# NFL Wide Receiver Data

Clay Jackson

DS-160

Spring 2024

# Introduction


Analyzing the stats of NFL wide receivers

Football is the biggest sport in America

Some of the biggest stars In the league are wide receivers

Millions of people now play fantasy football

# Introduction: Fantasy Football

- **Before the season, you draft NFL players to be on your team**
- **You earn points for how they perform**
  - **A reception = 1 point**
  - **10 yards gained = 1 point**
  - **A touchdown = 6 points**
- **Touchdowns worth a lot of points. We want to draft players who will score lots of touchdowns!**
- **However, touchdowns very volatile**
- **Can we predict a receiver’s touchdowns based on his underlying stats and metrics?**

# Project Objectives

**Dataset: Advanced Wide Receiver Stats (2021-2023)**

**Objectives:**

- **Collect and clean dataset**
- **Perform a thorough exploratory analysis**
- **Create an accurate model that predicts the number of touchdowns a wide receiver scores (x = statistics, y = touchdowns scored)**
- **Apply this model to fantasy football drafts**

# Data Source

- Data Source: FantasyPros
- Our Dataset: Advanced Wide Receiver Stats
- Dataset Structure
  - 28 columns, 590 rows
  - Every NFL receiver to have caught a pass in the last three years
  - Statistics and performance metrics
    - Continuous and discrete

# Dataset

```
1 dataset.head(3)
```

| | Year | PLAYER | Player ID | G | REC | YDS | Y/R | YBC | YBC/R | AIR | ... | CATCHABLE | DROP | RZ TGT | 10+ YDS | 20+ YDS | 30+ YDS | 40+ YDS | 50+ YDS | LNG | TD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2023 | Tyreek Hill (MIA) | 1 | 16 | 119 | 1799 | 15.1 | 1146 | 9.6 | 1847 | ... | 131 | 12 | 24 | 64 | 29 | 14 | 9 | 5 | 78 | 13 |
| 1 | 2023 | CeeDee Lamb (DAL) | 2 | 17 | 135 | 1749 | 13.0 | 1073 | 7.9 | 1726 | ... | 143 | 6 | 31 | 73 | 29 | 8 | 3 | 1 | 92 | 12 |
| 2 | 2023 | Amon-Ra St. Brown (DET) | 3 | 16 | 119 | 1515 | 12.7 | 847 | 7.1 | 1297 | ... | 127 | 8 | 23 | 60 | 24 | 6 | 3 | 1 | 70 | 10 |

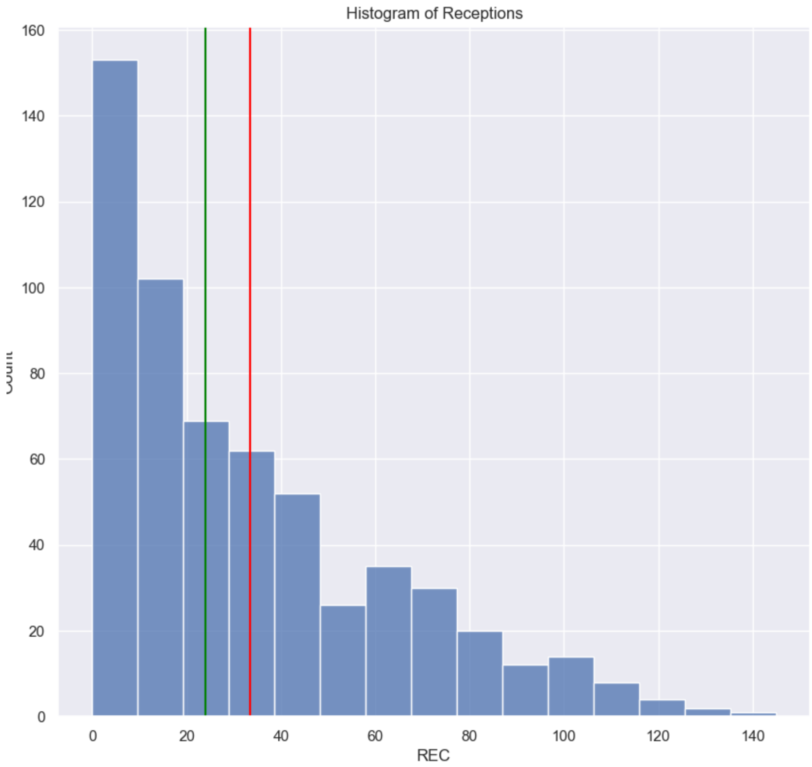
3 rows × 28 columns

# Exploratory Analysis: Data Exploration

```
In [7]:  1  dataset.isnull().sum()

Out[7]: Year          0
        PLAYER        0
        Player ID     0
        G             0
        REC           0
        YDS           0
        Y/R           0
        YBC           0
        YBC/R         0
        AIR           0
        AIR/R         0
        YAC           0
        YAC/R         0
        YACON         0
        YACON/R       0
        BRKTKL        0
        TGT           0
        % TM          0
        CATCHABLE     0
        DROP          0
        RZ TGT        0
        10+ YDS       0
        20+ YDS       0
        30+ YDS       0
        40+ YDS       0
        50+ YDS       0
        LNG           0
        TD            0
        dtype: int64
```

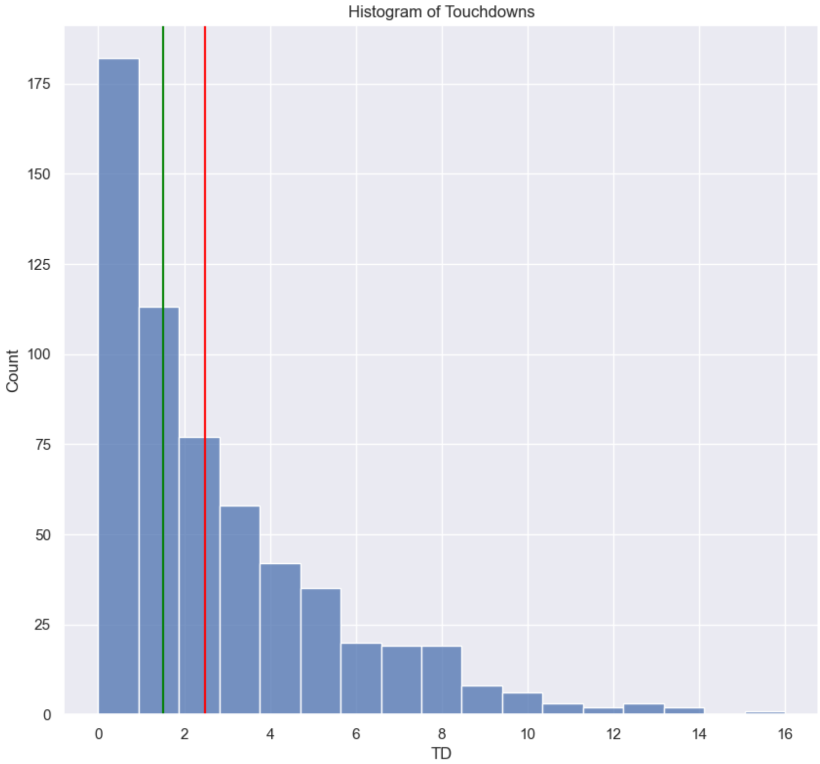
```
In [6]:  1  dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 590 entries, 0 to 589
Data columns (total 28 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Year       590 non-null    int64
 1   PLAYER     590 non-null    object
 2   Player ID  590 non-null    int64
 3   G          590 non-null    int64
 4   REC        590 non-null    int64
 5   YDS        590 non-null    int64
 6   Y/R        590 non-null    float64
 7   YBC        590 non-null    int64
 8   YBC/R      590 non-null    float64
 9   AIR        590 non-null    int64
 10  AIR/R      590 non-null    float64
 11  YAC        590 non-null    int64
 12  YAC/R      590 non-null    float64
 13  YACON      590 non-null    int64
 14  YACON/R    590 non-null    float64
 15  BRKTKL     590 non-null    int64
 16  TGT        590 non-null    int64
 17  % TM       590 non-null    float64
 18  CATCHABLE  590 non-null    int64
 19  DROP       590 non-null    int64
 20  RZ TGT     590 non-null    int64
 21  10+ YDS    590 non-null    int64
 22  20+ YDS    590 non-null    int64
```

# Exploratory Analysis: Distribution



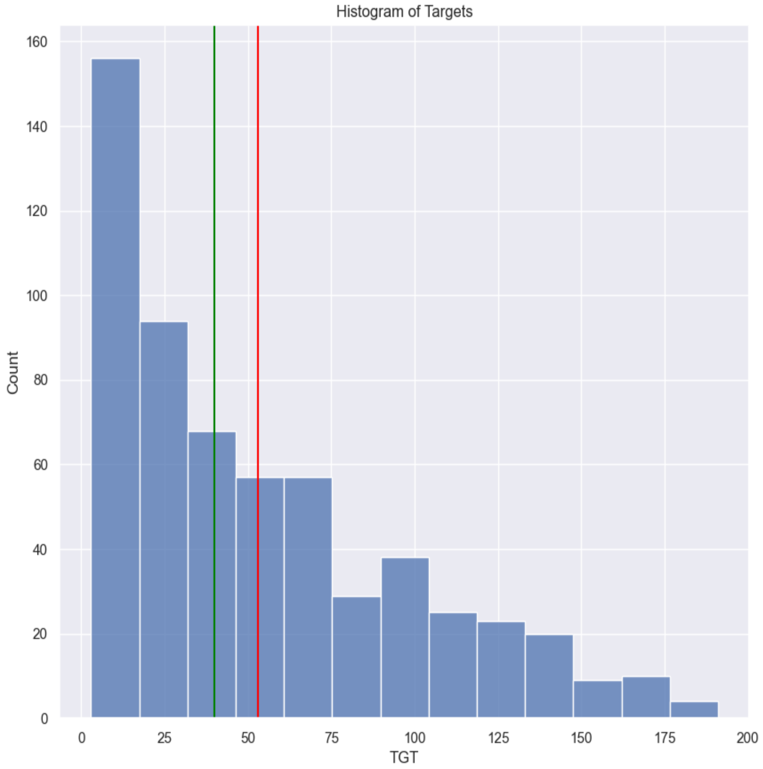

Mean: 33.52 Median: 24

Histogram: Receptions




Mean: 2.48 Median: 1

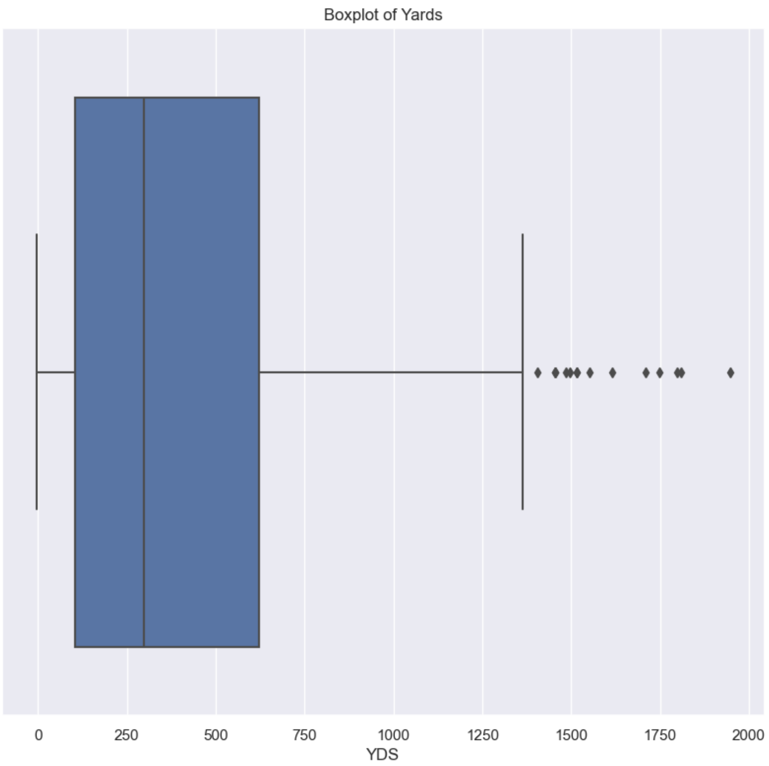
Histogram: Touchdowns



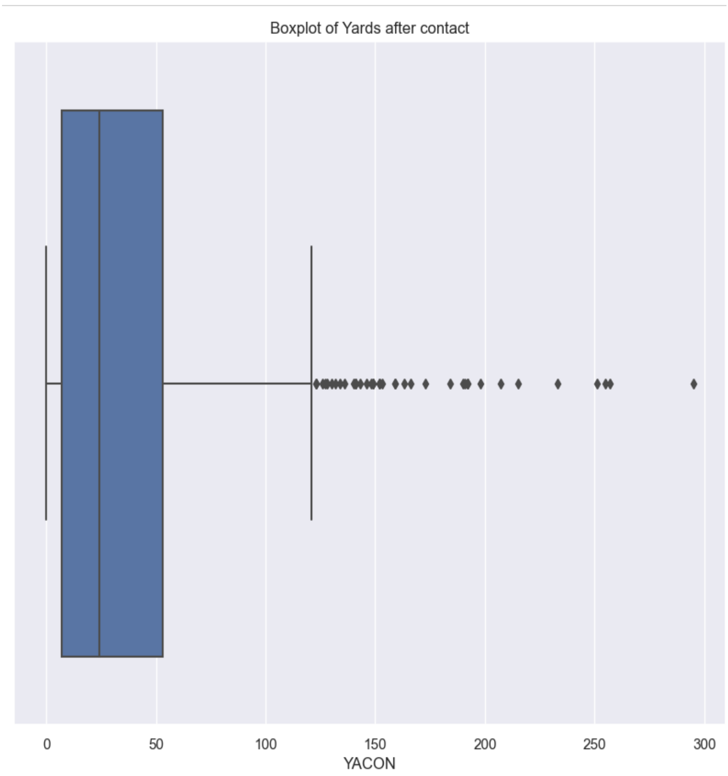

Mean: 53.06 Median: 40

Histogram: Targets

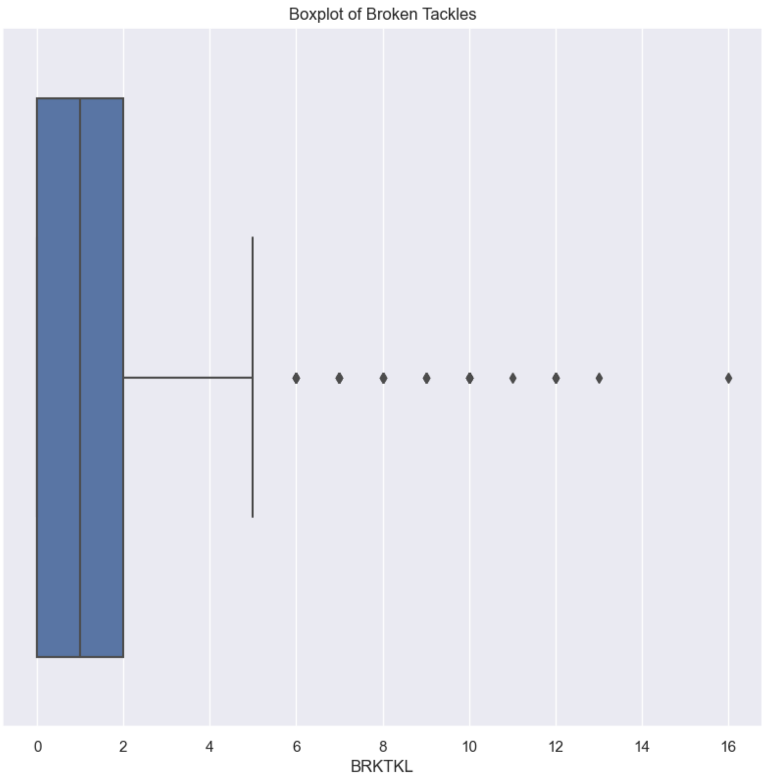
# Exploratory Analysis: Distribution




Boxplot: Yards




Boxplot: Yards after contact




Boxplot: Broken Tackles

# Exploratory Analysis: Distribution

- **Takeaways:**

Every receiver metric has larger mean than median

Every histogram/boxplot skewed to the right

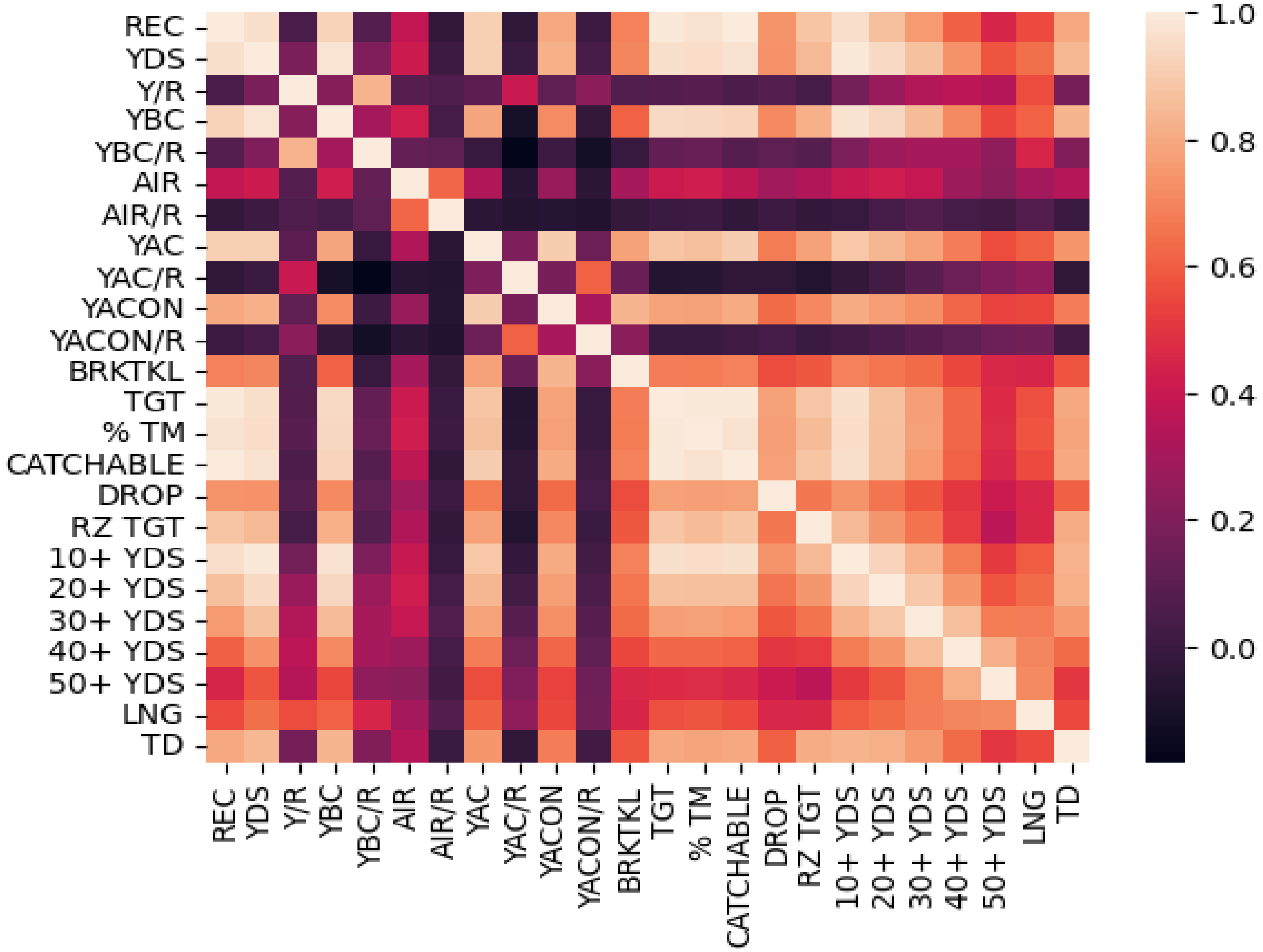
Positive outliers are skewing the data

Vast majority of NFL receivers put up mediocre statistics

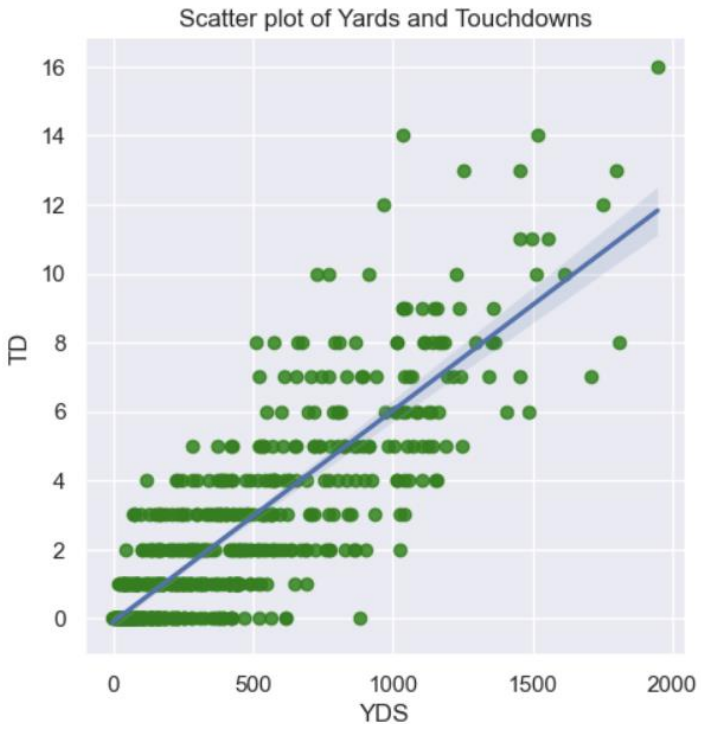
These outliers are the elite, or star, wide receivers in the NFL

# Exploratory Analysis: Correlation

- **Reminder: Our dependent variable is touchdowns scored**

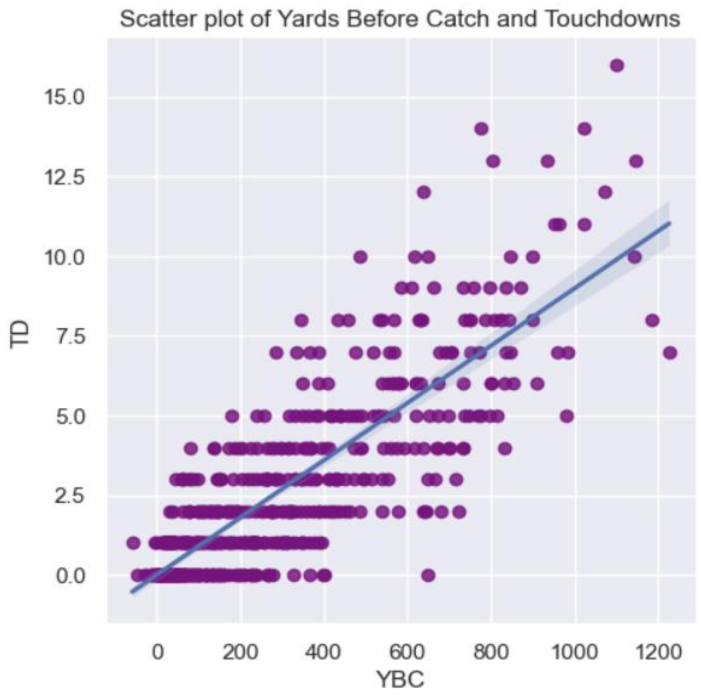


# Exploratory Analysis: Correlation



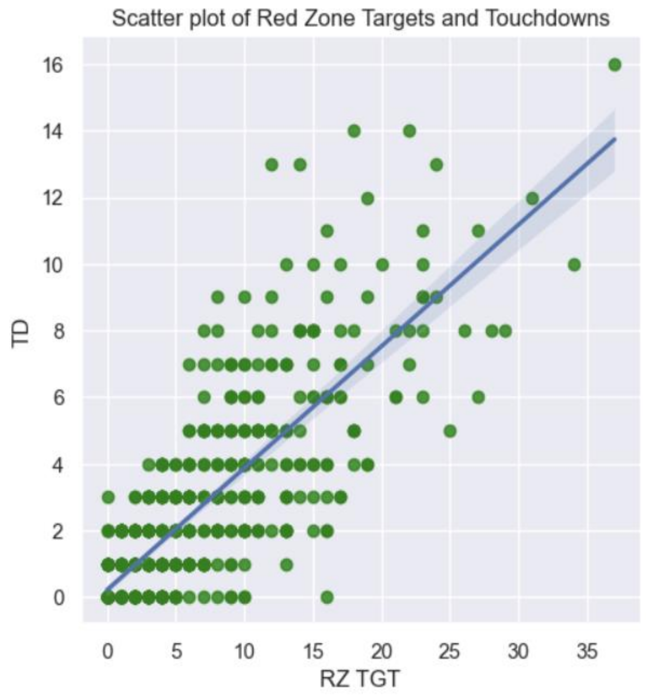

Scatterplot: YDS and TD’s

R-squared: .84




Scatterplot: YDS and TD’s

R-squared: .83

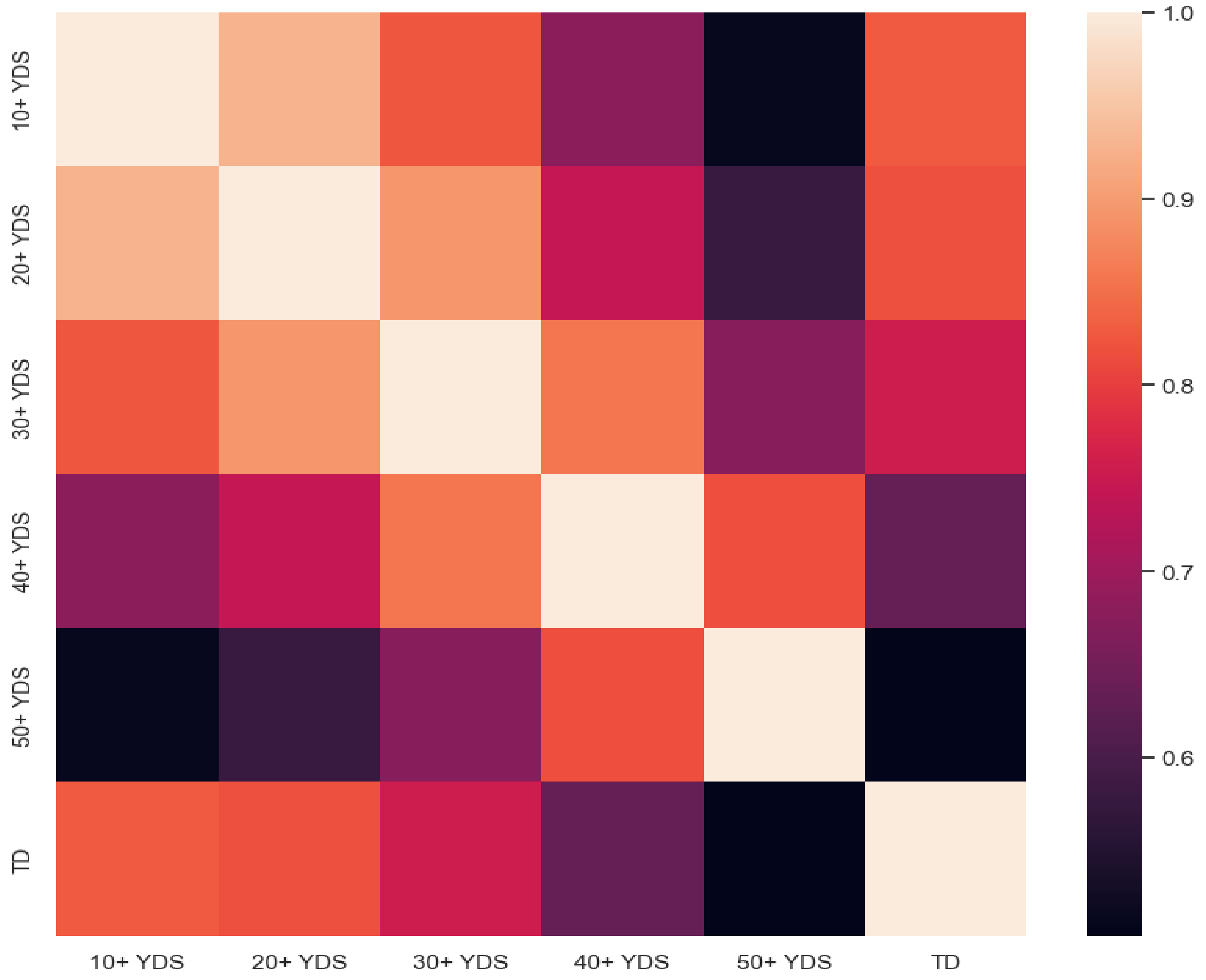



Scatterplot: YDS and TD’s

R-squared: .81

# Exploratory Analysis: Correlation

- **Focused in on the amount of plays a wide receiver had that went for 10, 20, 30, 40 or 50 yards:**



# Exploratory Analysis: Correlation

- **Takeaways:**

**Discovered 15 variables that have moderate to strong correlation with touchdowns score ( > .60)**

**"per reception" variables have much less correlation to touchdowns scored than overall production (ex. yards has a strong correlation to TD's but yards per reception does not).**

**The amount of 10 to 30 yard plays a receiver has correlates more to touchdowns than 40 to 50 yard plays**

# Creating a Model

- **Multiple Linear Regression Model**
- **Dropped Year, Name, Player ID**
- **Independent variables (x) is the 24 columns of advanced stats**
- **Dependent variable (y) is touchdowns scored**

**- y_test = actual value & y_pred = predicted value**

In [108]:
```
y_pred=regressor.predict(x_test.values)
y_pred[:5]
```
Out[108]:
```
array([[0.67721528],
       [1.63453385],
       [1.90175936],
       [1.31540547],
       [6.69949294]])
```

In [109]:
```
y_test.head()
```
Out[109]:

| | TD |
|---|---|
| **522** | 0 |
| **284** | 3 |
| **514** | 0 |
| **331** | 0 |
| **210** | 4 |

### 5b. Splitting the dataset into the Training set and Test set

In [104]:
```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test=train_test_split(x,y,
                                                  test_size=.20,
                                                  random_state=42)
```

### 5c. Training the Multiple Linear Regression model on the Training set

In [105]:
```
from sklearn.linear_model import LinearRegression
regressor=LinearRegression()
regressor.fit(x_train.values, y_train)
# fit for model training
```
Out[105]:
```
LinearRegression()
```

# Model Accuracy Results

- Raw variables for Model 1 & Model 2: All 24 variables
- Selected variables for Model 3 & Model 4: 15 variables with moderate to strong correlation with TD's scored
- Model 2 had the highest R-squared score

| Model Number | Parameters | R-squared | MSE | RMSE |
|---|---|---|---|---|
| 1 | All raw features with 80/20 split for train, and test | 0.708 | 1.752 | 1.32 |
| 2 | All raw features with 70/30 split for train, and test | 0.728 | 1.88 | 1.37 |
| 3 | Selected features with 80/20 split for train, and test | 0.707 | 1.756 | 1.325 |
| 4 | Selected features with 70/30 split for train, and test | 0.726 | 1.890 | 1.374 |

# Testing the Model

- **Tested all four models initially**
- **I then tested Model 2 on 75 receivers' 2023 season**
- **Compared its prediction to their actual number of TD's scored**
- **Model 2 was quite accurate**

```
# Tyreek Hill Real TD's: 13
regressor.predict([[16,119,1799,15.1,1146,9.6,1847,15.5,653,5.5,85,0.7,12,171,0.311,131,12,24,64,29,14,9,5,78
]])
```

```
]: array([[13.80082433]])
```

```
# CeeDee Lamb Real TD's: 12
regressor.predict([[17,135,1749,13,1073,7.9,1726,12.8,676,5,207,1.5,11,181,0.299,143,6,31,73,29,8,3,1,92
]])
```

```
]: array([[12.54852149]])
```

```
# Amon-ra St. Brown Real TD's: 10
regressor.predict([[16,119,1515,12.7,847,7.1,1297,10.9,668,5.6,159,1.3,10,164,0.286,127,8,23,60,24,6,3,1,70
]])
```

```
]: array([[9.85997767]])
```

# Fantasy Football Application: The Regression Candidates

- **Drake London**
  - **2023 TD's: 2**
  - **Model Prediction: 7.047**
  - **Based on his underlying metrics, London should have scored 5 more touchdowns, was unlucky, should regress positively next season**
  - **Likely undervalued in drafts, may want to target him in drafts**

- **Mike Evans**
  - **2023 TD's: 13**
  - **Model prediction: 8.35**
  - **Based on his underlying metrics, Evans should not have scored 13 TD's, got lucky, should regress next year**
  - **Likely overpriced in drafts, may want to avoid.**


# 2024 Underdog Fantasy Draft

## Draft board

| | BHEAFER | CBETZ716 | VERGE | HENKE | DANAUGHTYNINJA | STANSBURY9 | BCONBUCKS3 | SERIOUSGUY | CLIFFORDTHEBIGREDDO | OCEANMANN2021 | RTECHFOOTBALL | SUPERSUPREME |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 Christian McCaffrey RB - SF (1.1) | 2 Tyreek Hill WR - MIA (1.2) | 3 CeeDee Lamb WR - DAL (1.3) | 4 Ja'Marr Chase WR - CIN (1.4) | 5 Breece Hall RB - NYJ (1.5) | 6 Bijan Robinson RB - ATL (1.6) | 7 Amon-Ra St. Brown WR - DET (1.7) | 8 Justin Jefferson WR - MIN (1.8) | 9 A.J. Brown WR - PHI (1.9) | 10 Puka Nacua WR - LAR (1.10) | 11 Garrett Wilson WR - NYJ (1.11) | 12 Marvin Harrison Jr. WR - ARI (1.12) |
| 2 | 24 Josh Jacobs RB - GB (2.12) | 23 Chris Olave WR - NO (2.11) | 22 Nico Collins WR - HOU (2.10) | 21 Deebo Samuel WR - SF (2.9) | 20 Davante Adams WR - LV (2.8) | 19 De'Von Achane RB - MIA (2.7) | 18 Jonathan Taylor RB - IND (2.6) | 17 Drake London WR - ATL (2.5) | 16 Kyren Williams RB - LAR (2.4) | 15 Jahmyr Gibbs RB - DET (2.3) | 14 Saquon Barkley RB - PHI (2.2) | 13 Brandon Aiyuk WR - SF (2.1) |
| 3 | 25 Stefon Diggs WR - HOU (3.1) | 26 Michael Pittman Jr. WR - IND (3.2) | 27 DJ Moore WR - CHI (3.3) | 28 Mike Evans WR - TB (3.4) | 29 Derrick Henry RB - BAL (3.5) | 30 Malik Nabers WR - NYG (3.6) | 31 DK Metcalf WR - SEA (3.7) | 32 Jaylen Waddle WR - MIA (3.8) | 33 DeVonta Smith WR - PHI (3.9) | 34 Travis Etienne Jr. RB - JAX (3.10) | 35 Cooper Kupp WR - LAR (3.11) | 36 James Cook RB - BUF (3.12) |
| 4 | 48 C.J. Stroud QB - HOU (4.12) | 47 Rachaad White RB - TB (4.11) | 46 Keenan Allen WR - CHI (4.10) | 45 Tank Dell WR - HOU (4.9) | 44 George Pickens WR - PIT (4.8) | 43 Tee Higgins WR - CIN (4.7) | 42 Zay Flowers WR - BAL (4.6) | 41 Lamar Jackson QB - BAL (4.5) | 40 Travis Kelce TE - KC (4.4) | 39 Jalen Hurts QB - PHI (4.3) | 38 Josh Allen QB - BUF (4.2) | 37 Sam LaPorta TE - DET (4.1) |
| 5 | 49 Rome Odunze WR - CHI (5.1) | 50 Dalton Kincaid TE - BUF (5.2) | 51 Amari Cooper WR - CLE (5.3) | 52 Mark Andrews TE - BAL (5.4) | 53 Christian Kirk WR - JAX (5.5) | 54 Patrick Mahomes QB - KC (5.6) | 55 Isiah Pacheco RB - KC (5.7) | 56 Trey McBride TE - ARI (5.8) | 57 Rashee Rice WR - KC (5.9) | 58 Marquise Brown WR - KC (5.10) | 59 Terry McLaurin WR - WAS (5.11) | 60 Jayden Reed WR - GB (5.12) |
| 6 | 72 George Kittle TE - SF (6.12) | 71 Aaron Jones RB - MIN (6.11) | 70 David Montgomery RB - DET (6.10) | 69 Joe Burrow QB - CIN (6.9) | 68 Anthony Richardson QB - IND (6.8) | 67 Jordan Addison WR - MIN (6.7) | 66 Calvin Ridley WR - TEN (6.6) | 65 Alvin Kamara RB - NO (6.5) | 64 Kenneth Walker III RB - SEA (6.4) | 63 Diontae Johnson WR - CAR (6.3) | 62 Kyle Pitts TE - ATL (6.2) | 61 Joe Mixon RB - HOU (6.1) |
| 7 | 73 Brian Thomas Jr. WR - JAX (7.1) | 74 Kyler Murray QB - ARI (7.2) | 75 Evan Engram TE - JAX (7.3) | 76 Rhamondre Stevenson RB - NE (7.4) | 77 Brock Bowers TE - LV (7.5) | 78 James Conner RB - ARI (7.6) | 79 Jaxon Smith-Njigba WR - SEA (7.7) | 80 Chris Godwin WR - TB (7.8) | 81 DeAndre Hopkins WR - TEN (7.9) | 82 Christian Watson WR - GB (7.10) | 83 Raheem Mostert RB - MIA (7.11) | 84 Tony Pollard RB - TEN (7.12) |
| 8 | 96 Caleb Williams QB - CHI (8.12) | 95 Javonte Williams RB - DEN (8.11) | 94 Brock Purdy QB - SF (8.10) | 93 Zamir White RB - LV (8.9) | 92 Jake Ferguson TE - DAL (8.8) | 91 Mike Williams WR - NYJ (8.7) | 90 Xavier Worthy WR - KC (8.6) | 89 D'Andre Swift RB - CHI (8.5) | 88 Jordan Love QB - GB (8.4) | 87 Jaylen Warren RB - PIT (8.3) | 86 Najee Harris RB - PIT (8.2) | 85 Dak Prescott QB - DAL (8.1) |
| 9 | 97 Adonai Mitchell WR - IND (9.1) | 98 Jonathon Brooks RB - CAR (9.2) | 99 Trey Benson RB - ARI (9.3) | 100 Zack Moss RB - CIN (9.4) | 101 David Njoku TE - CLE (9.5) | 102 Jameson Williams WR - DET (9.6) | 103 Ladd McConkey WR - LAC (9.7) | 104 Nick Chubb RB - CLE (9.8) | 105 Brian Robinson Jr. RB - WAS (9.9) | 106 Curtis Samuel WR - BUF (9.10) | 107 Courtland Sutton WR - DEN (9.11) | 108 Austin Ekeler RB - WAS (9.12) |
| 10 | 120 Dalton Schultz TE - HOU (10.12) | 119 Gabe Davis WR - JAX (10.11) | 118 Romeo Doubs WR - GB (10.10) | 117 Gus Edwards RB - LAC (10.9) | 116 Jared Goff QB - DET (10.8) | 115 Justin Herbert QB - LAC (10.7) | 114 Chase Brown RB - CIN (10.6) | 113 Tua Tagovailoa QB - MIA (10.5) | 112 Tyler Lockett WR - SEA (10.4) | 111 Dallas Goedert TE - PHI (10.3) | 110 Jakobi Meyers WR - LV (10.2) | 109 Devin Singletary RB - NYG (10.1) |
| 11 | 121 Blake Corum RB - LAR (11.1) | 122 Keon Coleman WR - BUF (11.2) | 123 Trevor Lawrence QB - JAX (11.3) | 124 Rashid Shaheed WR - NO (11.4) | 125 Jayden Daniels QB - WAS (11.5) | 126 Tyjae Spears RB - TEN (11.6) | 127 Kirk Cousins QB - ATL (11.7) | 128 T.J. Hockenson TE - MIN (11.8) | 129 Khalil Shakir WR - BUF (11.9) | 130 Josh Downs WR - IND (11.10) | 131 Jahan Dotson WR - WAS (11.11) | 132 Jerry Jeudy WR - CLE (11.12) |
| | 144 | 143 | 142 | 141 | 140 | 139 | 138 | 137 | 136 | 135 | 134 | 133 |

# Reflection & Conclusion

**Problems**: Time it took to pull data and test model on 2023 receivers

**Model Limitations**: team success, quality of quarterback, physical traits (height, weight, speed)

**Tools**: Python, libraries (Pandas, Numpy, Seaborn, SKLearn, LinearRegression), Excel

**Conclusion**:

- Through exploratory analysis, I found distribution patterns and variables that correlate with touchdowns scored
- Based on his advanced statistics and metrics, a model can be created to predict a receivers touchdowns scored.
- Applied the model to fantasy football drafts: Found regression candidates and used this knowledge to determine our draft targets.

# Questions?