

Miami University

Business Intelligence and Data Visualization

ISA 401 - Final Project Report

11 - 04 - 2021

Comparing Twenty Years of Chicago White Sox Baseball Statistics to the 2020 Covid Season

Adam Bandola

Senior Business Analytics Major

Louie Raines

Senior Accounting Major & Business Analytics Minor

The Problem

The Chicago White Sox had a remarkable year during this 2020 season affected by Covid. While analyzing the data from the season and comparing it to the past 20 years of statistics, we were interested to see what aspects of the game might have improved their winning percentage. Aspects of their entire team might have changed, but we had hoped to identify some of the things the team did particularly well. We were also interested in seeing how the remainder of the 2020 season would have finished if the White Sox continued their strong performance. The team was having an incredible season with both overall wins and general statistics. If the White Sox continued to excel what would their statistics have looked like?

This analysis can aid the White Sox as well as any baseball programs interested in improving specific aspects of their game that would impact their ability as a team. The data shows which statistics helped to improve their season and which statistics had little to no effect. Another team can try to emulate what the White Sox did this year statistically in order to hopefully have a similar performance. The White Sox could also use this data to see which aspects of the game that need improvement. For instance, the team had more strikeouts in 2020 than in almost any year in the previous 20. Fixing problems like these could help the White Sox or a team like them to continue their positive trend.

How the Data has Changed

We are examining the idea that, if the 2020 Covid Season were to have been 162 games long, how the Chicago White Sox would have fared within their team, division and league. In the past, there has been a lot of analysis of baseball data, however, this year is much different than years in the past. With the onset of the coronavirus, the Major League Baseball association and MLB teams were forced to adapt. Examining how the statistics differed in a 2020 Covid year when comparing it to older seasons can help future teams.

Similar questions may have been asked with this type of data in the past. Major League Baseball has been heavily analyzed through various statistics and it has revolutionized the game. The major difference between our findings and others is that we are taking the statistics from the shortened season, applying those numbers to create what a full season may have looked like, and then analyzing that compared to previous non-Covid seasons. This will give us the ability to discover how the Covid season differed from a regular season and what type of statistics changed.

Data Extraction and Description

We used web-scraping techniques to import tables with information about pitching and batting statistics from *baseball-references.com* into R-Studio. Since the data came from a website that houses code for all teams within the MLB, it was comprehensive which allowed us to bypass the

steps of imputation for missing values. However, we wanted to showcase the code in a way that is easily digestible from a glance which is what much of the code is accomplishing.

The data transformation that took place was the removing and renaming of variables to be clear and concise about what was necessary and what it meant. Along with this was the need to double check whether there were missing values within the data, there were none. The trickiest part of the transformation was bringing the row for the year 2020 (which we dubbed as the Covid Season), into a separate data set where we could apply the 2.7 multiplier in order to get an accurate assessment of how the White Sox would have performed if they continued the same trend of statistics throughout the season. From there, we re-ordered the columns to make the application of the multiplier easier and then re-incorporated the Covid Season back into the data set that housed the seasons prior to 2020.

It is also important to note the importance of choosing the last two decades of data to observe. It is extremely unlikely that a team would have the same coaching staff, players and game plan for twenty years, which is the reasoning for which we based the season cutoff point.

Data was gathered from in-game events and then multiplied by 2.7 to get a realistic gauge of how the season would have played out if the 60 games played were turned into a 162. Baseball reference is known to track accurate data from all certified Major League Baseball games and, when checking the baseball reference data with similar data sources, it was shown to be accurate.

Potential Improvements

Improvements could be made throughout the model and in the future. For example, you could create a model, such as a linear regression, to predict the total number of wins based on the original data provided through *baseball-reference.com*. However, for the question we wanted to answer, our method provides an easily understood and simple way to discern how successful the White Sox could have been this past Covid Season.

Insights Drawn

After analyzing the data gathered, we came to see that the White Sox would have had their best season since they won the World Series in 2005, if they continued playing like they were. Their pitching performance was better in 2020 than any other year in the data and although the batting was not the best it was significantly better than most previous years.

It is important to keep in mind that the data represented is not perfect. Teams go through periods where they are unstoppable and also periods where they can't seem to get a win. However, with the assumption that the White Sox would continue to perform the same way they had been, the graphs and data clearly depict a very successful team that had a bright rest of the season.

Tableau Link

<https://public.tableau.com/profile/adam.bandola#!/vizhome/Bandola-RainesISA401FinalProject/TwentyYearsofWhiteSoxBaseball?publish=yes>