**Christian Acosta – DS710 Final Report**

**The Association Between MLB Team Twitter Activity and Stadium Attendance**

**PROJECT SUMMARY**

Many debate whether Twitter has the potential to boost brand recognition. Major League Baseball (MLB) teams use twitter to connect with their fans. They provide scoring updates, player profiles, community activities, promotions and more. Are there certain trends amongst the 30 MLB teams to suggest that there is a relationship between a team’s twitter activity and stadium attendance?

**PROCESS**

*Data Gathering & Manipulation in Python*

* A Twitter connection (REST API) was used to grab information on the 30 MLB team twitter accounts.
  + User information such as # of followers, status updates, and number of likes were gathered. This data would be compared to attendance data. Recent status update history was used to see recent tweet rate.
* Data was gathered from several sources in order to build a comprehensive MLB data set to connect with the Twitter data.
  + Historical attendance data from 2013-2016 and stadium capacity.
  + Team statistics such as win percentage.
  + TV market population for each team to determine the ratio of twitter followers in a market.
* Manual calculations such as percentage of stadium filled were used to bridge the gap between large and small market teams.

*Data Analysis in R*

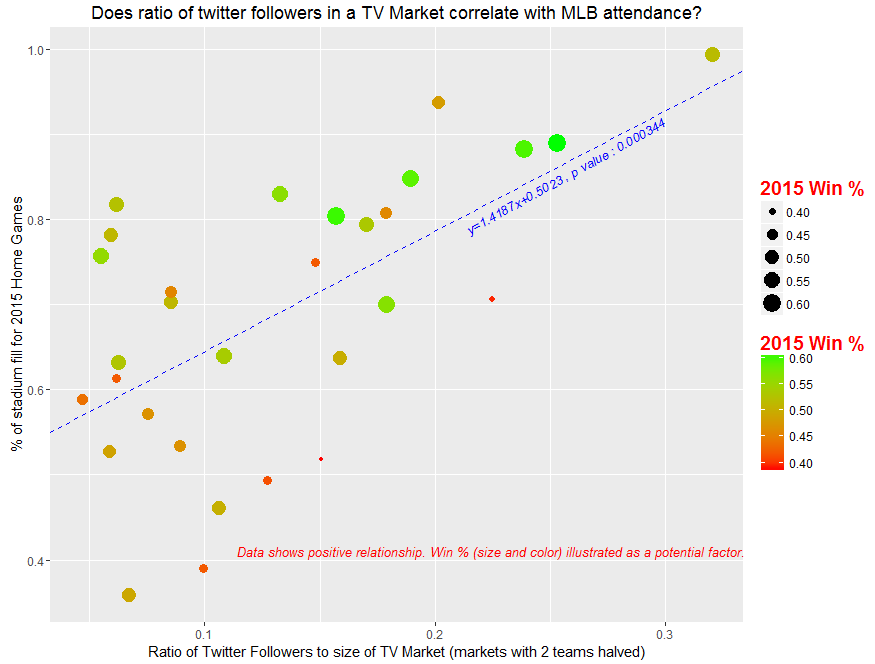
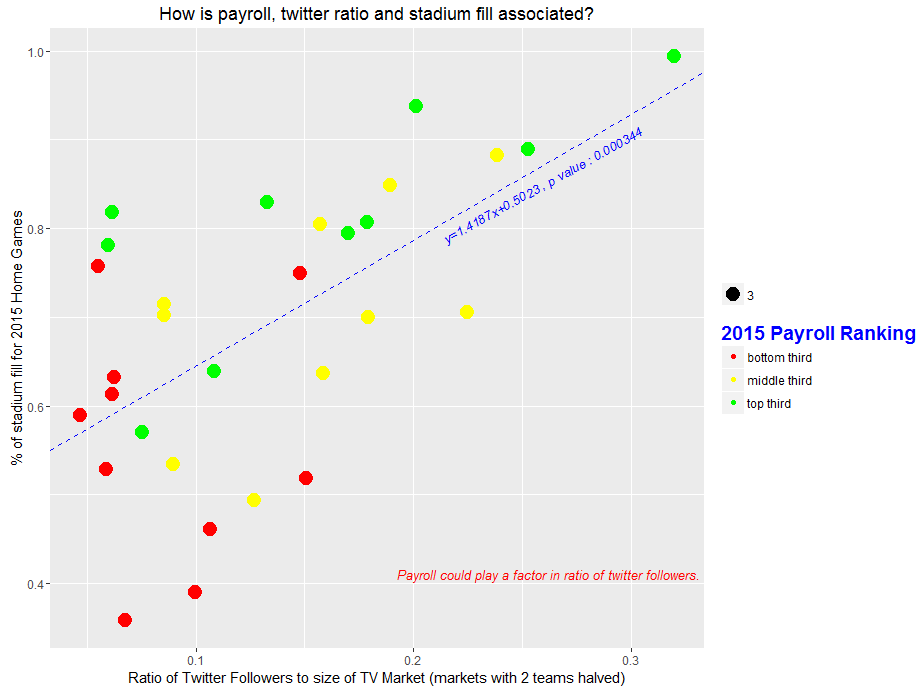
* Various aspects of each team’s twitter accounts were analyzed to investigate relationships with attendance.
  + *Average post rate: 1) since inception, 2) in recent history.* Does frequent contact help attendance?
  + *Number of team likes and % of stadium filled.* Likes are a way to connect directly with your fans on twitter; are they more likely to come to games then?
  + *The ratio between the number of twitter followers and the size of TV Market.* While all followers aren’t from team home markets, this metric can be used to analyze ‘twitter penetration’ in a market. It is assumed that a higher ratio means more *connected* fans (ex. 100k followers in a market of 1M gives .1).
* Efforts were taken to find any relationships and discover common threads in twitter and MLB data. For example: If the data shows higher twitter followers are related to higher attendance, is it because of the market size?

**FINDINGS AND CONCLUSIONS**

There is a clear correlation between the ratio of twitter followers in a TV Market and percentage of each stadium filled. However, there is not enough conclusive evidence to link number of likes or post rate to attendance. These findings suggest that twitter *saturation* in a market could lead to higher home game attendance. Analysis was conducted to link the independent variable (ratio of twitter followers in a market) to investigate any similarities with other, more traditional independent variables:

* A team’s win percentage can cause engagement to rise (figure 1) – denoted by the larger, greener circles. Additional work will need to be put in to ‘cancel’ out the impact of a strong record and separate the ‘noise.’
* A team’s payroll could be associated with a higher ratio of twitter followers in the TV Market (figure 2). This *could* be due to higher marketing budgets, or perhaps twitter users tend to attend more ball games. More investigation is required.

The higher the ratio of twitter users in an MLB market, the more filled stadiums get. However, it could be associated with several other factors. There’s no conclusive evidence that post rate or ‘like’ rate play a role in attendance. Therefore, a leading hypothesis to explain the results is that the % of followers in a market is a representation of the team’s engagement within said TV Market. This would explain why other factors in a team’s twitter account don’t seem to play a role in attendance. A future study could focus on associating this engagement ratio with other fan factors such as memorabilia sales, season ticket sales and appeal outside of each team’s home market.



[Figure 2 - left] shows a similar graph, but this time illustrating payroll ranking in thirds. This suggests that market engagement and payroll could be heavily-related.

[Figure 1 - left] shows the strong correlation between twitter followers in a TV market and how filled stadiums get. However, win % can play a key role. More investigation required.