ETL PROJECT Technical Report

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Title** | NBA STATS | **PROJECT NUMBER** |  |
| **Project Participants** | Tim Barger  Will Garrett  Derek Greeley | **DATE SUBMITTED** | 04/15/20 |
| **Reporting Period** | **From**: April 08, 2020  **To**: April 15, 2020 | **REPORT NUMBER** | 23 |
| **Invoice Number:** | | 123456789 | |
| **Funds Spend during reporting period** | | Stacks on Stacks on Stacks | |

Summary:

Play in the NBA has changed dramatically since 2003, following the success of Mike D’Antoini’s Fastbreak system, which utilized the passing ability Steve Nash, the athleticism of Amar’e Stoudemire and Shawn Marion, as well as the three-point shooting threat provided by Quentin Richardson.

Since that time, the NBA as a whole has moved away from the traditional half-court play that utilized mid-range isolation plays, and in the paint scoring. This shift was embodied by the success of the Stephen Curry led Golden State Warriors whose up tempo play and dead-eye three-point shooting made them three time champions, combined with a greater focus on quick points in the paint.

The data table created focuses on the statistics for each player from 2013-2017 focuses on the offensive output of each player by season. The table utilizes traditional statistics such as Points, Rebounds, Assists, and Blocks, as well player metric statistics such as Usage Rating (USG)[[1]](#endnote-1), Player Efficiency Rating (PER)[[2]](#endnote-2), and True Shooting Percentage (TS)[[3]](#endnote-3). The table also utilizes a number shooting specific statistics, focused on what kind of shots each player takes and how effective they are when taking those shots.

Detailed Accomplishments by Task

Task 1: Extract

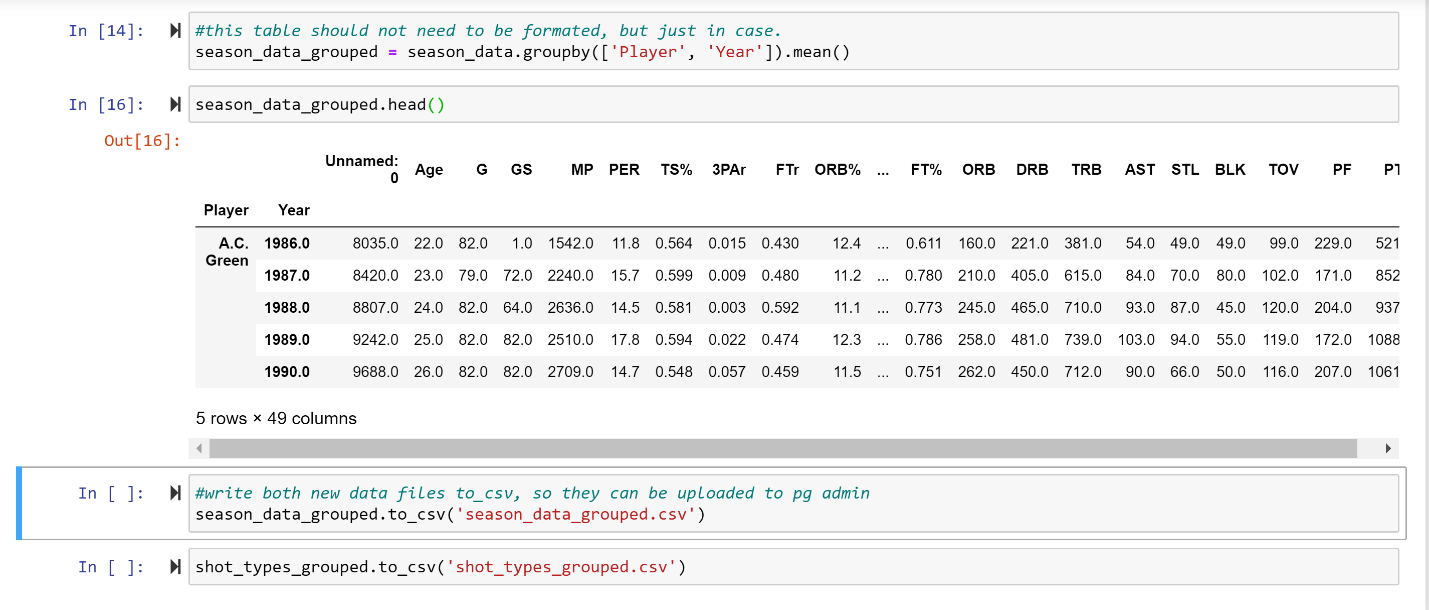
Our group began our search at basketball-reference.com, although we ended up utilizing two data sets from Kaggle: [NBA Players since 1950](https://www.kaggle.com/drgilermo/nba-players-stats), and [NBA Shot Types 2013-2019](https://www.kaggle.com/joshuawasserman/nba-shot-types-20132019). The first table provided us with box score statistics and player metrics from 1950 to 2017, by player by year. The second table, also provided information by player by year, although focused solely on shot types and percentages.

Task 2: Transform

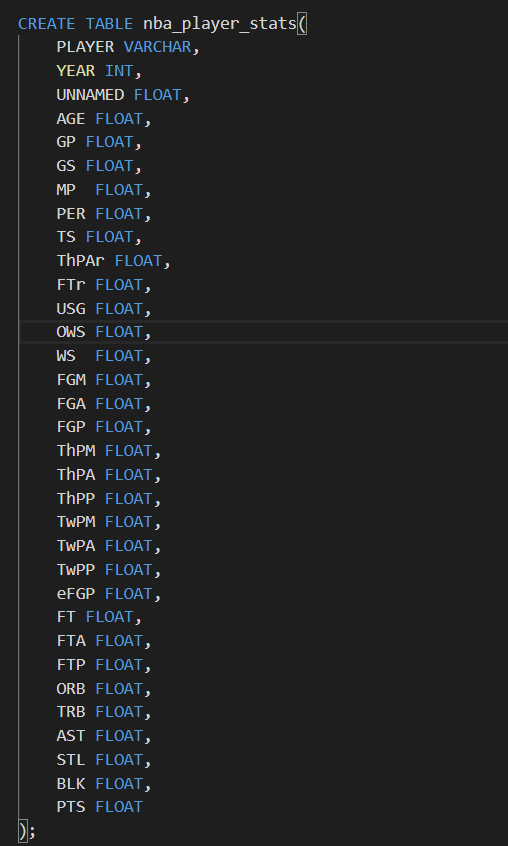
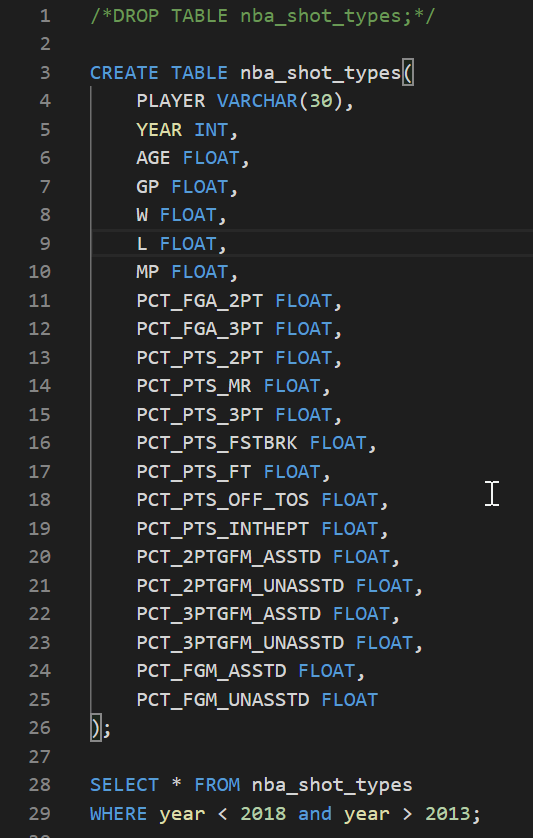
The clean up process began by reformatting the NBA Shot Types table by single year rather than by season span. A cell with the value of “2016-2017” was split with the “—” as the delimiter, and the column containing the first year of each season was deleted. In the NBA Players table columns were deleted if they were not pertinent to measuring offensive scoring. These columns included offensive rebounding percentage, defensive rebounding percentage, defensive win shares, value over replacement player, blocks per minute, turnovers, and personal fouls, among others.

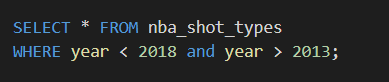
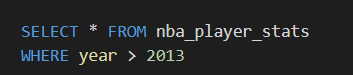
The data was then run through a pandas file, which used the groupby.mean function, which took the database of players and made sure there was one entry per year. Some players, like Deron Williams in 2017, were traded multiple times, and had multiple entries in the Shot Type database. In order to join them with the statistical data in NBA Players. Moreover, the using groupby.mean function organized the players alphabetically, allowing for a more intuitive table.



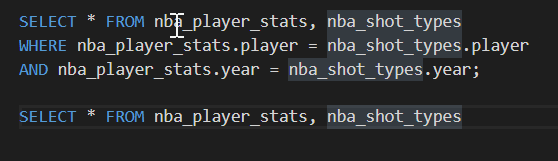


Following the clean up of the data, and the rechecking of the CSV files, the data was entered into SQL. The data was then put into tables.

After the CSVs were imported into these tables. The following queries were run to focus on the overlapping years of these two tables, 2014-2017.



Finally, to combine the tables, the following query was written.



Task 3: Load

The final database shows the entire scoring focused statistical line for each player by year with additional box score and metric statistics for greater context, thus providing clearer picture of each players induvial offensive capabilities and contributions.

This data base was chosen to help sports fans add depth to their arguments about the best player in the league those years, as well as demonstrate the value of underrated players compared to their overrated peers.

Appendix:

Explanation of Stats:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| GP | Games Played |  | FGM | Field Goals Made |  | FTA | Free Throws Attempted |
| GS | Games Started |  | FGA | Field Goals Attempted |  | FTP | Free Throw Percentage |
| MP | Minutes Played |  | FGP | Field Goal Perecentage |  | ORB | Offensive Rebounds |
| PER | Player Efficiency Rating |  | THPM | 3 Pointers Made |  | TRB | Total Rebounds |
| TS | True Shooting Percentage |  | THPA | 3 Pointers Attempted |  | AST | Assists |
| THPAR | 3 Point Attempt Rate |  | THPP | 3 Pointer Percentage |  | STL | Steals |
| FTR | Free Throw Attempt Rate |  | TWPM | 2 Pointers Made |  | BLK | Blocks |
| USG | Usage Percentage |  | TWPM | 2 Pointers Attempted |  | PTS | Points |
| OWS | Offensive Win Share |  | TWPP | 2 Pointer Percentage |  | PCT\_FGA\_2A | 2 Point Field Goals Attempted |
| WS | Win Share |  | EFGP | Effective Field Goal Percentage |  | PCT\_FGA\_3PT | 3 Point Field Goals Attempted |

|  |  |
| --- | --- |
| PCT\_PTS\_FSTBRK | % of Points on Fastbreak |
| PCT\_PTS\_FT | % of Points from Free Throws |
| PCT\_PTS\_OFF\_TOS | % of Points off of Turnovers |
| PCT\_PTS\_INTHEPT | % of Points in the Paint |
| PCT\_2PTGFM\_ASSTD | % of 2 Point Shots Assisted |
| PCT\_2PTGFM\_UNASSTD | % of 2 Point Shots Unassisted |
| PCT\_3PTGFM\_ASSTD | % of 3 Point Shots Assisted |
| PCT\_3PTGFM\_UNASSTD | % of 3 Point Shots Unassisted |
| PCT\_FGM\_ASSTD | % of Field Goals made Assisted |
| PCT\_FGM\_UNASSTD | % of Points Field Goals made Unassisted |

1. [Usage Rating](https://www.basketball-reference.com/about/glossary.html) [↑](#endnote-ref-1)
2. [Player Efficiency Rating](https://www.basketball-reference.com/about/per.html) [↑](#endnote-ref-2)
3. [True Shooting Percentage](https://en.wikipedia.org/wiki/True_shooting_percentage) [↑](#endnote-ref-3)