

Andrew Bauer

DATA 6560 – Sports Analytics

24 Nov 2025

Preprocessing Report & Data Dictionary

The Value of 3pt Shooting as Players Age

Pre-Processing Pipeline

1. Import Data from Basketballreference.com

- Dataset 1: 2024-2025 Advanced Player Data
 - o Organized by name, should start with Precious Achiuwa and end with Xavier Tillman Sr.
- Dataset 2: 2024-2025 Shooting Stats Player Data
 - o Organized by name, should start with Precious Achiuwa and end with Xavier Tillman Sr.
- Dataset 3: 2024-2025 Per Game Stats Player Data
 - o Organized by name, should start with Precious Achiuwa and end with Xavier Tillman Sr.

2. Combine Data & Align Datasets

- In one singular page on excel, combine the 3 data sets via copy and paste, with Advanced data first, then shooting data, then per game stats
 - o Checkpoint: First column should be RK and last column should be PTS

3. Refine Columns

- Delete columns E-T (should start with GS (Games Started), and end with DWS (defensive win shares))
- Delete WS/48 to BPM
- Delete FG% to 16-3P
 - o Checkpoint: Last column should still be PTS, and it should be column AO
- Delete from 2P to Awards
 - o Checkpoint: Last column should still be PTS, and it should be column Z
- Delete from 2P to PF
 - o Checkpoint: last column should still be PTS, and it should be column K

4. Eliminate Rows

- LeBron James
 - o Outlier that will ruin trends
- Any Player with Games played under 15 must be eliminated
 - o Minimal games can skew trends
- MVP Candidates (Nikola Jokic & Shai Gilgeous Alexander)
 - o Extremely high win shares can ruin trends

Checkpoint: dataset should consist of 461 players

Data Dictionary			
Column ID	Name	Format	Relevance
Rk	Rank	Number	Unique Identifier for Player
Player ▲	Player	Text	Each name is Unique in NBA
Age	Age	Number	Age for disitinguishing in trends
G	Games Played	Number	Determines prevelance of player on court
WS	Win Shares	Number	Metric being analyzed
VORP	Value over Replacement Player	Number	Metric being analyzed
3P	% of FG that are 3P shots	Percentage	Distinguishes shot selection
3P/Game	3Pts made per game	Number	Determines 3pt value
3PA	3Pt attempts per game	Number	Determines 3pt selection
3P%	3pt %	Percentage	Determines 3pt efficiency
PTS	Points per game	Number	Large contributior to WS