# QB Touchdown Prediction Model

By Garrett Hass

Data Science Project

August 1, 2022

Springboard

# The Issue:

We need to know how many touchdowns a QB will throw in game.

- The average TDs thrown per game:
  1.11
- What stats have a correlation?


- Stakeholders
- Can we predict how many a player will throw?

# The Solution to the Problem:

Create a model to predict how many TDs the Quarterback will throw.

| Player | Team | Gms | Att | Cmp | Pct | Yds | YPA | TD | TD% | Int | Int% | Lg | Sack | Loss | Rate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aaron Rodgers | GB | 16 | 552 | 371 | 67.2 | 4,295 | 7.8 | 39 | 7.1 | 8 | 1.4 | 73 | 51 | 293 | **108.0** |
| Peyton Manning | DEN | 16 | 583 | 400 | 68.6 | 4,659 | 8.0 | 37 | 6.3 | 11 | 1.9 | 71t | 21 | 137 | **105.8** |
| Robert Griffin III | WAS | 15 | 393 | 258 | 65.7 | 3,200 | 8.1 | 20 | 5.1 | 5 | 1.3 | 88t | 30 | 217 | **102.4** |
| Russell Wilson | SEA | 16 | 393 | 252 | 64.1 | 3,118 | 7.9 | 26 | 6.6 | 10 | 2.5 | 67 | 33 | 203 | **100.0** |
| Matt Ryan | ATL | 16 | 615 | 422 | 68.6 | 4,719 | 7.7 | 32 | 5.2 | 14 | 2.3 | 80t | 28 | 210 | **99.1** |
| Tom Brady | NE | 16 | 637 | 401 | 63.0 | 4,827 | 7.6 | 34 | 5.3 | 8 | 1.3 | 83t | 27 | 182 | **98.7** |
| Ben Roethlisberger | PIT | 13 | 449 | 284 | 63.2 | 3,265 | 7.3 | 26 | 5.8 | 8 | 1.8 | 82t | 30 | 182 | **97.0** |
| Drew Brees | NO | 16 | 670 | 422 | 63.0 | 5,177 | 7.7 | 43 | 6.4 | 19 | 2.8 | 80t | 26 | 190 | **96.3** |

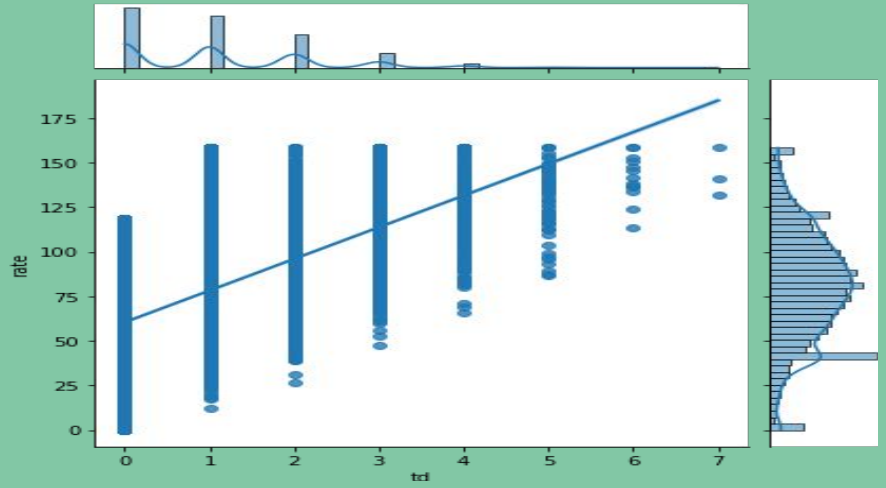
Our Dataset is a compilation of QB information from 1996-2016

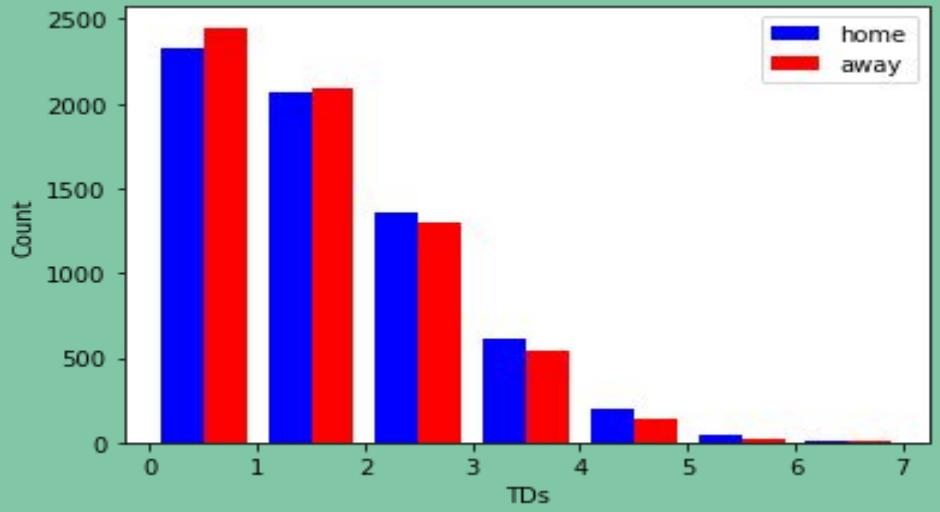
https://www.kaggle.com/datasets/speckledpingu/nfl-qb-stats

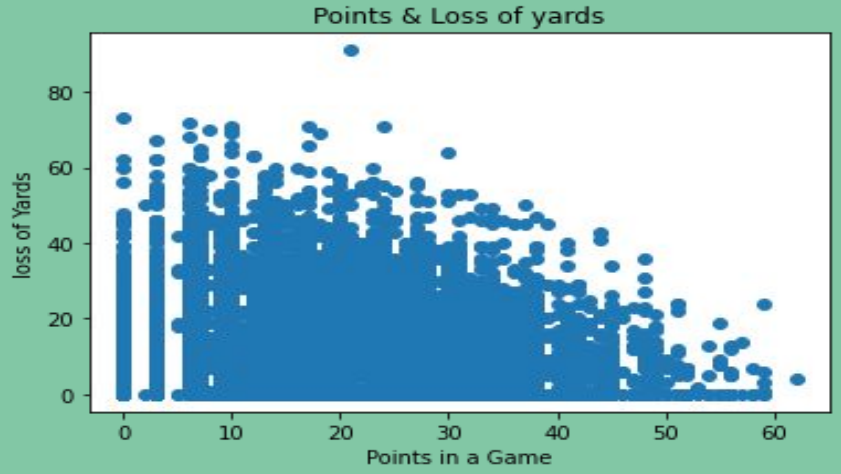
# Data Wrangling

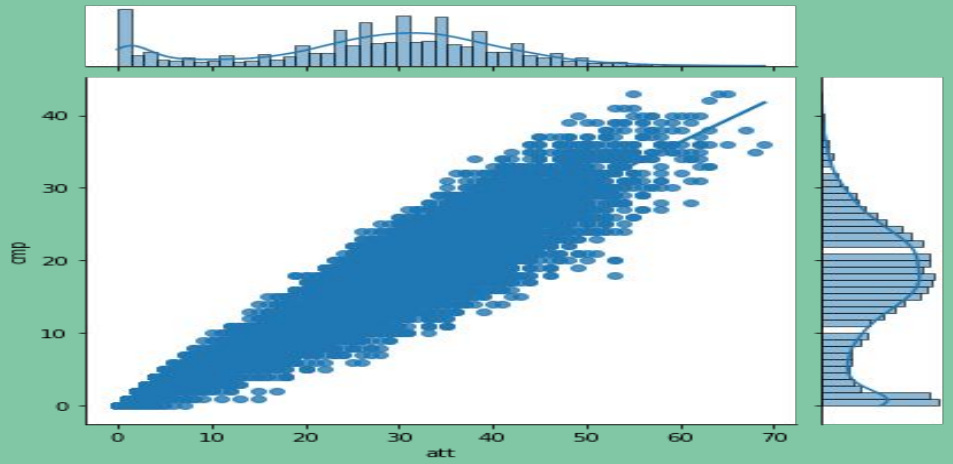
1. Started with 14 columns and 13,188 rows
2. Dropped certain rows
3. Add a few columns
4. NaN values
5. I ended up with 19 columns and 13,172 rows.
6. Target Variable is touchdowns

# EDA

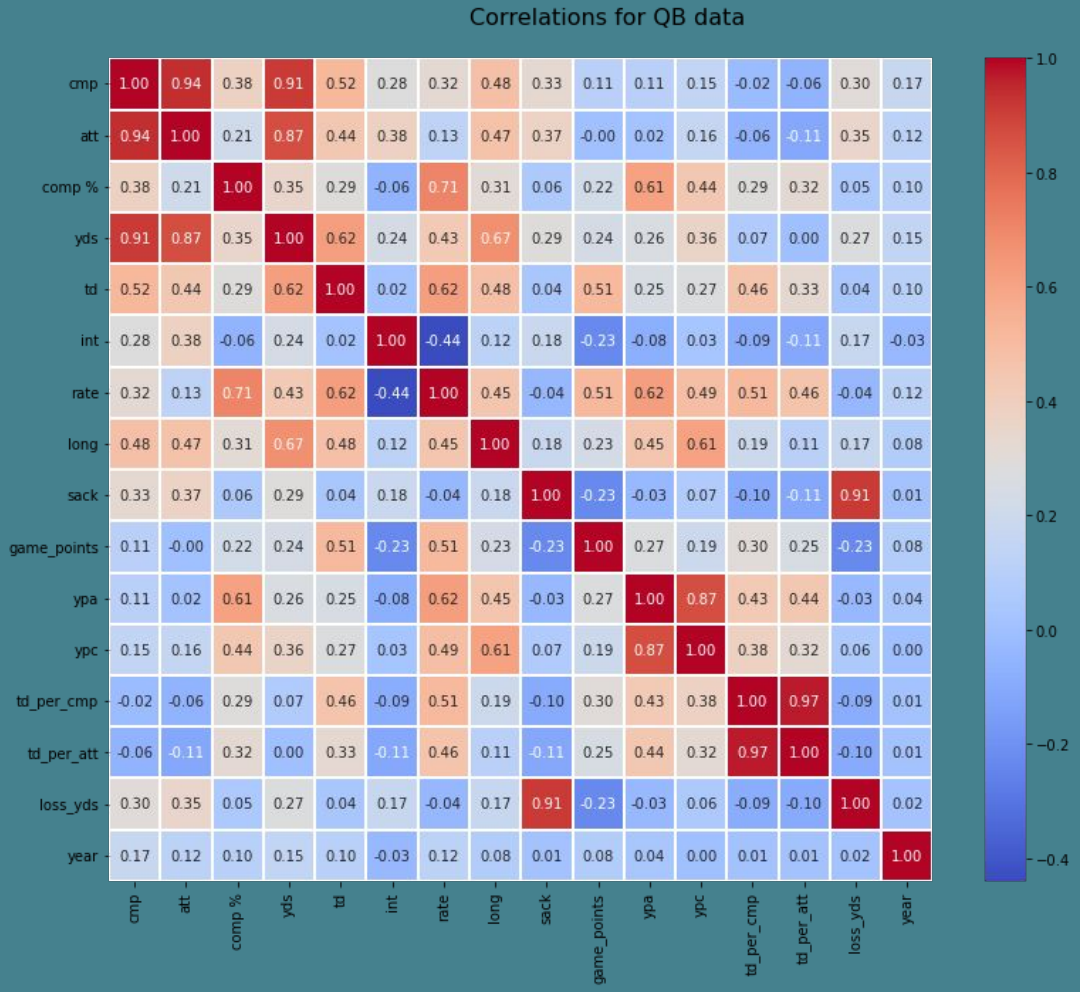

rate
ttd


home
away
Count
TDs


Points & Loss of yards
loss of Yards
Points in a Game


cmp
att

# EDA

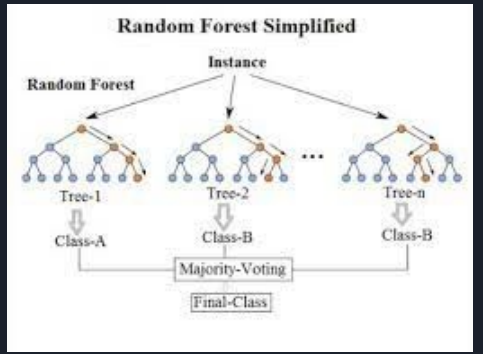

Correlations for QB data
cmp 1.00 0.94 0.38 0.91 0.52 0.28 0.32 0.48 0.33 0.11 0.11 0.15 -0.02 -0.06 0.30 0.17
att 0.94 1.00 0.21 0.87 0.44 0.38 0.13 0.47 0.37 -0.00 0.02 0.16 -0.06 -0.11 0.35 0.12
comp % 0.38 0.21 1.00 0.35 0.29 -0.06 0.71 0.31 0.06 0.22 0.61 0.44 0.29 0.32 0.05 0.10
yds 0.91 0.87 0.35 1.00 0.62 0.24 0.43 0.67 0.29 0.24 0.26 0.36 0.07 0.00 0.27 0.15
td 0.52 0.44 0.29 0.62 1.00 0.02 0.62 0.48 0.04 0.51 0.25 0.27 0.46 0.33 0.04 0.10
int 0.28 0.38 -0.06 0.24 0.02 1.00 -0.44 0.12 0.18 -0.23 -0.08 0.03 -0.09 -0.11 0.17 -0.03
rate 0.32 0.13 0.71 0.43 0.62 -0.44 1.00 0.45 -0.04 0.51 0.62 0.49 0.51 0.46 -0.04 0.12
long 0.48 0.47 0.31 0.67 0.48 0.12 0.45 1.00 0.18 0.23 0.45 0.61 0.19 0.11 0.17 0.08
sack 0.33 0.37 0.06 0.29 0.04 0.18 -0.04 0.18 1.00 -0.23 -0.03 0.07 -0.10 -0.11 0.91 0.01
game_points 0.11 -0.00 0.22 0.24 0.51 -0.23 0.51 0.23 -0.23 1.00 0.27 0.19 0.30 0.25 -0.23 0.08
ypa 0.11 0.02 0.61 0.26 0.25 -0.08 0.62 0.45 -0.03 0.27 1.00 0.87 0.43 0.44 -0.03 0.04
ypc 0.15 0.16 0.44 0.36 0.27 0.03 0.49 0.61 0.07 0.19 0.87 1.00 0.38 0.32 0.06 0.00
td_per_cmp -0.02 -0.06 0.29 0.07 0.46 -0.09 0.51 0.19 -0.10 0.30 0.43 0.38 1.00 0.97 -0.09 0.01
td_per_att -0.06 -0.11 0.32 0.00 0.33 -0.11 0.46 0.11 -0.11 0.25 0.44 0.32 0.97 1.00 -0.10 0.01
loss_yds 0.30 0.35 0.05 0.27 0.04 0.17 -0.04 0.17 0.91 -0.23 -0.03 0.06 -0.09 -0.10 1.00 0.02
year 0.17 0.12 0.10 0.15 0.10 -0.03 0.12 0.08 0.01 0.08 0.04 0.00 0.01 0.01 0.02 1.00
cmp att comp % yds td int rate long sack game_points ypa ypc td_per_cmp td_per_att loss_yds year
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4

# Regression Model Selection


- Linear Regression
- Lasso Regression
- Ridge Regression

The best one was:

- Random Forest Regression



# Hyperparameter Tuning

## Grid Search Cross Validation

- Improved the Random Forest Model by .0007 points
- From .9971 to .9978

# Main Ideas

- Random Forest Model was the best with an R2 Score of .9978
- Predicting TDs could help win more games
- Most important stat besides final score
- This will help to identify who start, bench or sign

# Future Recommendations

- Use models for other variables for every major stat
- Metrics on all 22 positions
- Offense, Defense, and Special Teams
- Situational Analytics