

Brianna Rawlings

Alec Schrock

Data Science Final Reflection

This final project involved the use of machine learning to analyze and predict a problem using the dataset. For this, we used MLB stats from a Kaggle dataset and was using the data to predict what stat is the best indicator for if a team will make the playoffs. The data on this Kaggle site is from the Baseball Reference, which is one of the leading baseball statistic databases, therefore trustworthy. The MLB keeps data like this as it can give many insights into many different areas that they can use to compare from year to year. For example, this dataset has features like total runs scored, win total, on base percentage, slugging percentage, batting average, and many more. These can all be used to help the MLB make models just like this project to predict how their teams are doing in certain areas and comparing teams to each other to see who is/could be the best. The MLB isn't the only organization to do this as many different sports, sports leagues, and sport organizations keep stats just like this dataset to use for predicting different analyzations.

During this project, some of the most challenging parts were finding the right data set to use and then finding the right visualizations to convey what we wanted to talk about. We had a dataset we were going to originally use that involved baseball statistics; however, it was going to involve multiple datasets that we would have had to merge and then clean and get rid of many null values. This led us to go with another dataset that didn't require as much cleaning and was only one dataset compared to multiple. Then when we got to this dataset, we knew what we wanted to do, but then figuring out the accurate visualizations that would convey what we wanted to share took some time to figure out and understand. The limitations we have with this dataset is that we are only looking at offensive statistics compared to defensive statistics. Another limitation we had was that our dataset started tracking very advanced stats that had only a couple years of data on it that would have been hard to have compared to the data that we had from previous years.

While doing an analysis like this one there are many decisions that are being made that go into it. As a Christian we have many principles that we are called to follow and practice when decisions-making. One being wisdom, as we are using knowledge, and our moral will to analyze the data thoroughly and decide respectfully. "If any of you lacks wisdom, you should ask God, who gives generously to all without finding fault, and it will be given to you" (James 1:5). When making decisions we should seek guidance from Him, as he can help us make the best decisions needed in many complex situations like data. As Christians, we should also be doing our work with honesty. We want our data to be as accurate as possible, that shows the truth and represents the data well. "Whatever you do, work at it with all your heart, as working for the Lord, not for human masters" (Colossians 3:23-24). We are called to work hard and do it with our heart, which brings honesty with this. Everything we do is for the Lord, therefore we need to do it with our hearts.

As a Data Evangelist working with data like ours, we are called to show the love of the Lord relay that through our work. As mentioned previously, we are to be honest and share our work accurately as it is to be represented well. Everything we do is for His glory, so by doing our work for Him, we show who He is to others through the act of service. When it comes to data-driven decision making as a Data Evangelist, we are called to share our knowledge and show how data can impact many different organizations. Data helps bridge the gap between unknown and know, so getting to share findings helps show better understandings.

