

NBA Player Efficiency

Sean Buchanan & Chris Brunet

Group 14

June 13th, 2023

Introduction

The NBA Player Efficiency program calculates NBA player's efficiency using publicly available tracking and scoring data. The tracking data includes speed and distance data for all players. When combined with the players weight, basic physics can be used to approximate the energy consumption of players for each game. The players can then be analyzed on a consumed power basis which isn't typically done.

Dataset

The chosen dataset is from the National Basketball Association website [1-3]. The dataset contains information on all players from the 2022-2023 and 2021-2022 regular seasons. The dataset is split into three main groups:

1. Traditional player scoring data [1]
2. Player biological data (e.g. height, weight, age) [2]
3. Player speed and distance data [3]

All data is collected on a per-game basis.

User Interface

The user interface for the program contains some general information about the program and how to use it at initial startup. Following the initial info, the user is prompted to enter a year between 2022 and 2023. Then the user is prompted enter a valid players name. If either the year or the players name is invalid the user will be informed and prompted to try again.

After valid input is entered, the program outputs:

1. The players stats for that year.
2. A summary of the league stats for that year.,
3. The top 5 most efficient players that year in terms of power consumed per point.
4. The top 5 least efficient players that year in terms of power consumed per point.
5. The total power output of each team from that year
6. A plot of the Avg power consumed per point vs player weight range
7. Some insights into the data.

Specifications

The program meets the given specifications by:

1. Merges 3 separate datasets (scoring, bios, speed and distance) with more than 10 columns and 200 rows.
2. No information is hardcoded.
3. The merged DataFrame is multi-indexed by Year, Team, and Player.
4. Multiple merge/join frames together and no rows or columns are duplicated.

5. The describe method is used to print aggregate stats for just the *interesting* columns.
6. Four columns are added to the dataset (average energy, average power, season power, power per point).
7. Aggregation computations on subsets of data are used throughout.
8. Masking operations, groupby, and pivot tables are used multiple times.
9. More than 2 user defined functions are used with doc strings and no global variables.
10. Clear instructions are given to the user upon startup.
11. Both a KeyError and a ValueError are used to catch invalid player names and years.
12. Dataset is exported to an Excel file, figure is plotted, and expected output screenshots are provided.
13. Code styling was followed.

References

- [1] *NBA Player Scoring Stats*, National Basketball Association, June 2023. [Online]. Available: <https://www.nba.com/stats/leaders>
- [2] *NBA Player Bio Stats*, National Basketball Association, June 2023. [Online]. Available: <https://www.nba.com/stats/players/bio>
- [3] *NBA Player Tracking Speed & Distance Stats*, National Basketball Association, June 2023. [Online]. Available: <https://www.nba.com/stats/players/speed-distance>