Christopher Evangelista

Exercise 6.1

Project Information: 2024 NBA Shot Efficiency Analysis

**Background:**

This analysis seeks to understand how various variables affect shot efficiency in the NBA.

This dataset contains all shots made in the 2024 NBA season. It includes attributes about the shots such as who made it, where it was made, and how it was made. I found the dataset in a github repository. The data was gathered from NBA.com and I verified the accuracy using ESPN.com.

**Dataset variables after cleaning:**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Data Type** | **Description** |
| season | category | Analyzing 1 season (2024) |
| game\_id | int64 | Unique game ID |
| game\_date | datetime64[ns] | The date of the game |
| home\_team | category | Home team |
| away\_team | category | Away team |
| team\_name | category | Team name of the player taking the shot |
| team\_abb | category | Team abbreviation of the player taking the shot |
| home\_game | bool | Boolean if player taking the shot is the Home team |
| opponent\_name | category | Opponent team name |
| opponent\_abb | category | Opponent team abbreviation |
| player\_name | object | Name of player taking the shot |
| position\_group | category | Position group of player taking the shot |
| position | category | Position of player taking the shot |
| shot\_made | bool | Boolean if the shot was made/went in |
| action\_type | category | What kind of shot was taken (ex. Jump shot/layup/dunk) |
| shot\_type | category | 3pt shot or 2pt shot |
| basic\_zone | category | Area where shot was taken on the court |
| zone\_name | category | High-level area where shot was taken on the court |
| zone\_range | category | 8 Ft increment range from the basket the shot was taken |
| loc\_x | float[pyarrow] | Location on sideline shot was taken |
| loc\_y | float[pyarrow] | Location on baseline shot was taken |
| shot\_distance | int8[pyarrow] | How far away the shot was taken in feet |
| quarter | int8[pyarrow] | The quarter of the game the shot was taken |
| mins\_left | int8[pyarrow] | How many whole minutes left in game the shot was taken |
| secs\_left | int8[pyarrow] | How many secs left rounded to the minute in the quarter the shot was taken\ |
| secs\_left\_in\_quarter | int16[pyarrow] | How many raw seconds left in the quarter the shot was taken |

Cleaning/Wrangling – Steps are outlined and explained in the Jupyter notebook

Data Limitations – The dataset does not include free throws.

Defining Questions –

* Who are the best teams/players at different shot types/ranges
* Who are the best clutch players (most consistent and high averages in the late minutes of a game)
* Is there a correlation between shot efficiency and home/away? Is there a home team advantage?
* Most improved team/player over the season