Individual Reflection

Wesley Chapman

Working on the project alone proved to be very different than it would have been if I were in a group. This was clear from the start, when I was picking my topic. Without having teammates, I was free to choose from anything without having to run the idea by other people. This was both good and bad, as it gave me more freedom, but no help in the process of topic selection. It ended up working out well, with my selection of baseball as a topic, something that I am very passionate about.

Throughout the project, working individually gave me the ability to work at my own pace, which was nice. I have never done a group coding project in the past, but I imagine that working collaboratively could be challenging in some ways, with coding being very detail oriented by nature. Again, here I was able to avoid possible hurdles that would have come working in a group. One place that I would have liked to have had assistance from a group was on the presentation itself, both in creating it and presenting it. I personally do not enjoy giving presentations or public speaking, so I definitely felt the lack of team members at that point.

My coding skills have advanced tremendously this semester, with it being the second and by far most in-depth data science course I've taken, so this project was a great opportunity for me to showcase my new knowledge, both with coding and working with data. Unlike with past projects, this provided me with a chance to choose any topic and data type as my focus. Working with something that I had interest in made it easier for me to continuously put in effort. I really wanted to come up with a model that was practical and effective at providing meaningful information on the topic, and by predicting batter OPS with multiple linear regression, I believe I accomplished that. Another aspect of this project that I

am proud of is the data collection. I feel that I was able to gather data on meaningful metrics, and without including any bias that would influence the final result of my model. Every statistic I used is somehow relevant to the model, and by filtering the years and minimum plate appearances of the players, I was able to avoid any skewed or misleading outcomes.

If I were to do this project again, I would try to incorporate even more statistics as independent variables to make my predictions stronger and more reliable. Although I used twelve variables in my model, I think it would have been even better to add more. I could have also attempted to include more visualizations to represent the model. I chose to use two, but I believe that more is always better when presenting statistical findings.

In fact, I think the most challenging part of the project was finding ways to visually represent my model. I was trying to come up with an effective way to show how predictive OPS was, and I chose to use a scatter plot, but I'm sure there are many other ways I could have also accomplished this. In the end, I am happy with the two visualizations I came up with. I remembered learning about pair plots, and I figured that would be a good way to show the individual correlations between OPS and the independent variables. Once I made that, I felt satisfied with the different ways I had put my project into graphs, covering multiple aspects of the model.

Overall, I am happy with how the project turned out. I feel good about the pace I worked at. At no point did I feel rushed or stressed about deadlines. It proved to be a great opportunity for me to display the new things I have learned throughout the semester.