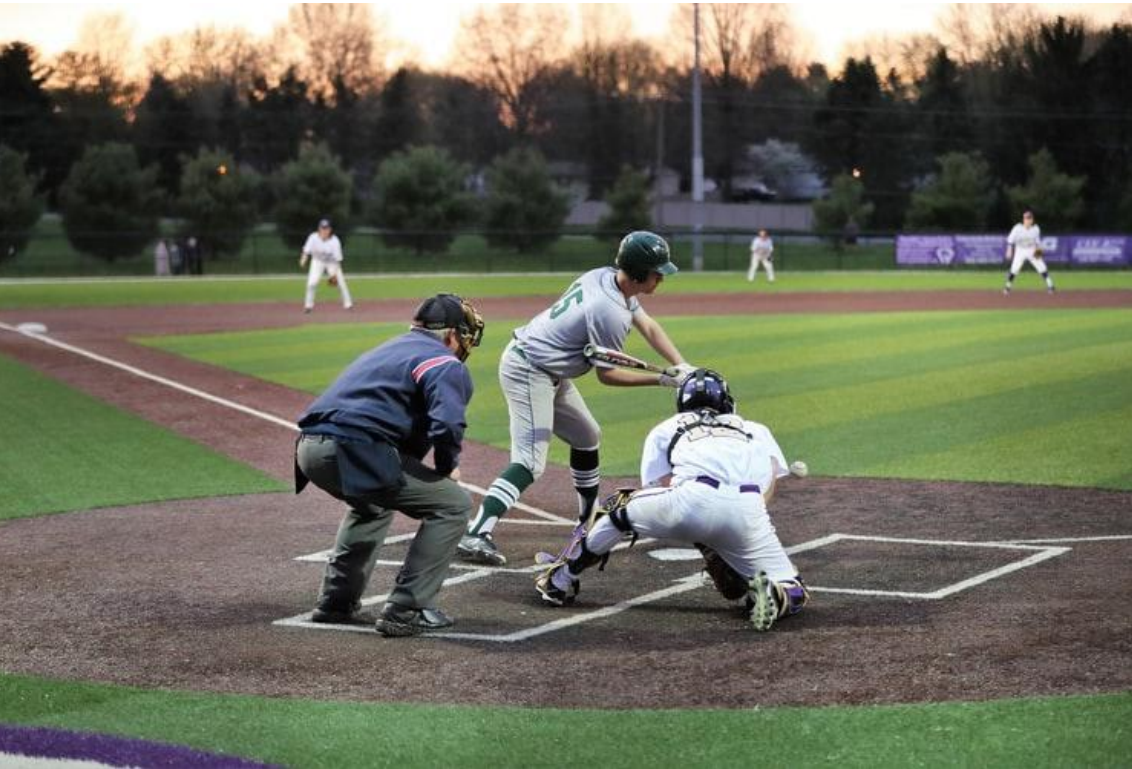


A photograph of a baseball pitcher in a white uniform with a red cap, captured in the middle of a pitching motion. The background shows a blurred crowd of spectators in a stadium. The entire image is overlaid with a semi-transparent orange filter.

Baseball Sabermetrics

Anthony Zurke

Introduction and Problem Statement



In baseball, a pitcher's success is heavily determined on the outcomes of a batted ball, and a swing and miss. Finding ways to increase success can come in many forms, velocity, spin rate, horizontal, and vertical break all play a key role.

I've used advanced sabermetrics from pitcher's from the past three seasons (2019-2021) to be able to assess how to use this data to find in-game success. To do this I used an unsupervised cluster model using KMeans

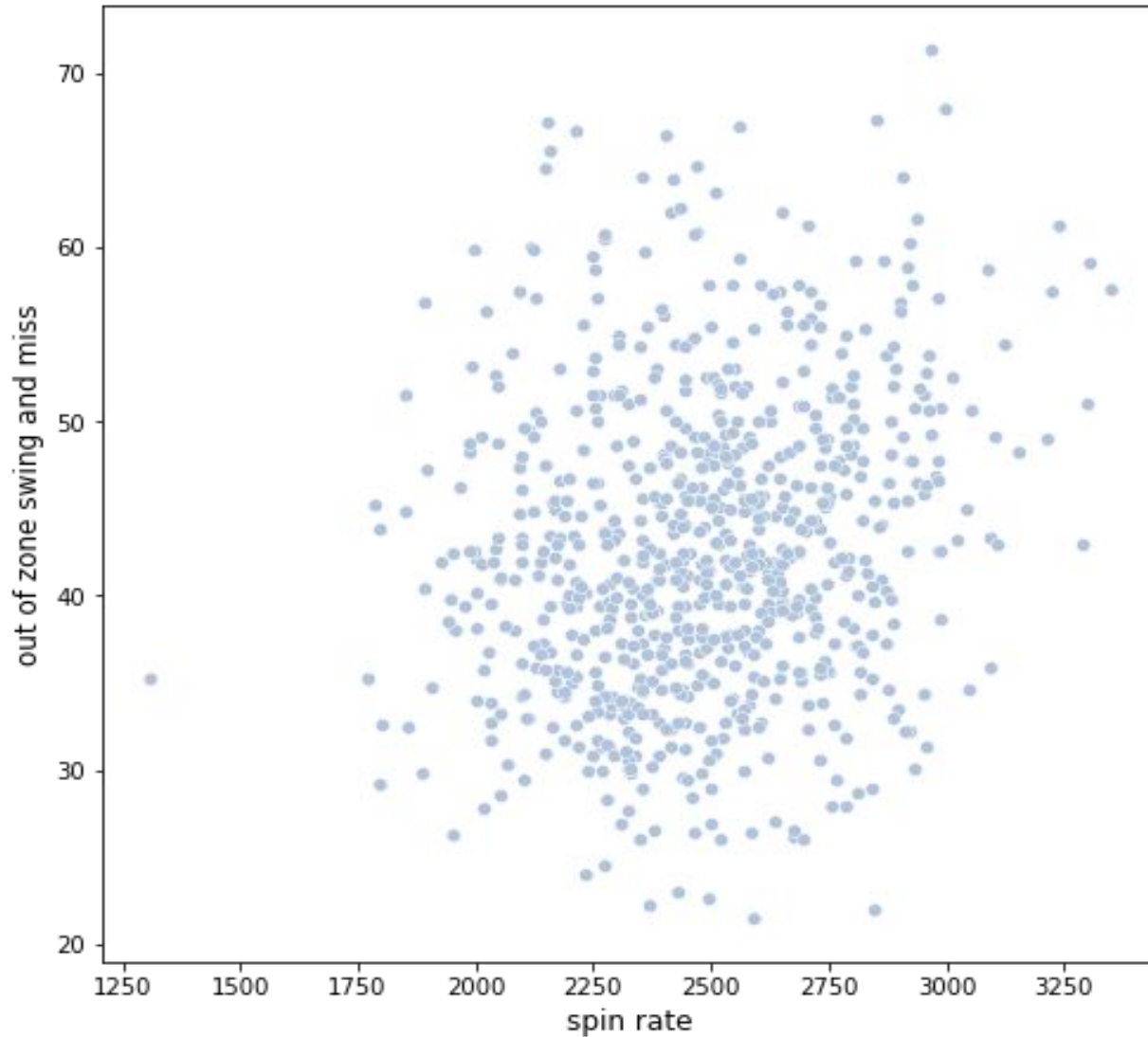
Data Gathering and Cleaning

- All pitchers used had a minimum batters faced of 200 for the 2019 and 2021 data and in the 2020 data minimum of 75 batters faced due to a shortened season,
- pitch types: 4 seam fastball, slider, curveball, changeup, and cutter.
- Imputed nulls with 0

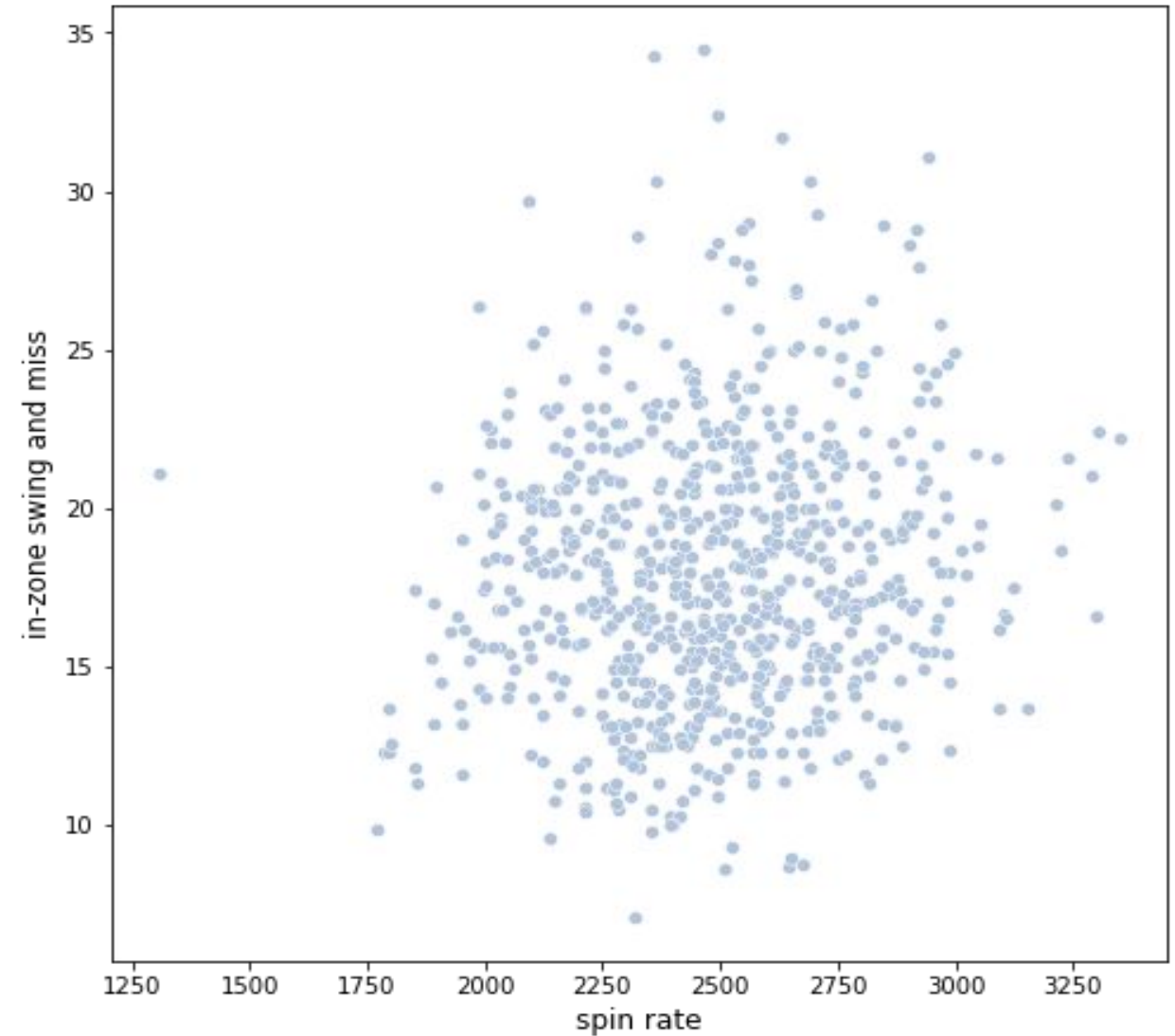


Fastball Velocity and Spin Rate

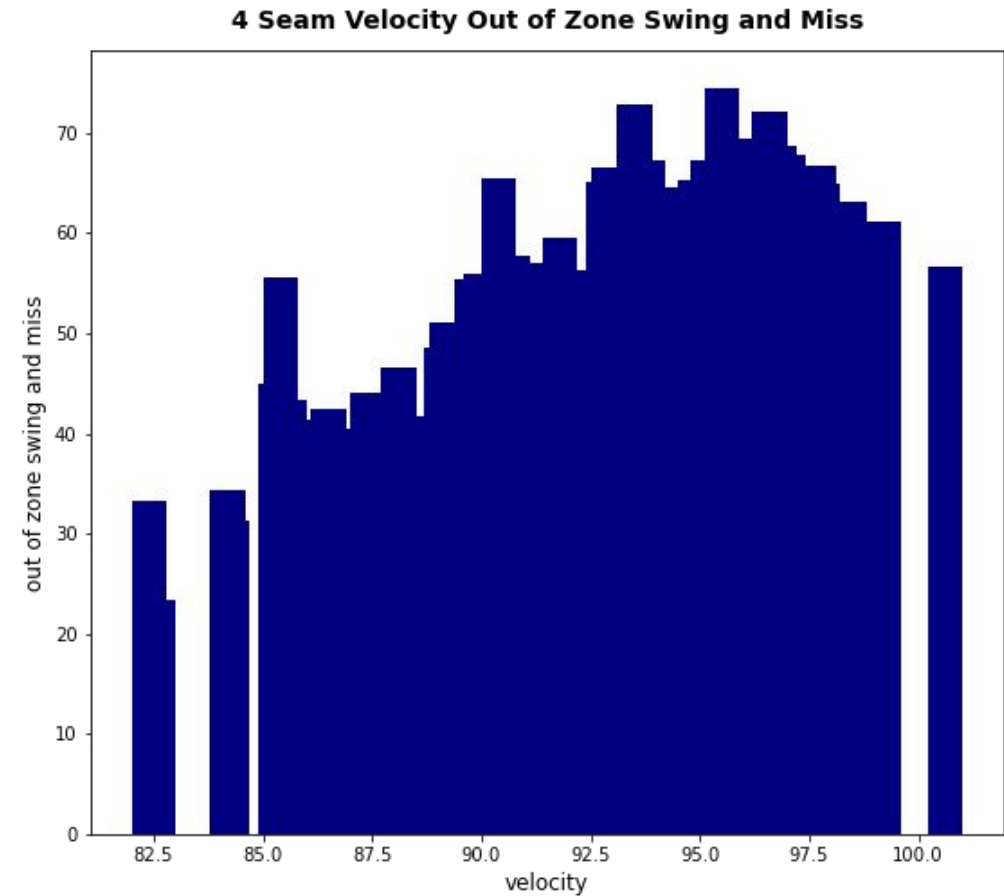
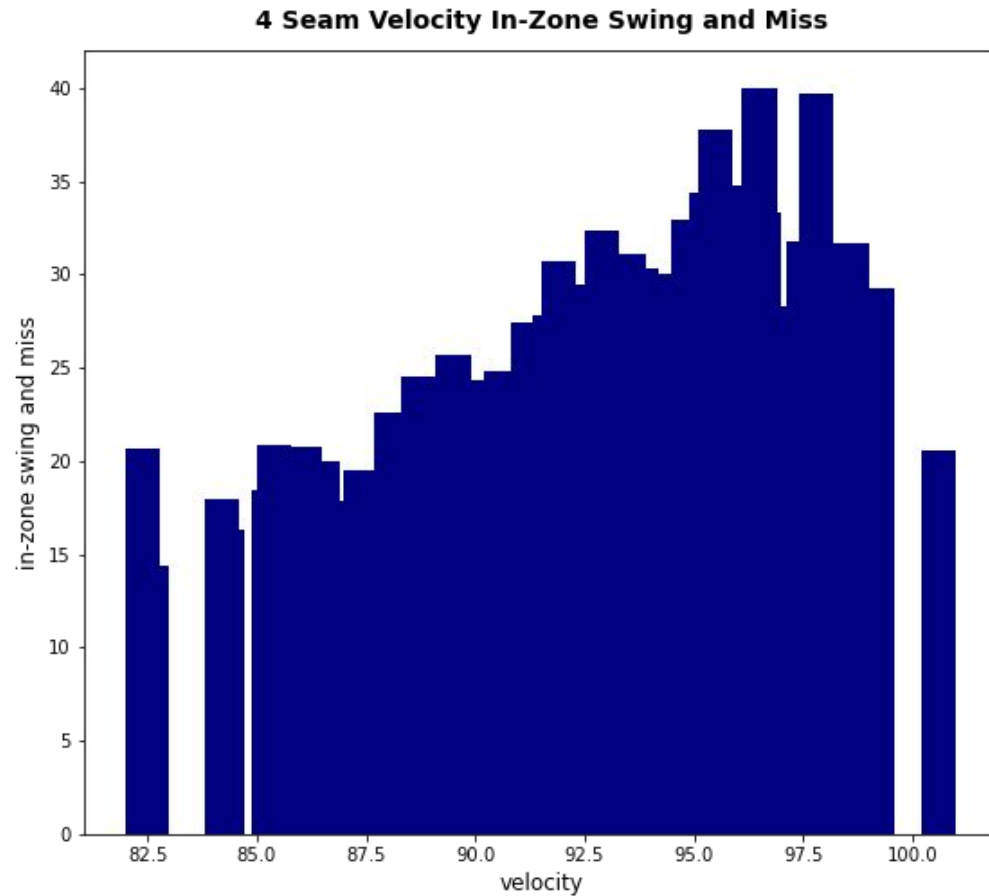
Curveball Spin Rate Out of Zone Swing and Miss



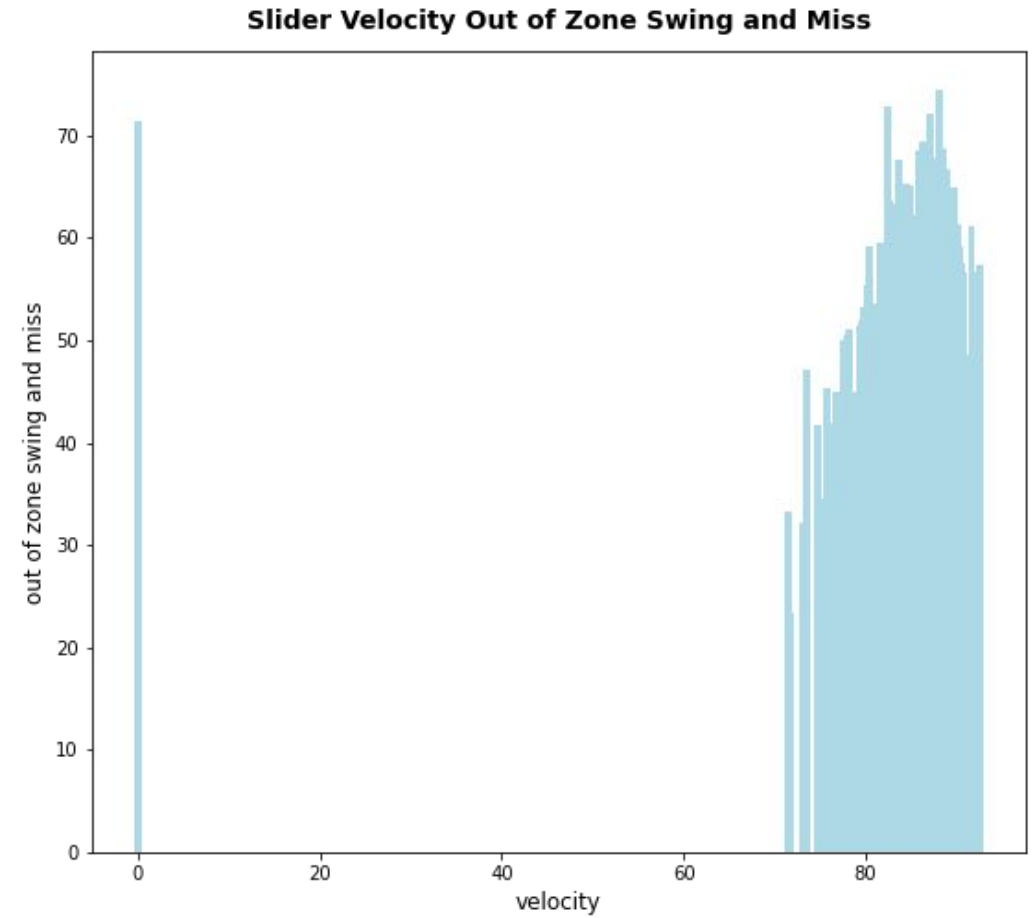
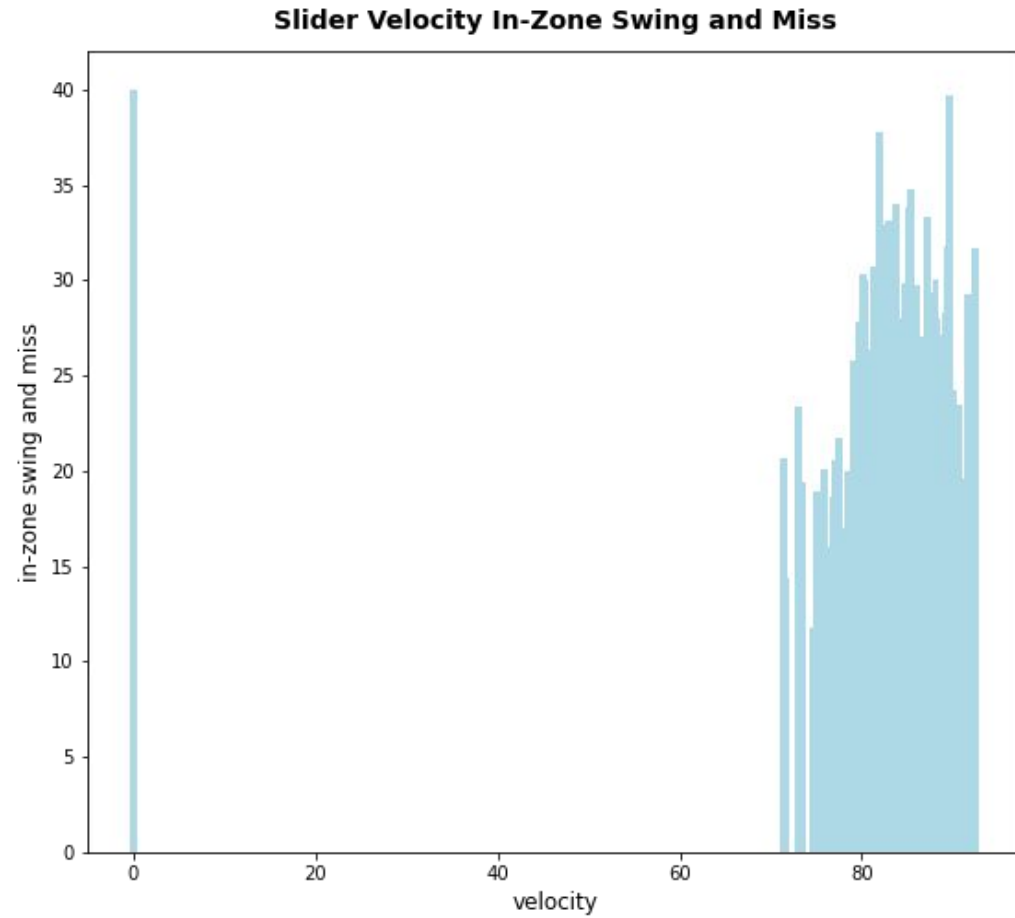
Curveball Spin Rate In-Zone Swing and Miss



Fastball Velocity and Spin Rate



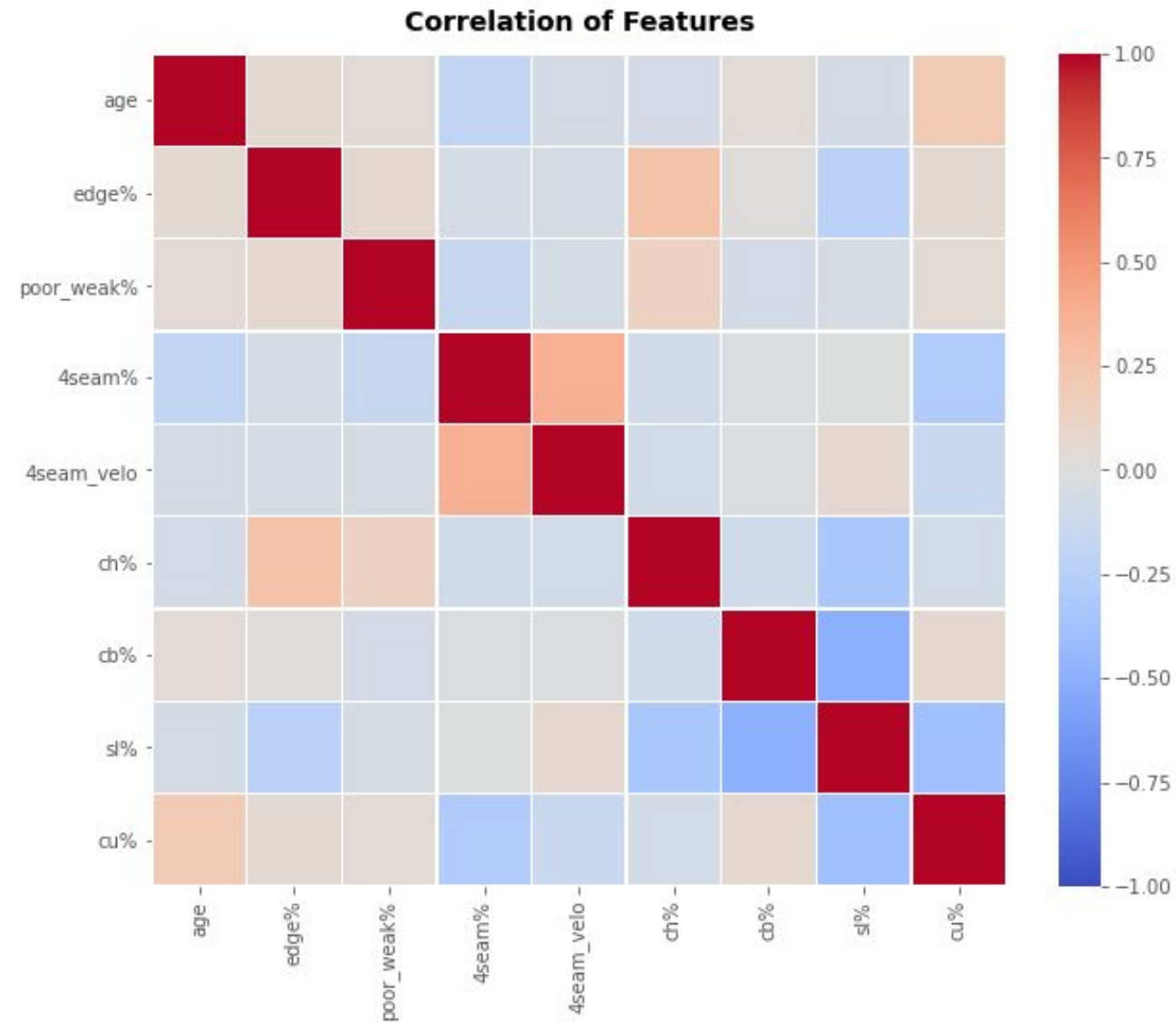
Slider Velocity and Spin Rate



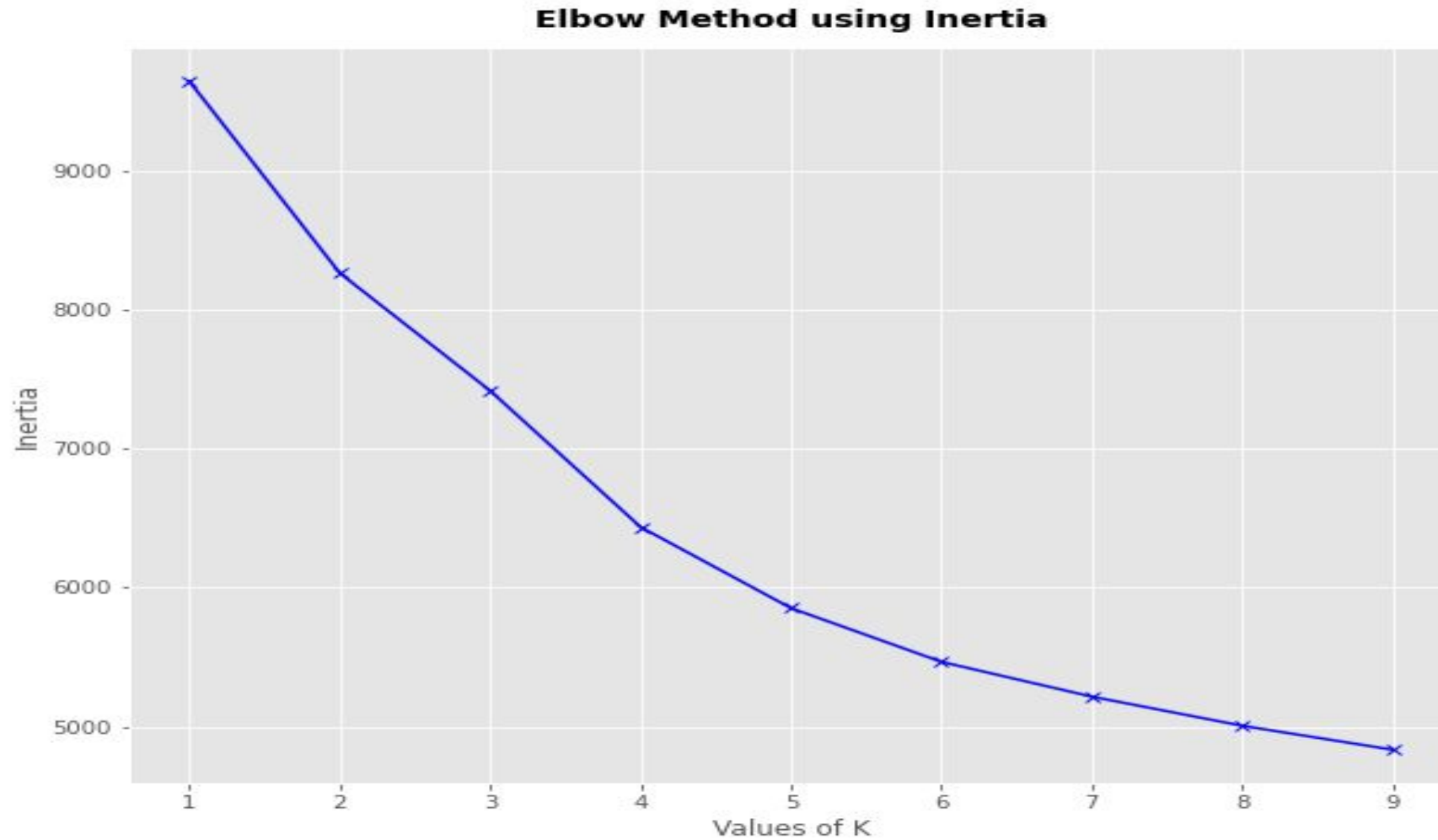
Slider Velocity and Spin Rate



Heatmap of Features

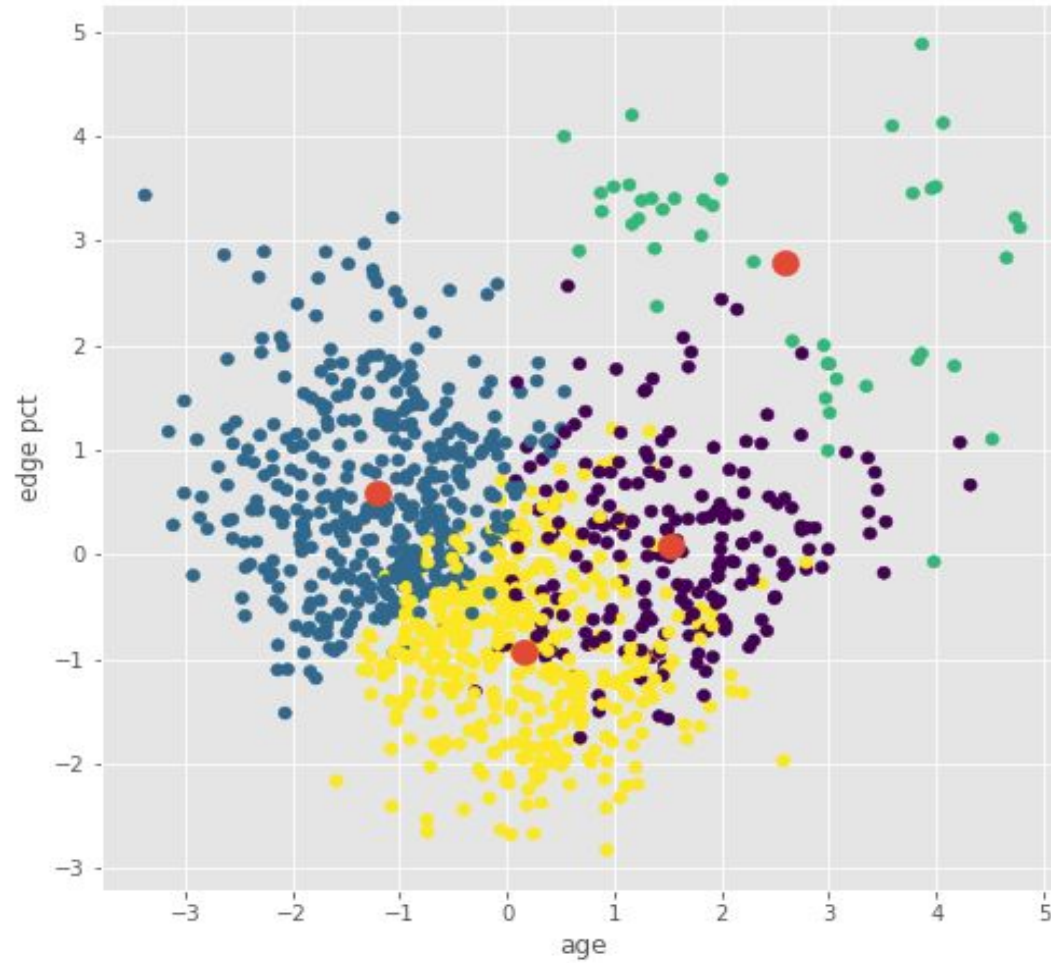


Determining Clusters

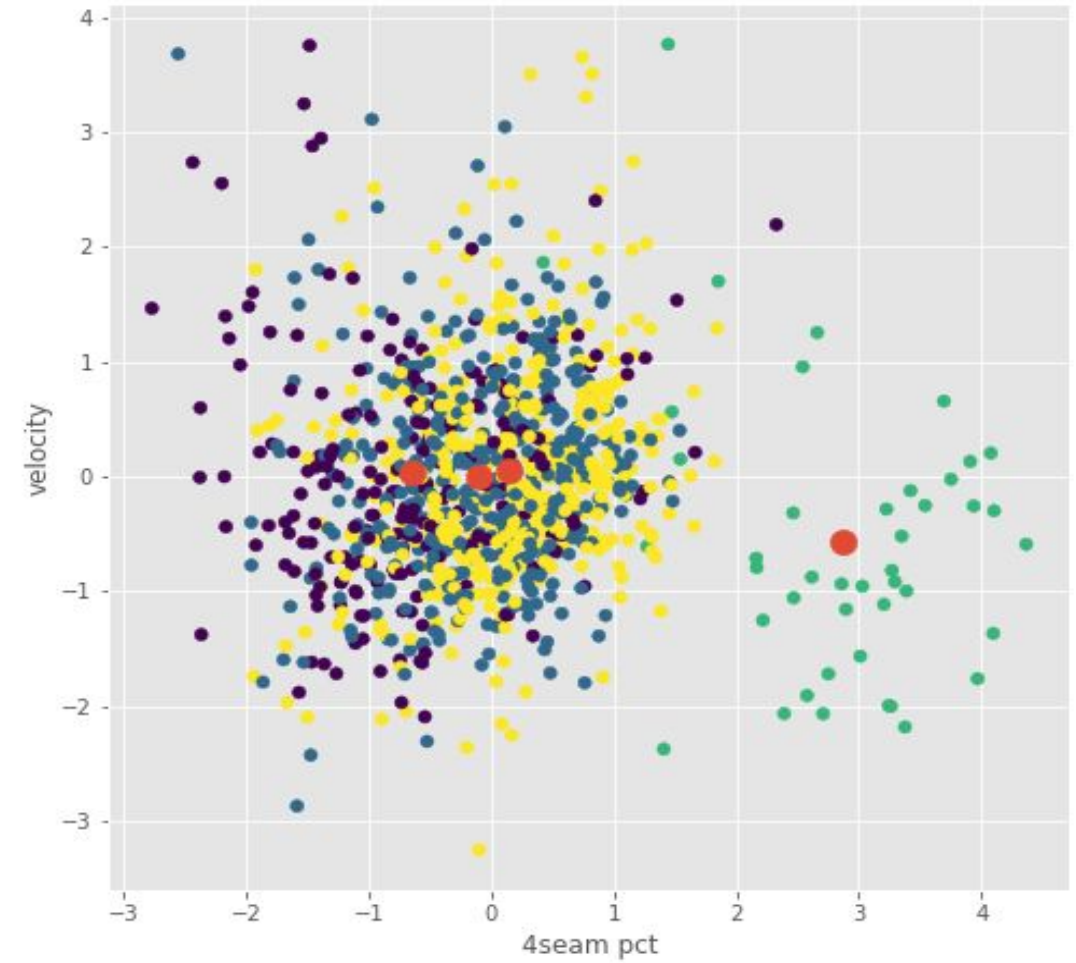


Model Performance

Cluster - Age and Edge Percentage

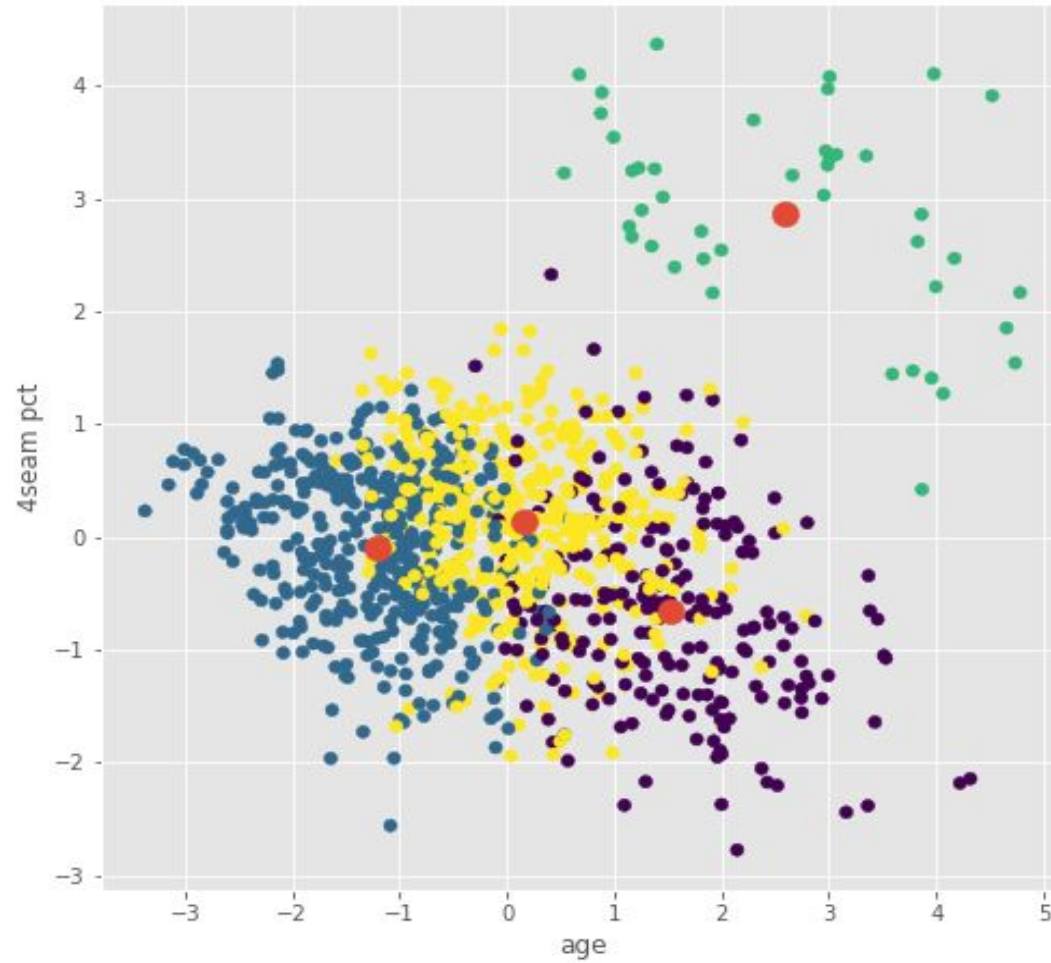


Cluster - 4seam Percentage and 4seam Velocity

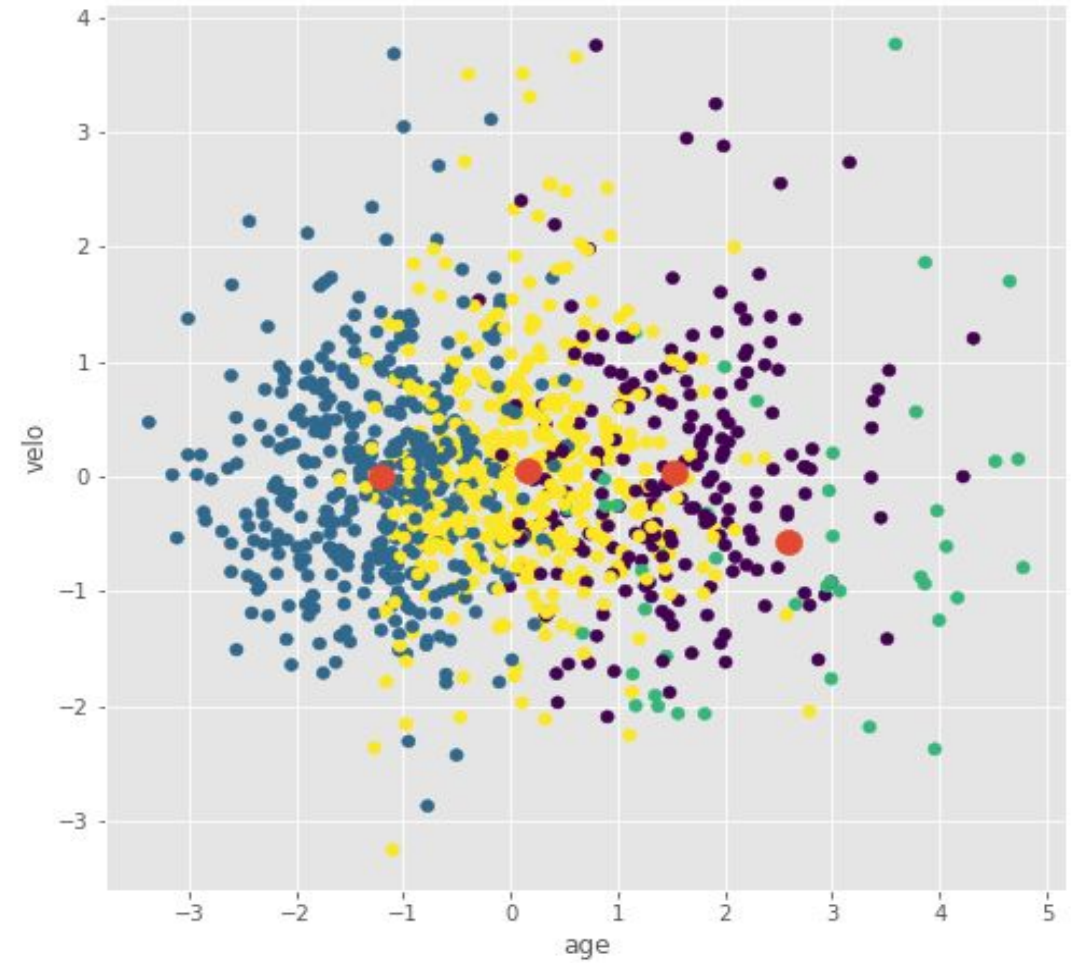


Model Performance

Cluster - Age and 4seam Percentage



Cluster - Age and 4seam Velocity



Model Evaluation

label	Cluster 1	Cluster 2	Cluster 3	Cluster 4
age	31.64	28.8	29.33	27.81
edge %	42.74	41.52	42.68	43.2
poor/weak %	4.42	4.07	4.67	4.27
4seam %	22.34	37	0	43.78
4seam velo	92.32	94.02	0	93.18
ch %	9.4	5.73	13.37	16.63
sl %	4.82	33.2	14.76	10.72
cu %	26.4	0.93	15.27	2.62
cb %	14.95	4.34	10.13	14.64
throws	0.68	0.79	0.53	0.7





Conclusion & Recommendations

The generalizations I made were to control what you can control. Instead of attacking hitters with their weaknesses, attack with your strengths.