved dataset from: <https://github.com/minhnguyen14797/nfl-2021-data-extraction/tree/main>

```
library(readr)
nfl_offense_data_2021 <- read_csv("nfl_offense_data_2021.csv")

## Rows: 712 Columns: 26
## -- Column specification -----
## Delimiter: ","
## chr (4): player_id, player, position, team
## dbl (22): pass_cmp, pass_att, pass_yds, pass_td, pass_int, pass_sacked, pass...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
df <- nfl_offense_data_2021
qb_data <- subset(df, df$position == 'QB')
print(qb_data)

## # A tibble: 83 x 26
##   player_id player  position team  pass_cmp pass_att pass_yds pass_td pass_int
##   <chr>      <chr>   <chr>   <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 RodgAa00 Aaron R~ QB      GNB      366      531     4115      37      4
## 2 DaltAn00 Andy Da~ QB      CHI      149      236     1515      8       9
## 3 MayfBa00 Baker M~ QB      CLE      253      418     3010     17     13
## 4 RoetBe00 Ben Roe~ QB      PIT      390      605     3740     22     10
## 5 GabbBl00 Blaine ~ QB      TAM        7       11       67      0      0
## 6 AlleBr00 Brandon~ QB      CIN       17       34      149      2      0
## 7 RypiBr00 Brett R~ QB      DEN        0        2        0      0      0
## 8 HoyeBr00 Brian H~ QB      NWE        9       11      227      1      0
## 9 BeatC.00 C.J. Be~ QB      JAX        2        2       33      0      0
## 10 NewtCa00 Cam New~ QB      CAR       69      126     684      4      5
## # i 73 more rows
## # i 17 more variables: pass_sacked <dbl>, pass_sacked_yds <dbl>,
## #   pass_long <dbl>, pass_rating <dbl>, rush_att <dbl>, rush_yds <dbl>,
## #   rush_td <dbl>, rush_long <dbl>, targets <dbl>, rec <dbl>, rec_yds <dbl>,
## #   rec_td <dbl>, rec_long <dbl>, fumbles_lost <dbl>, rush_yac <dbl>,
## #   rec_yac <dbl>, rec_drops <dbl>
```

## QB\_Data\_Classes

<https://www.nfl.com/news/nfl-qb-index-ranking-all-68-starting-quarterbacks-from-the-2022-nfl-season>  
This website contains the NFL ratings for the (current) top 68 quarterbacks in the 2022 season. I used these

results to assign each qb to a specific class: Highly Recommend, Recoomend, and Do Not Recommend (HR, R, DNR). These results will assist in training the data. Players that are ranked 1-20 are classified as 'HR', players that are ranked 21-40 are classified as 'R' and players that are ranked lower than 40 are classified as 'DNR.'

```
qb_data$Class <- NA
qb_data[qb_data$player == "Patrick Mahomes", "Class"] <- 'HR'
qb_data[qb_data$player == "Jalen Hurts", "Class"] <- 'HR'
qb_data[qb_data$player == "Joe Burrow", "Class"] <- 'HR'
qb_data[qb_data$player == "Josh Allen", "Class"] <- 'HR'
qb_data[qb_data$player == "Justin Herbert", "Class"] <- 'HR'
qb_data[qb_data$player == "Trevor Lawrence", "Class"] <- 'HR'
qb_data[qb_data$player == "Tua Tagovailoa", "Class"] <- 'HR'
qb_data[qb_data$player == "Jared Goff", "Class"] <- 'HR'
qb_data[qb_data$player == "Geno Smith", "Class"] <- 'HR'
qb_data[qb_data$player == "Tom Brady", "Class"] <- 'HR'
qb_data[qb_data$player == "Dak Prescott", "Class"] <- 'HR'
qb_data[qb_data$player == "Lamar Jackson", "Class"] <- 'HR'
qb_data[qb_data$player == "Justin Fields", "Class"] <- 'HR'
qb_data[qb_data$player == "Aaron Rodgers", "Class"] <- 'HR'
qb_data[qb_data$player == "Brock Purdy", "Class"] <- 'HR'
qb_data[qb_data$player == "Kyler Murray", "Class"] <- 'HR'
qb_data[qb_data$player == "Kirk Cousins", "Class"] <- 'HR'
qb_data[qb_data$player == "Daniel Jones", "Class"] <- 'HR'
qb_data[qb_data$player == "Ryan Tannehill", "Class"] <- 'HR'
qb_data[qb_data$player == "Jimmy Garoppolo", "Class"] <- 'HR'
qb_data[qb_data$player == "Jacoby Brissett", "Class"] <- 'HR'
qb_data[qb_data$player == "Jacoby Brissett", "Class"] <- 'R'
qb_data[qb_data$player == "Matthew Stafford", "Class"] <- 'R'
qb_data[qb_data$player == "Derek carr", "Class"] <- 'R'
qb_data[qb_data$player == "Andy Dalton", "Class"] <- 'R'
qb_data[qb_data$player == "Kenny Pickett", "Class"] <- 'R'
qb_data[qb_data$player == "Jameis Winston", "Class"] <- 'R'
qb_data[qb_data$player == "Marcus Mariota", "Class"] <- 'R'
qb_data[qb_data$player == "Deshaun Watson", "Class"] <- 'R'
qb_data[qb_data$player == "Mac Jones", "Class"] <- 'R'
qb_data[qb_data$player == "Russell Wilson", "Class"] <- 'R'
qb_data[qb_data$player == "Trey Lance", "Class"] <- 'R'
qb_data[qb_data$player == "Teddy Bridgewater", "Class"] <- 'R'
qb_data[qb_data$player == "Mitchell Trubisky", "Class"] <- 'R'
qb_data[qb_data$player == "Desmond Ridder", "Class"] <- 'R'
qb_data[qb_data$player == "Taylor Heinicke", "Class"] <- 'R'
qb_data[qb_data$player == "Carson Wentz", "Class"] <- 'R'
qb_data[qb_data$player == "Matt Ryan", "Class"] <- 'R'
qb_data[qb_data$player == "Sam Darnold", "Class"] <- 'R'
qb_data[qb_data$player == "Baker Mayfield", "Class"] <- 'R'
qb_data[qb_data$player == "Colt McCoy", "Class"] <- 'R'
qb_data$Class[is.na(qb_data$Class)] <- 'DNR'
print(qb_data)
```

```
## # A tibble: 83 x 27
```

	player_id	player	position	team	pass_cmp	pass_att	pass_yds	pass_td	pass_int	
	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	
##	1	RodgAa00	Aaron R~	QB	GNB	366	531	4115	37	4
##	2	DaltAn00	Andy Da~	QB	CHI	149	236	1515	8	9

```
## 3 MayfBa00 Baker M~ QB CLE 253 418 3010 17 13
## 4 RoetBe00 Ben Roe~ QB PIT 390 605 3740 22 10
## 5 GabbBl00 Blaine ~ QB TAM 7 11 67 0 0
## 6 AlleBr00 Brandon~ QB CIN 17 34 149 2 0
## 7 RypiBr00 Brett R~ QB DEN 0 2 0 0 0
## 8 HoyeBr00 Brian H~ QB NWE 9 11 227 1 0
## 9 BeatC.00 C.J. Be~ QB JAX 2 2 33 0 0
## 10 NewtCa00 Cam New~ QB CAR 69 126 684 4 5
## # i 73 more rows
## # i 18 more variables: pass_sacked <dbl>, pass_sacked_yds <dbl>,
## # pass_long <dbl>, pass_rating <dbl>, rush_att <dbl>, rush_yds <dbl>,
## # rush_td <dbl>, rush_long <dbl>, targets <dbl>, rec <dbl>, rec_yds <dbl>,
## # rec_td <dbl>, rec_long <dbl>, fumbles_lost <dbl>, rush_yac <dbl>,
## # rec_yac <dbl>, rec_drops <dbl>, Class <chr>
```

## QB Efficiency Stats (mutated data)

Adding the completion percentage and yards per completion statistics will assist in the model's precision.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
## filter, lag

## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union

qb_data_stats_1 <- mutate(qb_data, 'cmp percentage' = pass_cmp / pass_att)

qb_data_stats <- mutate(qb_data_stats_1, 'yds/cmp' = pass_yds / pass_cmp)

print(qb_data_stats)

## # A tibble: 83 x 29
##   player_id player position team pass_cmp pass_att pass_yds pass_td pass_int
##   <chr>      <chr>   <chr>   <chr>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 RodgAa00 Aaron R~ QB GNB 366 531 4115 37 4
## 2 DaltAn00 Andy Da~ QB CHI 149 236 1515 8 9
## 3 MayfBa00 Baker M~ QB CLE 253 418 3010 17 13
## 4 RoetBe00 Ben Roe~ QB PIT 390 605 3740 22 10
## 5 GabbBl00 Blaine ~ QB TAM 7 11 67 0 0
## 6 AlleBr00 Brandon~ QB CIN 17 34 149 2 0
## 7 RypiBr00 Brett R~ QB DEN 0 2 0 0 0
## 8 HoyeBr00 Brian H~ QB NWE 9 11 227 1 0
## 9 BeatC.00 C.J. Be~ QB JAX 2 2 33 0 0
## 10 NewtCa00 Cam New~ QB CAR 69 126 684 4 5
## # i 73 more rows
## # i 20 more variables: pass_sacked <dbl>, pass_sacked_yds <dbl>,
## # pass_long <dbl>, pass_rating <dbl>, rush_att <dbl>, rush_yds <dbl>,
## # rush_td <dbl>, rush_long <dbl>, targets <dbl>, rec <dbl>, rec_yds <dbl>,
## # rec_td <dbl>, rec_long <dbl>, fumbles_lost <dbl>, rush_yac <dbl>,
## # rec_yac <dbl>, rec_drops <dbl>, Class <chr>, `cmp percentage` <dbl>,
```

```
## # `yds/cmp` <dbl>
```

## Cleaning up the data

Remove variables and rows that are not necessary for machine learning and normalizing the data to improve analysis. (9 rows deleted)

```
QB_Num_Data <- subset(qb_data_stats, select = -c(player_id, player, position, team, rec_drops, rec_yac,
QB_Num_Data_clean <- na.omit(QB_Num_Data)
```

## Scaling Data

Scaling the data to reduce dimensions and allow for faster convergence.

```
scaled_data <- mutate(QB_Num_Data_clean, across(where(is.numeric), scale))
print(scaled_data)
```

```
## # A tibble: 74 x 16
##   pass_cmp[,1] pass_att[,1] pass_yds[,1] pass_td[,1] pass_int[,1]
##           <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1          1.31          1.20          1.36          2.06         -0.358
## 2        -0.0937        -0.0695        -0.159        -0.264          0.574
## 3          0.580          0.715          0.712          0.457          1.32
## 4          1.47          1.52          1.14          0.858          0.761
## 5         -1.01         -1.04         -1.00         -0.905         -1.10
## 6         -0.949        -0.940        -0.955        -0.745         -1.10
## 7         -1.00         -1.04        -0.909        -0.825         -1.10
## 8         -1.05         -1.08        -1.02        -0.905         -1.10
## 9         -0.612        -0.544        -0.643        -0.585         -0.171
## 10          1.03          1.14          1.03          1.26          0.201
## # i 64 more rows
## # i 11 more variables: pass_sacked <dbl[,1]>, pass_sacked_yds <dbl[,1]>,
## #   pass_long <dbl[,1]>, pass_rating <dbl[,1]>, rush_att <dbl[,1]>,
## #   rush_yds <dbl[,1]>, rush_td <dbl[,1]>, fumbles_lost <dbl[,1]>, Class <chr>,
## #   `cmp percentage` <dbl[,1]>, `yds/cmp` <dbl[,1]>
```

## Data Plot

```
library(tidyr)
library(ggplot2)
my_data_long <- pivot_longer(scaled_data, cols = -`Class`, names_to = "Variable", values_to = "Value")

ggplot(my_data_long, aes(x = `Class`, y = Value, color = Variable)) +
  geom_point() +
  labs(title = "Scatter Plot of QB Statistics",
       x = "Classes",
       y = "Player Statistic")
```

