MAR INVESTMENTS

THANK YOU!

ANNUAL CORPORATE TEAM BUILDING **REVEAL**

PLAY BALL!

We're heading to the fantasy league

CASH PRIZES

Weekly cash prizes for best performance

COMPETITION

Worst fantasy score has to complete an embarrassing task

GRAND PRIZE

Most points wins private suit premium club seats for final game

BATTERS

GLOSSARY

Runs Batted In (RBI) - # of points the whole team earned after the player hit the ball

SLUGGING % (SLG) - total # of bases a player records per at- bat

On Base % + SLG (OPS) - adds on base percentage and SLG. Meant to combined how well a hitter can reach base, how well he can hit for average, and power

TRIPLE CROWN = *RBI + Homeruns (HR) + Batting Average (BA)*

STEPS

1. Linear Regression

Used SciKit Learn to compare input variables (X) to Runs (Y)

2. Rank Averages per Category

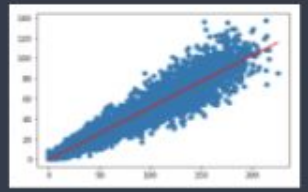
Remember: correlation does not equal causation!

HITS vs. RUNS

Dataset 1

ALL BATTERS 2010 - 2019

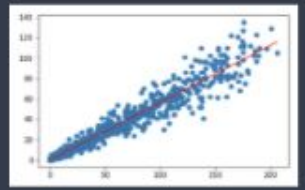
TOP 3: OPS, HITS, SLG



Dataset 2

ALL BATTERS 2019

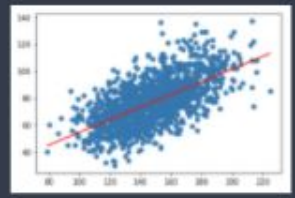
TOP 3: HITS, At Bats, RBIs



Dataset 3

ALL LEADS 2010 -2019

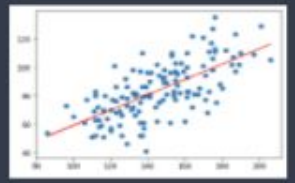
TOP 3: HITS, At Bats, RBIs



Dataset 4

LEADERS 2019

TOP 3: HITS, At Bats, RBIs



AVERAGE RANKINGS

Categories	Rank - Qualified Leaders (201	Rank - All Batters (Just 2019)	Rank - All Batters (2010 - 2019)	Rank - Qualified Leaders (Just 2019)	Rank - Average
On Base plus Slugging Percentage	1	11	13	1	6.5
Hits	2	1	1	2	1.5
Slugging Percentage	5	2	2	5	3.5
On Base Percentage	4	16	12	4	9
At Bats	5	2	2	5	3.5
Home Runs	6	5	8	6	6.25
Walks	7	6	6	7	6.5
RBIs	8	3	3	8	5.5
Doubles	9	4	5	9	6.75
Batting Average	10	15	15	10	12.5
Stolen Bases	11	14	4	11	10
Strike Outs	12	7	7	12	9.5
Caught Stealing	13	13	11	13	12.5
Triples	14	10	11	14	12.25
Hit By Pitcher	15	9	10	15	12.25
Age	16	18	18	16	17
Sacrifice Flys	17	8	9	17	12.75
Sacrifice Bunts	18	17	17	18	17.5
Games Played	19	19	19	19	19

Top 3 Inputs: SLG, RBIs, and Home Runs

MULTIPLE LINEAR REGRESSION MODELS

Also experimented with Lasso, Ridge, and ElasticNet Models

EVERYTHING

INPUTS: EVERYTHING

MSE: 0.0431

R2: 0.9762

TRIPLE CROWN

INPUTS: RBI
Homeruns
BA

MSE: 0.2227

R2: 0.8773

ARUN's MODEL

INPUTS: SLG
Homeruns
RBI

MSE: 0.2222

R2: 0.8776

FINAL RECOMMENDATION

Average Score of qualified leaders

SLUGGING %

0.44

Keep in mind: SLG deals only with hits and does not include walks or hits-by-pitches

HOMERUNS

20

RBIs

> 73

PITCHERS

MAIN FANTASY STATS

Wins (W)*

Earned Run Average (ERA)

Walks/Hits per Inning Pitched (WHIP)

Strikeouts (K)

Saves

**Wins is most likely to gain the most amount of points*

PROCESS

STEP ONE

Gather basic stats for MLB pitchers from 2010 - 2019

STEP TWO

Data exploration - Pearson method to determine best input/output relationship & seaborn regression plots

STEP THREE

Model Prediction for W and ERA

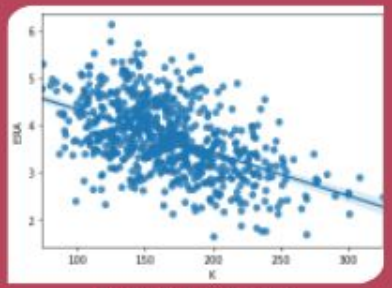
STEP FOUR

Repeat with Statcast Advanced Metrics

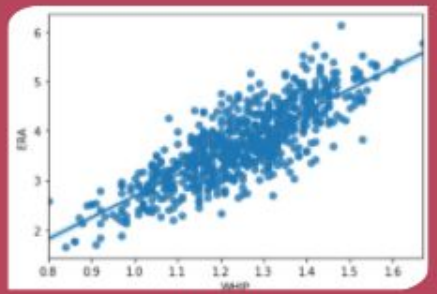
	W	K	ERA	WHIP	SV
W	1.000000	0.418542	-0.548234	-0.503120	-0.007094
K	0.418542	1.000000	-0.501607	-0.591046	-0.062235
ERA	-0.548234	-0.501607	1.000000	0.808146	-0.001620
WHIP	-0.503120	-0.591046	0.808146	1.000000	0.023214
SV	-0.007094	-0.062235	-0.001620	0.023214	1.000000

PRELIMINARY FINDINGS

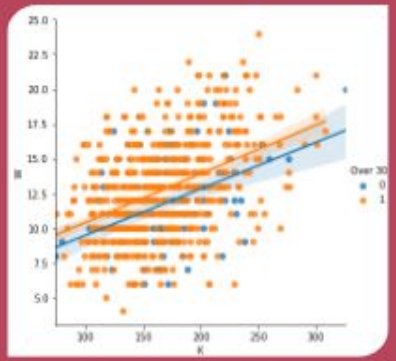
Basic Stats 2010 - 2019



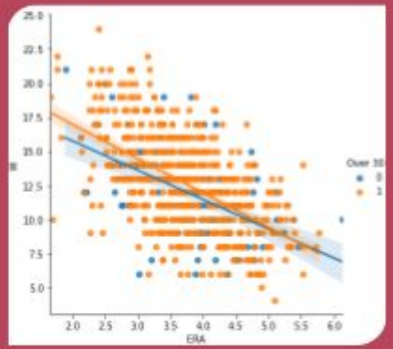
Strikeouts vs. ERA



WHIP vs. Wins



Strikeouts vs. Wins



ERA vs. Wins

W K and WHIP	**MSE:** 3.35 **R2:** 0.7862
ERA K and WHIP	**MSE:** 0.2178 **R2:** 0.7816
W K, ERA & WHIP	**MSE:** 3.349 **R2:** 0.7863

ADVANCED FINDINGS

01 Establish 'Basic Stats' Model

Simple Linear Regression Model of K per W

MSE: 6.739 R2: 0.6636

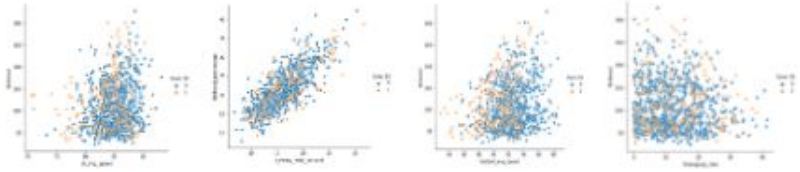
02 Explore Statcast Variables

Understand the data!

Strikeouts vs. Strikeout Percentage

03 Final Model Prediction

Used Multiple Linear Regression, Lasso, Ridge, ElasticNet



	Win	Strikeout_percentage	Earned Runs	ERA	Strikeout	hard_hit_percent	edge_percent
Win	1.000000	0.253435	0.540855	-0.378505	0.814596	-0.134194	0.124821
Strikeout_percentage	0.253435	1.000000	-0.217544	-0.523879	0.510139	-0.142782	-0.084108
Earned Runs	0.540855	-0.217544	1.000000	0.346306	0.589692	0.182454	0.092064
ERA	-0.378505	-0.523879	0.346306	1.000000	-0.361210	0.412973	-0.064386
Strikeout	0.814596	0.510139	0.589692	-0.361210	1.000000	-0.103727	0.062349
hard_hit_percent	-0.134194	-0.142782	0.182454	0.412973	-0.103727	1.000000	-0.079751
edge_percent	0.124821	-0.084108	0.092064	-0.064386	0.062349	-0.079751	1.000000

	Strikeout Percentage	Earned Runs	ERA	Strikeout	exit_velocity_avg	launch_angle_avg	sweet_spot_percent	hard_hit_percent	z_swin
Strikeout Percentage	1.000000	-0.437769	-0.564358	0.589794	-0.149719	0.180085	-0.021772	-0.081076	0.7990
Earned Runs	-0.437769	1.000000	0.627320	0.280287	0.258435	0.048074	0.123417	0.054975	-0.373
ERA	-0.564358	0.627320	1.000000	-0.376494	0.328525	0.054625	0.318504	0.322084	-0.445
Strikeout	0.589794	0.280287	-0.376494	1.000000	-0.101928	0.120226	-0.073292	-0.247755	0.4428
exit_velocity_avg	-0.149719	0.258435	0.328525	-0.101928	1.000000	-0.002741	0.065130	0.814238	-0.283
launch_angle_avg	0.180085	0.048074	0.054625	0.120226	-0.002741	1.000000	0.428780	-0.049433	0.4048
sweet_spot_percent	-0.021772	0.123417	0.318504	-0.073292	0.065130	0.428780	1.000000	0.146617	0.0633
hard_hit_percent	-0.081076	0.054975	0.322084	-0.247755	0.814238	-0.049433	0.146617	1.000000	-0.213
z_swing_miss_percent	0.799029	-0.373099	-0.445991	0.442855	-0.283237	0.404884	0.063310	-0.213046	1.0000
in_zone_percent	0.043840	0.092657	-0.048403	0.163191	0.040829	0.241819	0.047665	0.027510	0.1676
edge_percent	-0.099037	-0.125786	-0.055414	-0.174934	-0.234739	0.027575	0.021981	-0.166511	0.0413
Changeup_rate	-0.120324	-0.105934	0.081949	-0.265245	-0.206467	0.076081	0.091595	-0.152638	0.1342
ch_avg_speed	0.332839	-0.100947	-0.300782	0.320392	0.049204	-0.331159	-0.279277	0.044504	0.1426
Curveball_rate	0.115529	-0.030540	-0.012493	0.067458	0.115084	-0.129518	0.079523	0.169903	-0.079
cu_avg_speed	0.264707	-0.042372	-0.174456	0.222544	0.095656	-0.256455	-0.210334	0.091951	0.1377
fastball_rate	-0.239514	0.167919	0.047578	-0.033741	0.060757	0.005165	-0.058226	-0.013803	-0.199
fastball_avg_speed	0.426612	-0.116054	-0.252402	0.360907	0.019305	-0.190235	-0.244162	0.016319	0.3569

Model Results

K% based on *fastball speed & swing/miss %*

MODEL TEST SCORE: .624

Lasso model slightly better

K% based on *fastball average speed, swing/miss % & K*

MODEL TEST SCORE: .679

W based on *K%, FB speed, K, ERA, & swing/miss %*

MODEL TEST SCORE: .706

Lesson Learned:

Beginners should stick with what the experts tell you!

Q&A with the Team

MOST CHALLENGING

LEARNY ANYTHING?

MOST ENOYABLE

NEXT STEPS?

CLICK TO LAUNCH!