

NBA PROJECT

Data Collection and Preprocessing*

To conduct this analysis, we performed *web scraping* on the official *NBA website* (or a related statistics platform) to collect comprehensive player and team data. Here's an overview of the steps taken

Data Scraping .1

.Used Python's **requests** library to fetch HTML content from the NBA website -
Parsed the data using **BeautifulSoup (bs4)** to extract structured information (e.g., player -
.stats, team rosters, and performance metrics)

Data Cleaning .2

.Removed **duplicate entries** to ensure each player/team was represented once -
Handled **null/missing values** by either filling them with averages (for numerical data) or -
.dropping irrelevant rows
Applied **regular expressions (regex)** to standardize text fields (e.g., player names, team -
.abbreviations) and extract numerical values (e.g., percentages, ages)
Converted **column data types** (e.g., strings to integers for "Age" or "Points," cleaned "FG%" -
.into float values)

Exploratory Data Analysis (EDA) .3

Processed the cleaned data to generate visualizations (e.g., distributions, correlations, and -
.positional trends)

Analysis and Insights from NBA Player Data Visualizations

Age Distribution of NBA Players .1

The age distribution shows that the majority of NBA players are between **22.5 and 30 years old**, -
with a peak around **25–27.5 years**. Few players are above **35**, indicating that longevity in the -
.league is uncommon

Top 10 Players by Total Points .2

NBA PROJECT

Joel Embiid, **Luka Dončić, and **Damian Lillard* lead in total points scored. This highlights * -
.their offensive dominance and consistency in performance

Top 10 Teams with Most Players .3

Teams like *SAS (San Antonio Spurs)* and *DET (Detroit Pistons)* have the highest number of -
.players (close to 50), suggesting deep rosters or rebuilding phases with younger talent

Distribution of Player Positions .4

Point Guards (PG)* and *Shooting Guards (SG)* make up the largest share (~18–19% each), * -
while *Centers (C)* are the least common (~17%). This reflects the modern NBA's emphasis on
.guard-oriented play

Age vs. Total Points .5

Players aged *25–30* tend to score the most points, indicating peak performance during these -
.years. Productivity declines slightly after *30*, likely due to reduced athleticism

Shooting Percentages Distribution .6

Field Goal (FG%)* percentages are typically higher than *3-Point (3P%)*, as shots closer to the * -
.basket (2P%) are more efficient. Free Throw (FT%) consistency varies widely among players

Assists vs. Total Rebounds by Position .7

Point Guards (PG)* lead in assists, while *Centers (C)* dominate rebounds. This aligns with * -
.traditional positional roles, where guards facilitate offense and big men control the paint

Top 10 Teams by Average Points .8

Teams like *TOR (Raptors)* and *PHI (76ers)* rank highest in average points per game, reflecting -
.strong offensive systems or star-powered lineups

Shooting Percentages of Top 5 Players .9

Elite scorers like *Giannis Antetokounmpo* and *Kevin Durant* excel in *2P%* (inside scoring), -
. *%while *Steph Curry* and *Luka Dončić* balance efficiency across *FG%, **3P%, and **FT

NBA PROJECT

Correlation Heatmap of Player Statistics .10

:*Positive Correlations* -

.Points (PTS) strongly correlate with minutes played (MP) and field goal attempts -

.Assists (AST) and rebounds (TRB) are linked to position (e.g., guards vs. centers) -

:*Negative Correlations* -

Age shows slight negative trends with steals (STL) and blocks (BLK), suggesting younger players -

.are more defensively active

:*Notable Weak Correlations* -

.P% and age are unrelated, indicating shooting skill is age-independent3 -

Key Takeaways

.*Prime Age: NBA players peak offensively between **25–30 years* -

.Positional Trends*: Guards dominate ball-handling, while centers focus on rebounds* -

Team Strategies*: High-scoring teams (e.g., BOS, LAL) prioritize offensive firepower, while others * -

.(e.g., SAS) may emphasize roster depth

Shooting Efficiency*: Top players maintain high 2P% and FT%, with 3P% being a differentiator for * -

.guards

This data underscores the balance between age, position, and skill in shaping NBA performance.

.Teams leveraging these insights can optimize roster construction and player development

Storing Data in MongoDB

As part of our project workflow, we used MongoDB as a NoSQL database to store both the cleaned

.dataset and the generated visualizations

The cleaned CSV file was imported into a dedicated collection to allow efficient retrieval and analysis

.within the Streamlit app

Each visualization was saved as a Base64-encoded image and stored in a separate collection, along

.with its title and description

NBA PROJECT

This structure enabled us to dynamically retrieve and display the content in the app without having
.to regenerate plots on each load

Using MongoDB added flexibility and scalability to our application, making it easier to manage
.structured and semi-structured data together

:Steps to Build the Streamlit App

:Connecting to MongoDB .1

.We used pymongo to connect to our MongoDB database

.We retrieved both the cleaned dataset (CSV) and the stored visualizations from separate collections

:Data Preparation .2

.The data was converted into a Pandas DataFrame

We cleaned the DataFrame by removing the _id column and any duplicate rows using
.()drop_duplicates

:Building the Dashboard .3

:We created a user-friendly interface using Streamlit, with features like

NBA PROJECT

.A sidebar for CSV file download

.KPIs section showing total players, number of teams, and the top scorer

.Visualizations displayed with their respective insights and analysis

.Selectbox interaction to explore teams and player details dynamically

:Displaying Visualizations from MongoDB .4

.Pre-generated visualizations were retrieved from MongoDB in Base64 format

.They were decoded and displayed directly in the app along with descriptive analysis below each one

:Deployment

We successfully deployed the interactive dashboard using Streamlit, making it accessible as a standalone web app that allows users to explore NBA data, download the dataset, and view .meaningful statistics and visualizations
