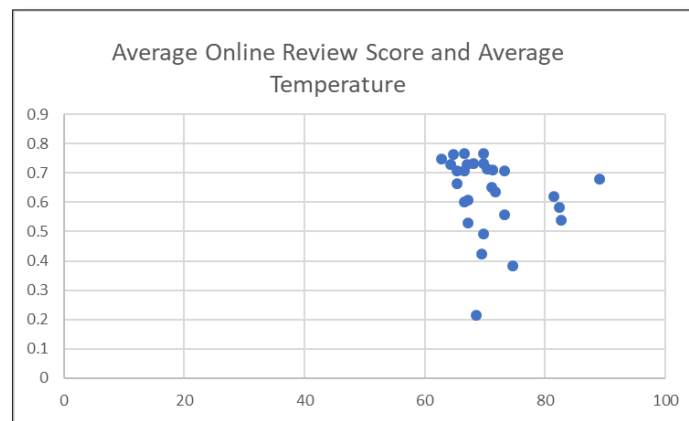


Assessment Report

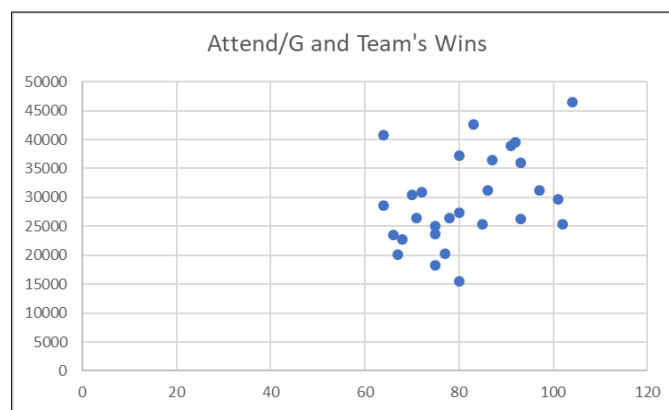
Joe Tucci

The overall quick assessment of my final dataset was a mixed bag of slightly correlated datapoints and datapoints that did not have all that much correlation at all. The first column in my dataset that stood out to me upon completing it was the Average Online Review Score column. I think my taking an average of the four categories from my intermediate dataset was still a good idea, but I don't think I did enough to it to make it fit within my dataset. Since it was on a -1 to +1 scale, every stadium was in the positive within about 0.5 points of each other from the worst to best.

The image below illustrates how the scatter plot essentially doesn't skew in any meaningful way.

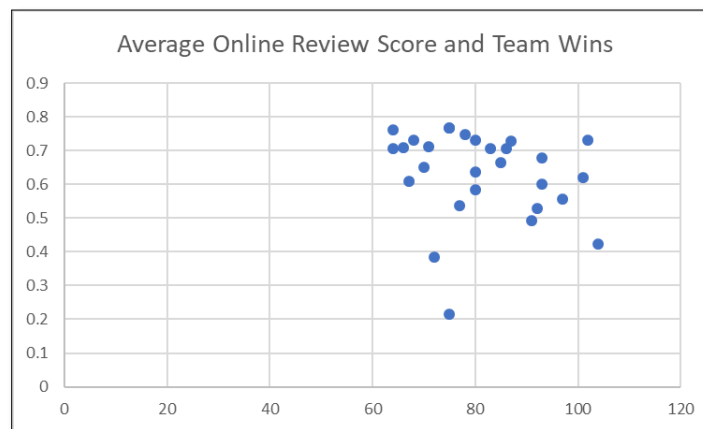


When comparing the average attendance per game with a team's number of wins in the 2017 season, it immediately jumped out at me that the team with the highest average attendance per game had the most wins of any team that year. Though, the team with the least number of wins was not the team with the lowest attendance per game average; that team won about 80 games that season.



When it came to weather, the more useful metric tended to be precipitation totals for the 6 month time span rather than average monthly temperatures. The temperatures generally average out to be much more similar than the precipitation totals. Maybe because baseball is played throughout the summer months, it's not useful to get the average temperature for the entire season. Based on what my dataset shows, I should have gotten an average temperature per individual month, and then compared it to a team's record in that given month. But in the end that would have required my entire dataset to change to revolve around a monthly time frame instead of a season time frame.

Generally, if fans had an overall good experience at a stadium (considering food, facilities, etc.), the team had at least 75-85 wins. Usually about 85-90 wins is good, and over 100 wins in a season is great for a baseball team. This data seemed pretty normal, with the outlier and exception being the team with the most wins rating in the bottom three for overall fan experience. This may also be due to the fact that the same team (the Los Angeles Dodgers) draw a huge crowd, which would potentially impact the fan experience rating while simultaneously impacting the team's win/loss record.



I think a lot of my data from my final dataset wound up showing very average results without showing strong correlation between any columns or datapoints. There were a lot weaker trends than I anticipated with weather and health code violations. Similar to my problem with taking average weather data for the entire season, I think most of this dataset could have benefited from either a more granular approach to look at each month, or create similar datasets for previous seasons other than the 2017 season to compare these points by year.