NBA Player Statistics Analysis by Decade

By: Cy Seeley, Arya Tatke, John Apel, Anthony Vo

Introduction

- Focus of this project was collecting and analyzing NBA player statistics from the 1970s to the 2020s
- Our goal was to compile a comprehensive dataset containing:
  - Player career stats- grouped by the decade in which players began their careers
  - Analyze trends in player performance metrics including points, rebounds, and assists over time while comparing domestic and foreign players' contributions

Questions to answer

1. How do key performance metrics (points, rebounds, assists) vary across decades for NBA players?
2. Are there trends in player performance or career longevity over time, and how do these differ between domestic and foreign players?
3. Which players or seasons stand out as exceptional within each decade, particularly among foreign players?
4. How has the contribution of foreign players to NBA performance metrics evolved over the decades?

Dataset

- The dataset consists of career statistics for a sample of NBA players who started their careers between 1970 and 2025, grouped by decade
- The dataset could be used in many different fields:
    - Sports analytics
    - Data science
    - Historical analysis
    - Business and marketing

Preprocessing

- Player sampling
- Data retrieval
- Decade labeling
- Foreign player identification and integration
- Data consolidation
- Error handling
- Data storage

Analysis Methods

- Data exploration
  - understand player stats and nationality distributions.
- Visualization
  - bar plots (top countries, scorers), box plots (points distribution), and line plots (decade trends)
- Statistical analysis
  - Calculated mean stats (PTS, REB, AST) for groups and decades, with group comparisons (foreign vs. all players).
- Outlier detection
  - Identified outliers in points distribution via box plots, highlighting performance variations.
- Trend analysis
  - Examined stat trends across decades (e.g., 1970s: 414.52 PTS, 2020s: 394.53 PTS) and player counts.
- Foreign player contribution
  - Analyzed foreign player representation (% per decade) and performance (e.g., Foreign: 536.20 PTS vs. Overall: 451.01 PTS).

Program overview

Key Features:

1. **Data Collection**: Samples 50 players/decade via nba_api, saves to nba_players_by_decade.csv.
2. **Foreign Analysis**: Filters non-U.S. players, compares stats.
3. **Visuals**: Bar/box/line plots for country representation, top scorers, decade trends.
4. **Insights**: Tracks stats (e.g., 1970s: 414.52 PTS, Foreign: 536.20 PTS) and foreign player growth.

Tools: Python, Pandas, Seaborn, Matplotlib, nba_api, Google Colab.

Output: CSVs (nba_players_by_decade.csv, foreign_vs_decade_players_avg.csv) and visualizations.

Challenges: API timeouts, deprecated Pandas warnings.

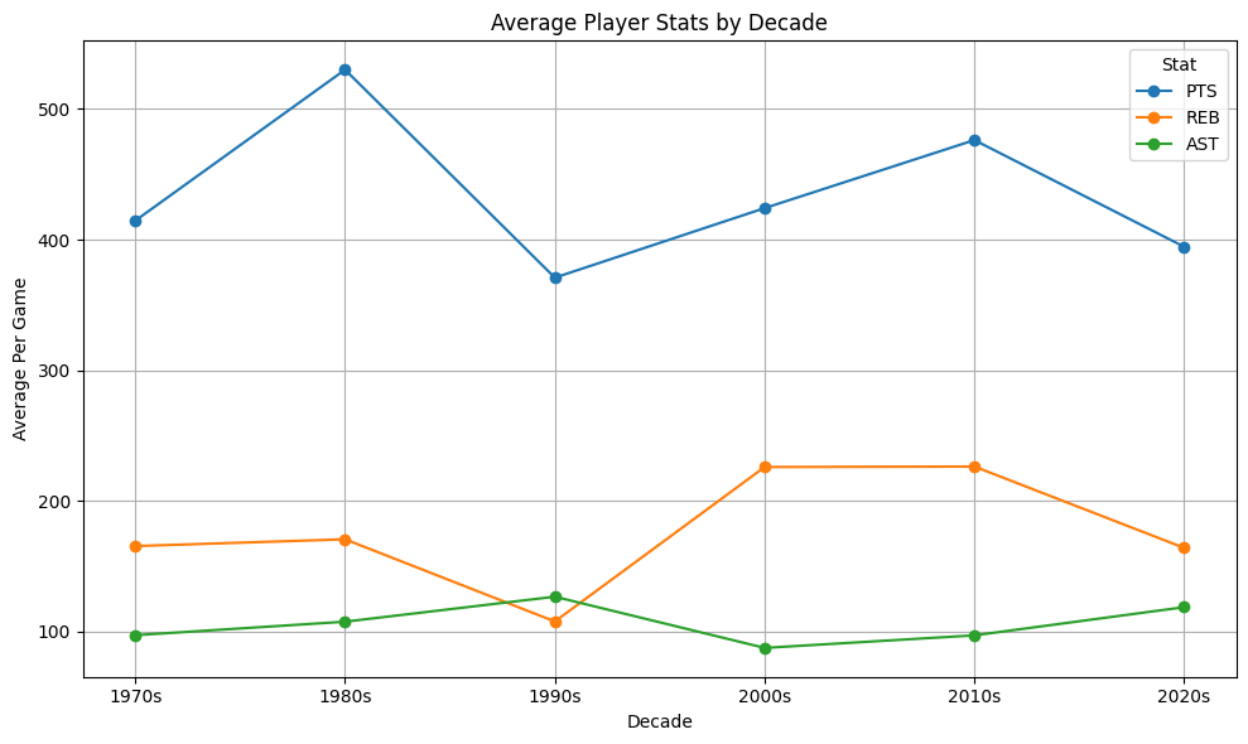
Results and Findings

- The script was able to collect NBA player stats from the 1970s to the 2020s
- A few of the observations include:
  - Average performance by decade
  - Average stats by decade
  - Points per game distributed by decade

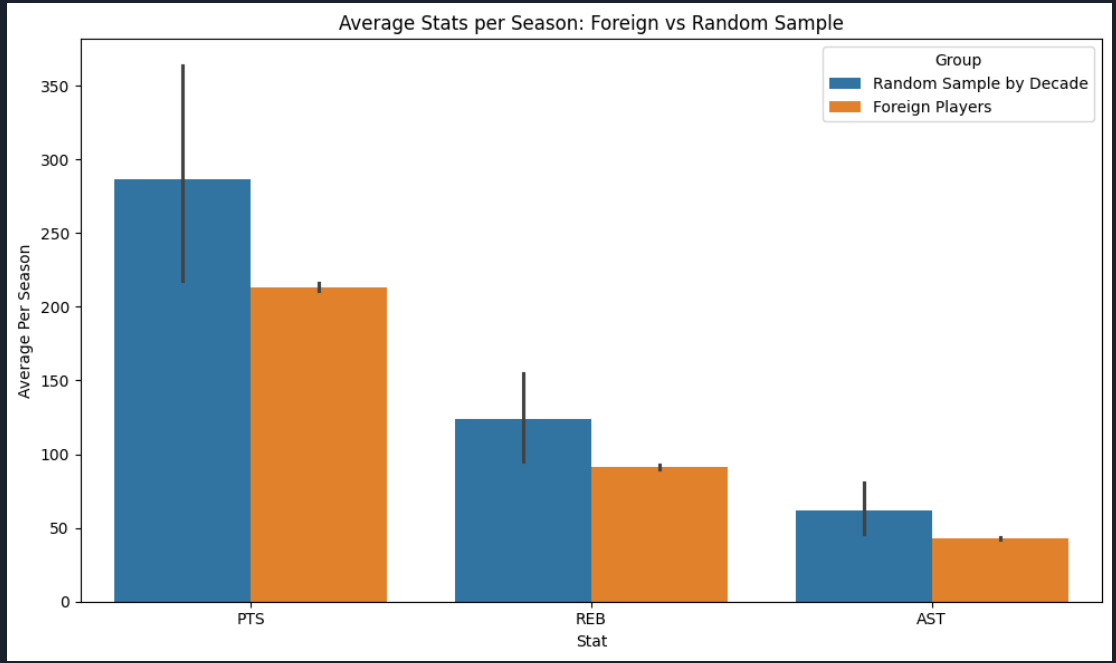
Findings:

- Performance trends: Points peaked in 1980s, rebounds in 2000s/2010s, assists in 1990s.
- Foreign players (e.g., Nowitzki, Divac) outperformed overall averages (536.20 PTS vs. 451.01).
- Foreign player presence grew significantly in 2010s/2020s.

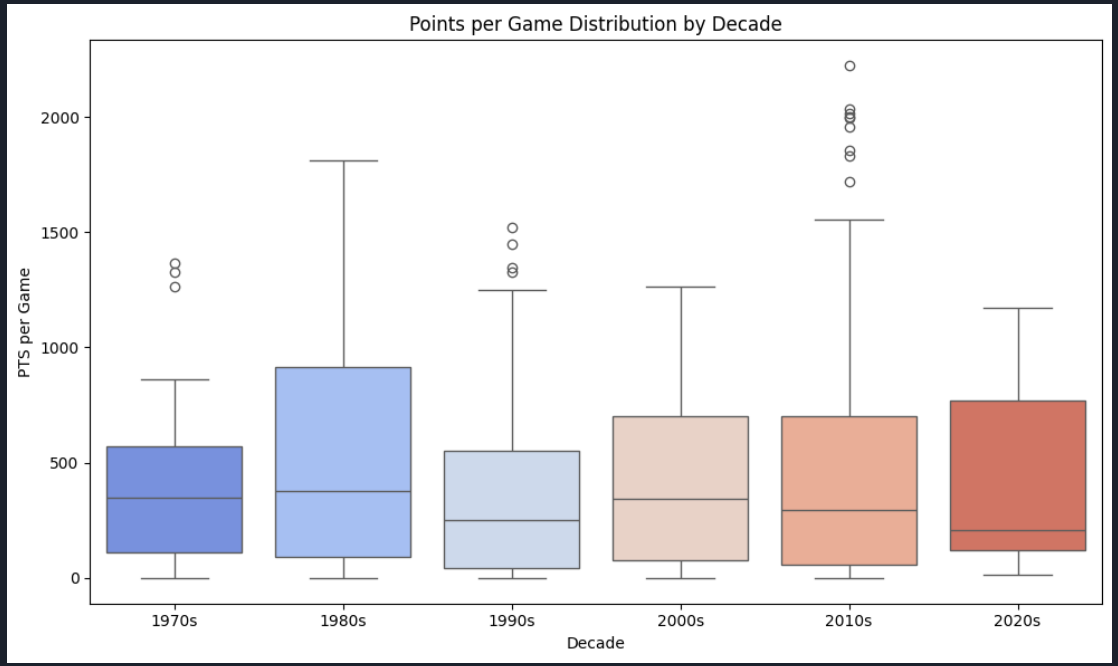
Results


Average Player Stats by Decade
Stat
PTS
REB
AST
Average Per Game
500
400
300
200
100
1970s
1980s
1990s
2000s
2010s
2020s
Decade

Results


Average Stats per Season: Foreign vs Random Sample
Group
Random Sample by Decade
Foreign Players
Average Per Season
0
50
100
150
200
250
300
350
PTS
REB
AST
Stat

Results


Points per Game Distribution by Decade
PTS per Game
2000
1500
1000
500
0
1970s
1980s
1990s
2000s
2010s
2020s
Decade