NFL Draft Combine Numbers and Rookie of the Year Winners

By: Sean Marren & Michael Merfeld

1. Introduction

Every NFL season comes around, and at the end of the season, there are 31 losers. At the end of the season, these 31 teams turn to the draft to find the players that will take them over the hump and make them winners. The first step in the NFL draft process is the NFL draft combine. We plan to use these combine numbers and examine the role that good combine numbers play in being the Rookie of the Year.

In this project, we will utilize data on draft combine results from players on Kaggle.com. As well as rookie of the year winners from Wikipedia.

2. Data

This project uses two primary places for our data. The first is on Kaggle where we collected data on draft combine results. The second is Wikipedia for information on rookie of the year winners.

2.1 Rookie of the Year Winners

To identify and collect our data, we consulted the Wikipedia article based on offensive rookie of the year winners. This page contains lists with each player's name, team, position, and year that they won. Each entity is accompanied by a hyperlink to each player's own Wikipedia page.

Based on the hyperlinks from the rookie of the year Wikipedia page, we wrote a scraping script that took the main data points of information we needed to perform our analysis. We scraped year, name, and team. This process produced a main data frame containing 65 data results. After, we cleaned the data to align it with our combine results from 2000-2018. This resulted in a final 19 observations. We included the scraping code in the R script file “final_project_everything.R”.

2.2 NFL Draft Combine Results

The Kaggle dataset is a collection of 6218 values. This is broken down by each players overall draft combine statistics when participating.

Kaggle divided the dataset into three main entities: player, position, and team they were drafted by. The data was downloaded into a csv file: “combine_data_since_2000_PROCESSED_2018-04-26.csv". After downloading the file, we removed unnecessary fields including height, weight, forty, bench, and other additional combine statistics.

After the initial data cleaning was completed, we wanted to decrease the number of results by eliminating players that had not been drafted after their combine. This eliminates our results from 6218 observations down to 3738 observations. The preprocessing code is included in our R script file “final_project_everything.R”.

2.3 Integrate Rookie of the Year Winner and NFL Draft Combine Results

Because both datasets have similar columns in common, we decided that the best way to consolidate is to make the player name the unique key method. 7 new variables were added to the data after integration was completed. This formed 3757 rows and 7 columns. We included an

inscription of each variable in table 1 below. The data integration and cleaning code is included in our R script file "final_project_everything.R".

Table 1. Data Dictionary

Column	Type	Description
Player	text	Name of the player
Pos	text	Position of the player
Year	numeric	Year player was drafted
Team	text	NFL team that drafted a player
Round	numeric	Round in which player was drafted
Pick	numeric	Pick that the player was drafted
ROY Winner	text	Player was awarded ROY that year

3. Analysis

The goal of this project is to see if there are any similarities between a player that participated in the NFL combine and someone winning Rooking of the Year.

3.1 Team that won the Most ROY Awards

What team produced the most Rookie of the Year winners from 2000-2018?

For this section we wanted to take the merged data and see if there are any significant outliers when it comes to Rookie of the Year winners. To do this, we create a table summary of just the Team and RoY winner from 2000-2018. From there, it was necessary to clean out any player that had not won the award before. This was crucial in determining a qplot for the necessary analysis.

Using a qplot, we sorted the geom by "col" and the xlab and ylab as "Team" and "RoY". After our qplot was uploaded by were able to determine that from 2000-2018 there were four teams that had more than one winner. The teams were: Denver Broncos, Minnesota Vikings, New York Giants, and St.Lous(LA) Rams. This was very interesting because each year a team only has a 1 in 32 chances of winning or a .03%.

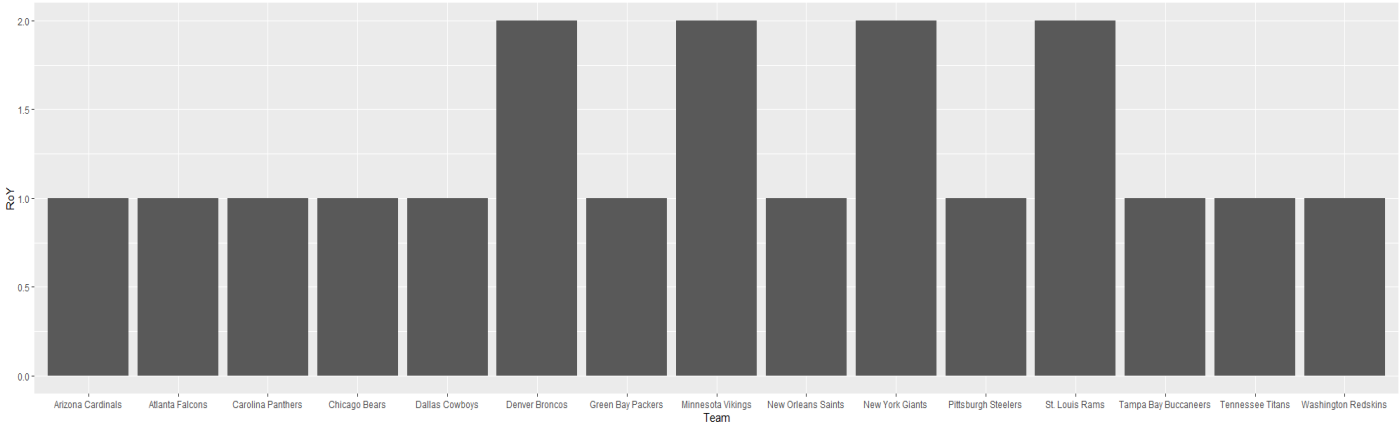
Below, we have added a picture to illustrate our findings as well as the table summary.

Table 2. Table Summary

Team	RoY
Denver Broncos	2
Chicago Bears	1
Arizona Cardinals	1
Pittsburgh Steelers	1

Tampa Bay Buccaneers	1
Tennessee Titans	1
Minnesota Vikings	2
Atlanta Falcons	1
St. Louis Rams	2
Carolina Panthers	1
Washington Redskins	1
Green Bay Packers	1
New York Giants	2
Dallas Cowboys	1
New Orleans Saints	1

Figure 1. Teams with RoY winner from 2000-2018



3.2 Correlation between what round and pick a player was drafted

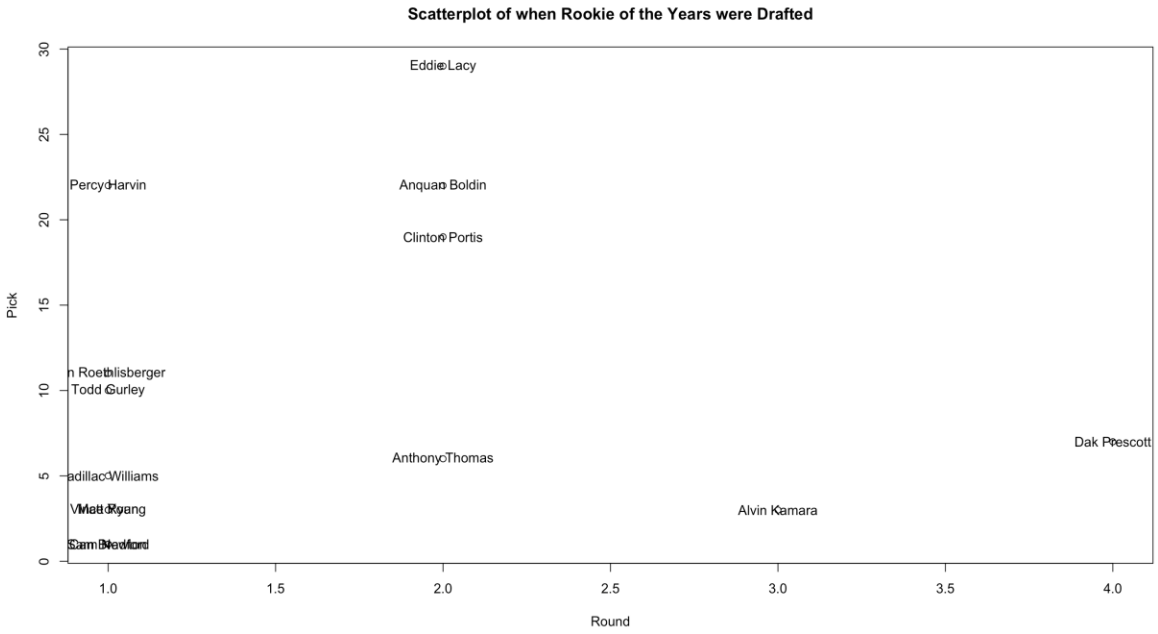
Is there a correlation between what round and pick a player is drafted by and winning RoY?

In this section, we wanted to see what the main factor was in what round a player is drafted and to see if there is any correlation to winning RoY. In order to do this, we created a scatterplot that would help determine at what point in the draft each RoY was drafted.

After careful analysis we were able to see that most of the RoY winners were picked in the first 2 rounds. This was interesting because we can see that the only outlier in this scatterplot was Dak Prescott who was picked in the fourth round. There were two clusters as well of players picked within the first 5 picks of the first round. This helped us in determining how a player will produce in their rookie year after being selected high in the draft.

This showed us that a player being selected within the first two rounds has a higher percent chance of winning RoY than someone who is drafted later.

Figure 2. Scatterplot of when Rookie of the Years were Drafted



3.3 Analysis of player positions compared to winning RoY

What position on offense gives a player the highest chance to win RoY?

For our final analysis we wanted to see what position on offense will produce the highest percent chance of winning RoY. In order to do this, we had to model our prediction similar to our first question. Before running the qplot, we had to clean out unnecessary data. First, we created a new table summary position_summary and kept only position and RoY winner. After getting rid of the unwanted data, we went on and created our bar plot.

After reviewing the information, we concluded that running back provides the most RoY winners from 2000-2018 with 9. The second highest was quarterback with 7. This showed that during the duration of our analysis a running back has a 47% chance of winning the award. For future predictions, it is safe to assume that in years to come a running back will most likely win the award.

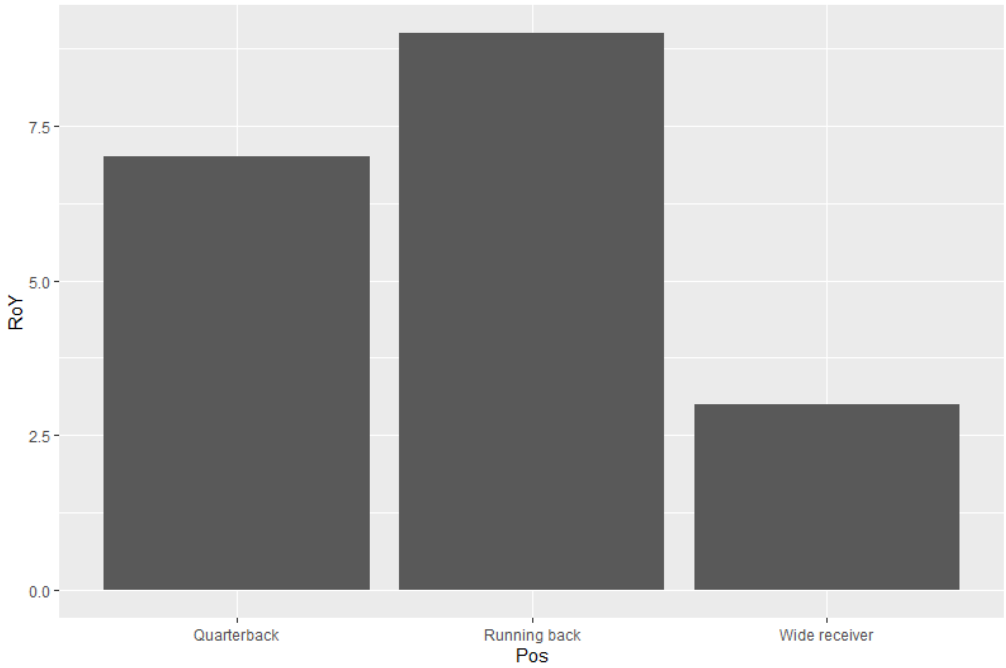
Below, we have included a table summary of our statistics on player position and them winning RoY. Also, we included the qplot of our findings.

Table 3. Position Summary

Position	Number of Winners
Running Back	9
Quarterback	7

Wide Receiver	3

Figure 3. Qplot of Position and Rookie of the Year Winners



We included our data in "final_project_everything.R"

4. Conclusion

In conclusion, we were able to find that certain teams have drafted more rookies of the years through a bar plot. Of the 15 teams who have drafted an offensive rookie of the year, 4 teams have drafted multiple. These teams include the New York Giants, the Denver Broncos, the Minnesota Vikings, and the St. Louis Rams, now the Los Angeles Rams. When it comes to the draft placement of rookies of the year, most of the winners have been drafted within the first 2 rounds, with Dak Prescott being an outlier. We were able to discover that most players were drafted within the first two rounds by using a scatter plot with the round being on the X axis, and the pick within the respective round on the y axis. Position wise, only 3 positions have ever won an award, with running backs having the most at 9 winners. For this, we used another bar graph to represent each Position. Through these graphs, and our findings, it can help NFL teams to learn more about how to discover their next great superstar.

Works Cited

https://www.kaggle.com/datasets/savvastj/nfl-combine-data?resource=download
https://en.wikipedia.org/wiki/Associated_Press_NFL_Rookie_of_the_Year_Award#cite_note-AP2018-66