Karen Chapski

MSIS 2629

Individual Project – MLB Attendance

Summary

Baseball is entertaining but at the end of the day it is a business as tickets can cost anywhere from an average of $15-$50 (https://www.reddit.com/r/baseball/comments/317808/serious_how_much_does_it_take_to_actually_sell/) per fan and every unsold ticket adds up not only in ticket sales, but also in potential revenue that the fan would spend inside the stadium. This amounts to almost $33M per year in lost sales for Chase Field (avg of 1.8M unsold seats a year – Retrosheet.com, avg ticket price from Reddit). Are there factors that can be controlled in order to maximize attendance and ticket sales?

By analyzing when baseball stadiums have the most unsold seats one can identify when baseball franchises have the most opportunity to maximize potential revenue through marketing, ticket price A/B testing, and promotions.

Link to Github project: https://github.com/kchapski/MSIS2629_projects.git

Link to Tableau Public: https://public.tableau.com/profile/karen4801#!/vizhome/Attendance_Baseball_stadiums/WhenCanBaseballTeamsIncreaseTicketSales

Documentation:

Purpose and Audience

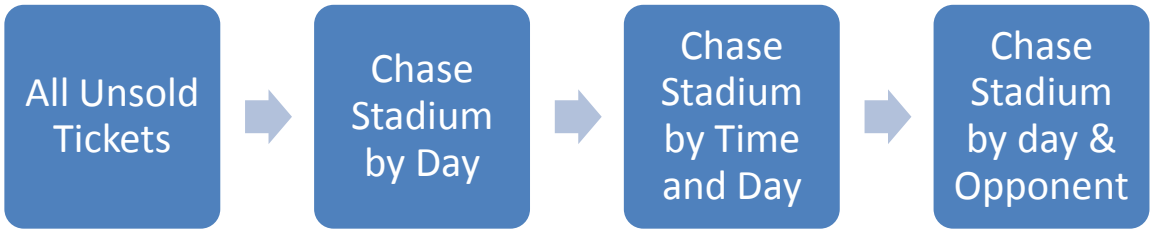
The purpose of this dashboard is to analyze when Major League Baseball Stadiums have the most unsold seats.

The audience for this dashboard would be the stadium marketing operations teams who would be designing programs to drive additional revenue opportunities to the franchise (and stadium) through ticket sales. They would be reporting to owners/managers—management who would be invested in any changes made.

As an operations manager for a Major League Baseball Stadium one would serve the following role:

- Help management understand when revenue is lost through empty seats – every ticket that goes unsold is lost opportunity to make additional revenue
- Set goals and expectations on what the average ticket sales should be
- Provide regular reporting on success and ways to improve ticket sales

Scope of Data used and Flow



To determine what stadiums to focus on, it had to be determined what stadiums had unsold seats. If stadiums (and teams) are selling out there is no need to see how they can optimize marketing and promotional efforts.

A multi-year snapshot was chosen so that there would be a horizon that spanned over one year and prevent any anomalies.

Chase Stadium was chosen for a deep dive as it was the stadium with the most unsold seats 2015-6.

The following questions were asked to explore revenue opportunity:

What stadiums have unsold seats?

Chase stadium had the largest number of unsold seats 2015-6.

Driving down to days of the week, further questions can be explored:

What days of the week have the most unsold seats?

What game times (day/night) have the most unsold seats?

Given night games have more opportunities, what games (vs specific opponents) have more unsold seats?

Based on these questions/findings what recommendations could be made? What actions could be taken?

The story format in Tableau helps guide the users attention.

The Dashboard

Assumptions: the audience is familiar with MLB ballparks, frequency and schedule of baseball games and the basic terminology of the game and teams.

This dashboard mainly uses a combination of bar charts to show differences in groups and sub groups.

The data and charts are used to explore data to find a recommendation and actionable insight around what times warrant promotional efforts to increase sales.

This dashboard can add value to a Major League Baseball organization by helping an operations team high light where they have the most opportunity to increase revenue by pinpointing what times (and days) they need to improve their engagement with fans.

At this point the data is static, working as a functional model for a dashboard that would also be useful with live updates that are checked at a regular cadence for improvements throughout improvements and testing.

Data Used and Data Prep:

Data sources:

Retrosheet: http://www.retrosheet.org/boxesetc/MISC/PKDIR.htm

Reddit: https://www.reddit.com/r/baseball/comments/317808/serious_how_much_does_it_take_to_actually_sell/

Formulas:

Unsold seats=[Max Attendance]-[Attendance]

Avg Ticket Price (for Chase Stadium)= $17.98

R was used to import and combine the 2015 and 2016 files on stadium data from Retrosheet.

Reddit data was pulled into Excel manually.

Next Steps/Recommendation

See Critique File for next steps.

Based on the data the recommendation is that games on Monday, Tuesday, and Friday have the most potential with games against Atlanta, Washington, Miami, and San Diego having the most average unsold tickets per visiting team. These games would give the greatest benefit as far as promotional tests and increase in revenue.

Project Critique

What I learned:

- Drowning in data: I did a lot of exploratory analysis before I decided on what attributes I should use in the final dashboard. It was really easy to drown and get lost in the amount of attributes that existed in the data sources
- Importance of definition prior to action and documentation: to prevent the drowning issue and getting confused, staying organized and defining clearly what each viz and formula is helps greatly in not getting confused down the line when the complexity increases. You can always go back to your documentation.
- Research your data: Initially I wanted to just use the stadium that had the most unsold seats which was Turner Field in Atlanta. I initially did a lot of work on graphing and looking into what days, teams, etc was related to that field. I ended up Google-ing it to check a date and found out that it was retired at the end of the 2016 and decided to exclude it from my data set.
- Choose data wisely: if there's no action, the data might not be worth adding and might not add value. Think about the value add. Specifically when I was looking about what might impact the seat sales I was looking at attributes like winning team. One may be able to guess stronger team, but little action can be taken on that attribute, it can't be predicted however popularity and overall visiting team can be looked at, which is included.
- Worksheet→dashboard→story→ while I have used tableau before I had not used the story feature to direct a user's attention. It is a useful tool, but there are nuances with formatting and graphics that can help to make it

more visually appealing and help with storytelling like the order in which the viz is created which is important

Next Versions:

There were a lot of different ideas I had for this project that if I had more time or different/more data I would implement for next versions if I were to pursue this project further:

- How do some of these patterns compare for other stadiums – do other stadiums having issues with unsold seats see the same day/time patterns?
- How can an operations and sales team utilize patterns and learnings from a stadium that sells out YOY to one that has a large number of unsold seats?
- Additional data sets:
    - Who is attending by zip code: where are fans coming from? This data could help target specific audiences or reach out to new ones
    - Ticket price by game: further analysis for ticket price variation by game. How does ticket price variation impact attendance/sales? (as well as % of season tickets)
- Which kinds of seats are not selling (IE cheap seats, sections, visitor vs home) – action: how can these be rebranded to bring fans to these areas?
- Using regression to look at correlation of unsold seats and other attributes to see what impacts sales
- How do the presence of certain players – visiting or home team—impact attendance?

What I like

- To be as clear and direct as possible I used the Tableau story structure, it helped also be streamlined in my thoughts about the message I was trying to get across.
- Bar charts seemed to be the most effective way in showing the message about times, days, and cost of not selling seats

What I don't like

- Although not always necessary I would like to make the dashboard have more sophisticated graphics, dig deeper into the data

- It might be hard to tell that the drilled down cards are for Chase Stadium specifically (mentioned in the cards, but not in the embedded viz’s)
- In the team analysis, highlighting the cells above 30k would help draw the users attention to specific cells.