The Keenfolks - Data Engineer Challenge

ESPN – NBA predictor Web-Application

In this assessment you will define the architecture for the ETL of a ML based tool that predicts which basketball players in the NBA make the All-Star team in any given year. We will be analyzing your data wrangling abilities as well as your capabilities as a data engineering practitioner. This means we will look more closely at your code and your reasoning rather than the modeling and the complexity of the architecture.

The file you will be working with is a .ipynb, feel free to add any additional cells and expand on the answers you want under each question-header! The dataset you will be working with contains data for all NBA basketball players who played between 1950 and 2017, the variable descriptions are below.

To submit the assessment, please upload your code as a public repository on your GitHub and share the link. You must delete the repository after 14 days. Good luck and have fun!

Variable Descriptions:

Year - The season the stats were recorded from
Player - The name of the player
Pos – The player's position
Age - Self-explanatory
Tm - Their team
G - Games played

PER
Player Efficiency Rating: PER sums up all a player's positive accomplishments, subtracts the negative accomplishments, and returns a per-minute rating of a player's performance

TS%
True shooting percentage: True shooting percentage is a measure of shooting efficiency that takes into account field goals, 3-point field goals, and free throws.

ORB%
Offensive Rebound Percentage: Offensive rebound percentage is an estimate of the percentage of available offensive rebounds a player grabbed while he was on the floor.

DRB%
Defensive rebound percentage: Defensive rebound percentage is an estimate of the percentage of available defensive rebounds a player grabbed while he was on the floor.

TRB%
Total Rebound Percentage: Total rebound percentage is an estimate of the percentage of available rebounds a player grabbed while he was on the floor.

AST%
Assist Percentage: Assist percentage is an estimate of the percentage of teammate field goals a player assisted while he was on the floor.

STL%
Steal Percentage; Steal Percentage is an estimate of the percentage of opponent possessions that end with a steal by the player while he was on the floor.

BLK%
Block Percentage: Block percentage is an estimate of the percentage of opponent two-point field goal attempts blocked by the player while he was on the floor.

TOV%
Turnover Percentage: Turnover percentage is an estimate of turnovers per 100 plays.

WS
Win Shares; an estimate of the number of wins contributed by a player

BPM
Box Plus/Minus; A box score estimate of the points per 100 possessions that a player contributed above a league-average player, translated to an average team.

VORP
Value Over Replacement Player; a box score estimate of the points per 100 TEAM possessions that a player contributed at an above-average level

FG%
Field Goal Percentage (percentage of total 2 point and 3 point shots that went in, excluding free throws)

3P%
3-Point Field Goal Percentage

2P%
2-Point Field Goal Percentage

eFG%
This statistic adjusts for the fact that a 3-point field goal is worth one more point than a 2-point field goal. For example, suppose Player A goes 4 for 10 with 2 threes, while Player B

goes 5 for 10 with 0 threes. Each player would have 10 points from field goals, and thus would have the same effective field goal percentage (50%).

FT%
Free Throw Percentage

TRB
Total Rebounds (available since the 1950-51 season)

AST - Total Assists
STL - Total Steals
BLK - Total Blocks
TOV - Total Turnovers
PTS - Total Points