MLB Attendance

Baseball's Best Fan Bases

Michael Aronson, Jonathan Finch, Cody Justus, David Parker

What factors drive modern-day MLB attendance and what teams outperform attendance expectations?

Washington University in St. Louis
Olin Business School

MAJOR LEAGUE BASEBALL

INTRODUCTION

Executive Summary

This study first and foremost seeks to determine what factors drive MLB attendance in the 21st century. Using these factors as a baseline, we then took an empirical approach in determining which fan bases outperform expectations and thus are inherently strong fan bases.

Motivation

- Game attendance, and associated ticket sales, is critically important in Major League baseball, a result of the 162 game season, far longer than any of the other major professional sports leagues
- Widespread debate on which teams have the "best" fan bases
- Numerous "rankings," but typically qualitative in nature or based on a single season's numbers
- Opportunity to fill this gap and take a data-driven approach to answering this age-old question

Past Research

- Relatively little research on the topic of fan loyalty at the team level
- Existing research has focused on the psychology behind and drivers of extreme fandom at the individual level
  1. Richard Kolbe and Jeffrey James determined that primary drivers are developing a passion for the team at a young age from the father and the feeling of a "community of fans."
  2. Alan Tapp concurred regarding the importance of community and added that fans can become easily frustrated with a team and stop supporting them

Sources:

1. "An Identification and Examination of Influences That Shape the Creation of a Professional Team Fan" Kolbe, Richard H., James, Jeffrey D., *International Journal of Sports Marketing & Sponsorship*
2. "The loyalty of football fans-- We'll support you evermore?." Tapp, Alan, *Journal Of Database Marketing & Customer Strategy Management*

Preview of Results

From our analysis, several factors revealed themselves to be significant including win percentage, team payroll, and the age of the team. Below are the MLB franchises that were determined to have the strongest and weakest fan bases:

Strongest fan bases	Weakest fan bases
San Francisco Giants	Cleveland Indians
Los Angeles Angels	Washington Nationals
Milwaukee Brewers	Chicago White Sox
Boston Red Sox	Atlanta Braves
Houston Astros	Toronto Blue Jays

METHODOLOGY

Variables of Interest [Source]

Dependent variable: Attendance as % of capacity [ESPN]

- Control for differences in stadium size

Independent Variables:

1. Average Ticket Price (not including luxury suites) [Rodney Fort]
   - Natural log transformation to convert to percent changes
2. Payroll [USA Today]
   - Natural log transformation to convert to percent changes
3. Win% [baseballreference.com]
4. # of Pro Teams in Market [ESPN]
   - NFL, MLB, NBA, NHL, MLS
5. Founding Year* [mlb.com]
   - Separated into three categories (Founded: 1870-1900, Founded: 1901-1960, Founded: 1961-1990)
6. Time Since Last Championship [ESPN]
7. Playoffs?* [baseballreference.com]
8. American League?* [ESPN]

*Indicator Variable

Summary Statistics

Attendance:
Mean: 30,448 fans | Std. Dev.: 8,543
Range: 10,038 (Marlins ’02) - 53,069 (Yankees ‘08)

Ticket Price:
Mean: $23.17 | Std. Dev.: 9.10
Range: $9.33 (Twins ’00) - $72.97 (Yankees ’09)

Payroll:
Mean: $81,796,445 | Std. Dev.: 37,337,589
Range: $15m (Marlins ’06) - $241m (Dodgers ’14)

Win %:
Mean: 50% | Std. Dec.: 7.09%
Range: 26.5% (Tigers '03) - 71.6% (Mariners '01)

Number of Pro Teams in Market:
Mean: 4.47 teams | Std. Dev.: 2.19
Range: 2 (Baltimore) - 10 (New York)

Data Limitations

- Time span: only covers 15 years, excluding a large portion of baseball’s long history
- Singular focus: only covers the MLB, excluding the other major professional sports leagues
- Variables: excludes certain (hard to measure) variables such as weather in a given region/city

METHODOLOGY

Correlation Tests

	Attendance as % of Capacity	Win %	ln(Ticket Price)	ln(Payroll)	# of Pro Teams in Market	Founded: 1870-1900	Founded: 1901-1960	Founded: 1961-1990	Time Since Last Championship	Playoffs?	AL?
Attendance as % of Capacity	1.00										
Win %	0.44	1.00									
ln(Tcket Price)	0.63	0.17	1.00								
ln(Payroll)	0.67	0.36	0.68	1.00							
# of Pro Teams in Market	0.32	0.26	0.35	0.44	1.00						
Founded: 1870-1900	0.28	0.12	0.07	0.16	-0.02	1.00					
Founded: 1901-1960	0.07	0.14	0.28	0.16	0.18	-0.36	1.00				
Founded: 1961-1990	-0.14	-0.15	-0.13	-0.10	-0.09	-0.43	-0.43	1.00			
Time Since Last Championship	0.09	-0.14	0.12	-0.07	-0.07	0.08	-0.05	0.18	1.00		
Playoffs?	0.34	0.70	0.14	0.27	0.18	0.13	0.14	-0.20	-0.16	1.00	
AL?	-0.11	0.05	0.10	0.05	0.06	-0.57	0.64	0.06	-0.10	0.03	1.00

Results/Implications

- Ran correlation tests in order to test for potential issues of collinearity
- Payroll and ticket price were shown to be partially correlated, but due to their importance to the interpretation of our model (strongly correlated with attendance), both variables were kept, and ultimately both proved to be statistically significant
- AL indicator was removed due to collinearity with the year founded variables

ANALYSIS

Modeling Technique: Multiple Linear Regression

Initial Regression Results:

Observations	450	
R^2	0.6041	

	Coefficient	*Standard Error*
Intercept	-2.5714	0.2881
Win %	0.5948*	0.1134
Ln(Ticket Price)	0.1916*	0.0237
Ln(Payroll)	0.1284*	0.0188
# of Pro Teams in Market	0.0011	0.0029
Founded: 1870-1900	0.0613*	0.0205
Founded: 1901-1960	-0.0254	0.0204
Founded: 1961-1990	0.0018	0.0193
Time Since Last Championship	0.0008*	0.0003
Playoffs?	0.0204	0.0174

*p-value<0.05

Key Results:

- Statistically significant and strong positive coefficients for Win% and Ln(Payroll)
- Significant and positive coefficient for Founded: 1870-1900
- (Surprisingly) positive coefficient for Ln(Ticket Price)

Implications:

- Improvements in on-field performance and increases in payroll, which can be viewed as a proxy for ownership's commitment to the team's success, significantly increase attendance
- Baseball's most historic teams, those founded between 1870-1900, have an inherent attendance advantage
- Lowering ticket price is not an effective means of increasing attendance and is unlikely to compensate for poor on-field performance

ANALYSIS

Iterative Process

Step 1: Correlation Analysis
[Check for collinearity]

- Removed American League variable (high correlation with Year Founded variables)

Step 2: Initial Regression
[Check for statistical significance]

- Removed # of Pro Teams in Market variable
- Removed Founded (1901-1960 & 1961-1990) variables
- Removed Playoffs variable

Steps 3 & 4: Final Regression & Predictive Model

- With variables removed, ran a final regression
- The coefficients from this regression were used to build our predictive model

Predictive Model

Attendance = -2.6252 + 0.6623*Win% + 0.1829*Ln(Ticket Price) + 0.1311*Ln(Payroll) + 0.0711*Founded:1870-1900 + 0.0008*Time Since Last Championship + ε

Empirical Approach

The above model predicts a team's attendance based on all of the listed factors. The error term (ε) thus captures any other variables that could explain attendance but are not included in the model. In our effort to include the majority of the measurable variables that could drive attendance, we believe the error term would effectively capture the unmeasurable effect of the inherent strength of fan bases. Under this belief, the difference between our prediction and the actual attendance would provide a metric for measuring this inherent factor. The results of this empirical approach are shown on the following slide.

CONCLUSIONS

Difference Between Predicted & Actual Attendance (Highest to Lowest)

San Francisco Giants	14.78%
Los Angeles Angels	13.77%
Milwaukee Brewers	11.90%
Boston Red Sox	8.34%
Houston Astros	8.26%
Colorado Rockies	6.48%
St. Louis Cardinals	5.39%
Detroit Tigers	5.32%
San Diego Padres	5.26%
Minnesota Twins	3.67%
Chicago Cubs	2.99%
Arizona Diamondbacks	2.11%
Pittsburgh Pirates	1.79%
Kansas City Royals	0.05%
Texas Rangers	-0.20%
Miami Marlins	-1.12%
Baltimore Orioles	-2.19%
Oakland Athletics	-2.50%
Seattle Mariners	-4.13%
Cincinnati Reds	-4.28%
Philadelphia Phillies	-4.68%
Los Angeles Dodgers	-4.77%
New York Mets	-4.79%
Tampa Bay Rays	-4.87%
New York Yankees	-6.54%
Cleveland Indians	-7.73%
Washington Nationals	-9.09%
Chicago White Sox	-9.46%
Atlanta Braves	-11.21%
Toronto Blue Jays	-12.55%

Insights

- Many teams with positive differentials are consistent with common perceptions of teams with strong fan bases (Giants, Red Sox, Cardinals, Tigers, and Cubs)
- Some of the more surprising teams (Brewers, Astros, Twins, Pirates) most likely show up in our list for two reasons:
  - (1) Teams with smaller stadiums perform better in our study, considering our dependent variable is percentage of stadium capacity filled.
  - (2) Our model predicts that these teams would have weaker attendance due to their performance with respect to our independent variables (win percentage, payroll, etc.), but they have been able to exceed the notably low expectations.
- Similarly, several teams in the bottom half of the rankings stand out as surprises, namely the Dodgers and Yankees
  - These teams are subject to the flip side consequences of reasons (1) and (2) above. They have well above average size stadiums and thus are at a disadvantage within our model with respect to percentage of capacity filled instead of total attendance.
  - Furthermore, based on these teams' consistently strong performance (i.e. playoffs, win percentage, payroll), our model predicts extremely high attendance numbers that might be unachievable in the real world.
- **In sum, we believe our model serves as a credible basis to examine the relative strengths of fan bases for MLB teams**

Potential Extensions

- Extend list of independent variables to try to capture additional explanatory power for some of the more surprising findings (new stadium, median income, market size, etc.)
- Expand to other pro sports and look for trends within cities