WARP Vs Salary:

HOW TO WIN IN MAJOR LEAGUE BASEBALL

Jeff Ragauskis, Tyson Herrin,
Damien Franco, Will Partain

TAKE ME OUT
TO THE BALLGAME...

⚾ Confession time

⚾ A little bit of history

⚾ CBA

⚾ WARP...What is it good for?
Absolutely everything.

THE MAIN EVENT

$$bWAR = (P_{runs} - A_{runs}) + (A_{runs} - R_{runs})$$

The term $P_{runs} - A_{runs}$ may be calculated from the first five factors, and the other term from the remaining factor.[12]

Batting runs depends on weighted Runs Above Average (wRAA), weighted to the offense of the league, and is calculated from wOBA.[13]

$$wRAA = \frac{wOBA - .320}{1.25} * (AB + BB + HBP + SF + SH)$$

where

$$wOBA = \frac{(\alpha_1 * uBB + \alpha_2 * HBP + \alpha_3 * 1B + \alpha_4 * 2B + \alpha_5 * 3B + \alpha_6 * HR + \alpha_7 * SB - \alpha_8 * CS)}{(AB + BB - IBB + HBP + SF)}$$

VS

DATA SOURCES

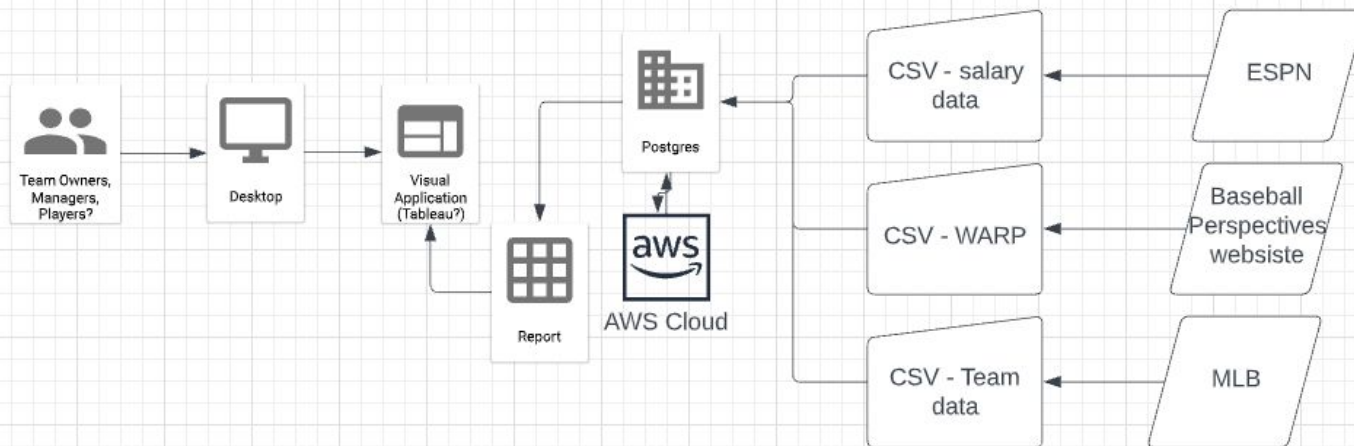- 🥎 MLB.com
- 🥎 Baseball Prospectus
- 🥎 MLB Stats API

QUESTIONS TO BE ANSWERED

- ⚾ Does increasing payroll lead to success?
- ⚾ Is WARP more effective when it comes to building a championship team?
- ⚾ Are there other variables that may affect the success of a team?

FLOWCHART

Does WARP Predict MLB Team Playoff Appearence Better Than Overall Team Salary?

ETL PROCESS

Stats Data Set

season	fullName	team_id	player_id	team_name	assists	putOuts	errors
2000	Rod Barajas	109	150148	Arizona Diamondbacks	0	24	0
2000	Danny Bautista	109	110712	Arizona Diamondbacks	0	1	0
2000	Danny Bautista	109	110712	Arizona Diamondbacks	2	28	0
2000	Alex Cabrera	109	294697	Arizona Diamondbacks	6	102	0
2000	Alex Cabrera	109	294697	Arizona Diamondbacks	0	2	0

Player Data Set

Player	Pos	Salary	Pct	WARP	WARP/$M	$/WARP	TEAM	YEAR
Mo Vaughn	1B	11166667	0.16	1.84	0.16	6070787.0	Angels	2000
Tim Salmon	RF	6000000	1.01	6.05	1.01	991179.0	Angels	2000
Ken Hill	RP	5600000	NaN	-1.40	NaN	NaN	Angels	2000
Tim Belcher	RP	4600000	NaN	-0.19	NaN	NaN	Angels	2000
Kent Bottenfield	RP	4000000	NaN	-0.14	NaN	NaN	Angels	2000

```
complete_df.groupby(['player_id','player','pos','team_id','team_name','season'],as_index=False).agg({
    'salary':'first','pct':'first','warp':'first','warpmil':'first',
    'dollarwarp':'first','assists': np.sum,'putouts': np.sum, 'errors': np.sum, 'chances': np.sum, 'fiel
    'double_plays': np.sum,'triple_plays':np.sum, 'throwing_errors': np.sum, 'ground_outs': np.sum, 'air
    'runs': np.sum, 'doubles': np.sum, 'triples': np.sum, 'home_runs': np.sum, 'strike_outs': np.sum, 't
    'intentional_walks': np.sum, 'average': np.mean, 'at_bats': np.sum, 'obp': np.mean, 'slg': np.mean,
```

player_id	player	pos	team_id	team_name	season	salary	pct	warp	warpmil	...	wild_pitches	pickoffs	pitched_total_bases	pitches_per_inning
451661	Josh Fields	RP	117	Astros	2014	506500	NaN	-0.63	NaN	...	0.0	0.0	70.0	17.38
451661	Josh Fields	RP	117	Astros	2015	516700	1.24	0.64	1.24	...	1.0	0.0	59.0	16.97
451661	Josh Fields	RP	117	Astros	2016	900000	0.02	0.01	0.02	...	0.0	0.0	30.0	17.74
451661	Josh Fields	RP	117	Astros	2016	900000	0.02	0.01	0.02	...	0.0	0.0	41.0	17.04

MACHINE LEARNING

- **Principal Component Analysis**
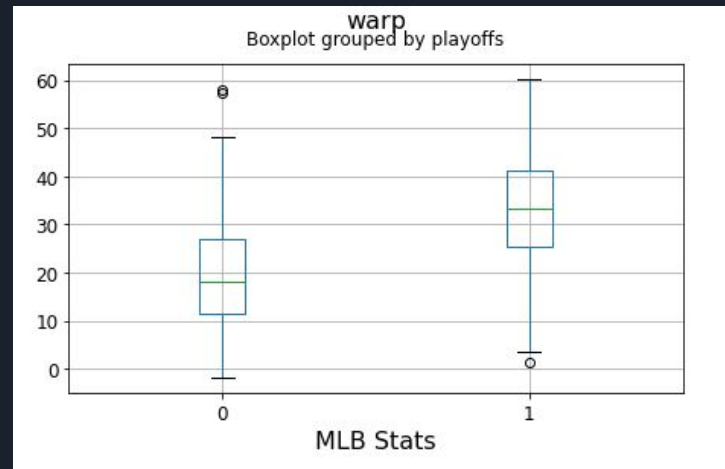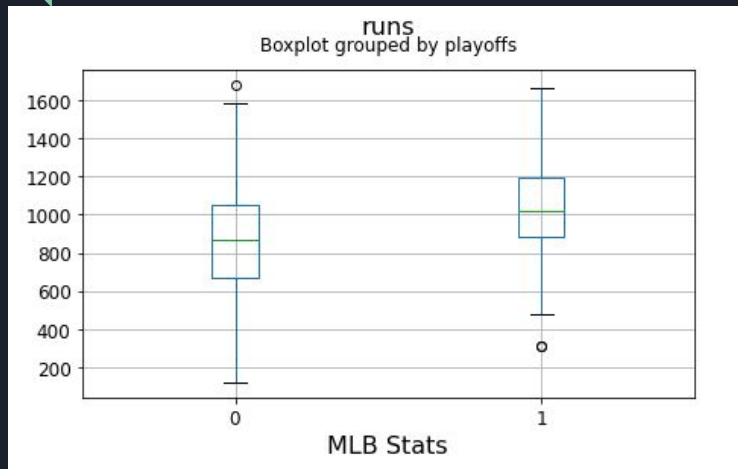 - Used to condense large data set into smaller subset
- **Random Forest Classifier**
  - Produces better results
  - Cannot predict results outside of the dataset
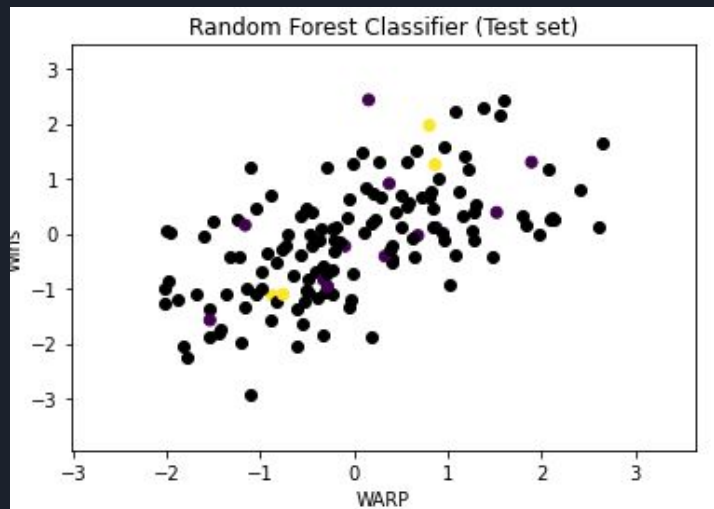
PCA
& EXPLAINED VARIANCE RATIO

[1.58691121 1.36183634 1.02149516 0.64960702 0.33188507]
[0.32047581 0.27502207 0.20629036 0.13118776 0.067024]
[0.32047581 0.59549787 0.80178824 0.932976   1.   ]

RANDOM FOREST CLASSIFIER

⚾ **X Variable - Team Salary (WARP, ERA, Runs)**

⚾ **Y Variable - Wins**

Random Forest Classifier (Test set)

```
warp: (0.5215964158321654)
team_payroll: (0.4784035841678347)
```

```
runs: (0.5017080327645135)
team_payroll: (0.49829196723548647)
```

WHAT DID WE LEARN?

- � Does increasing payroll lead to success?
- � Is WARP more effective when it comes to building a championship team?
- � Are there other variables that may affect the success of a team?