MLB Team Attendance and Relocation (Technical Presentation)

Chris Cochrane

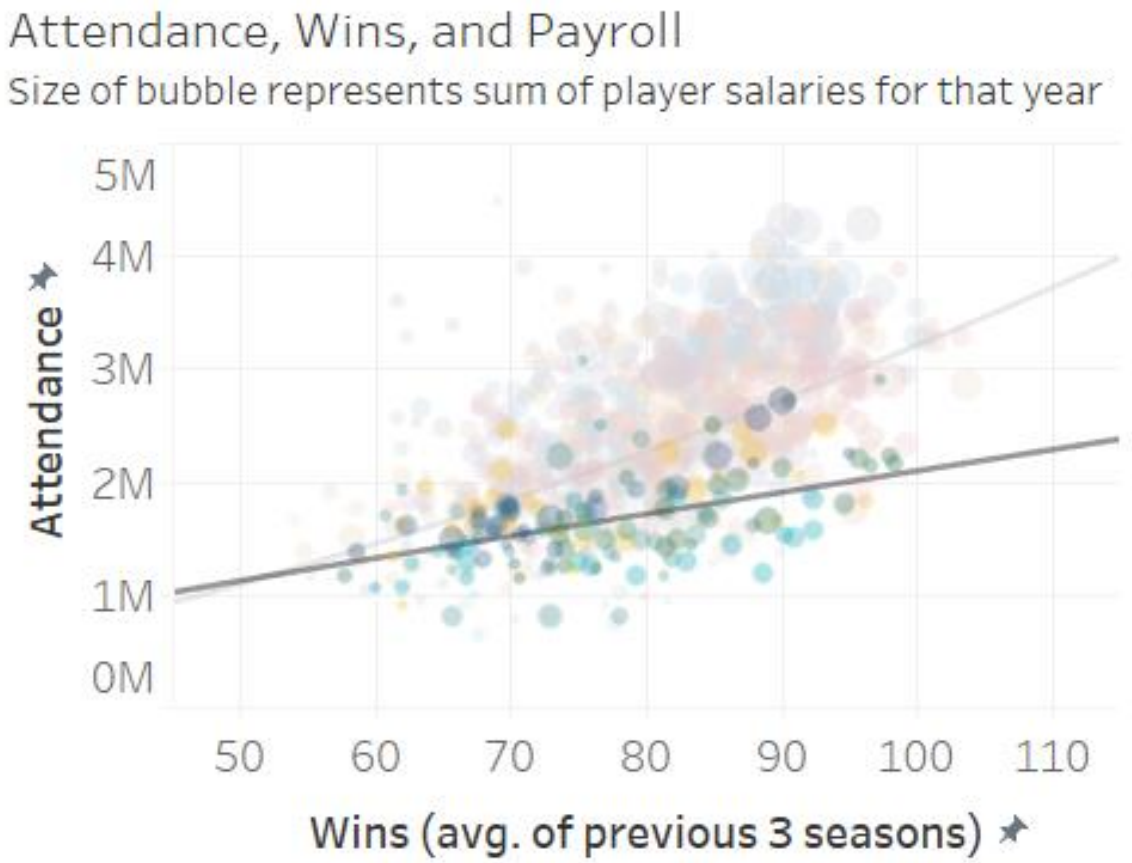
By creating a regression model for team attendance on a season basis, which, if any, MLB team would benefit the most from relocating to a new city and which city would maximize their attendance?

Major League Baseball (MLB) teams generate between 30-45% of their revenues from ticket sales, concessions, and parking. Unfortunately, there are some ballclubs that almost always have below-average attendance figures.	**Which teams have been having poor attendance figures in the last 10-15 years?**
It is assumed that the more a team wins, the more fans will come to the game. However, there have been teams that have successful seasons and still fail to bring in average-to-above-average sized crowds.	**What teams have had winning seasons but still below average attendance? What is causing there poor ticket sales?**
If geography, or local demographics are factors in a team's attendance, which factors are they?	**Which teams would benefit the most from moving to another city and which city would maximize their attendance?**

https://www.statista.com/statistics/193408/percentage-of-ticketing-revenue-in-the-mlb-since-2006/

Through a combination of descriptive statistics and inferential statistics, a linear regression model was created in order to choose the ideal city for a struggling MLB team to move to.

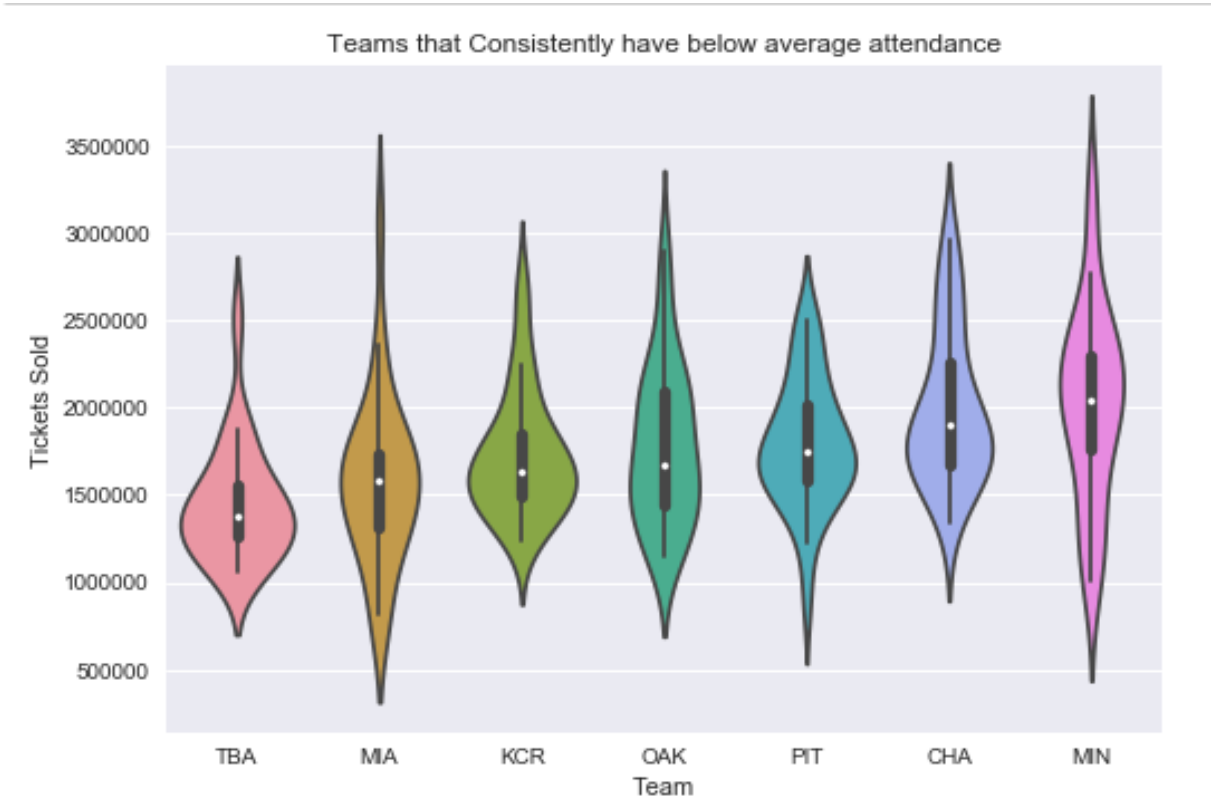
- By using descriptive statistics, five MLB teams (Tampa, Miami, Kansas City, Oakland, and Pittsburgh) have been chosen as relocation candidates due to attendance issues.
- Three variables related to an individual team and three others related to the city a team plays in were used to create a linear regression model to predict annual attendance.
- After combining the regression coefficients and test data and looking at hypothetical team relocations, it was determined Tampa Bay would have the highest percent growth in attendance by moving to Nashville with 21%.



https://www.kaggle.com/seanlahman/the-history-of-baseball/
MLB Salaries - MLB Baseball

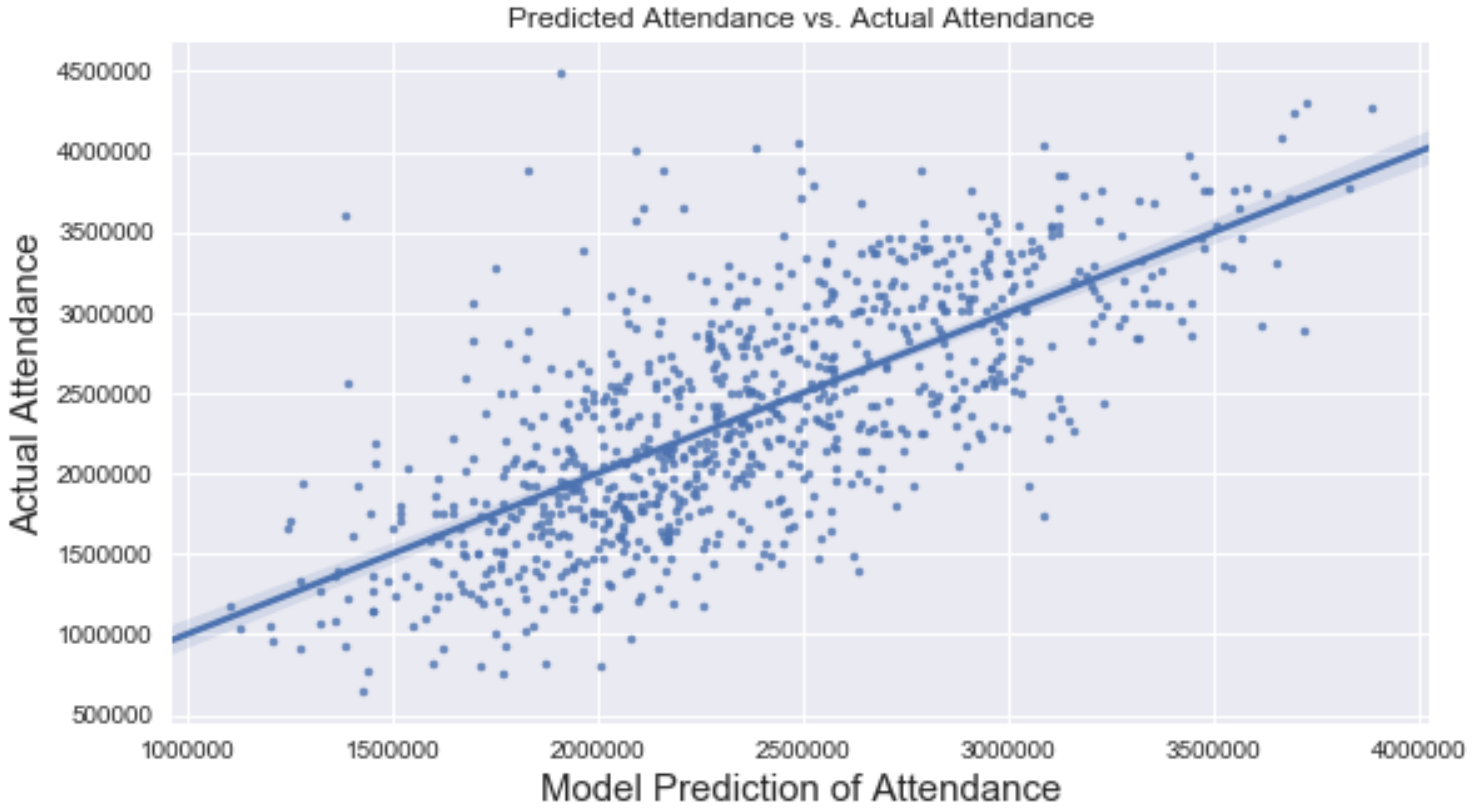
By using descriptive statistics, five MLB teams (Tampa, Miami, Kansas City, Oakland, and Pittsburgh) have been chosen as relocation candidates due to attendance issues.

- In regards to median annual attendance figures since 1990, these five teams had the lowest for the MLB. Chicago and Minnesota's 75th quartiles are the highest among the teams, and so are their 25th quartiles, therefore we will not analyze them for relocation.
- By looking at a violin plot, Tampa and Kansas City seem to have the least variation in the last 30 years for attendance. Their interquartile ranges are smaller than the other teams.
- The long plot for Miami indicates that there is some probability, although small, for the team to have very high attendance of over 3.5 million. Unfortunately, they also have a chance to have a season of below 0.5 million. The lowest attendance for any team since 1990, was Montreal in 2004 (0.75 million). They were relocated to Washington, DC the following year.



https://www.kaggle.com/seanlahman/the-history-of-baseball/
MLB Salaries - MLB Baseball

Three variables related to an individual team and three others related to the city a team plays in were used to create a linear regression model to predict annual attendance.



- Variables that had significant correlations with attendance were selected, keeping in mind to disregard any variables that were strongly correlated with each other to avoid multicolinearity.
- The model has an R-squared of .50, indicating it explains half of the variance of yearly attendance in the MLB.

https://www.kaggle.com/seanlahman/the-history-of-baseball/
MLB Salaries - MLB Baseball

Coefficients Explained

Dep. Variable:	attendance	R-squared:	0.500
Model:	OLS	Adj. R-squared:	0.497
Method:	Least Squares	F-statistic:	145.3
Date:	Thu, 16 Apr 2020	Prob (F-statistic):	1.57e-127
Time:	18:38:57	Log-Likelihood:	-12780.
No. Observations:	878	AIC:	2.557e+04
Df Residuals:	871	BIC:	2.561e+04
Df Model:	6		
Covariance Type:	nonrobust		

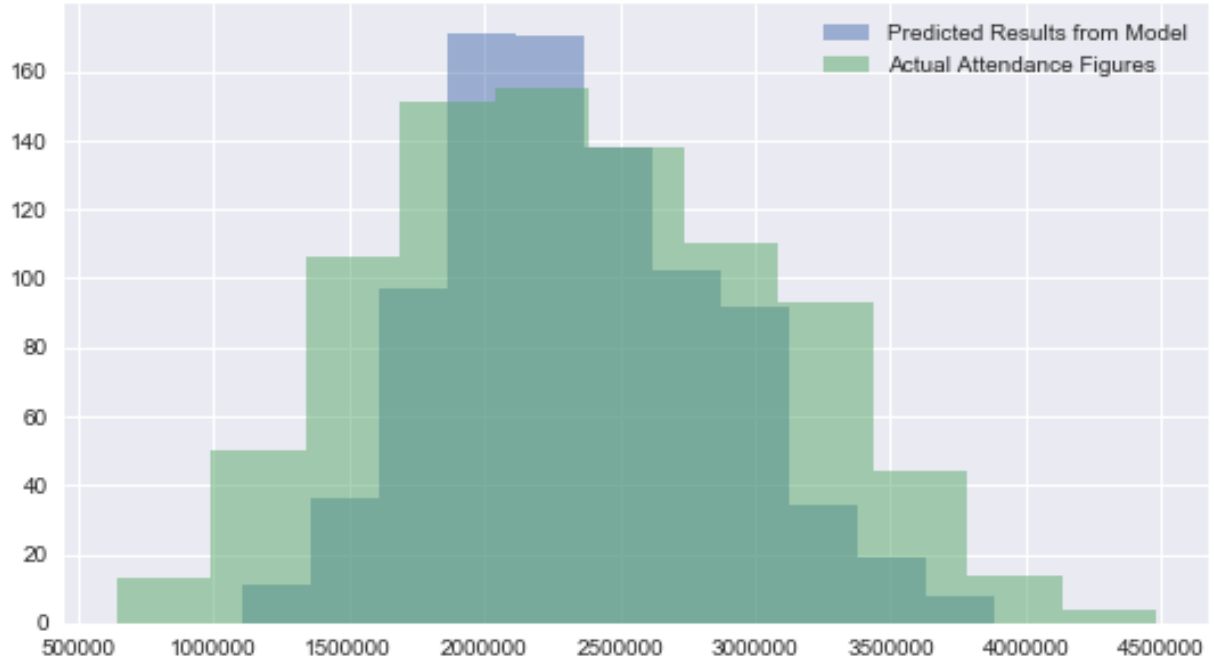
	coef	std err	t	P>|t|	[0.025	0.975]
const	-9.337e+05	1.71e+05	-5.456	0.000	-1.27e+06	-5.98e+05
rollingwins	3.197e+04	2100.374	15.219	0.000	2.78e+04	3.61e+04
payroll	4064.0588	459.012	8.854	0.000	3163.160	4964.958
parkscore	6.397e+05	6.96e+04	9.194	0.000	5.03e+05	7.76e+05
big4teams	4.316e+04	9637.123	4.479	0.000	2.43e+04	6.21e+04
abstemp	-2.171e+04	5101.539	-4.256	0.000	-3.17e+04	-1.17e+04
milestonextpark	494.5929	108.638	4.553	0.000	281.370	707.816

rollingwins	For every 1 more win a team averages over the previous 3 seasons, they can expect 31,970 more fans in the current year
payroll	For every 1 million in player payroll, a team can expect 4,064 more fans that year.
parkscore	For every 1 percentile increase a stadium increases it's ranking (parkscore), the team can expect 6,397 more fans. The best park would expect to bring in 639,700 more fans than the worst park.
big4teams	For every pro sports team in the city, a MLB team can expect 43,160 more fans that year.
abstemp	For every 1 degree a city's average temperature during baseball season strays from 70 F, a team can expect to have 21,710 less fans that year.
milestonextpark	For every 1 mile away from the next closest MLB team, a team will expect 495 more fans that year.

https://www.kaggle.com/seanlahman/the-history-of-baseball/
MLB Salaries - MLB Baseball

Three variables related to an individual team and three others related to the city a team plays in were used to create a linear regression model to predict annual attendance.

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.128e-16	0.024	4.71e-15	1.000	-0.047	0.047
rollingwins	0.4089	0.027	15.219	0.000	0.356	0.462
payroll	0.2526	0.029	8.854	0.000	0.197	0.309
parkscore	0.2372	0.026	9.194	0.000	0.187	0.288
big4teams	0.1267	0.028	4.479	0.000	0.071	0.182
abstemp	-0.1097	0.026	-4.256	0.000	-0.160	-0.059
milestonextpark	0.1291	0.028	4.553	0.000	0.073	0.185

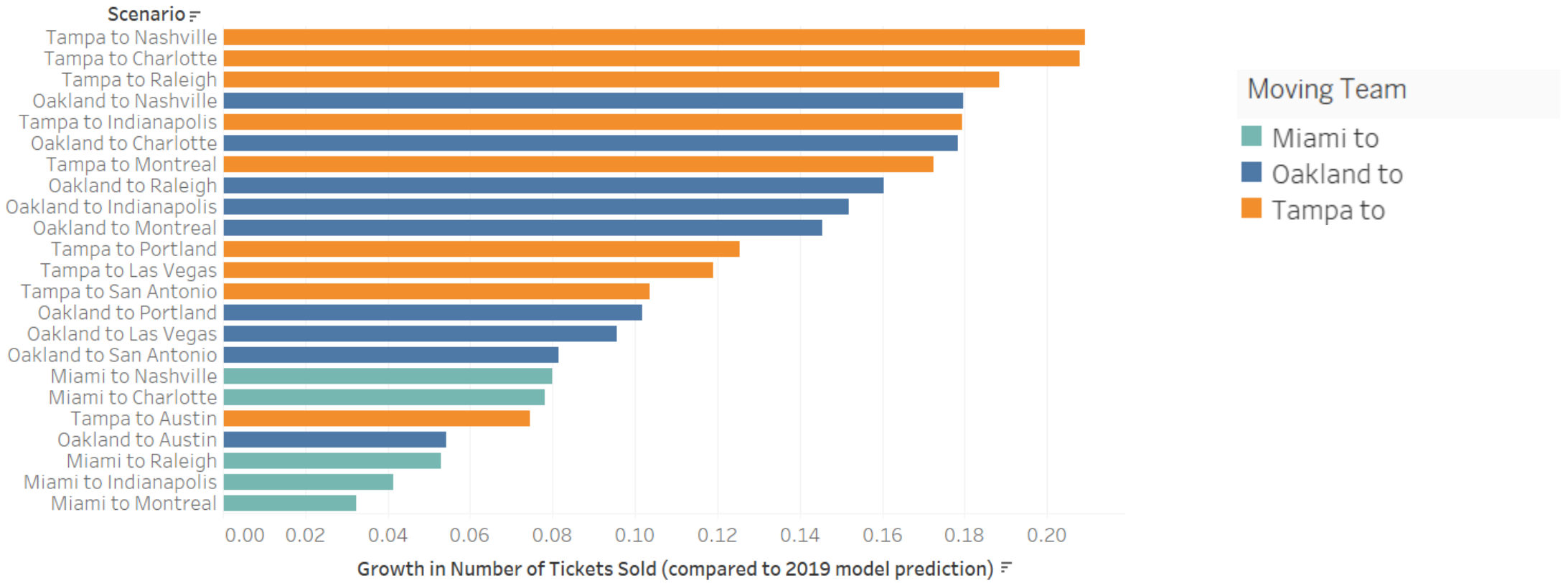


- After standardizing the model, it is determined the amount of wins over the last 3 years has the greatest effect on attendance among the tested variables. Player salaries (payroll) and ballpark quality were the next highest.

- The model does a fairly good job at predicting attendance, although it predicts less highs and lows than the actual results show in the histogram. The model's spread is smaller.

https://www.kaggle.com/seanlahman/the-history-of-baseball/
MLB Salaries - MLB Baseball

After combining the regression coefficients and test data and looking at hypothetical team relocations, it was determined Tampa Bay would have the highest percent growth in attendance by moving to Nashville with 21%.



- When inputting test data for relocation sites, a new parkscore of .5 was added for new cities while old city remained their current parkscore. Kept same win totals and payrolls.
- Nashville and Charlotte scored the highest for best relocation sites.
- Kansas City and Pittsburgh were determined to be in the best possible city when compared to relocation cities
- Nashville and Charlotte scored high due to their weather being near 70 degrees, being about 250 miles from the next nearest MLB team, and having other pro sports teams in the city already.

https://www.kaggle.com/seanlahman/the-history-of-baseball/
MLB Salaries - MLB Baseball