Bertrand Bailey III
09/24/20
SMU – Data Science Boot Camp 2020

ETL Project Report

Extract:

The data I used for this project came from Kaggle.com. I used 2 datasets, one of NBA Players and their salaries and another of NBA Teams and their net worth. The data was formatted as CSV files. I was able to successfully upload the project to Jupyter Notebook to complete the transformation.

Transform:

For the NBA Players file, I needed to filter out the position column as it was not needed. I also had to rename the columns as player_name, team_name, and player_salary.

For the NBA Teams file, I had to filter and rename the columns as team_name and team_worth. From there, I proceeded to load the data.

Load:

The final database in PostgreSQL consisted of a players and teams table with player_name as the primary key for the players table and team_name as the primary key for the teams table. I chose a relational and structured database. My goal was to compare an NBA players yearly salary to the net worth of the team they play for. I found this study to be very interesting and I will continue to perform queries and research.