ETL- NBA Basketball Statistics (2015-17)

By Shivang Patel

*Project Overview*

Statistics within the NBA play of significant role on player’s careers and how teams evaluate payroll. Basketball players build a legacy through how well they perform on the court. Every year data about players performance is provided by the NBA. In this project, I have created a database that combines players statistics from the last 3 years. With this database one can analyze player trends per year, consistency, and/or player performance over a span of 3 years.

*Extraction*

Player statistics were taken from <https://www.basketball-reference.com>. Statistics for the 2016-2017 NBA season were web scraped directly through the source (HTML) using Python and Pandas library. The statistics for the 2015-2016 season were in the form of a csv provided by the source. Lastly, statistics for the 2014-2015 NBA season were web scraped directly through the source (HTML) using Python and Pandas.

*Transform*

Player dataset were generally consistent throughout each data. However, minor cleaning and transformation were needed. Player ranks were on a per season basis, therefore, column needed to be dropped. Within concatenating the data, the data for two seasons included titles and column names. Rows were cleaned.

*Load*

Complete dataset was loaded using HTML and CSV files onto my Jupyter Notebook into the same file, as an Excel file.