G.O.A.T.S

An Elementary Adventure in Statistical Analysis

By:
Justin Wallander

Goals:

- Evaluate the greatest statistical performances of All-Time across multiple sports
- Determine whether or not the MVP for a given season generally has one of the best performances for that season/ in general

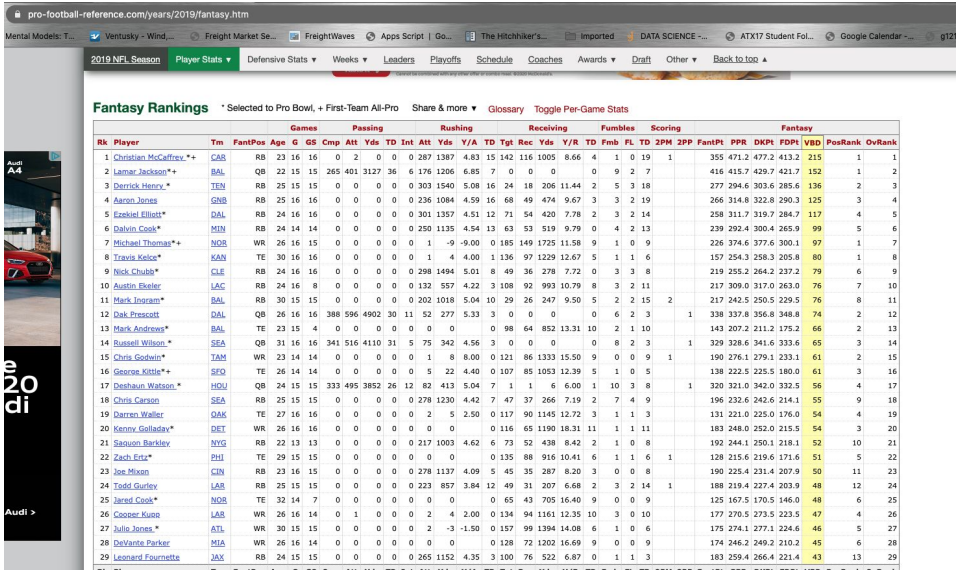
Challenges:

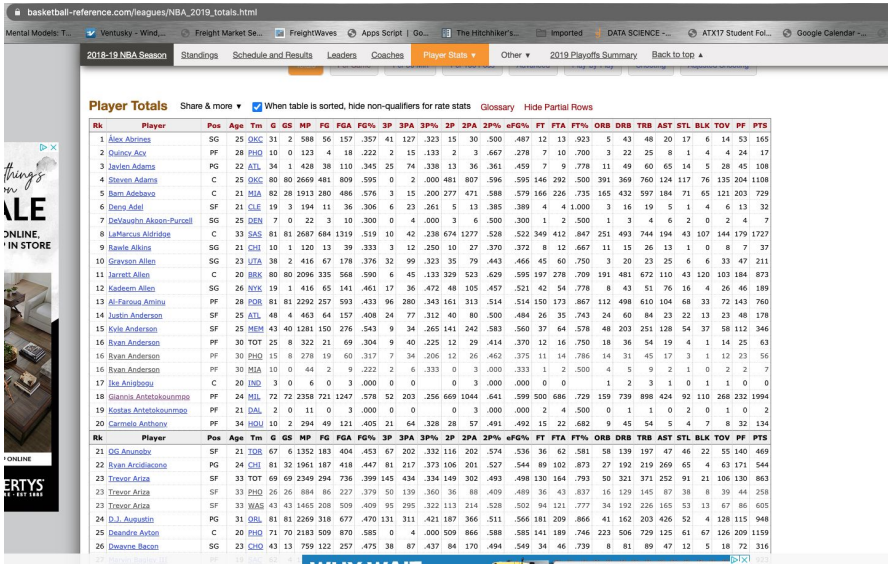
- Figuring out what data to run hypothesis tests on
- Keeping things simple/ focusing on the simple things

Data:

https://www.pro-football-reference.com/years/2019/fantasy.htm

https://www.basketball-reference.com/leagues/NBA_2019_totals.html

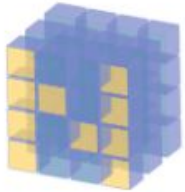




Tech:

SciPy

Seaborn



NumPy

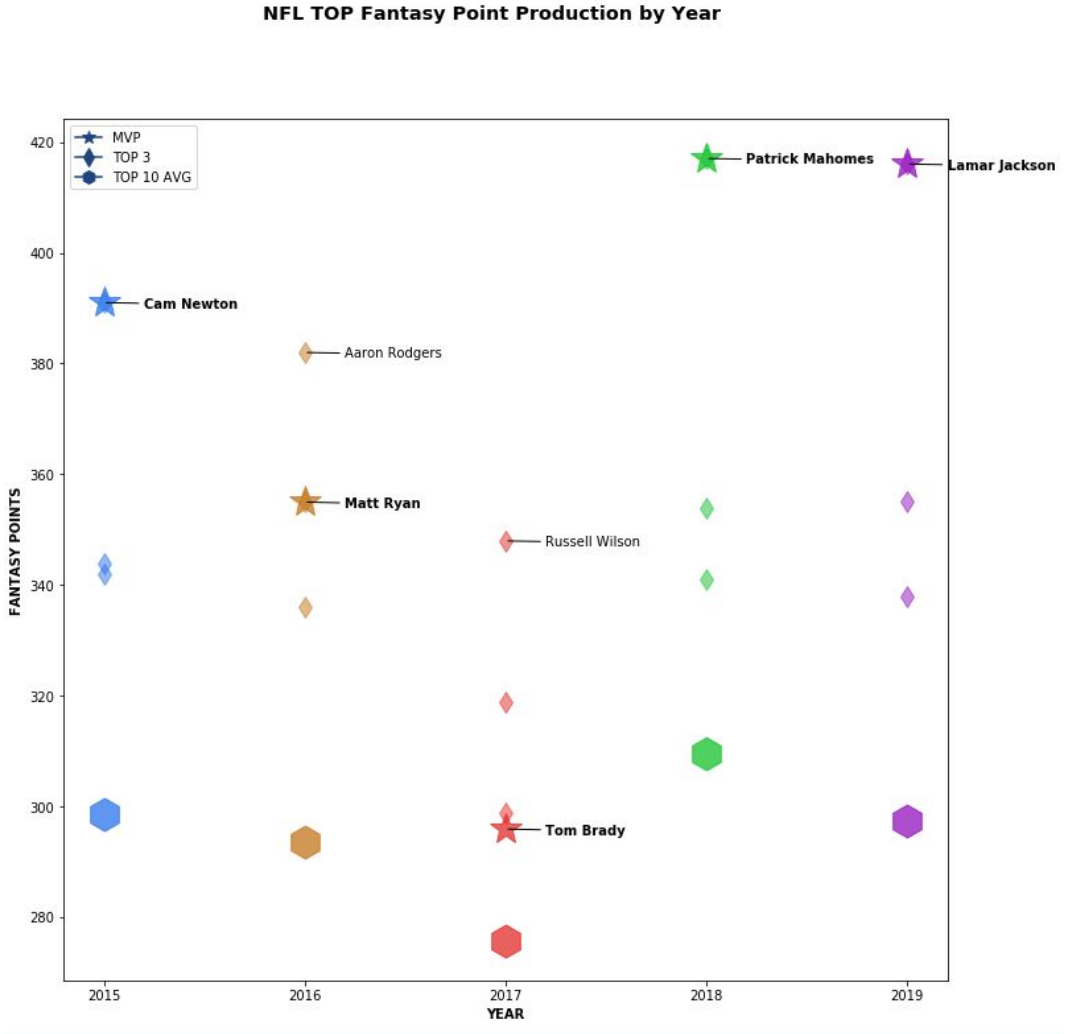
Beautiful Soup

matplotlib

Experiment Design:

- Performance metric : Total Fantasy Point (basic formula) per season for each player who played 50% or greater games dating back to 1970 for NFL (as far back as database went) and 1977 for NBA (first year after NBA/ABA merger)
- Hypothesis test 1:
  - Null Hypothesis = MVP Player fantasy point production is in the top 10 percent of performances for that year 99% of the time.
  - Alternate Hypothesis = MVP Player fantasy point production is in the top 10 percent of performances for that year less than 99 % of the time.

Hypothesis Test 1:

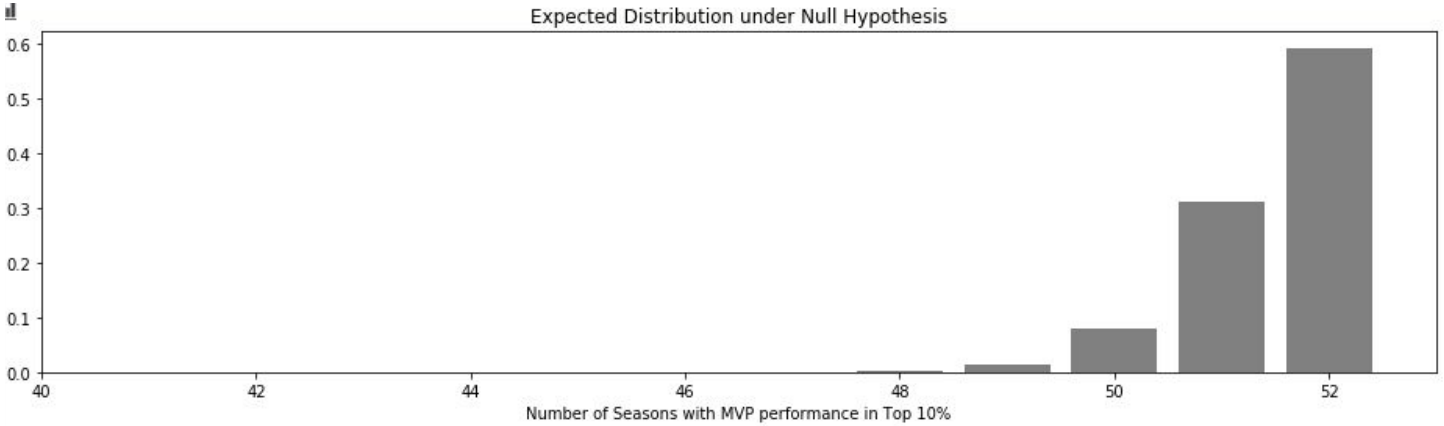


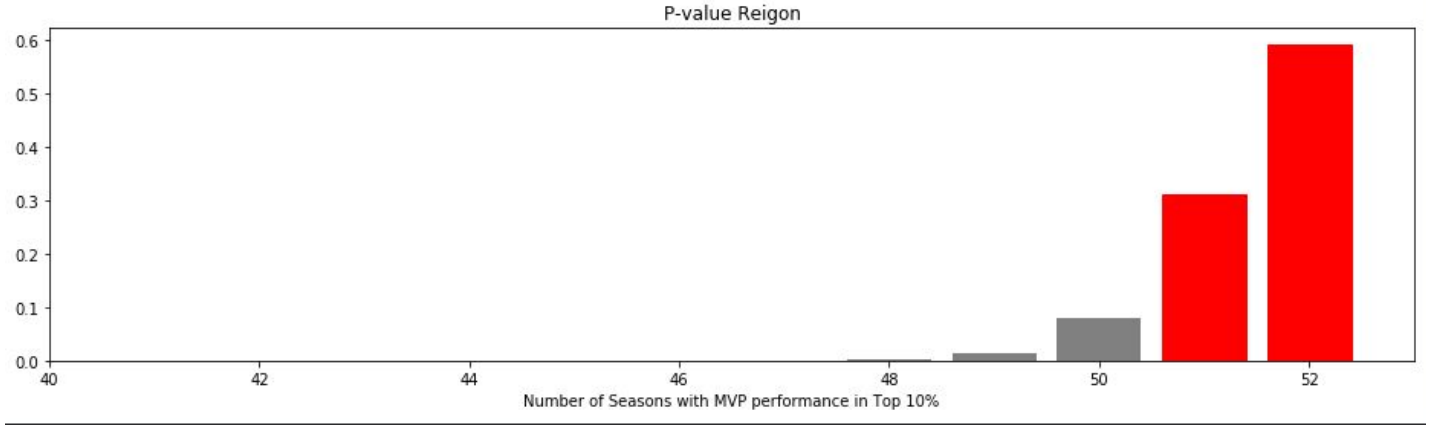
Hypothesis Test 1: Binomial Testing

P-value = 0.9

Alpha = 0.05

Result: I do not reject the null hypothesis



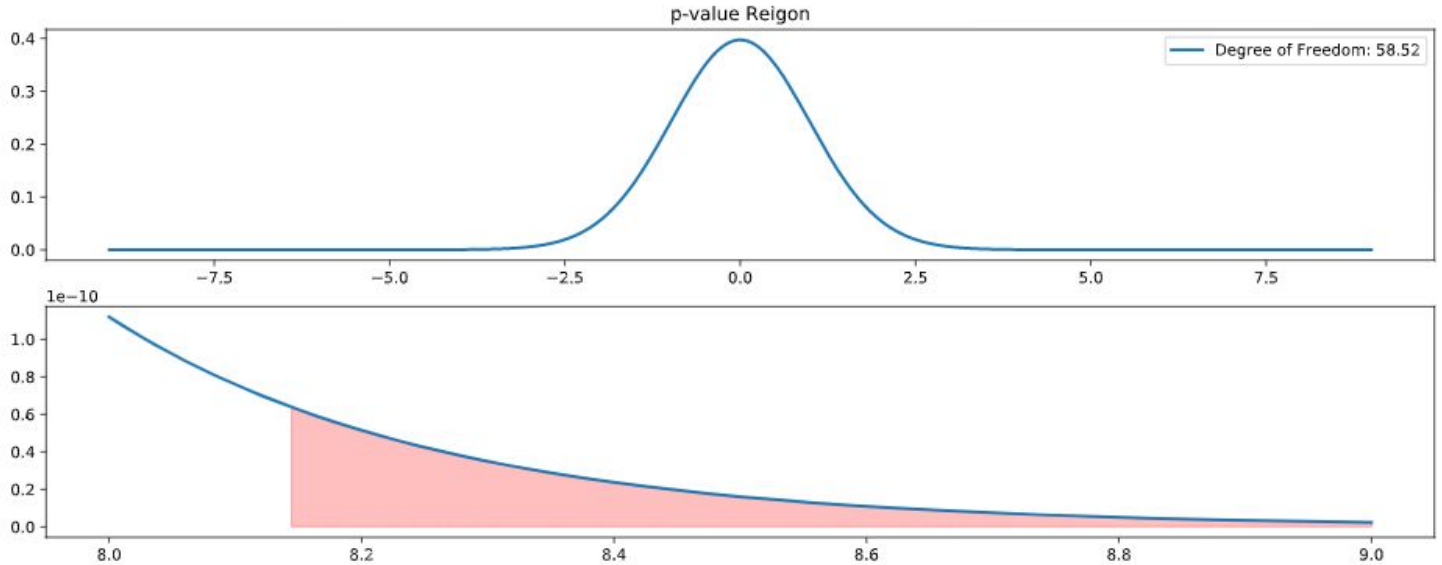


Experiment Design:

- Performance metric : Total Fantasy Point (basic formula) per season for each player who played 50% or greater games dating back to 1970 for NFL (as far back as database went) and 1977 for NBA (first year after NBA/ABA merger)
- Hypothesis test 2:
  - Null Hypothesis = There is no difference between mean mvp performance and the mean performance of the top 10 percent.
  - Alternate Hypothesis = There is a difference, mean MVP performances are greater than the mean Top 10 percent.

Hypothesis Test 2: Welsh's T-Test

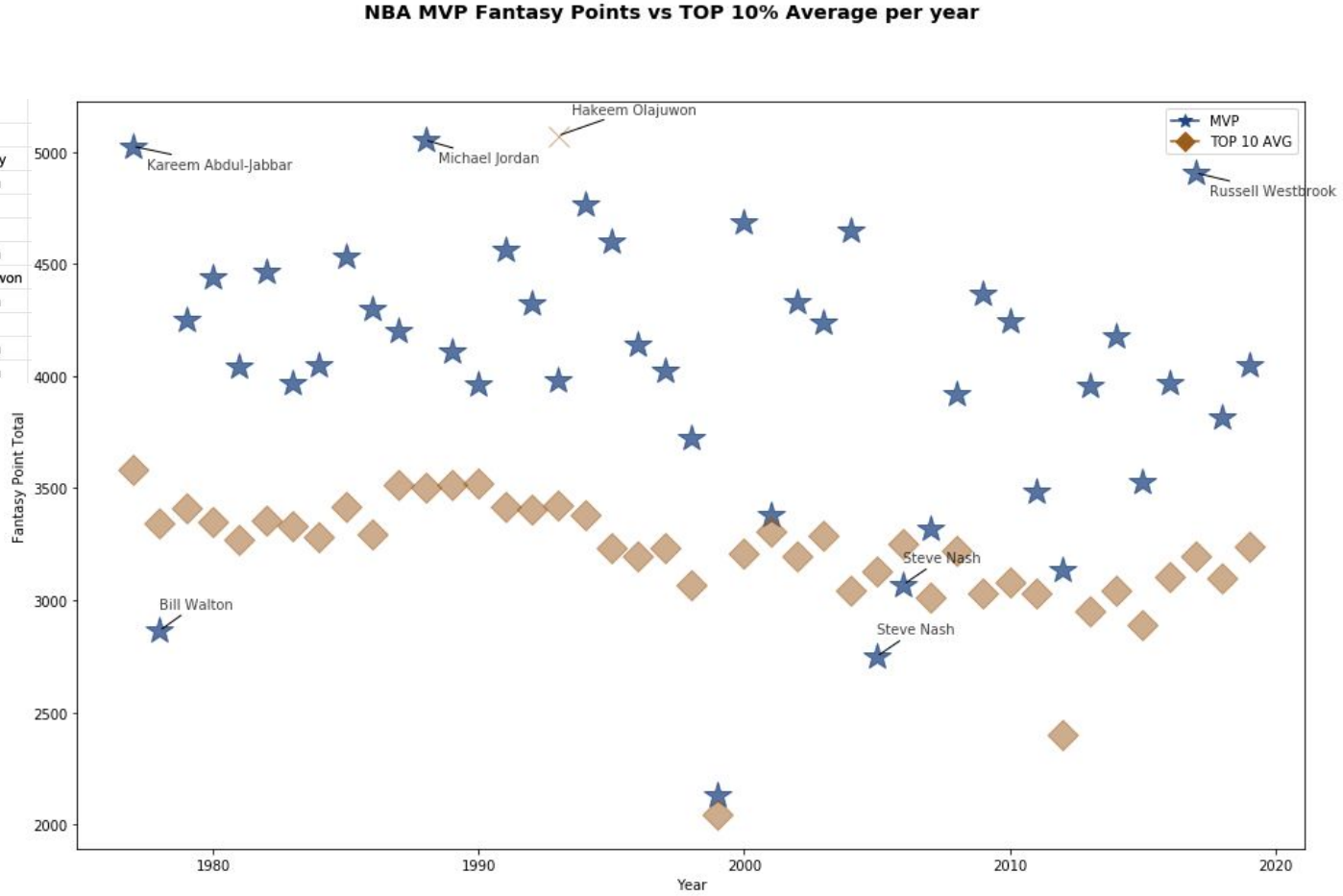
- Test Stat = -8.14 Degrees of Freedom = 58.52 Alpha = 0.05
- The P-value for H0, different average performance between MVP and Top 10 percent: 0.000000000033
- The P-value for average MVP performance being greater than average top 10 percent performance: 0.000000000017
- I reject the Null Hypothesis and can conclude that the Average MVP performance is greater than the average performance in the Top 10%



Hypothesis Test 2:

Top 10 Performances:

Player:	Year:	FantPt:	MVP:	
Hakeem Olajuwon	1993	5072.1	False	MVP: Charles Barkley
Hakeem Olajuwon	1990	5058.8	False	MVP: Magic Johnson
Michael Jordan	1988	5052.3	True	
Kareem Abdul-Jabbar	1977	5024.5	True	
Michael Jordan	1989	4997.4	False	MVP: Magic Johnson
David Robinson	1994	4939.5	False	MVP: Hakeem Olajuwon
Michael Jordan	1987	4933.5	False	MVP: Magic Johnson
Russell Westbrook	2017	4905.8	True	
Michael Jordan	1990	4805.5	False	MVP: Magic Johnson
Hakeem Olajuwon	1989	4793.5	False	MVP: Magic Johnson



Next Steps:

- Add MLB and College Sports
- Use different Performance Metrics/ Points of Comparison
- Rework web scraper to build a more robust database of stats
- ULTIMATE STEP: Use that database to build a machine learning model to help with fantasy drafts

Thanks!

Email : justin.wallander@gmail.com

Github: https://github.com/justin-wallander

LinkedIn: https://www.linkedin.com/in/justin-wallander/