Data Slackers

How Does an NBA Team's Performance Affect Evaluation?

Summary

Extract

We used Basketball-Reference to download CSV files containing team season stats (Team Per Game Stats) and advance stats (Miscellaneous Stats). These data tables gave us insight as to how well a team performed throughout the 2018-2019 season in all statistical categories.

We scraped Forbes to extract team evaluations and other monetary stats with the use of Beautiful Soup and Splinter. A for-loop was used to loop through each team's website and all team data were appended to lists, and compiled into one dataframe with the use of Pandas. The table was then saved to a CSV file for the Transform step.

Transform

For the Forbes data cleaning, we had to convert the columns with VARCHAR, TEXT and BLOB values to INT and BIGINT values. Basketball-Reference data cleaning included removal of %, Billion and Million signs, and decimal points, as well as removal the averages row to maintain consistency with the Forbes data.

Load

When we loaded all tables into the MySQL schema, we encountered data issues associated with invisible characters in rank column of the Basketball-Reference data sets, so we copied the numbers without the invisible characters into a column called ID. When loading the Forbes data, eight rows were consistently missing.

After analyzing through the Import Wizard step, we noticed the missing rows all had high team values, so we changed INT values to BIGINT values. To maintain referential integrity, we created Primary Keys and Foreign Keys in our tables. We converted the TEXT and BLOB values to VARCHAR because text cannot become a primary key.