Final Step 2

Brian Reppeto

2023-08-6

Markdown Basics

Introduction

Fantasy football is a popular online game where participants act as team managers and draft real players from the National Football League (NFL) to form fantasy teams. The performance of the actual stats of the players in real-life games determines the fantasy team's score. Participants compete against one another on a weekly head to head basis and these scores determine a weekly winner. One of the most important aspects of fantasy football is the drafting phase, which takes place before the NFL regular season starts. Participants select players based on their projected performance and rankings. I will use the redraft type of league along with a snake draft type of draft. Additionally, I will use the standard scoring system, which is 1 point per 25 passing yards, 4 points per passing touchdown, 1 point per 10 rushing or receiving yards, 6 points per rushing or receiving touchdown, -2 points per fumble lost or interception.

The issue I want to research is being able to help predict fantasy football drafting rankings. Specifically, drafting quarterbacks, running backs, wide receivers, and tight ends. The goal is to develop a model or methodology that can help accurately predict the ending rankings of NFL players for fantasy football drafts. This project would be of particular interest to fantasy football participants, as it can help them better draft a fantasy teams for their season. By leveraging data science techniques, I can analyze historical player performance and other relevant data to make informed predictions for the upcoming season's draft rankings.

Research questions

1. What statistical features of NFL players historical performances are most indicative of their fantasy football draft rankings?
2. Can we identify any patterns or trends in player performance over the past few seasons that might help predict future draft rankings (rookies will be ignored for this evaluation)?
3. Is there a relationship between the SOS of a team and the points scored by the players for its team?
4. Are there specific positions (quarterback, running back, wide receiver, etc.) that are more valuable to draft in a fantasy league before others?
5. Is there a relationship between an NFL teams winning percentage and the fantasy points scored by its players?

Approach

1. Data Collection: Gather historical player performance data, fantasy football draft rankings data, team data, and any other relevant datasets.
2. Conduct data analysis to understand the data, identify any missing values or outliers, and gain insights into player performance and rankings trends.
3. Select or create relevant features that may impact draft rankings (i.e. player statistics from previous seasons, team performance metrics (Strength of schedule), etc.)
4. Create a heat map to show the most valuable selections for a snake type draft.

5. Provide a recommendation for the most effective method to help predict fantasy football drafting rankings.

How approach addressed the problem

By combining the multiple approaches listed above I will be able to compare individual performances along with other factors to help provide valuable insights into a ranking and value list for fantasy football managers. This will all them to build competitive teams for the fantasy season. This will also allow me to group the players into clusters.

This project will not be fully operational as all cuts for the NFL season will not have happened by the time this project is complete.

Cleaning data

Fantasy stats

The data had to have positions added to the data field, in addition to the fantasy year. Additionally, each position file had to be downloaded separately and combined into one one fantasy stats table per position for 2019-2022. Moreover, the fields also had to have the team for each player added.

Team standings

The data had to be put into a tibble format removing the division and conference. Additionally, the teams moving on to the playoffs had to have the signifying symbols removed from the team name. Moreover, the fields also were standardized by unbolding, removing the underlines, and hyperlinks.

Team offensive stats

The offensive team stats had to have the fields standardized by unbolding, removing the underlines, and hyperlinks.

Strength of schedule (SOS)

The SOS data set had to have the date added to the file. Each file was downloaded separately and appended to each file. The file is as of 2/28 for the year that the season started (2019-2022). This allows for the final SOS to have been calculated.

Data

```r
head(QB_stats)
```

```
## # A tibble: 6 x 21
##    Rank Player      Current_team   CMP   ATT   PCT   YDS   Y.A    TD   INT SACKS
##   <dbl> <chr>       <chr>        <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1 Josh Allen~ Buffalo Bil~   409   646  63.3  4407   6.8    36    15    26
## 2     2 Justin Her~ Los Angeles~   443   672  65.9  5014   7.5    38    15    31
## 3     3 Tom Brady ~ NO TEAM FA     485   719  67.5  5316   7.4    43    12    22
## 4     4 Patrick Ma~ Kansas City~   436   658  66.3  4828   7.3    37    13    28
## 5     5 Matthew St~ Los Angeles~   404   601  67.2  4886   8.1    41    17    30
## 6     6 Aaron Rodg~ New York Je~   366   531  68.9  4115   7.7    37     4    30
## # i 10 more variables: ATT.1 <dbl>, YDS.1 <dbl>, TD.1 <dbl>, FL <dbl>, G <dbl>,
## #   FPTS <dbl>, FPTS.G <dbl>, ROST <dbl>, Fantasy_yr <dbl>, Posit <chr>
```

```r
head(RB_stats)
```

```
## # A tibble: 6 x 21
##    Rank Player      Current_team   ATT   YDS   Y.A    LG  X20.    TD   REC   TGT
##   <dbl> <chr>       <chr>        <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1 Jonathan T~ Indianapoli~   332  1811   5.5    83    30    18    40    51
## 2     2 Austin Eke~ Los Angeles~   206   911   4.4    28     3    12    70    94
## 3     3 Joe Mixon ~ Cincinnati ~   292  1205   4.1    32     7    13    42    48
## 4     4 Najee Harr~ Pittsburgh ~   307  1200   3.9    37     8     7    74    94
## 5     5 James Conn~ Arizona Car~   202   752   3.7    35     4    15    37    39
## 6     6 Ezekiel El~ NO TEAM FA     237  1002   4.2    47     5    10    47    65
## # i 10 more variables: YDS.1 <dbl>, Y.R <dbl>, TD.1 <dbl>, FL <dbl>, G <dbl>,
## #   FPTS <dbl>, FPTS.G <dbl>, ROST <dbl>, Fantasy_yr <dbl>, Posit <chr>
```

```r
head(WR_stats)
```

```
## # A tibble: 6 x 20
##    Rank Player      Current_team   REC   TGT   YDS   Y.R    LG  X20.    TD   ATT
##   <dbl> <chr>       <chr>        <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1 Cooper Kup~ Los Angeles~   145   191  1947  13.4    59    57    16     4
## 2     2 Deebo Samu~ San Francis~    77   121  1405  18.2    83    53     6    59
## 3     3 Ja'Marr Ch~ Cincinnati ~    81   128  1455  18      82    49    13     7
## 4     4 Justin Jef~ Minnesota V~   108   167  1616  15      56    45    10     6
## 5     5 Davante Ad~ Las Vegas R~   123   169  1553  12.6    59    37    11     0
## 6     6 Mike Evans~ Tampa Bay B~    74   114  1035  14      46    30    14     1
## # i 9 more variables: YDS.1 <dbl>, TD.1 <dbl>, FL <dbl>, G <dbl>, FPTS <dbl>,
## #   FPTS.G <dbl>, ROST <dbl>, Fantasy_yr <dbl>, Posit <chr>
```

```r
head(TE_stats)
```

```
## # A tibble: 6 x 20
##    Rank Player      Current_team   REC   TGT   YDS   Y.R    LG  X20.    TD   ATT
##   <dbl> <chr>       <chr>        <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     1 Mark Andre~ Baltimore R~   107   153  1361  12.7    43    27     9     1
## 2     2 Travis Kel~ Kansas City~    92   134  1125  12.2    69    20     9     2
## 3     3 Dalton Sch~ Houston Tex~    78   104   808  10.4    32     8     8     0
## 4     4 George Kit~ San Francis~    71    94   910  12.8    48    17     6     3
## 5     5 Rob Gronko~ NO TEAM FA      55    88   802  14.6    42    20     6     0
## 6     6 Dawson Kno~ Buffalo Bil~    49    71   587  12      53    14     9     0
## # i 9 more variables: YDS.1 <dbl>, TD.1 <dbl>, FL <dbl>, G <dbl>, FPTS <dbl>,
## #   FPTS.G <dbl>, ROST <dbl>, Fantasy_yr <dbl>, Posit <chr>
```

- footballdb (2023a)

```r
tail(SOS)
```

```
## # A tibble: 6 x 7
##   ranking Team         rating    Hi   Low  Last  Year
##     <dbl> <chr>         <dbl> <dbl> <dbl> <dbl> <dbl>
## 1      28 Minnesota      -0.6     3    28    27  2022
## 2      29 Indianapolis   -0.7    10    30    28  2022
## 3      30 Carolina       -0.8    23    32    30  2022
## 4      31 Houston        -0.9    16    32    31  2022
## 5      32 Atlanta        -1      15    32    32  2022
## 6       0 NO TEAM FA      0       0     0     0  2022
```

- Strength of schedule (2023b)

```r
head(team_stand_22)
```

```
## # A tibble: 6 x 7
##   Team                     W     L     T   Pct    PF    PA
##   <chr>                <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Buffalo Bills           13     3     0 0.813   455   286
## 2 Miami Dolphins           9     8     0 0.529   397   399
## 3 New England Patriots     8     9     0 0.471   364   347
## 4 New York Jets            7    10     0 0.412   296   316
## 5 Cincinnati Bengals      12     4     0 0.75    418   322
## 6 Baltimore Ravens        10     7     0 0.588   350   315
```

Sumamry stats

```r
library(dplyr)
# Group data by Player and calculate summary statistics
QB_summary_data <- QB_stats %>%
  group_by(Player,Posit, Current_team) %>%
  summarize(Avg_fant_pts = round(mean(FPTS))) %>%
  arrange(desc(Avg_fant_pts))
```

```
## `summarise()` has grouped output by 'Player', 'Posit'. You can override using
## the `.groups` argument.
```

```r
print(QB_summary_data)
```

```
## # A tibble: 173 x 4
## # Groups:   Player, Posit [173]
##    Player                Posit Current_team         Avg_fant_pts
##    <chr>                 <chr> <chr>                       <dbl>
##  1 Josh Allen (BUF)      QB    Buffalo Bills                 383
##  2 Patrick Mahomes II KC QB    Kansas City Chiefs            369
##  3 Justin Herbert (LAC)  QB    Los Angeles Chargers          343
##  4 Tom Brady FA          QB    NO TEAM FA                    322
##  5 Lamar Jackson (BAL)   QB    Baltimore Ravens              315
##  6 Aaron Rodgers (NYJ)   QB    New York Jets                 314
##  7 Kyler Murray (ARI)    QB    Arizona Cardinals             302
##  8 Russell Wilson (DEN)  QB    Denver Broncos                298
##  9 Kirk Cousins (MIN)    QB    Minnesota Vikings             296
## 10 Joe Burrow (CIN)      QB    Cincinnati Bengals            292
## # i 163 more rows
```

```r
library(dplyr)
# Group data by Player and calculate summary statistics
RB_summary_data <- RB_stats %>%
  group_by(Player,Posit, Current_team) %>%
  summarize(Avg_fant_pts = round(mean(FPTS))) %>%
  arrange(desc(Avg_fant_pts))
```

```
## `summarise()` has grouped output by 'Player', 'Posit'. You can override using
## the `.groups` argument.
```

```r
print(RB_summary_data)
```

```
## # A tibble: 398 x 4
## # Groups:   Player, Posit [398]
##    Player                Posit Current_team         Avg_fant_pts
##    <chr>                 <chr> <chr>                       <dbl>
```

```
##  1 Derrick Henry (TEN)   RB    Tennessee Titans              259
##  2 Dalvin Cook (FA)      RB    NO TEAM FA                    226
##  3 Jonathan Taylor (IND) RB    Indianapolis Colts            223
##  4 Austin Ekeler (LAC)   RB    Los Angeles Chargers          217
##  5 Nick Chubb (CLE)      RB    Cleveland Browns              215
##  6 Aaron Jones (GB)      RB    Green Bay Packers             211
##  7 Najee Harris (PIT)    RB    Pittsburgh Steelers           205
##  8 Josh Jacobs (LV)      RB    Las Vegas Raiders             204
##  9 Alvin Kamara NO       RB    New Orleans Saints            201
## 10 Ezekiel Elliott (FA)  RB    NO TEAM FA                    201
## # i 388 more rows
```

```r
library(dplyr)
# Group data by Player and calculate summary statistics
WR_summary_data <- WR_stats %>%
  group_by(Player,Posit, Current_team) %>%
  summarize(Avg_fant_pts = round(mean(FPTS))) %>%
  arrange(desc(Avg_fant_pts))
```

```
## `summarise()` has grouped output by 'Player', 'Posit'. You can override using
## the `.groups` argument.
```

```r
print(WR_summary_data)
```

```
## # A tibble: 626 x 4
## # Groups:   Player, Posit [626]
##    Player                 Posit Current_team         Avg_fant_pts
##    <chr>                  <chr> <chr>                       <dbl>
##  1 Justin Jefferson (MIN) WR    Minnesota Vikings             216
##  2 Davante Adams (LV)     WR    Las Vegas Raiders             207
##  3 Tyreek Hill (MIA)      WR    Miami Dolphins                195
##  4 Ja'Marr Chase (CIN)    WR    Cincinnati Bengals            190
##  5 Stefon Diggs (BUF)     WR    Buffalo Bills                 186
##  6 Cooper Kupp (LAR)      WR    Los Angeles Rams              179
##  7 Mike Evans (TB)        WR    Tampa Bay Buccaneers          170
##  8 A.J. Brown (PHI)       WR    Philadelphia Eagles           168
##  9 CeeDee Lamb (DAL)      WR    Dallas Cowboys                164
## 10 Jaylen Waddle (MIA)    WR    Miami Dolphins                163
## # i 616 more rows
```

```r
library(dplyr)
# Group data by Player and calculate summary statistics
TE_summary_data <- TE_stats %>%
  group_by(Player,Posit, Current_team) %>%
  summarize(Avg_fant_pts = round(mean(FPTS))) %>%
  arrange(desc(Avg_fant_pts))
```

```
## `summarise()` has grouped output by 'Player', 'Posit'. You can override using
## the `.groups` argument.
```

```r
print(TE_summary_data)
```

```
## # A tibble: 346 x 4
## # Groups:   Player, Posit [346]
##    Player               Posit Current_team         Avg_fant_pts
##    <chr>                <chr> <chr>                       <dbl>
##  1 Travis Kelce (KC)    TE    Kansas City Chiefs            186
```

```
##   2 Mark Andrews (BAL)   TE    Baltimore Ravens               142
##   3 George Kittle (SF)   TE    San Francisco 49ers            121
##   4 Darren Waller (NYG)  TE    New York Giants                109
##   5 Jared Cook (FA)      TE    NO TEAM FA                     100
##   6 T.J. Hockenson (MIN) TE    Minnesota Vikings               93
##   7 Hunter Henry (NE)    TE    New England Patriots            89
##   8 Dallas Goedert (PHI) TE    Philadelphia Eagles             88
##   9 Pat Freiermuth (PIT) TE    Pittsburgh Steelers             88
##  10 Mike Gesicki (NE)    TE    New England Patriots            87
## # i 336 more rows
```

```r
all_players <- rbind(QB_summary_data, RB_summary_data, WR_summary_data, TE_summary_data)

print(all_players)
```

```
## # A tibble: 1,543 x 4
## # Groups:   Player, Posit [1,543]
##    Player                Posit Current_team         Avg_fant_pts
##    <chr>                 <chr> <chr>                       <dbl>
##  1 Josh Allen (BUF)      QB    Buffalo Bills                 383
##  2 Patrick Mahomes II KC QB    Kansas City Chiefs            369
##  3 Justin Herbert (LAC)  QB    Los Angeles Chargers          343
##  4 Tom Brady FA          QB    NO TEAM FA                    322
##  5 Lamar Jackson (BAL)   QB    Baltimore Ravens              315
##  6 Aaron Rodgers (NYJ)   QB    New York Jets                 314
##  7 Kyler Murray (ARI)    QB    Arizona Cardinals             302
##  8 Russell Wilson (DEN)  QB    Denver Broncos                298
##  9 Kirk Cousins (MIN)    QB    Minnesota Vikings             296
## 10 Joe Burrow (CIN)      QB    Cincinnati Bengals            292
## # i 1,533 more rows
```

```r
all_sum_team <- all_players %>%
  group_by(Current_team) %>%
  summarize(ttl_Avg_fant_pts =sum(Avg_fant_pts))

print(all_sum_team)
```

```
## # A tibble: 33 x 2
##    Current_team       ttl_Avg_fant_pts
##    <chr>                         <dbl>
##  1 Arizona Cardinals               974
##  2 Atlanta Falcons                 760
##  3 Baltimore Ravens               1212
##  4 Buffalo Bills                  1337
##  5 Carolina Panthers               978
##  6 Chicago Bears                  1013
##  7 Cincinnati Bengals             1082
##  8 Cleveland Browns               1064
##  9 Dallas Cowboys                  996
## 10 Denver Broncos                  954
## # i 23 more rows
```

```r
sos_avg_rank <- SOS %>%
  group_by(Team) %>%
  summarize(avg_sos  = round(mean(ranking)))
# the higher the ranking the harder the schedule
```

```r
print(sos_avg_rank)
```

```
## # A tibble: 33 x 2
##    Team       avg_sos
##    <chr>        <dbl>
##  1 Arizona         13
##  2 Atlanta         18
##  3 Baltimore       11
##  4 Buffalo         12
##  5 Carolina        19
##  6 Chicago         14
##  7 Cincinnati      18
##  8 Cleveland       18
##  9 Dallas          18
## 10 Denver          18
## # i 23 more rows
```

```r
all_sum_team <- all_sum_team %>% mutate(avg_sos_rank = sos_avg_rank$avg_sos)

print(all_sum_team)
```

```
## # A tibble: 33 x 3
##    Current_team       ttl_Avg_fant_pts avg_sos_rank
##    <chr>                         <dbl>        <dbl>
##  1 Arizona Cardinals               974           13
##  2 Atlanta Falcons                 760           18
##  3 Baltimore Ravens               1212           11
##  4 Buffalo Bills                  1337           12
##  5 Carolina Panthers               978           19
##  6 Chicago Bears                  1013           14
##  7 Cincinnati Bengals             1082           18
##  8 Cleveland Browns               1064           18
##  9 Dallas Cowboys                  996           18
## 10 Denver Broncos                  954           18
## # i 23 more rows
```

```r
library(lm.beta)
model_1 <- lm( ttl_Avg_fant_pts ~ avg_sos_rank, data = all_sum_team)
summary(model_1)
```

```
##
## Call:
## lm(formula = ttl_Avg_fant_pts ~ avg_sos_rank, data = all_sum_team)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1807.5  -579.2   -94.3   408.4  7117.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3463.19     740.38   4.678  5.4e-05 ***
## avg_sos_rank  -134.16      43.61  -3.077  0.00435 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1486 on 31 degrees of freedom
## Multiple R-squared:  0.2339, Adjusted R-squared:  0.2092
## F-statistic: 9.466 on 1 and 31 DF,  p-value: 0.004351
```

```
standardized_betas_model_1 <- lm.beta(model_1)
print(standardized_betas_model_1)
```

```
##
## Call:
## lm(formula = ttl_Avg_fant_pts ~ avg_sos_rank, data = all_sum_team)
##
## Standardized Coefficients::
##  (Intercept) avg_sos_rank
##           NA    -0.483653
```

Datsets

Strength of schedule (as of 9/1 each year):

https://www.teamrankings.com/nfl/ranking/schedule-strength-by-other?date=2022-09-01

Fantasy scoring points by two types of systems:

https://www.footballdb.com/fantasy-football/index.html

Team standings: https://www.nfl.com/standings/division/2022/REG

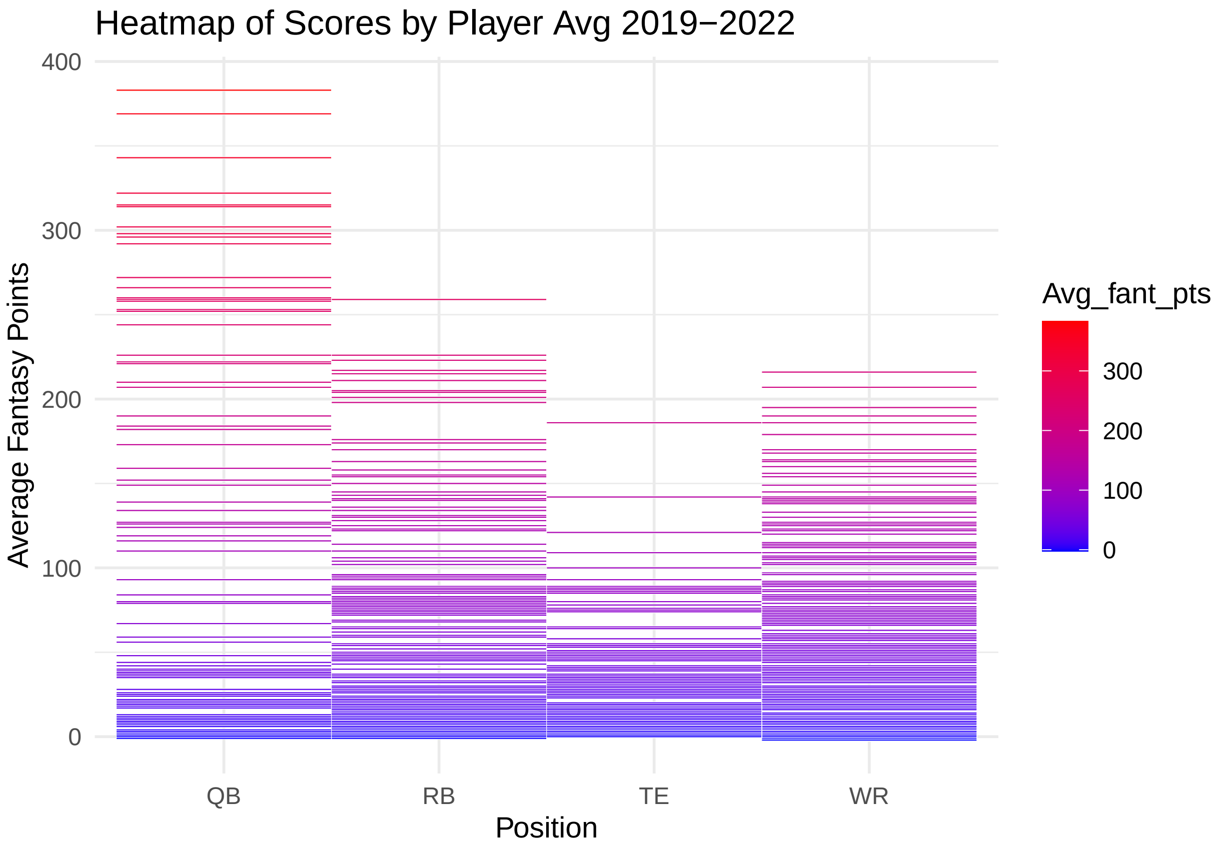
Required Packages

'tidyverse' 'ggplot2' 'dplyr' 'lm.beta'

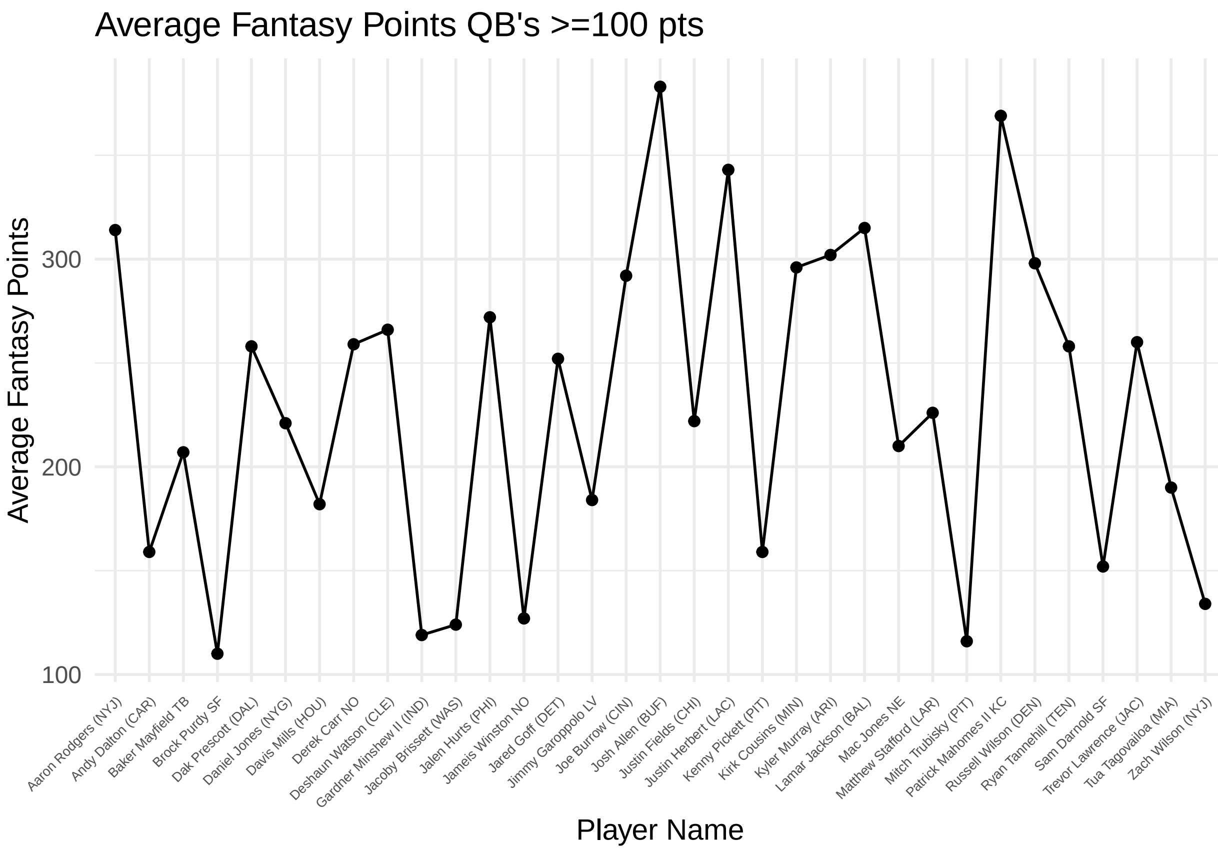
Plots and tables needed

1. Line plots to visualize trends in player performance over time.
2. Tables to display the top-performing players based on historical fantasy points.
3. Heat map to help understand which positions are more important to draft due to outliers in scoring by position.

Plots




```
QB_filtered_players <-
  subset(QB_summary_data, Current_team != "NO TEAM FA" &
           Avg_fant_pts >= 100)
RB_filtered_players <- subset(RB_summary_data, Current_team != "NO TEAM FA")
WR_filtered_players <- subset(WR_summary_data, Current_team != "NO TEAM FA")
TE_filtered_players <- subset(TE_summary_data, Current_team != "NO TEAM FA")
```



Question for future steps

1. How can I incorporate additional data, such as team dynamics (coaching issues, player issues), weather conditions, or player off-field behavior, to improve the model's predictions?
2. Can I build an ensemble model that combines the strengths of multiple algorithms to achieve better prediction accuracy?
3. How can I continuously update the model during the NFL season to account for emerging trends and player performance changes?
4. Can I use this research for in-season predictors?
5. Review the correlations and adjust for outliers, also adjust for historical teams.
6. Once historical teams added to player then review correlation for winning percentages to player FFB points totals.
7. Provide a recommendation for the most effective method to help predict fantasy football drafting rankings.

References

2023a. https://www.footballdb.com/fantasy-football/index.html.

———. 2023b. https://www.teamrankings.com/nfl/ranking/schedule-strength-by-other?date=2020-02-28.