Drivers of Major League Baseball Game Attendance

Daniel Freeman, Laurie Harris, Timothy McWilliams
File Organization and Database Management, Southern Methodist University
MSDS 7330, Section 403
August 23, 2017

Abstract: In professional sports, there are a number of factors that determine a team's attendance total for a given year. In this exploratory study, we analyze several of these factors, including win-loss record, runs scored, runs allowed, home runs and market size, via ranking methods for each of Major League Baseball's 30 teams. Our analyses show that winning percentage is the primary driver of attendance while runs scored also plays an important role in bringing fans out to the ballpark. Runs allowed, home runs and market size are less influential in driving attendance.

1. Introduction

This paper examines annual home game attendance totals for all Major League Baseball (MLB) teams from 1969 through 2016. We identify maximum attendance for each team and study the relationship between that and the team's win-loss record for the season. We also investigate potential covariates such as home runs, total runs scored and total runs allowed, as well as population counts for each U.S.-based team's metropolitan statistical area (MSA).

2. Existing Research

Baseball is a game for which many statistical measurements are collected. There is frequent review of many team management decisions and how they affect performance outcomes. Many studies, such as those made available by the Society for American Baseball Research [1], focus on the correlations between ballpark attendance and ticket prices or team performance and player payrolls. In addition, we found the 2015 article published by www.fivethirrtyeight.com [2] which asserted that payroll is a significant driver of a team's win total to be influential in our decision to study this topic.

Our analysis will remove traditional discussions of team salaries and ballpark ticket prices and focus on the association between team performance and market size on paid attendance. We find our approach to studying this topic to be both interesting and unique.

3. Data Specifications

Analysis was performed using a MySQL relational database on a Windows 10 operating system. Unless otherwise noted, all charts and tables were produced in Microsoft Office. Data files were obtained through easily accessible website links. All source data files and analysis queries are located at the following link: www.github.com/llh1000/MSDS7330.

3.1 Data Sources

Attendance: We sourced yearly attendance totals for each team from www.ballparksofbaseball.com [3]. Through the period of review[1], some teams have played at multiple facilities. For the purposes of this analysis, we identified the top attendance by team, regardless of changes in facility. (Data file size ~ 3 KB)

Team Performance: For each team, for each season, we obtained win-loss records from www.baseball-reference.com [4] and runs scored [5], runs allowed [6] and home runs [5] from www.mlb.com. (Win-loss data file size ~ 22 KB, runs data file size ~ 87 KB)

Population Data: Population counts were obtained from the U.S. Census Bureau [7, 8] for the period of 2000-2016 for each of the MSA's in which the 29 U.S.-based teams are located. We did not collect population data for the Toronto metropolitan area, since we were unable to find historical population figures prior to 2013 for the Canadian metropolitan areas. Thus, the Toronto Blue Jays are excluded from our population ranking analysis. Furthermore, we did not use population counts prior to 2000 due to the Census Bureau's re-definition of MSA for the 2000 decennial census; including data prior to 2000 would have caused major trend breaks in each of the MSA population time series. (Data file size ~ 20 KB)

3.2 Data Processing

Data file sizes were very manageable, as noted above, and system resources were not impacted in any way that would impede the analysis.

4. Analysis

4.1 Maximum Attendance

For each of the 30 teams, we identified a single season to represent the year for which the highest attendance total was observed at the team's home ballpark. Although our period of review extended from 1969 through 2016, we observed the maximum attendance records for most teams occurring in the more recent decades. Figure 1 shows the number of teams and the decade of their maximum attendance record.

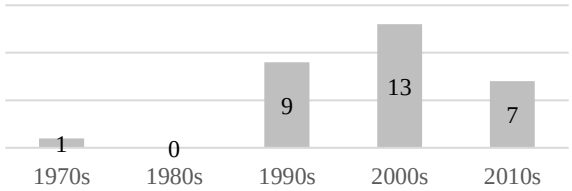

Figure 1. Number of Teams with Record Attendance by Decade

[1] All seasons of play are included for teams who entered MLB after 1969.

Beyond the noticeable concentration of maximum attendance records occurring in the later decades, we wanted to determine if there was a heavy concentration of maximum attendance records in any specific years. There did not appear to be any unusual distribution of maximum attendance records by year as seen in Figure 2 below.

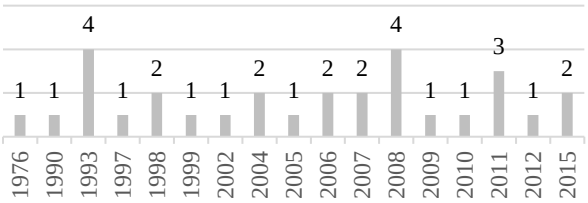

Figure 2. Number of Teams with Record Attendance by Year

4.2 Win-Loss Records and Maximum Attendance

Using the information about the maximum attendance records, we identified whether those observations occurred during a season in which the home team had a particularly successful season.

For the purposes of this analysis we identified the most successful seasons as one of the top ten seasons by winning percentage. We found that for 17 of the 30 MLB teams (57%), the maximum attendance record was observed during one of the team's top ten winning seasons. Table 1 shows the 17 teams' season, attendance, winning percentage and win-loss rank.

Table 1. Teams with Maximum Attendance during a Top Ten Winning Season

Team	Season	Attendance	Winning Percentage	Win-Loss Rank
Milwaukee Brewers	2011	3,071,373	59%	1
Philadelphia Phillies	2011	3,680,718	63%	1
Atlanta Braves	1993	3,884,720	64%	2
Cincinnati Reds	1976	2,629,708	63%	2
Pittsburgh Pirates	2015	2,498,596	60%	2
Chicago Cubs	2008	3,299,840	60%	3
Cleveland Indians	1999	3,468,436	60%	3
Kansas City Royals	2015	2,708,549	59%	3
Oakland Athletics	1990	2,900,217	64%	3
Seattle Mariners	2002	3,542,938	57%	3
Toronto Blue Jays	1993	4,057,947	59%	4
Minnesota Twins	2010	3,223,640	58%	5
Texas Rangers	2012	3,460,280	57%	5
Baltimore Orioles	1997	3,711,132	60%	6
Houston Astros	2004	3,087,872	57%	6
Chicago White Sox	2006	2,957,414	56%	8
San Diego Padres	2004	3,040,046	54%	8

For those teams that had a maximum attendance record outside of one of their top ten seasons, we wanted to explore if there were other potential factors that would intuitively explain a maximum attendance observation during such a season.

Expansion and Relocation: Five teams saw maximum attendance during their inaugural MLB season. In 1993, the Colorado Rockies and Miami Marlins had maximum attendance records, as did the Arizona Diamondbacks and Tampa Bay Rays in 1998. The Washington Nationals also saw their maximum attendance record in 2005, their first year following relocation and renaming from the Montreal Expos.

Winning Seasons; Not Highest Performance: Six other teams saw maximum attendance during a season that was not considered one of their top ten seasons, but was still a winning season. For the teams listed in Table 2 below, we note that these are all long-time, well-established teams.

Table 2. Teams with Maximum Attendance during a Winning Season (outside of top ten win-loss rank)

Team	Season	Attendance	Winning Percentage	Win-Loss Rank
Boston Red Sox	2009	3,062,699	58%	11
Los Angeles Angels	2006	3,406,790	55%	13
Los Angeles Dodgers	2007	3,856,753	50%	37
New York Mets	2008	4,047,404	55%	12
New York Yankees	2008	4,298,655	55%	28
San Francisco Giants	2011	3,387,303	53%	21

Other Considerations: Of the 30 MLB teams, 17 saw maximum attendance during one of their top ten seasons. For five teams, it was observed during their inaugural league season and for six teams it was observed outside of a top ten year but within a winning season. The two remaining teams are the 2007 St. Louis Cardinals and the 2008 Detroit Tigers. We suspect the St. Louis case is due to a residual effect of excitement following the team's 2006 World Series championship. The Detroit case is not as easily explained. Although the 2008 Detroit Tigers only won 45% of their games and finished last in their division, the team saw a maximum attendance total set that year of 3,202,654. However, the offseason acquisitions of all-stars Miguel Cabrera and Dontrelle Willis resulted in high expectations for the Tigers in 2008 [9], which likely boosted ticket sales.

4.3 Runs Scored, Runs Allowed, Home Runs and Maximum Attendance:

We reviewed runs scored totals for each team for each season in order to compare the runs scored rank for each team to that team's maximum attendance year. Of the 30 MLB teams, the maximum attendance record was associated with one of their top ten most successful seasons by runs scored for 13 teams (43%), as shown in Table 3. Seventeen of the 30 MLB teams (56%) had maximum attendance records during a year that was not considered one of their top ten seasons for runs scored.

Table 3. Maximum attendance record and top ten runs scored ranked

Team	Season	Attendance	Runs Scored	Runs Scored Rank
Chicago Cubs	2008	3,299,840	855	1
Cleveland Indians	1999	3,468,436	1,009	1
Cincinnati Reds	1976	2,629,708	857	2

Chicago White Sox	2006	2,957,414	868	3
San Diego Padres	2004	3,040,046	768	4
Toronto Blue Jays	1993	4,057,947	847	5
Houston Astros	2004	3,087,872	803	6
Baltimore Orioles	1997	3,711,132	812	6
New York Mets	2008	4,047,404	799	6
Seattle Mariners	2002	3,542,938	814	7
Minnesota Twins	2010	3,223,640	781	8
Boston Red Sox	2009	3,062,699	872	8
Detroit Tigers	2008	3,202,654	821	9

We also reviewed runs allowed totals for each team for each season in order to compare the runs allowed rank for each team to that team's maximum attendance year. The maximum attendance record was associated with one of the top ten most successful seasons by runs allowed for 7 teams (23%). Twenty-three teams (76%) had maximum attendance records during a year that was not considered one of their top ten seasons for runs allowed.

We also examined home run totals for each season for each team in the same way that we did for runs scored and runs allowed. The maximum attendance record was associated with one of the top ten most successful seasons by home runs for 8 teams (26%). Twenty-two teams (73%) had maximum attendance records during a year that was not considered one of their top ten seasons for home runs.

When looking at ranked runs scored, runs allowed and home runs against the maximum attendance for each team it is clearly evident that runs scored is more closely related to the maximum attendance. Figure 3, a chart that we produced in R, displays the teams in Table 3 and compares their runs scored, runs allowed and home run ranks. From this figure we can say that the maximum attendance year for the 30 MLB teams is more closely related to the runs scored ranks than the runs allowed or home run ranks.

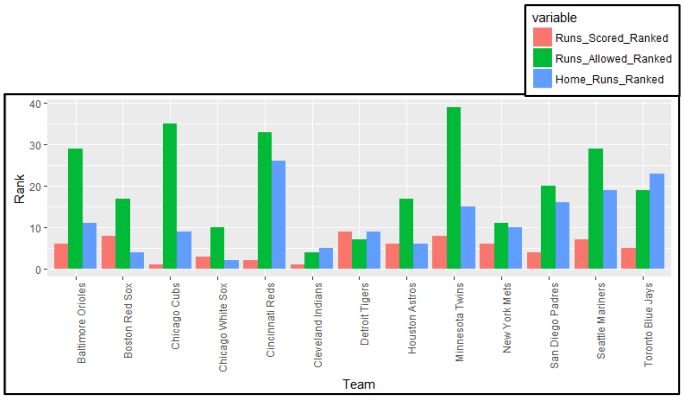

Figure 3. Comparison of Runs Scored, Runs Allowed and Home Run Ranks for Teams' Maximum Attendance Year

4.4 Population and Maximum Attendance

For each of the 29 U.S.-based teams, we compared their ranks for average annual attendance with their ranks for average annual MSA population for each of the following time periods: 2000-2016 (entire period for which we have data), 2000-2004 (first half of 2000's decade), 2005-2009 (second half of 2000's decade) and 2010-2016 (2010's decade to date).

For the period 2000-2016, large-market teams the New York Yankees (1st), Los Angeles Dodgers (2nd) and Chicago Cubs (5th) rank in the top five by average attendance (Table 4). However, as shown in Table 4, the St. Louis Cardinals (3rd), San Francisco Giants (4th), Boston Red Sox (7th) and Colorado Rockies (10th) all rank in the top 10 by average attendance, despite ranking no higher than 13th in average MSA population. For the most part, this can be attributed to these teams' zealous fan bases, especially those of the Cardinals, Giants and Red Sox, all of which have won multiple World Series titles during this period.

Table 4. Average Attendance Rank vs. Average Population Rank (2000-2016, Top 15 Teams)

Team	Attendance Rank	Population Rank
New York Yankees	1	1
Los Angeles Dodgers	2	3
St. Louis Cardinals	3	21
San Francisco Giants	4	15
Chicago Cubs	5	5
Los Angeles Angels	6	3
Boston Red Sox	7	13
New York Mets	8	1
Philadelphia Phillies	9	8
Colorado Rockies	10	24
Texas Rangers	11	7
Seattle Mariners	12	18
Atlanta Braves	13	12
Milwaukee Brewers	14	29
Detroit Tigers	15	14

Conversely, Table 5 shows that the Astros (16th), White Sox (22nd) and Marlins (28th) fall in the bottom half by average attendance yet rank in the top 10 by average MSA population. During the 2000-2016 period, these teams suffered through several losing seasons, although the Marlins and the White Sox won the World Series in 2003 and 2005, respectively.

Table 5. Average Attendance Rank vs. Average Population Rank (2000-2016, Bottom 14 Teams)

Team	Attendance Rank	Population Rank
Houston Astros	16	9
San Diego Padres	17	20
Baltimore Orioles	18	23
Arizona Diamondbacks	19	17
Minnesota Twins	20	19
Cincinnati Reds	21	26
Chicago White Sox	22	5
Cleveland Indians	23	27
Pittsburgh Pirates	24	25
Washington Nationals	25	11
Oakland Athletics	26	15
Kansas City Royals	27	28
Miami Marlins	28	10
Tampa Bay Rays	29	22

During the first half of the 2000s decade, seven teams – including the second-ranked Seattle Mariners – in the top 10 by average attendance ranked outside of the top 10 by average MSA population. It is unclear exactly what the causes were for that many teams in smaller markets to rank that highly in attendance. The only apparent factor is that the 2001 Seattle Mariners won a Major League record-tying 116

games, which likely gave a spike to their attendance numbers during that and subsequent years.

Table 6. Average Attendance Rank vs. Average Population Rank (2000-2004, Top 15 Teams)

Team	Attendance Rank	Population Rank
New York Yankees	1	1
Los Angeles Dodgers	2	3
St. Louis Cardinals	3	21
San Francisco Giants	4	15
Chicago Cubs	5	5
Los Angeles Angels	6	3
Boston Red Sox	7	13
New York Mets	8	1
Philadelphia Phillies	9	8
Colorado Rockies	10	24
Texas Rangers	11	7
Seattle Mariners	12	18
Atlanta Braves	13	12
Milwaukee Brewers	14	29
Detroit Tigers	15	14

During the second half of the 2000s decade, the St. Louis Cardinals ranked 3rd in average attendance – behind only the Yankees and Dodgers – despite ranking only 21st in average MSA population. This is likely attributed to the Cardinals' on-field success, including winning the 2006 World Series, as well as the rise of the team's superstar player Albert Pujols. As compared to the first half of the 2000s decade, the Red Sox, who won the World Series in 2004 and 2007, move up from 11th to 9th by average attendance while the Giants slide back from 3rd to 8th.

Table 7: Average Attendance Rank vs. Average Population Rank (2005-2009, Top 15 Teams)

Team	Attendance Rank	Population Rank
New York Yankees	1	1
Los Angeles Dodgers	2	3
St. Louis Cardinals	3	21
New York Mets	4	1
Los Angeles Angels	5	3
Chicago Cubs	6	5
Philadelphia Phillies	7	8
San Francisco Giants	8	15
Boston Red Sox	9	13
Houston Astros	10	9
Milwaukee Brewers	11	29
Detroit Tigers	12	14
Chicago White Sox	13	5
Atlanta Braves	14	12
San Diego Padres	15	20

Over the first seven years of the 2010s decade, the Dodgers overtook the Yankees as the team to draw the most fans while the Tigers entered the top 10 in attendance being ranked 17th in population. The recent, consistent on-field success of both the Dodgers and the Tigers is likely the primary driver for the uptick in their attendance rankings. Each team won four straight division titles during the decade.

Table 8: Average Attendance Rank vs. Average Population Rank (2005-2009, Top 15 Teams)

Team	Attendance Rank	Population Rank
Los Angeles Dodgers	1	3
New York Yankees	2	1
St. Louis Cardinals	3	22
San Francisco Giants	4	14
Boston Red Sox	5	13
Los Angeles Angels	6	3
Chicago Cubs	7	5
Philadelphia Phillies	8	9
Texas Rangers	9	7
Detroit Tigers	10	17
Colorado Rockies	11	24
Milwaukee Brewers	12	29
Minnesota Twins	13	19
New York Mets	14	1
Washington Nationals	15	10

5. Conclusion

The one common theme that comes from our analyses is that team performance measures, specifically win-loss record and runs scored, are strongly correlated with attendance rankings. This is borne out from the fact that the maximum attendance year for 23 of the 30 MLB teams occurred during a winning season or during a season in that team's top 10 by winning percentage. Also, offensive scoring overall tends to be a better indicator of attendance than home runs or runs allowed. Lastly, our MSA population analysis shows that, while large-market teams such as the Yankees and Dodgers have built-in advantages when it comes to attendance, small-market teams can help close the gap by fielding winning rosters.

That being said, this paper should serve as only the start of a more in-depth research study. Further research should focus on statistically estimating the impact of winning percentage, offensive and defensive scoring measures and market size on attendance. In addition, a team's level of all-star talent and its playoff / championship success should be taken into account.

References

[1] http://sabr.org/research/what-do-your-fans-want-attendance-correlations-performance-ticket-prices-and-payroll-factor

[2] https://fivethirtyeight.com/features/dont-be-fooled-by-baseballs-small-budget-success-stories/

[3] http://www.ballparksofbaseball.com/baseball-ballpark-attendance/

[4] https://www.baseball-reference.com/leagues/MLB/index.shtml

[5] http://mlb.mlb.com/stats/sortable.jsp#elem=%5Bobject+Object%5D&tab_level=child&click_text=Sortable+Team+pitching&game_type=%27R%27&season=2017&season_type=ANY&league_code=%27MLB%27&sectionType=st&statType=pitching&page=1&ts=1503459516150

[6] http://mlb.mlb.com/stats/sortable.jsp#elem=%5Bobject+Object%5D&tab_level=child&click_text=Sortable+Team+pitching&game_type=%27R%27&season=2017&season_type=ANY&league_code=%27MLB%27&sectionType=st&statType=pitching&page=1&ts=1503459516150

[7] https://www2.census.gov/programs-surveys/popest/datasets/2000-2009/metro/totals/

[8] https://www2.census.gov/programs-surveys/popest/datasets/2010-2016/metro/totals/

[9] http://www.dacubs-project.com/?p=206