This project was a lot of fun! While I struggled at times to analyze the data to support my initial question, I learned a lot about pulling data, merging data, filtering data, and plotting data. Not to mention, I learned a bit about baseball as well! To review, my initial goal was to review how switch-hitters performed against left and right-handed pitchers compared to how left and right-handed hitters performed against left and right-handed pitchers! This question presented a lot of data and a lot of different ways to look at the data.

Pulling data in from a CSV and doing some basic cleaning was probably the easiest portion of this project! Where I ran into some trouble here was working with indexes. I read my CSV and set an index as I read it in. I also did some groupby operations, which also resulted in an index on my dataframe. At this point in the course, I was not quite prepared to work with indexes and it caused some frustration. But shortly after turning in Milestone 2, we covered the groupby function, so I believe this set me up for success the rest of the course!

The most exciting part of my project was doing the web scraping. I had to build a URL for each player I wanted to gather stats for as they were not contained on a single page. Also, the table was not easily accessible. After additional research on the web, I determined I had to use a headless web browser so that underlying javascript could execute and generate the HTML so that I could read it! With more experience, I could probably have made this more efficient but for now I was happy with how it turned out.

One area in which I was disappointed with my project was the API portion. Major League Baseball has an API to pull stats from. And there are some existing Python libraries that read from this API. But the API is undocumented (at least publicly). And the existing libraries did not pull the data I was looking for. Major League Baseball also provides a stats application that allows you to search and return a myriad of stat combinations. And you can download this as a CSV. So while I used this stats application and it somewhat provided the data I wanted, it was a bit of a letdown as it didn’t seem to be a true API call. To get the data I truly wanted from the MLB stats API, I would have had to reverse engineer it without any documentation. While fun, would have taken too much time for this project.

Another great aspect of this project was the visualization. While reading the project information sheet, I must have missed the part where the last milestone included visualizations as I tried to include some visualizations throughout the other milestones. While this gave me a leg up on the last milestone, it caused some consternation during the other milestones. But that is part of the process! Though I struggled at times in the earlier milestones with the visualizations, it forced me to read documentation and do additional research which has hopefully given me a better understanding of some of the visualization libraries in python. I am also thankful that this this course and this project finally helped me get over the hump with Jupyter Notebooks. There was a time where I did not particularly care for them and struggled to use them. But now I’m much more comfortable and can clearly the see the benefits!

While I learned a significant amount about Python, Pandas, Beautiful Soup, and a whole host of other things, I actually learned a bit about baseball. From my data gathering and visualizations, it appears that switch-hitting is slightly beneficial and not just a gimmick. I would say that switch-hitters tend to hit better against left-handed pitchers than left-handed hitters. And they do slightly better against right-handed pitchers than right-handed hitters. While I’m not sure that it is a significant difference, I would tell an up and coming youngster that if they struggle hitting against the opposite handed pitcher to try switch hitting (with a lot of practice of course!).