

Drivers of MLB Attendance: Is Winning Worth it?

An examination of the influence of various factors on Attendance in Major League Baseball

Will Antink - Avery Pike - Tyler Gomez - Jerry McGlade

Who Cares: Why is MLB Attendance important?

- Interested Parties:
 - MLB Commissioner's Office
 - Front Offices of MLB Teams
- Emblematic of overall revenue and popularity of baseball
- Influences decisions around team construction and payroll as well as league rules, policies, and actions taken to grow the game

"If you go back and you look in a decade, those teams that win 54% of the time always wind up in the postseason. And they, more often than not, wind up in the World Series. So there's your bigger picture process. Nobody wants to hear the goal this year is, 'We're going to win 54% of the time.' Because sometimes 54% is -- one year, you're going to win 60%, another year you're going to win 50%. It's whatever it is. But over time, that type of mindset gets you there ... If what you're doing is focusing year-to-year on, 'what do we have to do to win the World Series this year?' You might be one of the teams that's laying in the mud and can't get up for another decade."

- Jerry Depoto: GM of the Seattle Mariners

MLB Rule Changes, Industry Assumptions:

- Rule Changes in 2023 designed to improve run scoring and shorten games
 - Pitch Clock (adjusted again this year)
 - Limited pickoff attempts
 - Ban of the Shift
 - Other Recent Rule Changes include
- These rule changes suggest that MLB believes:
 - Strong offenses draw more fans than strong defenses
 - Shorter games bring in more fans
 - More action on the basepaths/runners on base draws more fans
- MLB has also expanded the playoffs twice in the last 10 years, allowing more teams who are not necessarily the best compete for a title

Background Research:

- [Factors that influence Major League and Minor League Baseball - 3241 Words | Research Paper Example \(ivypanda.com\)](#)
 - Many factors have been shown to directly affect MLB game attendance, such as local household income levels, concession prices, and marketing of team rivalries
- [Top Prospects and Minor League Baseball Attendance - Seth R. Gitter, Thomas A. Rhoads, 2011 \(sagepub.com\)](#)
 - Prospects who have not yet played in the MLB can affect attendance
 - High quality minor league prospects tend to bring in lots of viewers from MiLB
- [\(PDF\) Team Roster Turnover and Attendance in Major League Baseball. \(researchgate.net\)](#)
 - Higher roster turnover correlates to lower attendances the following season
 - According to this study, a 1% increase in roster turnover can lead to a 0.4-0.7% decrease in attendance the following season for a given team

Primary Questions and Project Goals

- Does offensive play have a greater impact on per game attendance than defensive play?
- Does it make sense for teams to aim for 'just good enough', or are there consistent marginal returns for more investment in players?
- How big of a role does local population play on game attendance?
- What factors beside population have the greatest impact on per game attendance?
- Can we predict a team's season attendance given their in games statistical projections?
- How should MLB front offices spend in free agency if they want to increase attendance?

Where We Sourced Our Data

- Out of game stats such as attendance and gametime are from Baseball Reference
- In-game and advanced stats are from Fangraphs
- Data on metro area populations came from MacroTrends
- 35 columns across 510 rows of clean data



Our Data

...	1	Rk	Team	W	L	WL	R	RA	Rdiff	Home	Road	tYear	Attendance	Attend/G	BatAge	PAge
	0	1	Atlanta Braves	104	58	0.642	5.8	4.4	1.4	52-29	52-29	Atlanta Braves2023	3191505	39401	27.9	29.9
	1	2	Baltimore Orioles	101	61	0.623	5.0	4.2	0.8	49-32	52-29	Baltimore Orioles2023	1936798	23911	27.2	28.4
	2	3	Los Angeles Dodgers	100	62	0.617	5.6	4.3	1.3	53-28	47-34	Los Angeles Dodgers2023	3837079	47371	30.9	28.1
	3	4	Tampa Bay Rays	99	63	0.611	5.3	4.1	1.2	53-28	46-35	Tampa Bay Rays2023	1440301	17781	26.8	28.5
	4	5	Milwaukee Brewers	92	70	0.568	4.5	4.0	0.5	49-32	43-38	Milwaukee Brewers2023	2551347	31498	27.7	29.3
	5	6	Houston Astros	90	72	0.556	5.1	4.3	0.8	39-42	51-30	Houston Astros2023	3052347	37683	28.8	29.2
	6	7	Philadelphia Phillies	90	72	0.556	4.9	4.4	0.5	49-32	41-40	Philadelphia Phillies2023	3052605	37686	28.4	29.8
	7	8	Texas Rangers	90	72	0.556	5.4	4.4	1.0	50-31	40-41	Texas Rangers2023	2533044	31272	28.3	30.4
	8	9	Toronto Blue Jays	89	73	0.549	4.6	4.1	0.5	43-38	46-35	Toronto Blue Jays2023	3021904	37307	28.8	30.6
	9	10	Seattle Mariners	88	74	0.543	4.7	4.1	0.6	45-36	43-38	Seattle Mariners2023	2690418	33215	27.8	27.4
	10	11	Minnesota Twins	87	75	0.537	4.8	4.1	0.7	47-34	40-41	Minnesota Twins2023	1974124	24372	28.5	29.0
	11	12	Arizona Diamondbacks	84	78	0.519	4.6	4.7	-0.1	43-38	41-40	Arizona Diamondbacks2023	1961182	24212	27.4	28.5
	12	13	Miami Marlins	84	78	0.519	4.1	4.5	-0.4	46-35	38-43	Miami Marlins2023	1162819	14356	29.4	27.6
	13	14	Chicago Cubs	83	79	0.512	5.1	4.5	0.6	45-36	38-43	Chicago Cubs2023	2775149	34261	28.4	29.6
	14	15	San Diego Padres	82	80	0.506	4.6	4.0	0.6	44-37	38-43	San Diego Padres2023	3271554	40390	28.4	30.7
	15	16	Cincinnati Reds	82	80	0.506	4.8	5.1	-0.2	38-43	44-37	Cincinnati Reds2023	2038302	25164	26.8	27.7
	16	17	New York Yankees	82	80	0.506	4.2	4.3	-0.2	42-39	40-41	New York Yankees2023	3269016	40358	28.5	29.1
	17	18	San Francisco Giants	79	83	0.488	4.2	4.4	-0.3	45-36	34-47	San Francisco Giants2023	2500153	30866	28.5	30.0
	18	19	Detroit Tigers	78	84	0.481	4.1	4.6	-0.5	37-44	41-40	Detroit Tigers2023	1612876	19912	27.4	27.7
	19	20	Boston Red Sox	78	84	0.481	4.8	4.8	0.0	39-42	39-42	Boston Red Sox2023	2672130	32989	28.6	30.0
	20	21	Cleveland Guardians	76	86	0.469	4.1	4.3	-0.2	42-39	34-47	Cleveland Guardians2023	1834068	22643	26.7	26.1
	21	22	Pittsburgh Pirates	76	86	0.469	4.3	4.9	-0.6	39-42	37-44	Pittsburgh Pirates2023	1630624	20131	27.3	27.6
	22	23	New York Mets	75	87	0.463	4.4	4.5	-0.1	43-38	32-49	New York Mets2023	2573555	31772	28.9	31.9
	23	24	Los Angeles Angels	73	89	0.451	4.6	5.1	-0.6	38-43	35-46	Los Angeles Angels2023	2640575	32600	28.6	28.0
	24	25	St. Louis Cardinals	71	91	0.438	4.4	5.1	-0.7	35-46	36-45	St. Louis Cardinals2023	3241091	40013	27.5	29.7

Our Data

#A-S	#a-tA-S	Est. Payroll	Time	Season	TeamAbb	TG	HR	RTotal	SB	K.	ISO	AVG	OBP	wOBA	wRC.	WAR	city.x	Population
8	16	\$194,197,500	02:42:00	2023	ATL	162	307	947	132	0.2062730	0.2251206	0.2756834	0.3439552	0.3593281	124.70284	40.3310896	Atlanta	6106000
4	10	\$82,758,114	02:47:00	2023	BAL	162	183	807	114	0.2237465	0.1663330	0.2545951	0.3207083	0.3204568	105.41290	24.0359798	Baltimore	2355000
5	12	\$227,091,667	02:43:00	2023	LAD	162	249	906	105	0.2145902	0.1973208	0.2574222	0.3402404	0.3411893	116.29866	32.3573717	LosAngeles	12534000
4	7	\$75,441,212	02:38:00	2023	TBR	162	230	860	160	0.2303699	0.1850844	0.2598439	0.3314388	0.3352108	117.93957	32.3094620	Tampa	2977000
2	11	\$138,288,760	02:40:00	2023	MIL	162	165	728	129	0.2320842	0.1455626	0.2395098	0.3192295	0.3086803	91.85584	18.1681236	Milwaukee	1455000
3	9	\$240,388,766	02:45:00	2023	HOU	162	222	827	107	0.1984647	0.1778337	0.2588468	0.3311928	0.3316407	111.80440	27.2326505	Houston	6707000
2	12	\$241,362,606	02:44:00	2023	PHI	162	220	796	141	0.2386016	0.1820971	0.2557300	0.3270877	0.3294203	105.48895	22.1396231	Philadelphia	5785000
6	11	\$248,537,867	02:40:00	2023	TEX	162	233	881	79	0.2245480	0.1896336	0.2627346	0.3371465	0.3396940	114.10619	32.3496756	Arlington	6574000
5	16	\$211,190,269	02:45:00	2023	TOR	162	188	746	99	0.2094855	0.1607335	0.2558432	0.3286657	0.3239747	106.97842	25.7162399	Toronto	6372000
3	9	\$128,155,663	02:43:00	2023	SEA	162	210	758	118	0.2585067	0.1703636	0.2421818	0.3213363	0.3192701	106.66413	25.4725303	Seattle	3519000
2	8	\$137,798,640	02:42:00	2023	MIN	162	233	778	86	0.2659592	0.1841866	0.2433959	0.3257612	0.3266612	108.74735	24.8086404	Minneapolis	2990000
4	8	\$115,247,571	02:43:00	2023	ARI	162	166	746	166	0.2036251	0.1582046	0.2500000	0.3216634	0.3169704	97.37705	20.4183641	Phoenix	4717000
2	15	\$114,351,500	02:38:00	2023	MIA	162	166	666	86	0.2129737	0.1458978	0.2588685	0.3163757	0.3121733	93.95988	11.1328793	Miami	6265000
3	10	\$162,918,250	02:43:00	2023	CHC	162	196	819	140	0.2236334	0.1666061	0.2541788	0.3298404	0.3257697	104.42706	24.1078056	Chicago	8937000
2	15	\$236,200,139	02:46:00	2023	SDP	162	205	752	137	0.2121359	0.1695982	0.2436586	0.3291633	0.3232177	107.04318	26.1657435	SanDiego	3319000
1	3	\$77,877,833	02:47:00	2023	CIN	162	198	783	190	0.2421308	0.1702128	0.2493181	0.3269106	0.3246395	97.62472	14.2907294	Cincinnati	1775000
2	12	\$259,417,008	02:42:00	2023	NYN	162	219	673	100	0.2389084	0.1705805	0.2267518	0.3037062	0.3044883	94.24732	14.9383197	NewYork	18937000
2	11	\$177,920,416	02:39:00	2023	SFG	162	174	674	57	0.2447908	0.1485588	0.2348485	0.3120474	0.3041781	92.85549	14.2172798	SanFrancisco	3328000
1	5	\$119,236,836	02:38:00	2023	DET	162	165	661	85	0.2422697	0.1456736	0.2358525	0.3048137	0.2998402	88.58722	10.9673681	Detroit	3521000
1	7	\$181,282,500	02:45:00	2023	BOS	162	182	772	112	0.2222222	0.1659475	0.2583603	0.3240169	0.3236913	99.21710	13.0694294	Boston	4344000
2	9	\$70,114,729	02:39:00	2023	CLE	162	124	662	151	0.1873360	0.1313260	0.2501360	0.3131678	0.3015157	92.29197	15.2214764	Cleveland	1764000
2	5	\$72,407,500	02:40:00	2023	PIT	162	159	692	117	0.2404336	0.1527932	0.2391787	0.3145308	0.3084180	90.35013	13.7190484	Pittsburgh	1702000
2	13	\$208,427,344	02:45:00	2023	NYM	162	215	717	118	0.2197820	0.1693082	0.2379265	0.3162606	0.3149904	100.77711	18.9046202	NewYork	18937000
3	13	\$218,537,055	02:44:00	2023	LAA	162	231	739	72	0.2480065	0.1809073	0.2452177	0.3168624	0.3199747	101.19558	15.8959003	LosAngeles	12534000

The Models We Used

Simple Linear Regression

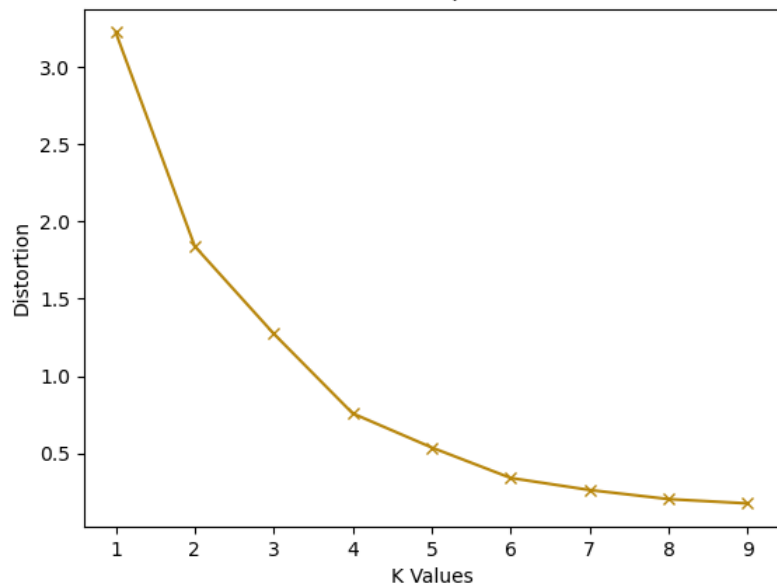
- We wanted to see if there might be an easier or simpler way to accomplish our goal
- Offensive Model
 - Used Runs Scored to predict Attendance - R^2 0.11
 - After trying a combination of different variables, we eventually utilized On Base Percentage to predict Attendance - R^2 0.23
- Defensive Model
 - Used Runs Against to predict Attendance - R^2 0.08
- Could not get the results we wanted from these models, likely due to their simplicity, as well as the fact that the data was still relatively raw.

K-Means Clustering

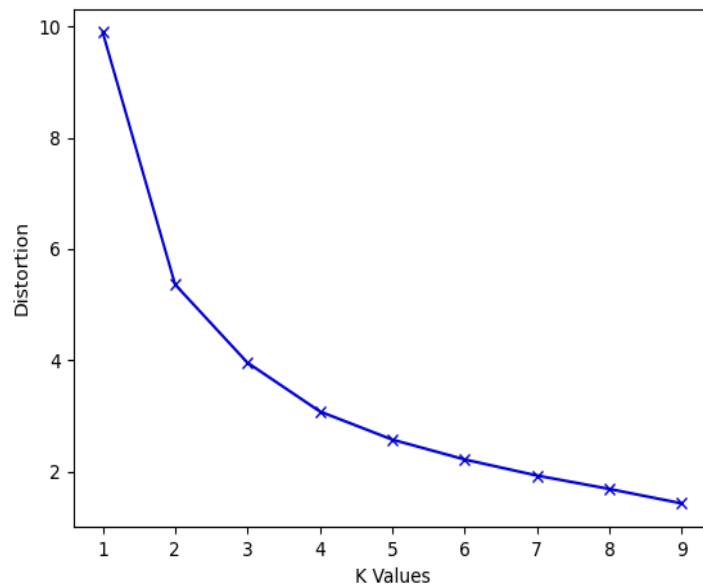
- We wanted to define thresholds for the population of metro areas around each MLB team's stadium and the contender status of individual teams based on wins to facilitate easier multiple linear regression
- Before we did that though, we needed to determine the appropriate number of clusters for each of these groups
- We used two methods to determine the appropriate number of clusters: the elbow method and silhouette scoring

K-Means Clustering: Elbow Method

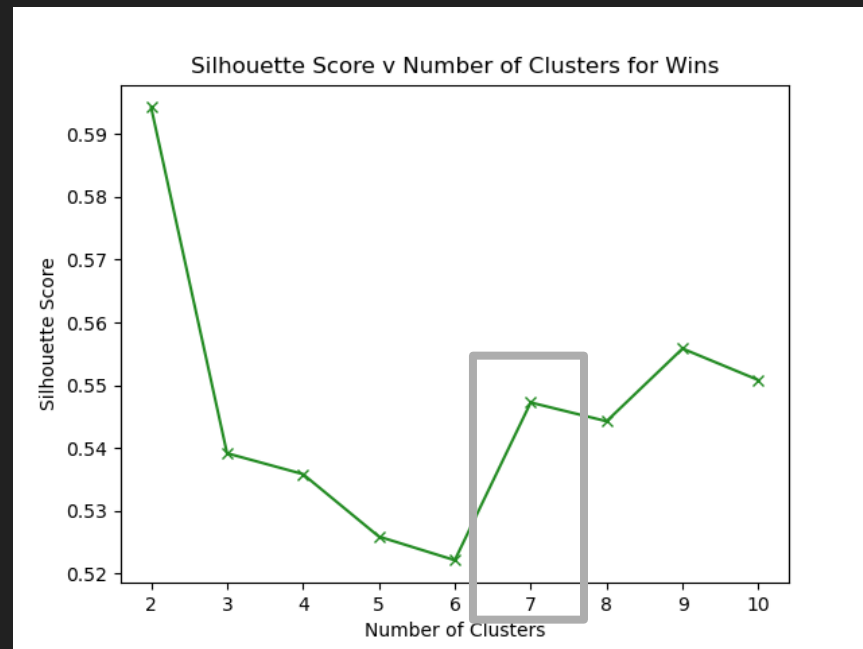
Elbow Method for Population Clusters



Elbow Method for Win Clusters

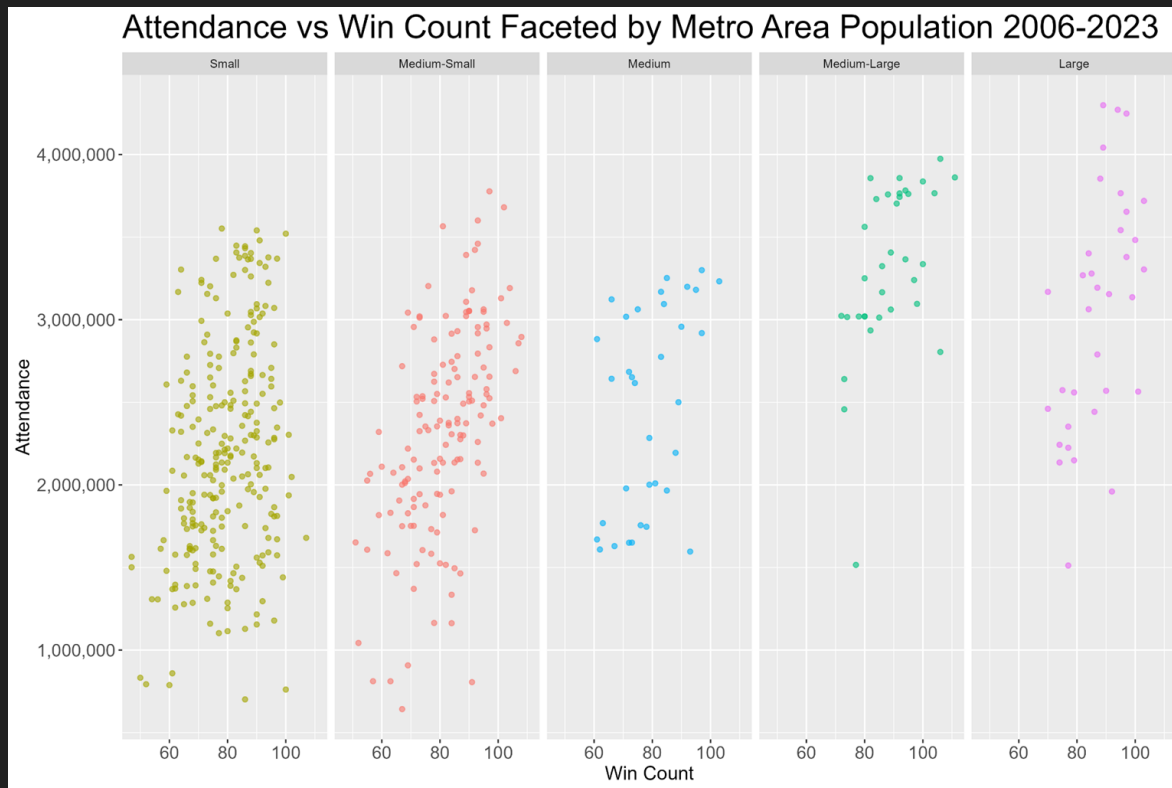


K-Means Clustering: Silhouette Scores



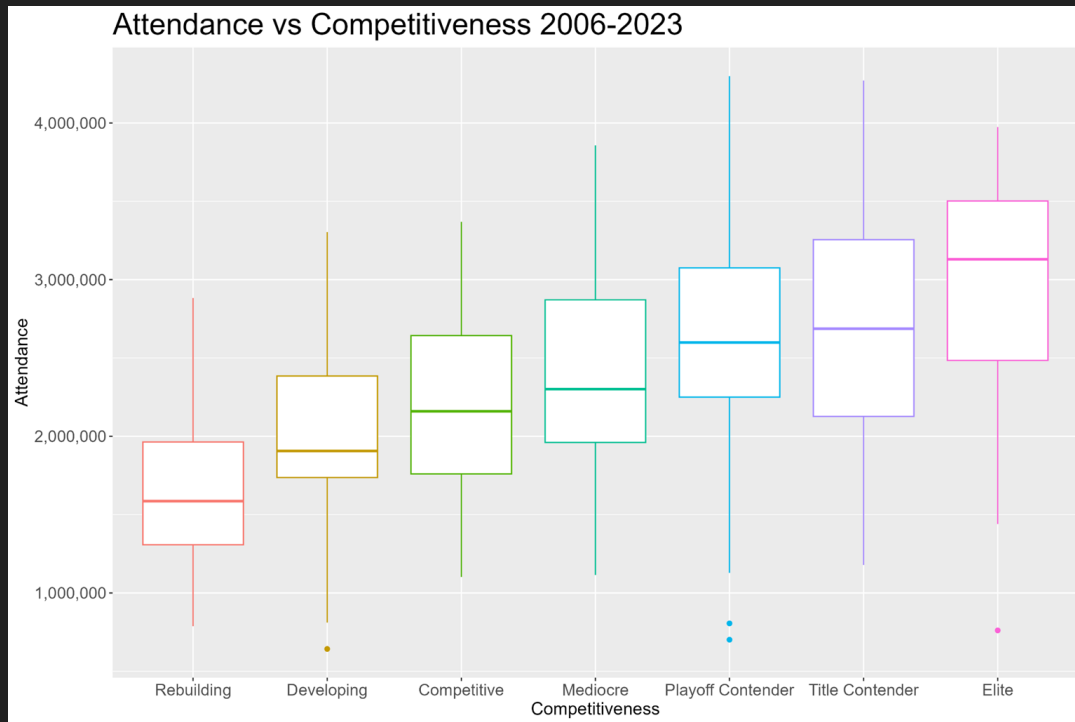
K-Means Classification (Population Levels)

- This graph shows attendance vs win count faceted by population around stadiums



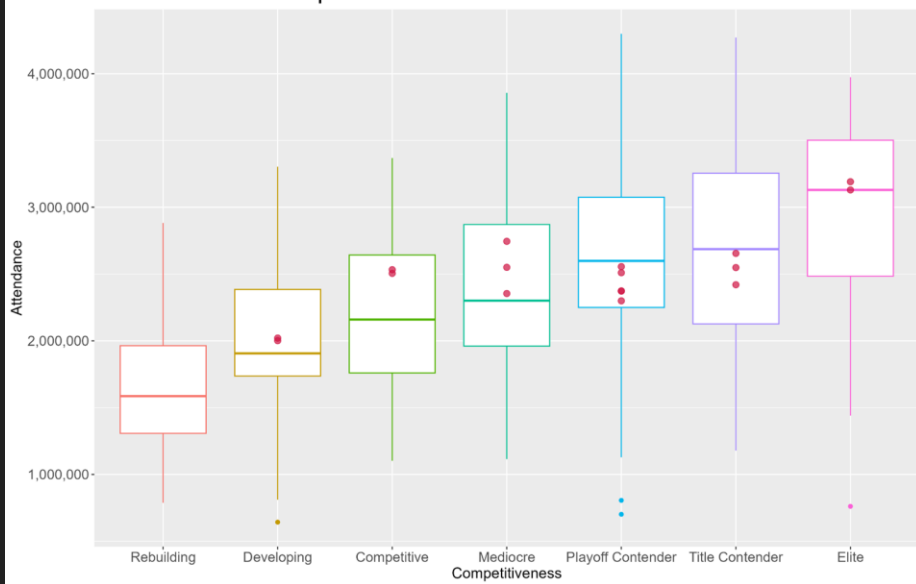
Win Totals and Team Investment (Classified)

- This graph plots attendance vs team competitiveness
- For teams in the Playoff Contender and Title Contender, there is not a large difference between their medians, indicating that at the higher level, investing in team performance produces a small boost in attendance, unless you can push the team to the elite level

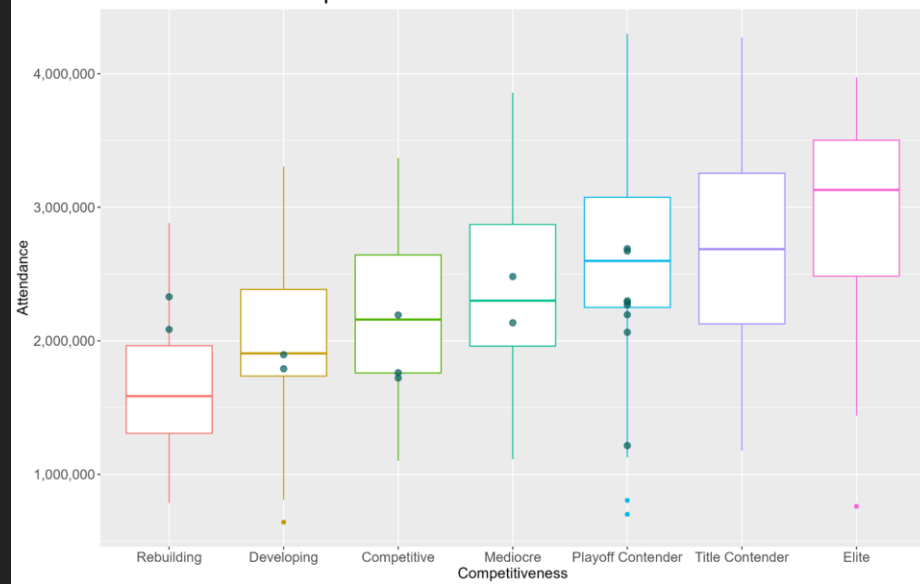


Case Study: Atlanta Braves and Seattle Mariners

Attendance vs Competitiveness with Atlanta Braves 2006-2023

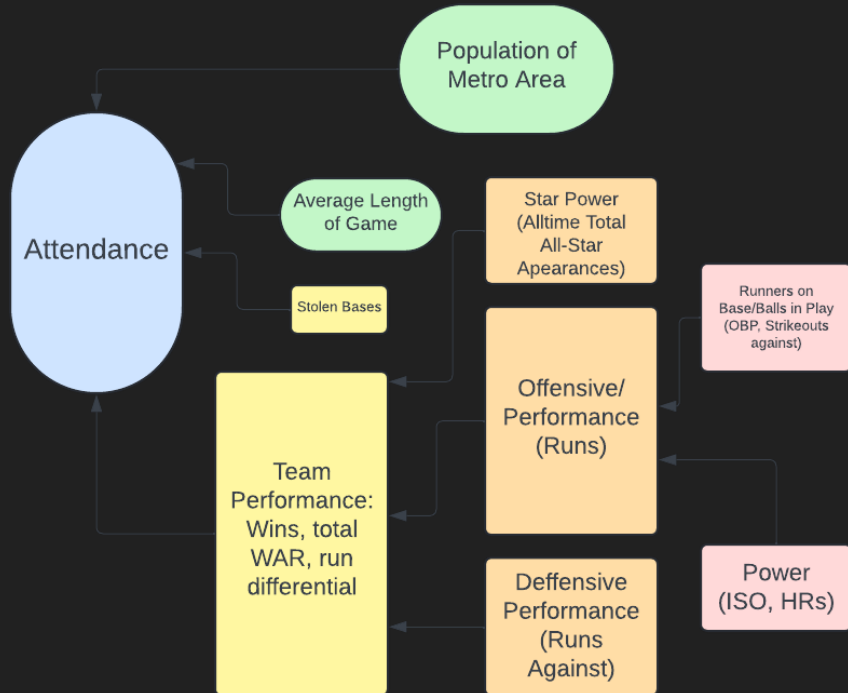


Attendance vs Competitiveness with Seattle Mariners 2006-2023



Feature Selection for Regression Models

- Used a variety of methods to select appropriate features for model
 - Domain knowledge
 - Understanding of multicollinearity
 - Correlation Matrix
- The graphic on the right shows a visual representation of some of the issues caused by multicollinearity

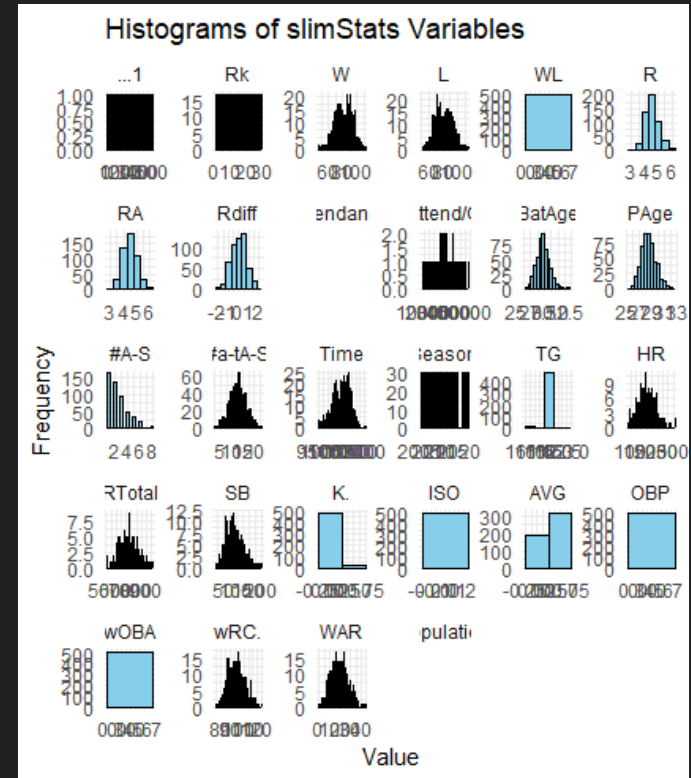


Correlation Matrix

	W	R	#a-tA-S	Rk	L	WL	RA	Rdiff	BatAge	PAGE	#A-S	HR	RTotal	SB	K	ISO	AVG	OBP	wOBA	wRC.	WAR	Population	Attendance
W	1.000000	0.598129	0.523908	-0.972901	-0.999743	0.999924	-0.709319	0.939405	NaN	NaN	0.638213	0.411221	0.601933	NaN	NaN	0.443257	NaN	0.539844	0.574046	0.694340	0.787654	NaN	0.454468
R	0.598129	1.000000	NaN	-0.579554	-0.597639	0.597636	NaN	0.662475	NaN	NaN	0.464513	0.669415	0.997958	NaN	NaN	0.760924	0.656164	0.829521	0.913154	0.727238	0.677080	NaN	NaN
#a-tA-S	0.523908	NaN	1.000000	-0.506384	-0.524218	0.523954	NaN	0.537596	0.412276	0.404007	0.505135	NaN	NaN	NaN	NaN	NaN	NaN	0.427604	0.425478	0.436392	0.490762	NaN	0.529812
Rk	-0.972901	-0.579554	-0.506384	1.000000	0.972866	-0.972974	0.689855	-0.910728	NaN	NaN	-0.636945	NaN	-0.582641	NaN	NaN	0.418073	NaN	-0.525382	-0.553458	-0.669772	-0.765465	NaN	-0.444030
L	-0.999743	-0.597639	-0.524218	0.972866	1.000000	-0.999928	0.709620	-0.939181	NaN	NaN	-0.638587	-0.410253	-0.601045	NaN	NaN	0.442725	NaN	-0.540106	-0.573588	-0.693690	-0.788067	NaN	-0.453437
WL	0.999924	0.597636	0.523954	-0.972974	-0.999928	1.000000	-0.709821	0.939384	NaN	NaN	0.638518	0.410727	0.601256	NaN	NaN	0.442891	NaN	0.539700	0.573606	0.693871	0.787814	NaN	0.453718
RA	-0.709319	NaN	NaN	0.689855	0.709620	-0.709821	1.000000	-0.728040	NaN	NaN	-0.438417	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	-0.499883	NaN	NaN
Rdiff	0.939405	0.662475	0.537596	-0.910728	-0.939181	0.939384	-0.728040	1.000000	NaN	NaN	0.646474	0.435025	0.667417	NaN	NaN	0.478963	NaN	0.592012	0.626338	0.743122	0.836628	NaN	0.454272
BatAge	NaN	NaN	0.412276	NaN	NaN	NaN	NaN	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	-0.442867	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.481435
PAGE	NaN	NaN	0.404007	NaN	NaN	NaN	NaN	NaN	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.478674
#A-S	0.638213	0.464513	0.505135	-0.636945	-0.638587	0.638518	-0.438417	0.646474	NaN	NaN	1.000000	NaN	0.464636	NaN	NaN	NaN	NaN	0.424077	0.468880	0.560244	0.615391	NaN	0.411121
HR	0.411221	0.669415	NaN	NaN	-0.410253	0.410727	NaN	0.435025	NaN	NaN	NaN	1.000000	0.671364	NaN	NaN	0.964895	NaN	NaN	0.556925	0.557255	0.476682	NaN	NaN
RTotal	0.601933	0.997958	NaN	-0.582641	-0.601045	0.601256	NaN	0.667417	NaN	NaN	0.464636	0.671364	1.000000	NaN	NaN	0.762354	0.657474	0.831015	0.914340	0.729759	0.680214	NaN	NaN
SB	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
K	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	-0.442867	NaN	NaN	NaN	NaN	NaN	1.000000	NaN	-0.695876	-0.521348	NaN	NaN	NaN	NaN	NaN
ISO	0.443257	0.760924	NaN	-0.418073	-0.442725	0.442891	NaN	0.478963	NaN	NaN	NaN	0.964895	0.762354	NaN	NaN	1.000000	NaN	0.427667	0.663669	0.590253	0.534435	NaN	NaN
AVG	NaN	0.656164	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.657474	NaN	-0.695876	NaN	1.000000	0.828894	0.793582	0.487247	0.450136	NaN	NaN
OBP	0.539844	0.829521	0.427604	-0.525382	-0.540106	0.539700	NaN	0.592012	NaN	NaN	0.424077	NaN	0.831015	NaN	-0.521348	0.427667	0.828894	1.000000	0.927738	0.668301	0.654707	NaN	0.434960
wOBA	0.574046	0.913154	0.425478	-0.553458	-0.573588	0.573606	NaN	0.626338	NaN	NaN	0.468880	0.556925	0.914340	NaN	NaN	0.663669	0.793582	0.927738	1.000000	0.759195	0.703049	NaN	0.406616
wRC.	0.694340	0.727238	0.436392	-0.669772	-0.693690	0.693871	NaN	0.743122	NaN	NaN	0.560244	0.557255	0.729759	NaN	NaN	0.590253	0.487247	0.668301	0.759195	1.000000	0.859277	NaN	NaN
WAR	0.787654	0.677080	0.490762	-0.765465	-0.788067	0.787814	-0.499883	0.836628	NaN	NaN	0.615391	0.476682	0.680214	NaN	NaN	0.534435	0.450136	0.654707	0.703049	0.859277	1.000000	NaN	0.419931
Population	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.000000	0.407611
Attendance	0.454468	NaN	0.529812	-0.444030	-0.453437	0.453718	NaN	0.454272	0.481435	0.478674	0.411121	NaN	NaN	NaN	NaN	NaN	NaN	0.434960	0.406616	NaN	0.419931	0.407611	1.000000

Feature Scaling

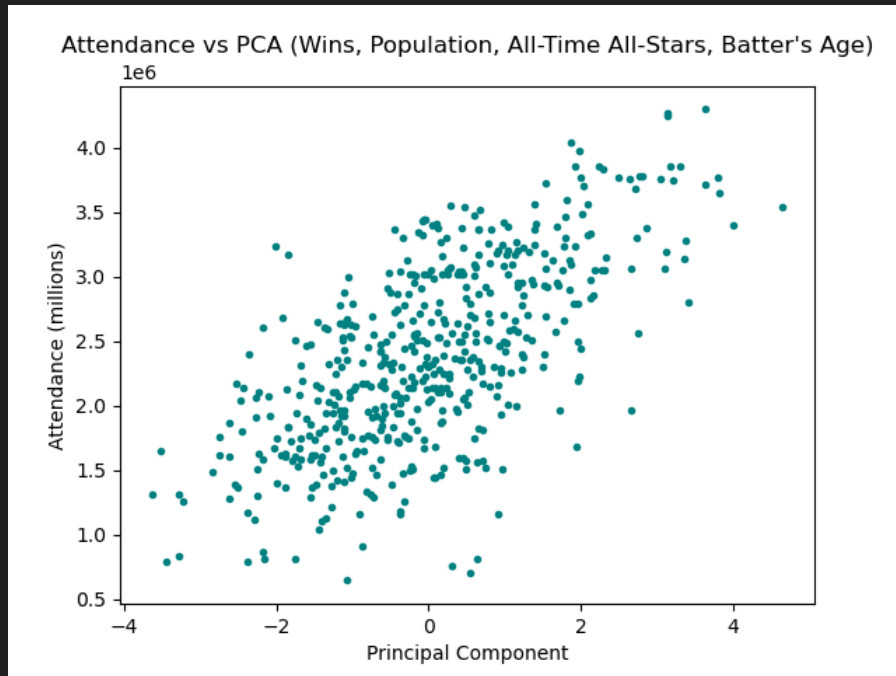
- In order to use some of our modeling techniques feature scaling was necessary to normalize the data
- Fortunately, the vast majority of our data was normally distributed
- This allowed for simple Z-score standardization of our data



* Example pre-processing EDA visualization, not intended to display results. Further visualizations showed normal distribution of metrics we used that are poorly visualized on this graphic

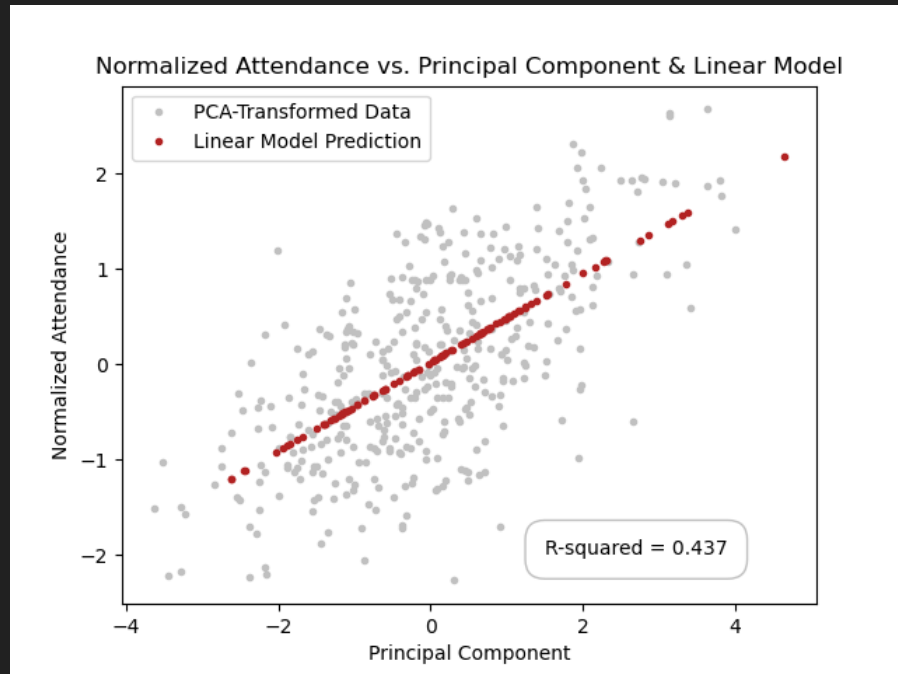
Principal Component Analysis

- PCA is our method of choice for dimensionality reduction
- Using the correlation matrix, we determined the features that had the highest correlation with attendance
- The features we used for our PCA were population, wins, all-time All-Stars, and batter's age
- We decomposed these four features down into a single principal component, which we used to perform regression



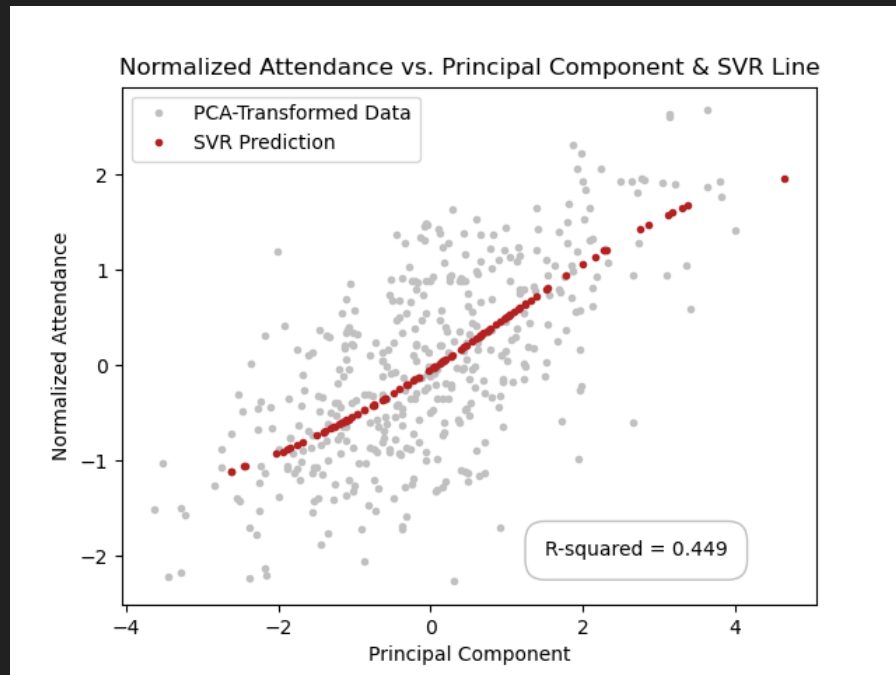
PCA-Transformed Linear Regression

- Wanted to revisit linear regression with the PCA-transformed data
- Got much better results with normalized data, and a wider variety of features
- R-squared of 0.437, while still low, was better than the initial regression model



Support Vector Regression

- Tried a handful of kernels, had best results with RBF
- Looped through combinations of C and γ , eventually getting our highest accuracy with $C=20$ and $\gamma=0.04$
- R-squared of 0.449 while less than ideal makes sense given the large amount of noise in attendance data



Our Model of Choice

- The best model we built for predicting attendance was our Support Vector Regression model utilizing Principle Component data
- While the R-squared wasn't as big as hoped, the model is still moderately strong, and outperformed the other regression models
- While the classification models with K-Means gave us some interesting observations, our SVR was by far most effective in predicting attendance
- Given the similar accuracies of our SVR and PCA-transformed linear regression models, the latter might be preferable if complexity/interpretability is a concern.

Conclusions

- Clustering models show evidence that there is a plateau around the “54%” mark
 - Also, benefits teams hunting for expanded playoff starts
- Simple linear regression and the correlation matrix suggests that there is correlation from a number of rule change targets
 - Stolen bases
 - Balls in play (also affected attendance more than runs or power related statistics)
 - Game time
- Our final model shows that total number of all star appearances and player age have a large impact on attendance suggesting that name recognition has a considerable effect on attendance that may be undervalued by MLB front offices

Primary Questions and Project Goals Revisited

- Does offensive play have a greater impact on per game attendance than defensive play?
 - **Yes, but only marginally per our basic linear models**
- Does it make sense for teams to aim for 'just good enough', or are there consistent marginal returns for more investment in players?
 - **At the Playoff Contender and Title Contender levels, there is little difference in attendance. Attendance does increase at the elite level**
- How big of a role does local population play on game attendance?
 - **Population does have an effect on game attendance, though this feature is hard to gauge accurately**

Primary Questions and Project Goals Revisited

- What factors beside population have the greatest impact on per game attendance?
 - **Per the correlation matrix and SVR, batter's age, wins, and all-time All-Stars have the greatest impact on attendance, indicating that name recognition is a factor in attendance**
- Can we predict a team's season attendance given their in games statistical projections?
 - **Our best model is SVR using a PCA decomposed from win count, batter's age, population and all-time All-Stars, which should allow us to predict somewhat accurately, though this can be improved with more non-collinear features**
- How should MLB front offices spend in free agency if they want to increase attendance?
 - **MLB front offices should invest in player's with higher name recognition as this seems to have the greatest effect on game attendance**

Limitations of Our Data

- MLB attendance data is inherently noisy
 - There are only 30 teams and many teams have unique factors affecting their attendance that can be difficult to capture (stadium quality, team history, local interest in baseball). This also makes it difficult to account for outliers
 - Public interest in MLB goes has varied for a litany of reasons including things like negative press causing attendance dips after the steroid era, and the fallout of the pandemic
- There is no perfect way to capture the geographic and demographic influences on team fan bases and potential attendees
 - Some teams share geographic markets whereas others have entire regions to themselves
 - Some teams have more of a national audience and better branding than others
- Similarly there is no perfect way to capture access to ballparks and the influence of surrounding communities
 - Neighborhoods and access to stadiums are variable
 - Public transport and traffic can play a role in people attending games

Avenues for Future Research

- To address many of the concerns with our data, many more advanced models could be implemented to more accurately account for the large variance in MLB attendances
 - Creating a complex model analyzing population and potential fans in surrounding regions that could account for larger issues
 - Finding more ways to adjust for league wide trends created by non baseball related influences
 - Creating a model to better address stadium quality and accessibility issues
- With the knowledge that established players appear to draw more fans, further research could be done on name recognition that could also incorporate roster turnover