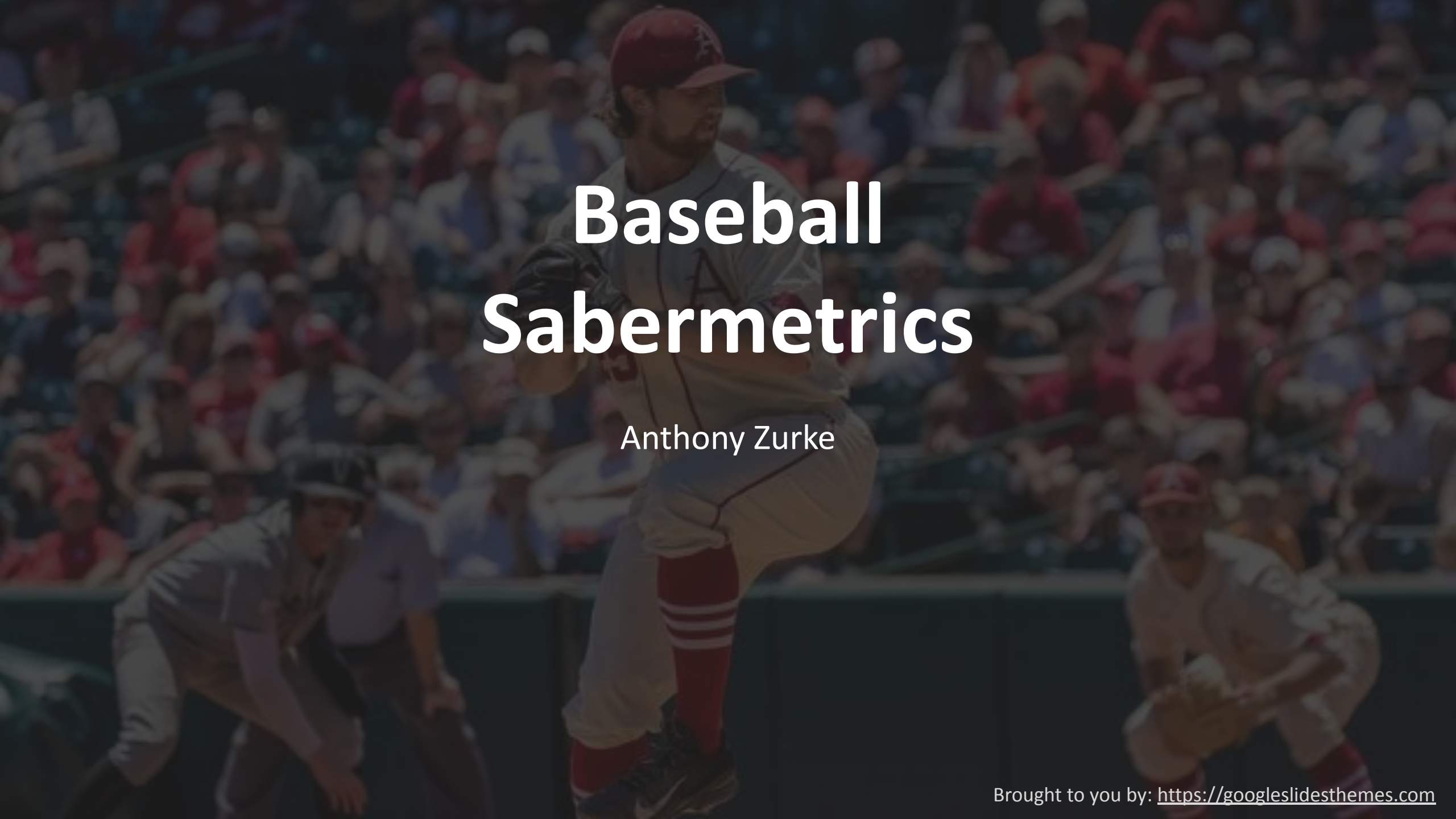
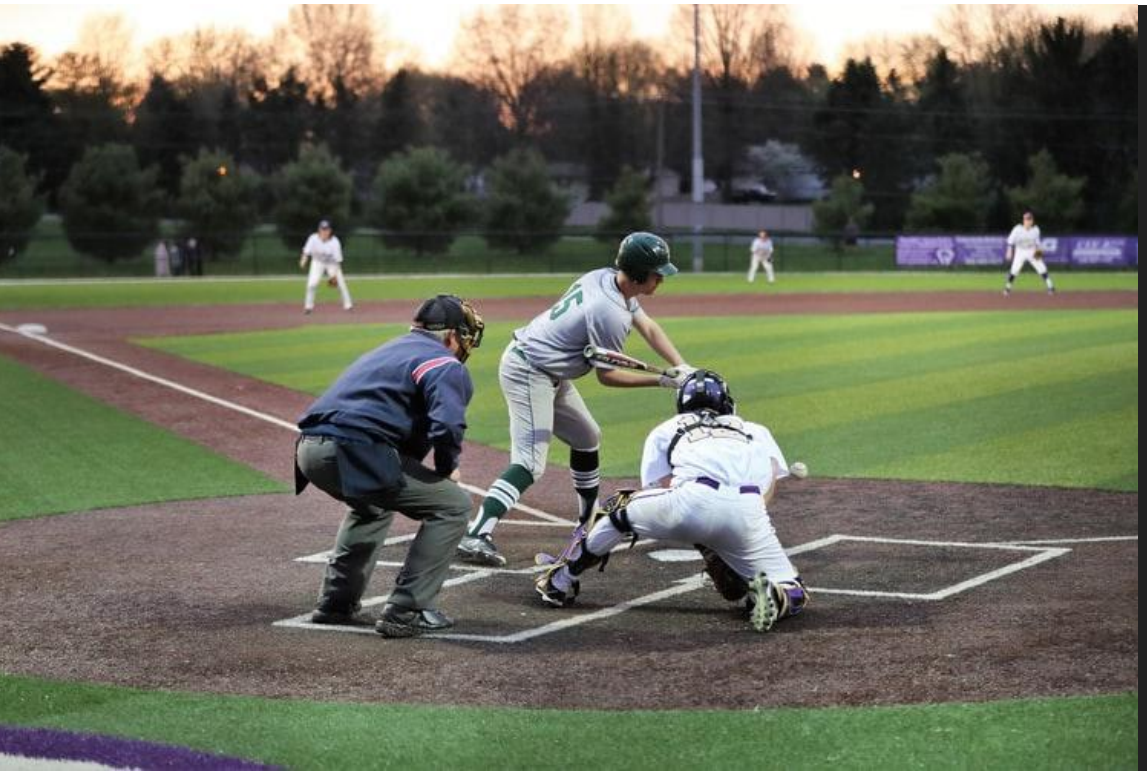


Baseball Sabermetrics

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

Tunneling

Data Gathering and Cleaning

STATCAST | Custom Leaderboard
Google Cloud

CURRENT: CUSTOM LEADERBOARDS

This leaderboard allows you to create custom shareable leaderboards and charts based on selected columns. Click the "Custom Columns" button to view selectable columns.

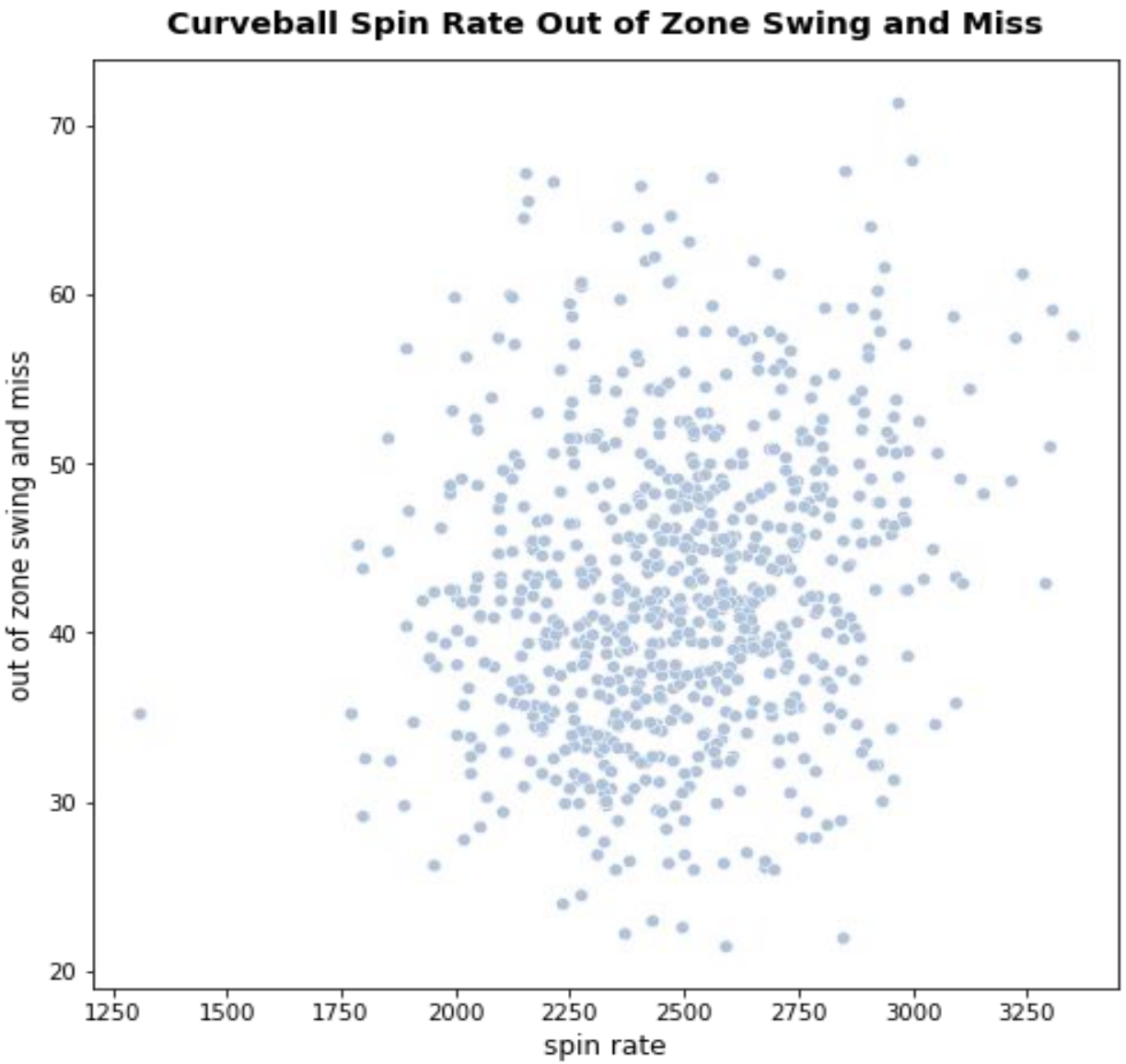
PITCHERS | YEAR (2021) | MINIMUM PA (50) | CUSTOM COLUMNS (8) | Update | Download CSV | Create Chart

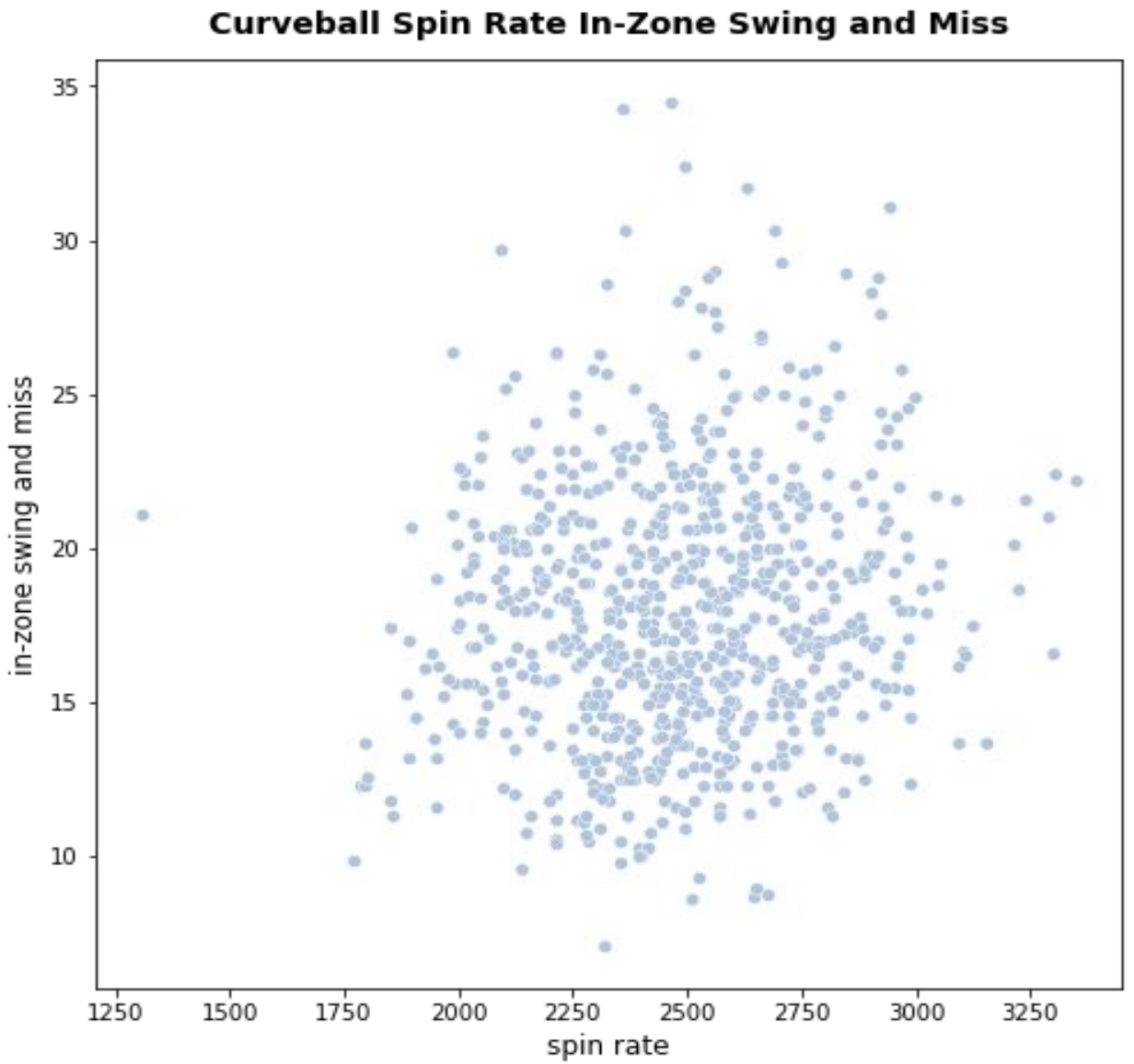
Rk.	Player	Year	xBA	xSLG	xwOBA	xOBP	xISO	Avg EV (MPH)	Avg LA (°)	Barrel%
1	deGrom, Jacob	2021	.145	.246	.190	.179	.101	87.7	13.7	6.6
2	Warren, Art	2021	.138	.201	.202	.234	.063	85.8	19.9	2.6
3	Hendriks, Liam	2021	.165	.309	.215	.190	.143	89.8	19.4	9.7
4	Burnes, Corbin	2021	.179	.254	.218	.228	.075	85.5	8.4	3.1
5	Loaisiga, Jonathan	2021	.184	.249	.225	.240	.065	84	1.5	3.1
6	Clase, Emmanuel	2021	.200	.262	.226	.246	.062	86.4	0.2	1.6

Data Gathering and Cleaning

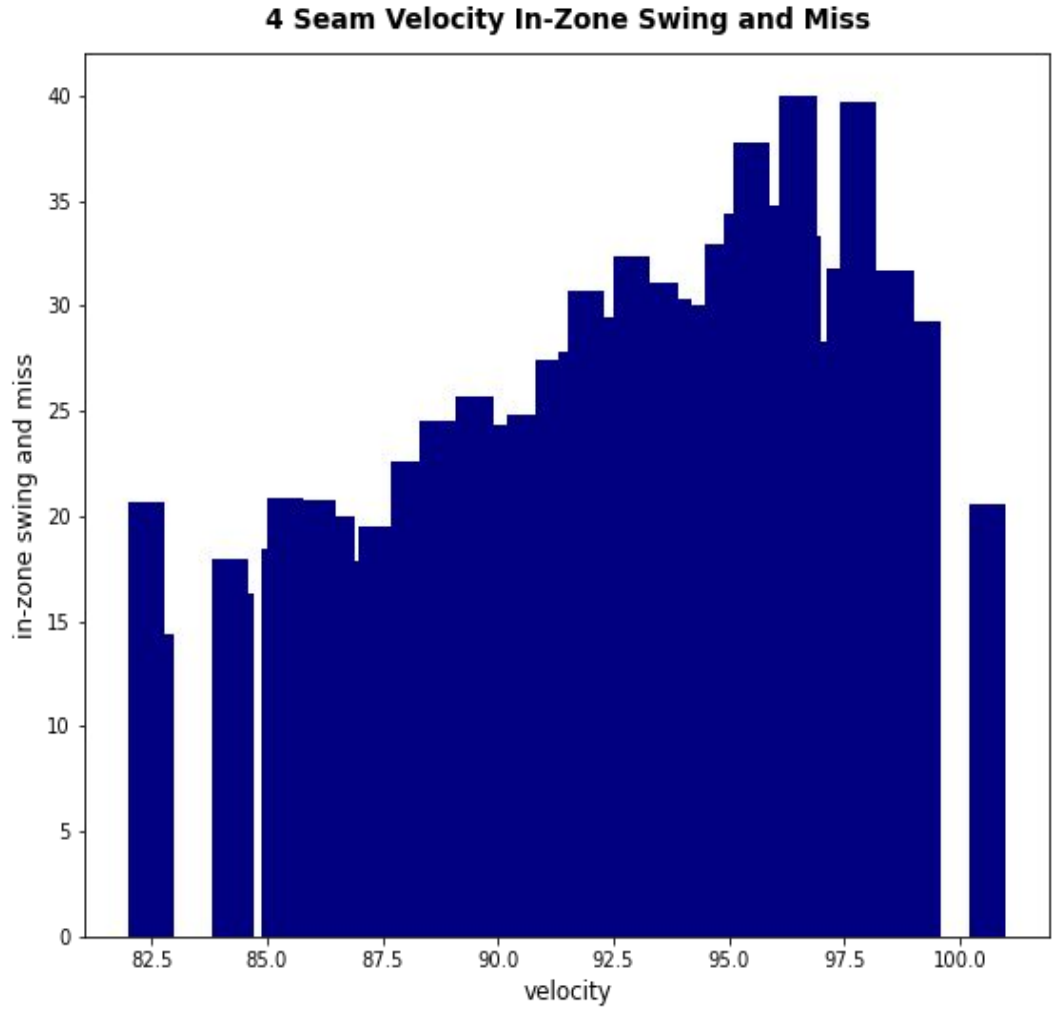
- All pitchers used had a minimum batters faced of 200 for the 2019 and 2021 data and in the 2020 data minimum of 75 batters faced due to a shortened season,
- pitch types: 4 seam fastball, slider, curveball, changeup, and cutter.
- Imputed nulls with 0

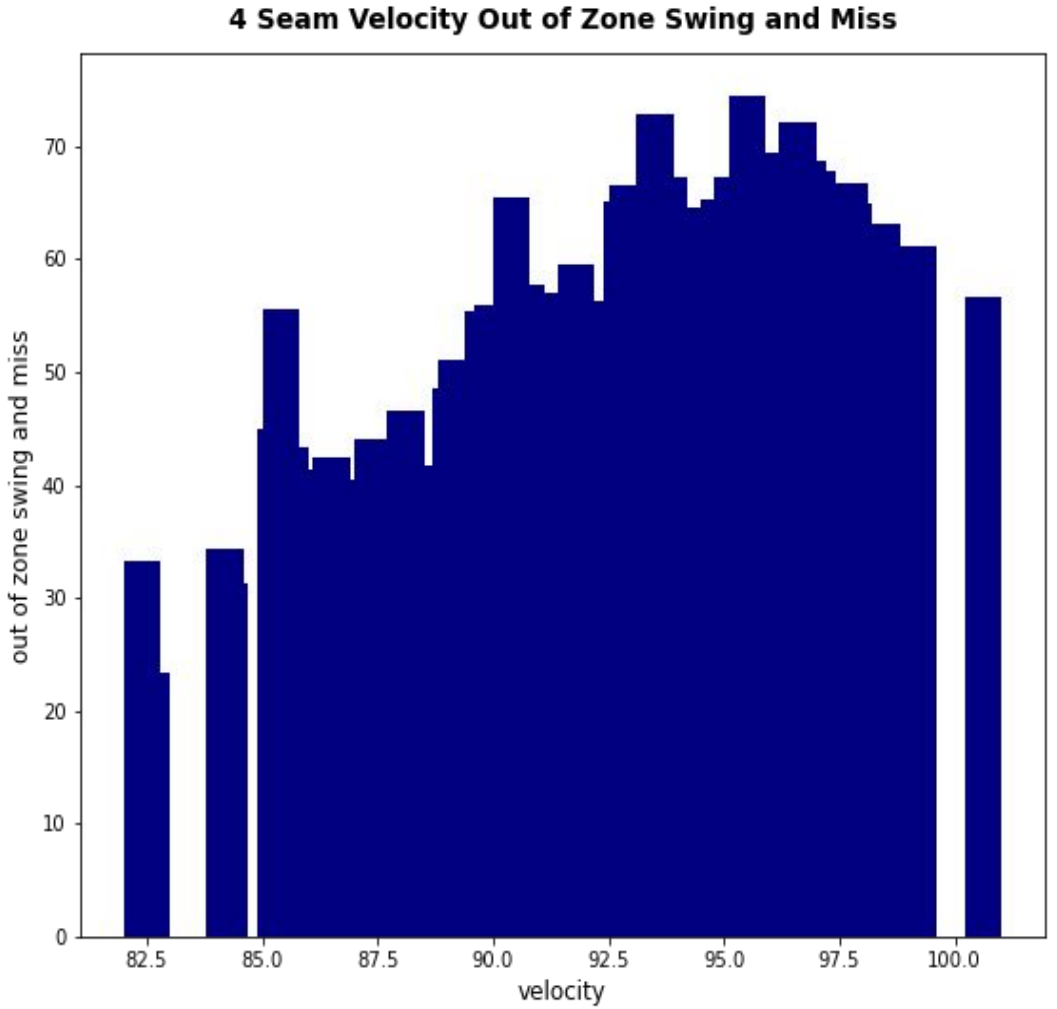
Curveball Spin Rate Swing and Miss



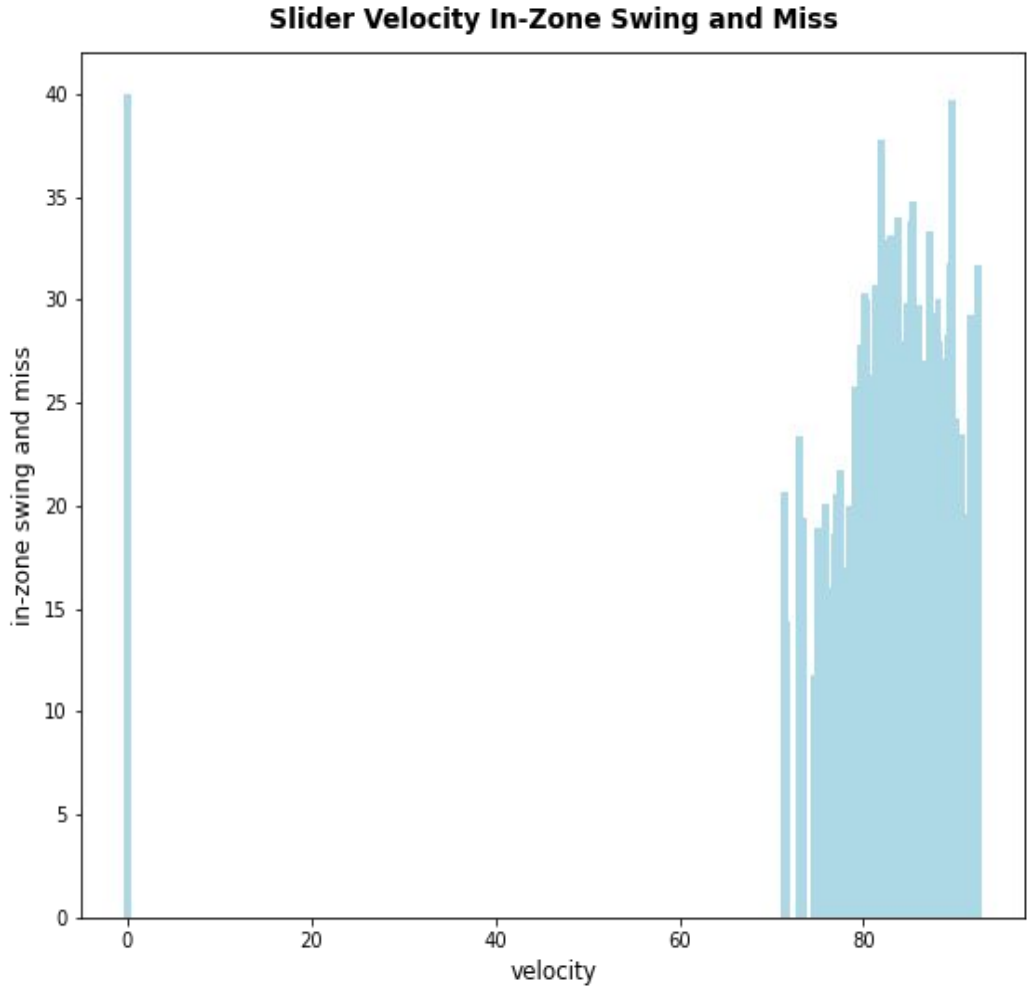


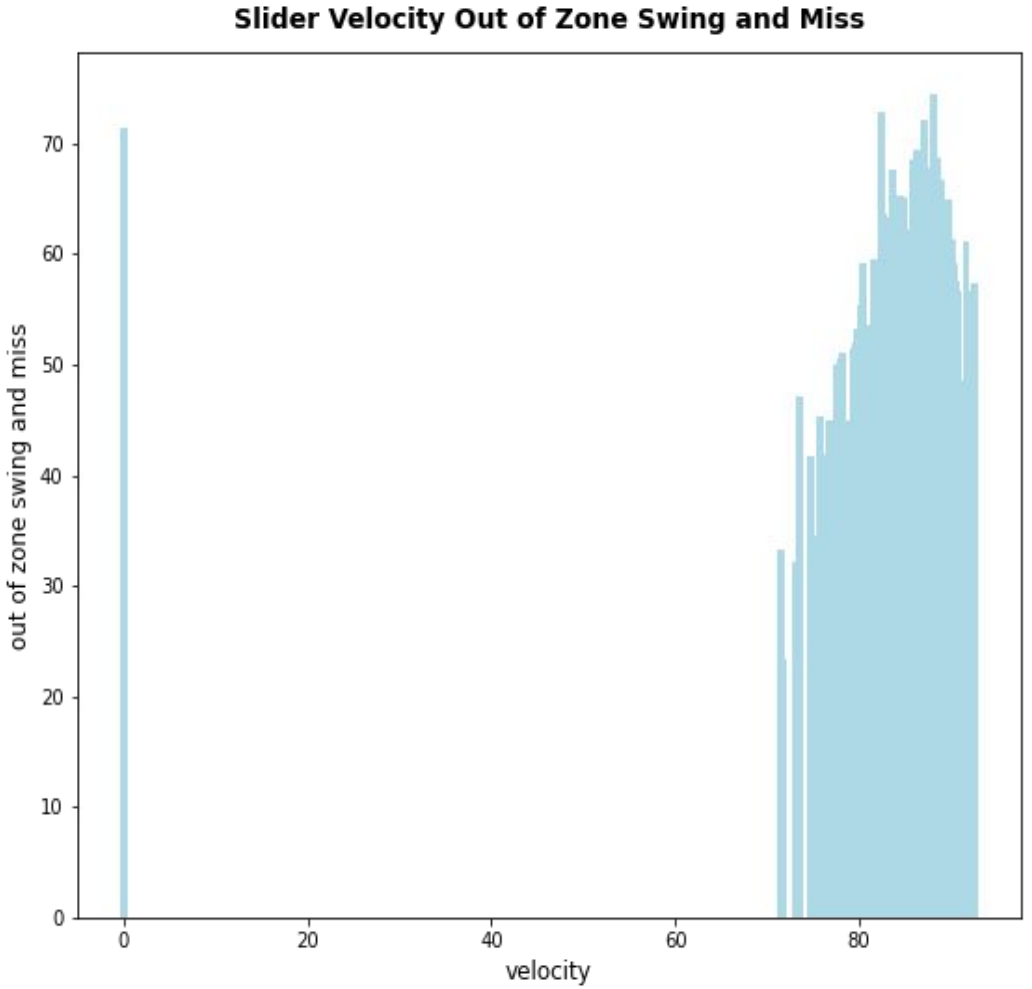
Fastball Velocity Swing and Miss



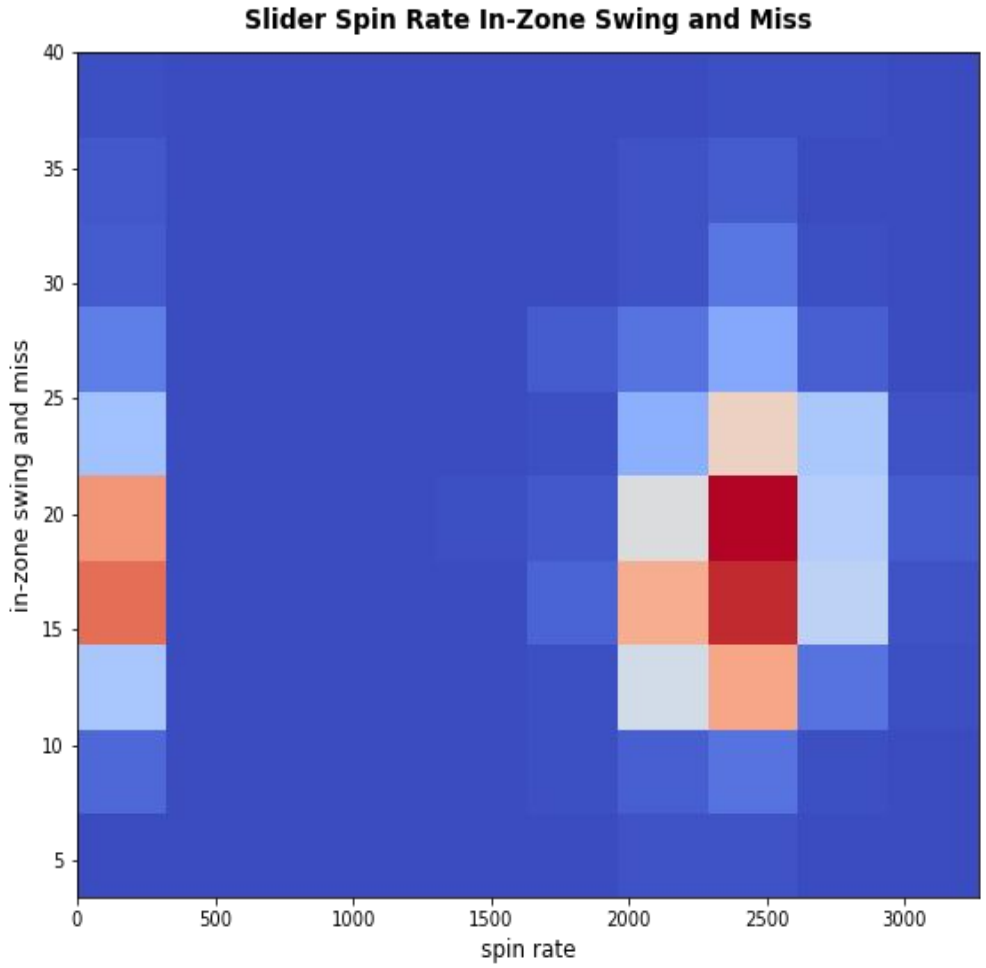


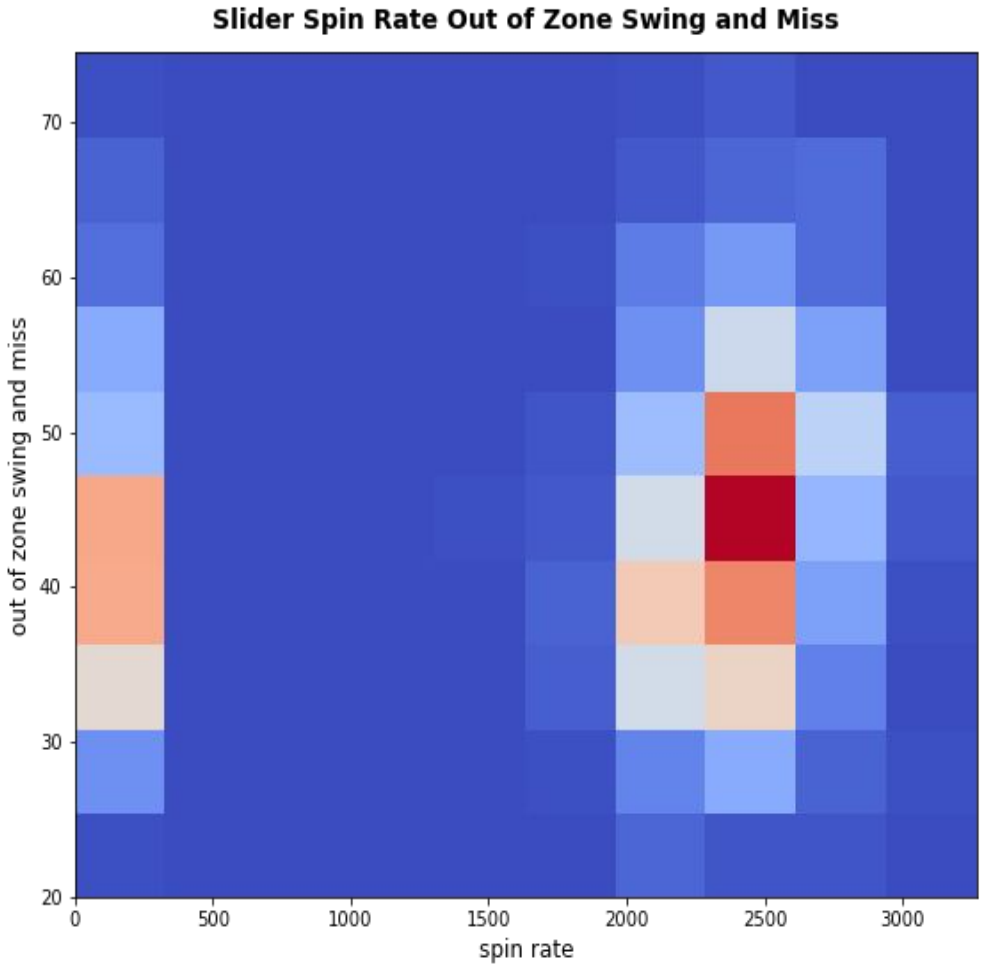
Slider Velocity Swing and Miss





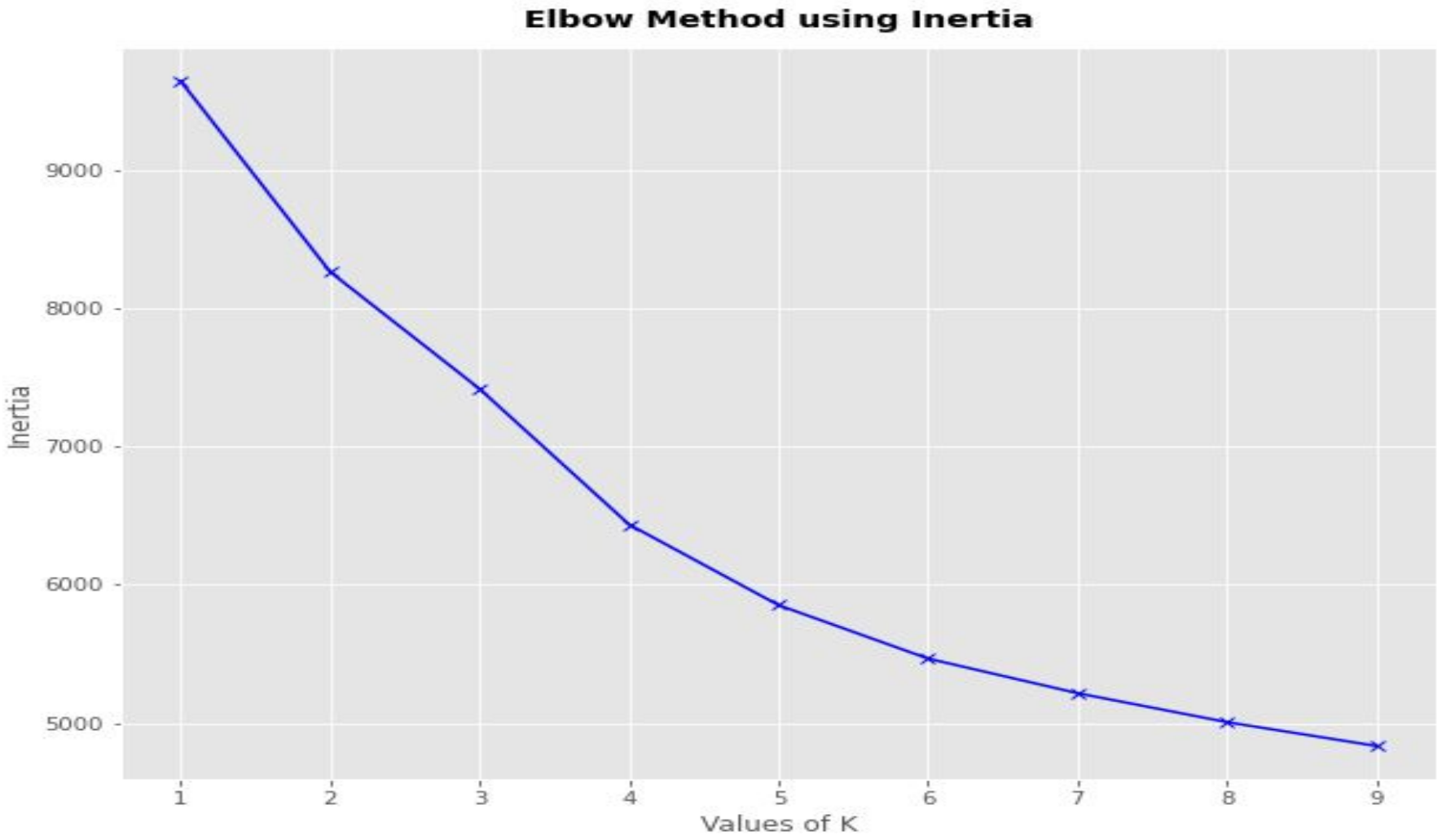
Slider Spin Rate Swing and Miss



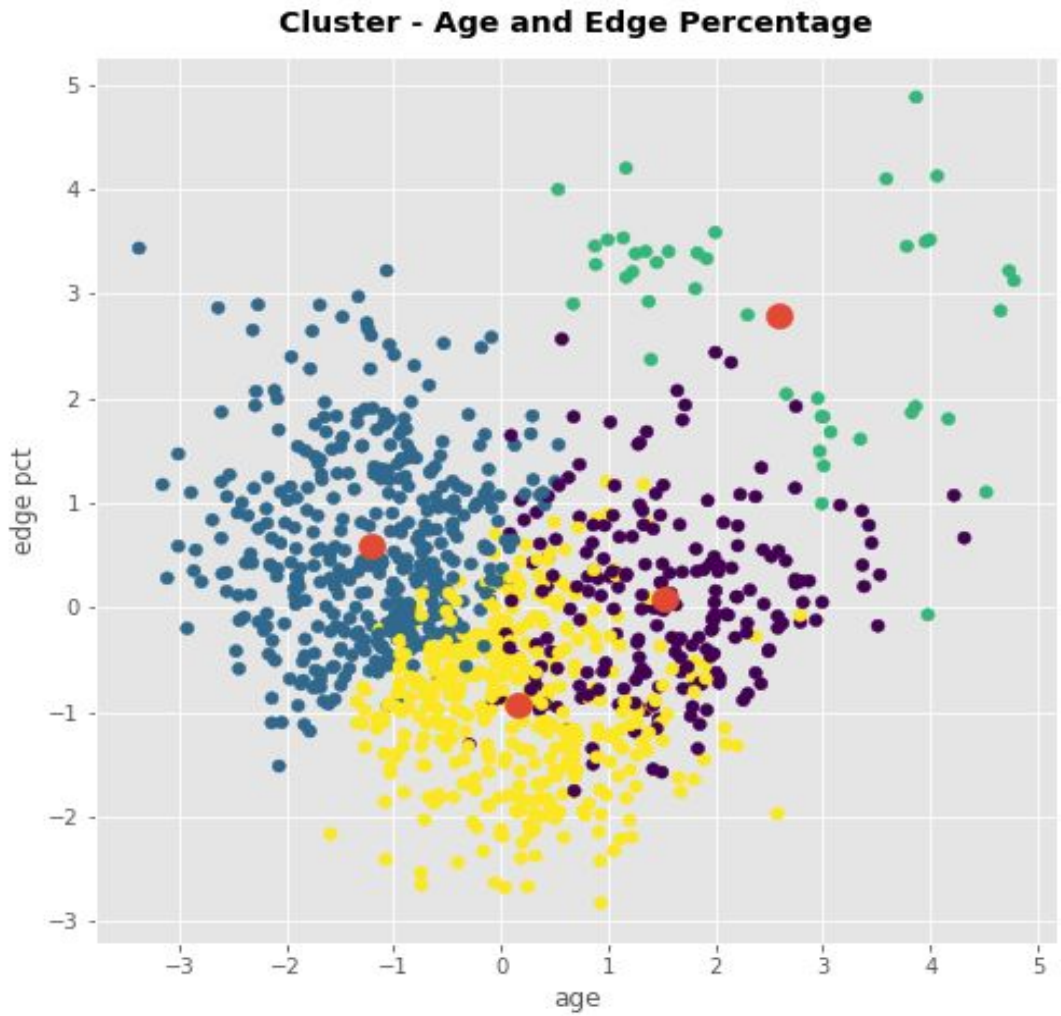


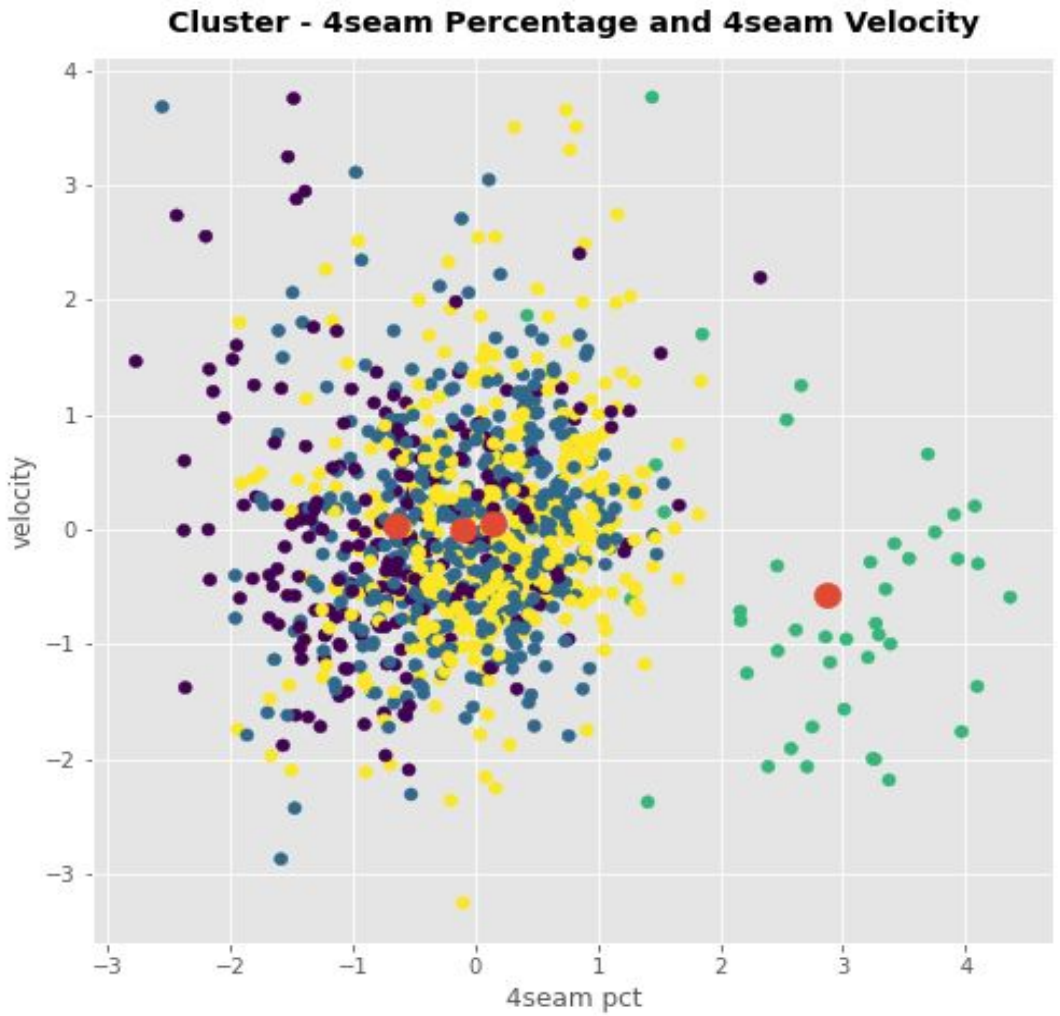
Heatmap of Features

Determining Clusters for Model

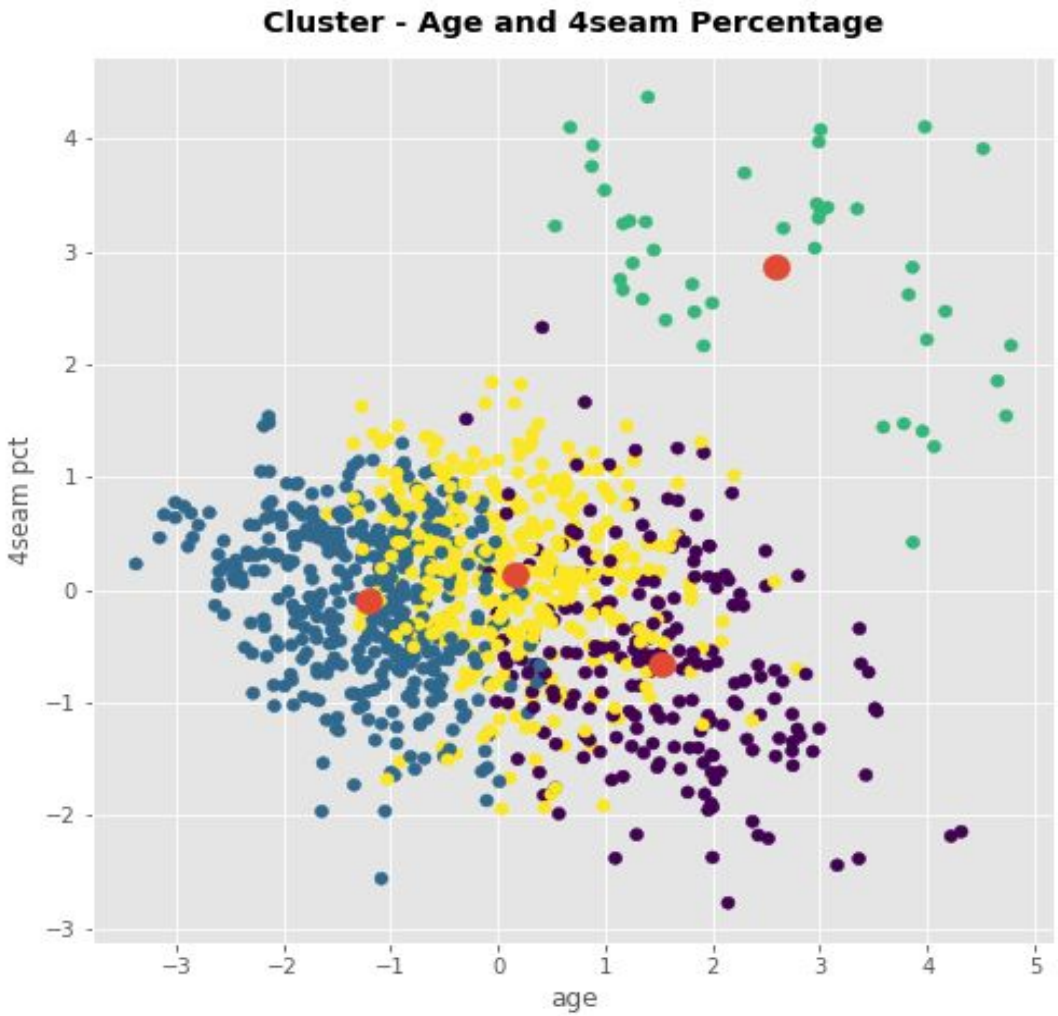


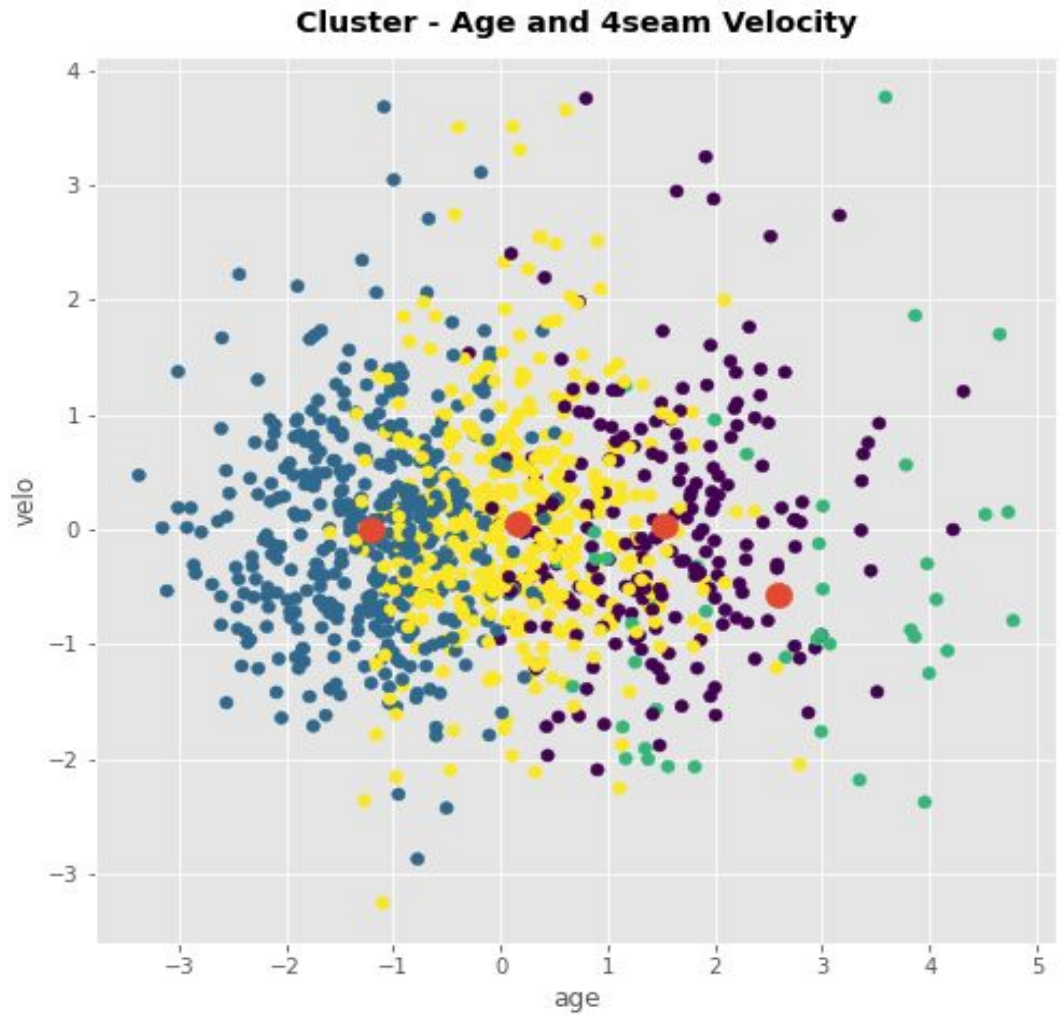
Model Performance





Model Performance





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.