Sana Khan
IST 718

Lab 3

The purpose of this assignment was to suggest a salary for the Syracuse Football coach based on relevant data. The data that was used includes the University Name, Coach, Conference, Graduation Rates, Stadium size and Win/Loss ratio.

The steps taken were to merge the data from the Coaches.csv with the data that was found through the following websites:
Graduation Rates : https://web3.ncaa.org/aprsearch/gsrsearch
Stadium Size: https://www.collegegridirons.com/comparisons-by-capacity/
Win/Loss : https://www.teamrankings.com/ncf/trends/win_trends/

The data was then cleaned by removing non-integer or string characters and converting any numeric columns from a string to a float. I then merged all of the data together using the College name as the join ID. Not all of the colleges were in all of the files so there was some expected data loss, but I ran into an issue where a lot more data was lost because the join did not work properly. I was unable to understand why the join didn't work for some of the colleges that were in more than one data set which resulted in about 50 rows not being included.

I then did some data analysis and plotted variables like total pay against the conference, graduation rate, and stadium capacity. The conference with the most range in total pay is the SEC, along with Big 10 and ACC. The highest paid coach is at Alabama which is in the SEC and the lowest paid is at Akron in the MEC. Stadium size also has an impact on total pay, as schools with larger stadiums having a higher pay for the coach. Graduation rates have a weaker association with total pay but show a positive relationship.

I used a correlation heat map to pick the variables that have the most influence on total pay, which are capacity, and PCT. It makes sense that these variables would have an influence on total pay given that a larger stadium can produce more revenue than a smaller stadium.

I then used a linear regression model to predict the salary for Syracuse head coach at $2,370,929. If Syracuse was part of the Big 10 the predicted salary would be $2,971,211 and $2,678,799 as part of the ACC.