For my final project, I used NFL data for wide receivers (WR) across multiple datasets to compare 2019 season performance with fantasy data performance and 2020 projections. I had to learn several new skills in Python to accomplish this. While in previous courses I had to connect to a csv file and build a Python dataframe, I had to learn how to clean up the data source to make sure it was in a format that was consistent across all 3 datasets on the primary key, which was Player name. This included developing code to remove special characters or extra spaces to ensure there was an accurate join across the tables. I also learned how to parse data out of a field to split the fields and then derive a new field based on the parsed data.

Web scraping was a new technique for connecting to data for me, so I had to learn how to find a url of a site and scrape the relevant information using the Beautiful Soup package. Luckily my site wasn’t too complicated to the data I needed for NFL WR stats, but there was still some data clean-up that was needed to get it in a format that matched the CSV data. The web page had receiving data for all positions, but I only wanted to look at WR data, so I had to subset the data once I had it in a dataframe. I also learned how to derive a calculated field (Catch % Rank) from the dataset. Finally, I had to update some headers to make sure they matched other datasets and made sense when looking at the data in data table.

Connecting to API data was a new concept for me this quarter as well, so using the “read\_json” function I was able to get the data I needed from the API into a dataframe so I could further wrangle the data to be able to join to other datasets. I then did similar cleaning as csv and web scraping data sets.

Then across all 3 datasets, I had to learn how to export the data from the Python dataframes into a database. I used sqllite because it has built in functionality with Python, plus I wasn’t building huge databases, so it was easy to work with. I exported the individual tables, plus I joined the data within sqllite across tables and created new tables as well. In addition, I exported the sqllite tables into csv just to understand how that process works, as it makes it easy to share data with others who don’t have connection to your local SQL database.

Finally, with the datasets now wrangled, cleaned and inserted into a database, I was able to build my visualizations. I used Tableau because I am familiar with it at work and knew it was easy to connect to whatever dataset I needed, whether it be a sqllite database in this case or a csv file or whatever was needed to visualize the data. I was able to build some more advanced visualizations than I currently can in Python, like adding team logos and doing clustering on a scatterplot, but looking forward how to build similar views in Python as well. Overall this project was challenging yet fun, as I learned a variety of new techniques on how to connect to data within Python, clean my dataset, and visualize data across multiple datasets. I look forward to building on these skills in upcoming classes.