NBA PLAYER STATISTICAL ANALYSIS AND PREDICTION PROJECT

COLLABORATORS

JACOB EVANS
KARAN ANAND
MERT OZTOP
PRATIK PUROHIT
PRIYA MARINGANTI

OVERVIEW

This project aims to develop a basketball player statistics analysis and prediction system using machine learning techniques with NBA player datasets. The system utilizes historical player data to provide insights into player performance and predict player stats for the upcoming season, by leveraging the power of machine learning algorithms and the comprehensive NBA player statistics dataset.

Objectives 

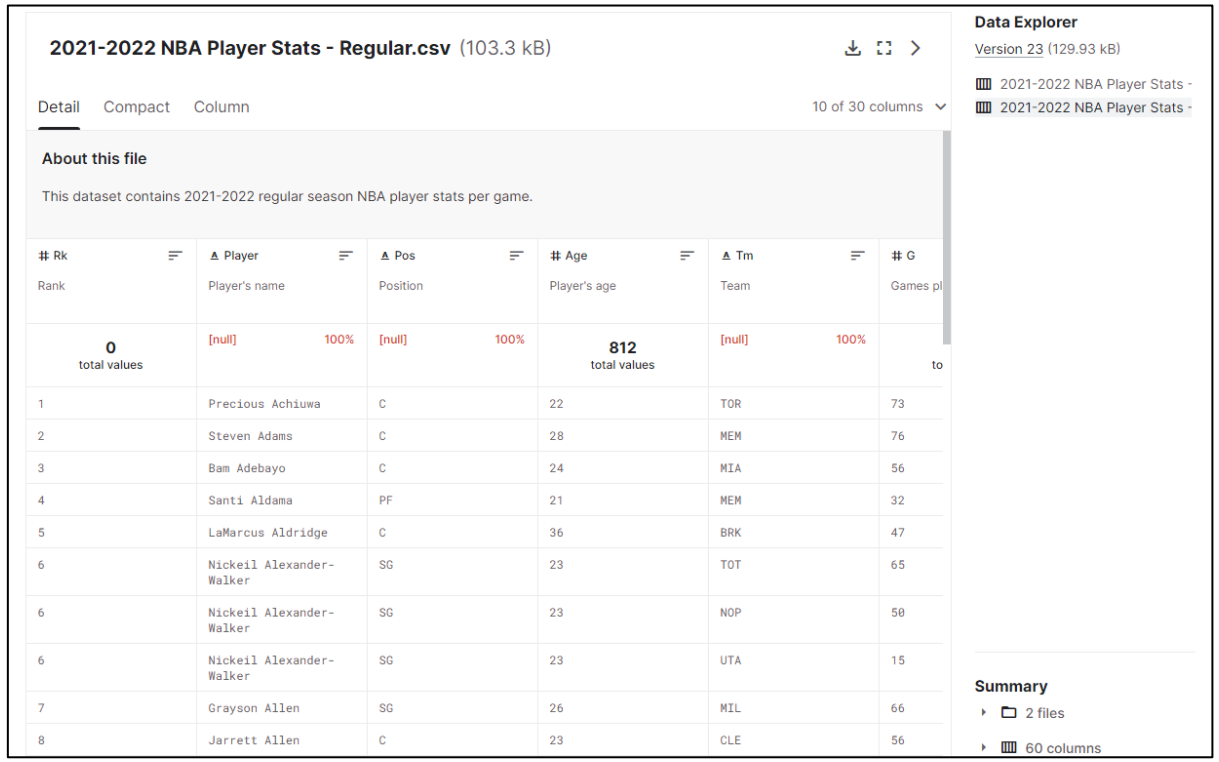
- Predict basketball player stats for the upcoming season based on historical data
- Leverage player stats from the previous year to forecast player performance
- Identify suitable machine learning algorithms for scalability, accuracy, and interpretability in predicting player performance
- Evaluate the performance of the prediction system and ensure its effectiveness in real-world scenarios, such as team selection and player scouting

METHODOLOGY

Methodology 

1. Data Collection: Gather a large dataset of NBA player statistics from the Kaggle dataset
2. Data Preprocessing: Clean, normalize, encode, and engineer features from the NBA dataset
3. Exploratory Data Analysis: Gain insights, patterns, and analyze feature distributions
4. Model Training and Evaluation: Experiment with various machine learning algorithms, fine-tune models for accurate predictions
5. User Interface Development: Create an intuitive interface for users to input player data and view predictions
6. Testing and Validation: Ensure the accuracy, robustness, and scalability of the prediction system

DATA SOURCE - KAGGLE



2021-2022 NBA Player Stats - Regular.csv (103.3 kB)

Detail | Compact | Column | 10 of 30 columns

About this file

This dataset contains 2021-2022 regular season NBA player stats per game.

Rk Rank	Player Player's name	Pos Position	Age Player's age	Tm Team	G Games pl
0 total values	[null] 100%	[null] 100%	812 total values	[null] 100%	to
1	Precious Achiuwa	C	22	TOR	73
2	Steven Adams	C	28	MEM	76
3	Bam Adebayo	C	24	MIA	56
4	Santi Aldama	PF	21	MEM	32
5	LaMarcus Aldridge	C	36	BRK	47
6	Nickeil Alexander-Walker	SG	23	TOT	65
6	Nickeil Alexander-Walker	SG	23	NOP	50
6	Nickeil Alexander-Walker	SG	23	UTA	15
7	Grayson Allen	SG	26	MIL	66
8	Jarrett Allen	C	23	CLE	56

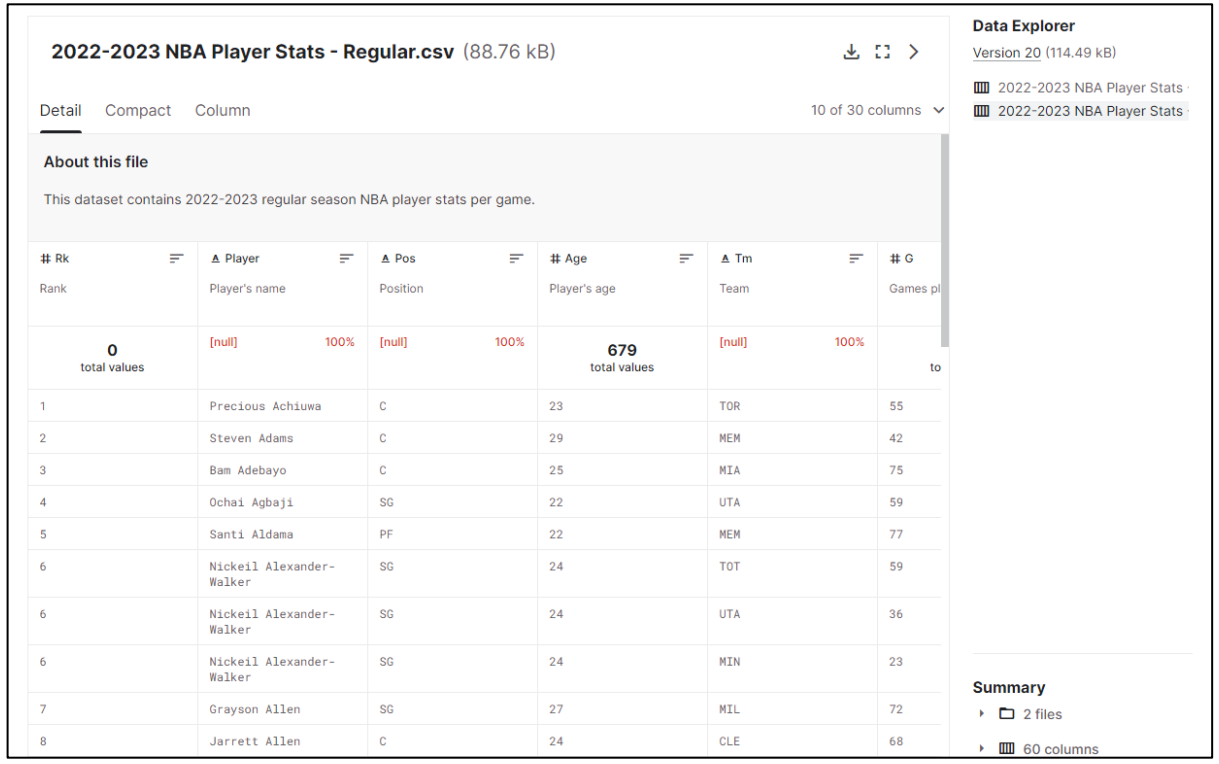
Data Explorer

Version 23 (129.93 kB)

2021-2022 NBA Player Stats -
2021-2022 NBA Player Stats -

Summary

2 files
60 columns



2022-2023 NBA Player Stats - Regular.csv (88.76 kB)

Detail | Compact | Column | 10 of 30 columns

About this file

This dataset contains 2022-2023 regular season NBA player stats per game.

Rk Rank	Player Player's name	Pos Position	Age Player's age	Tm Team	G Games pl
0 total values	[null] 100%	[null] 100%	679 total values	[null] 100%	to
1	Precious Achiuwa	C	23	TOR	55
2	Steven Adams	C	29	MEM	42
3	Bam Adebayo	C	25	MIA	75
4	Ochai Agbaji	SG	22	UTA	59
5	Santi Aldama	PF	22	MEM	77
6	Nickeil Alexander-Walker	SG	24	TOT	59
6	Nickeil Alexander-Walker	SG	24	UTA	36
6	Nickeil Alexander-Walker	SG	24	MIN	23
7	Grayson Allen	SG	27	MIL	72
8	Jarrett Allen	C	24	CLE	68

Data Explorer

Version 20 (114.49 kB)

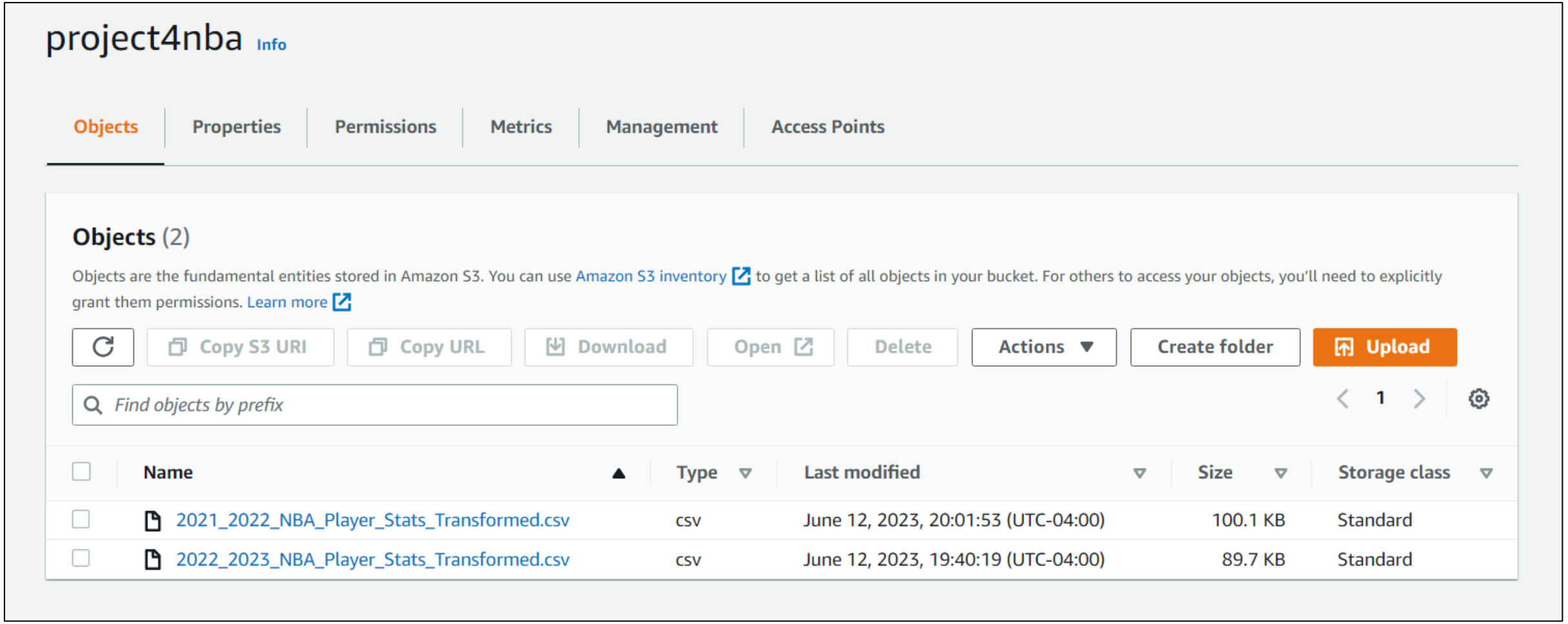
2022-2023 NBA Player Stats -
2022-2023 NBA Player Stats -

Summary

2 files
60 columns

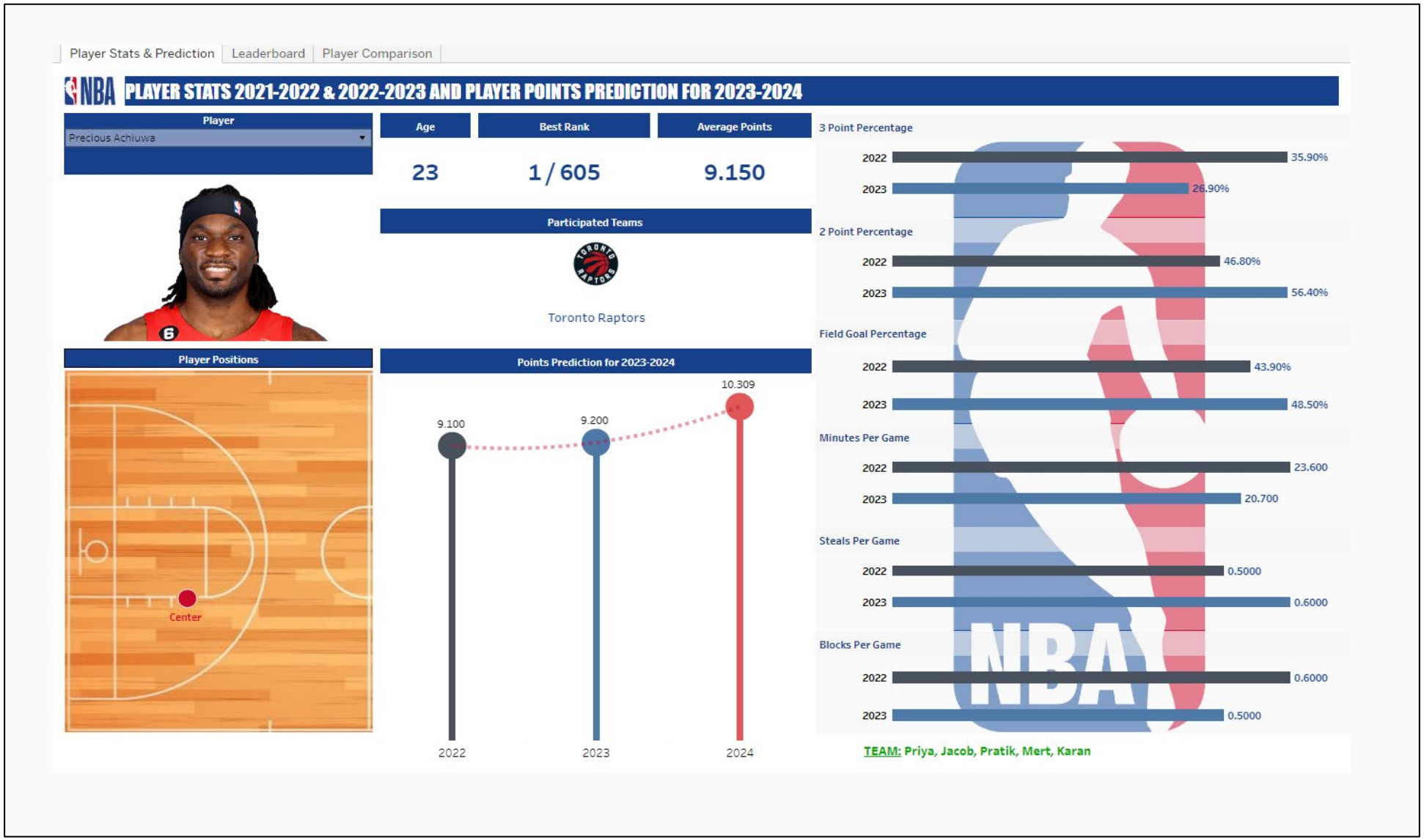
Retrieving last two seasons' data from Kaggle

Data on s3 AWS

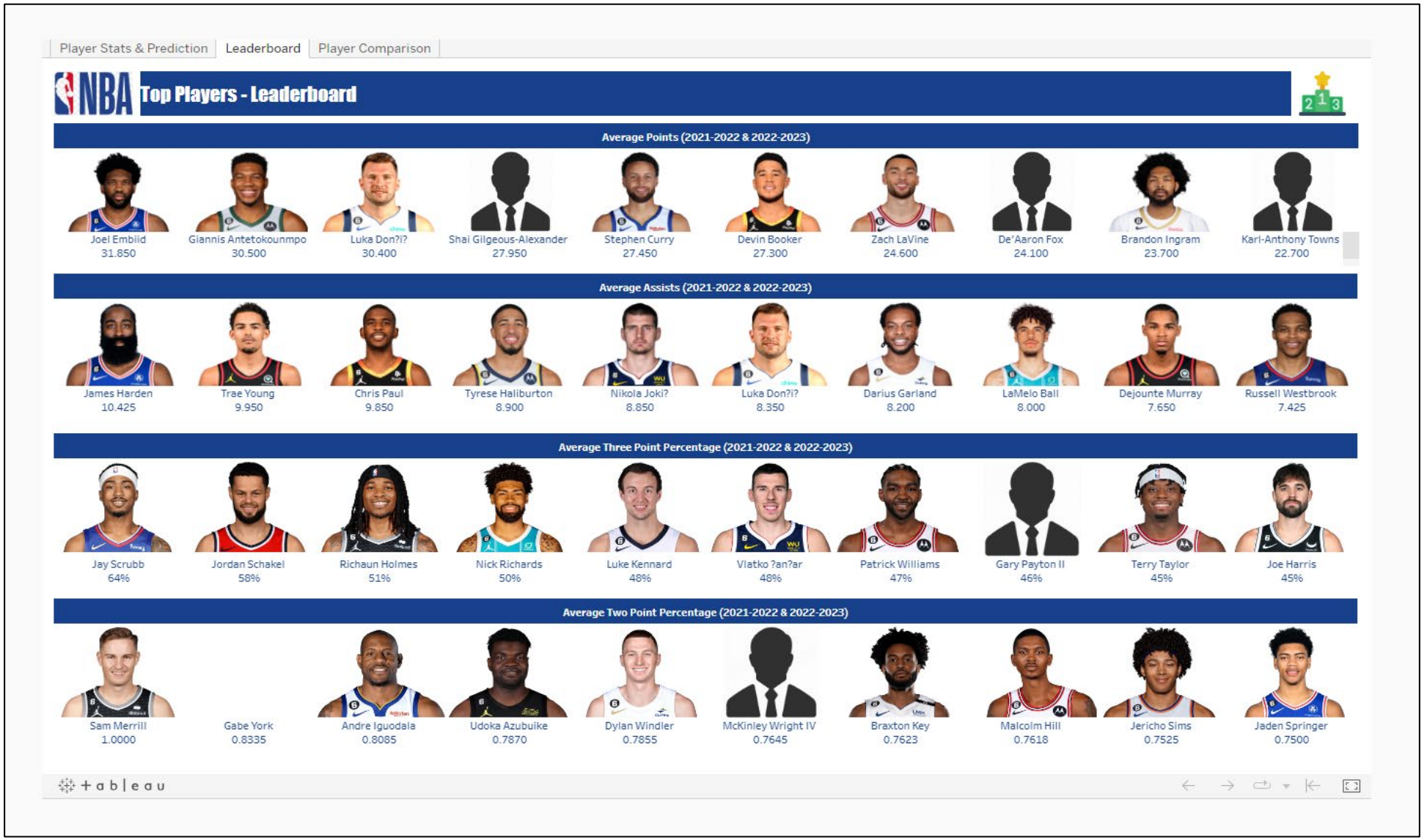


Uploading the data sets on AWS for the main code

DASHBOARD 1


Player Stats & Prediction
Leaderboard
Player Comparison
NBA PLAYER STATS 2021-2022 & 2022-2023 AND PLAYER POINTS PREDICTION FOR 2023-2024
Player
Precious Achiuwa
Age
23
Best Rank
1 / 605
Average Points
9.150
Participated Teams
Toronto Raptors
Player Positions
Center
Points Prediction for 2023-2024
9.100
9.200
10.309
2022
2023
2024
3 Point Percentage
2022 35.90%
2023 26.90%
2 Point Percentage
2022 46.80%
2023 56.40%
Field Goal Percentage
2022 43.90%
2023 48.50%
Minutes Per Game
2022 23.600
2023 20.700
Steals Per Game
2022 0.5000
2023 0.6000
Blocks Per Game
2022 0.6000
2023 0.5000
TEAM: Priya, Jacob, Pratik, Mert, Karan

DASHBOARD 2


Player Stats & Prediction
Leaderboard
Player Comparison
NBA Top Players - Leaderboard
Average Points (2021-2022 & 2022-2023)
Joel Embiid 31.850
Giannis Antetokounmpo 30.500
Luka Don?i? 30.400
Shai Gilgeous-Alexander 27.950
Stephen Curry 27.450
Devin Booker 27.300
Zach LaVine 24.600
De'Aaron Fox 24.100
Brandon Ingram 23.700
Karl-Anthony Towns 22.700
Average Assists (2021-2022 & 2022-2023)
James Harden 10.425
Trae Young 9.950
Chris Paul 9.850
Tyrese Haliburton 8.900
Nikola Joki? 8.850
Luka Don?i? 8.350
Darius Garland 8.200
LaMelo Ball 8.000
Dejounte Murray 7.650
Russell Westbrook 7.425
Average Three Point Percentage (2021-2022 & 2022-2023)
Jay Scrubb 64%
Jordan Schakel 58%
Richaun Holmes 51%
Nick Richards 50%
Luke Kennard 48%
Vlatko ?an?ar 48%
Patrick Williams 47%
Gary Payton II 46%
Terry Taylor 45%
Joe Harris 45%
Average Two Point Percentage (2021-2022 & 2022-2023)
Sam Merrill 1.0000
Gabe York 0.8335
Andre Iguodala 0.8085
Udoka Azubuike 0.7870
Dylan Windler 0.7855
McKinley Wright IV 0.7645
Braxton Key 0.7623
Malcolm Hill 0.7618
Jericho Sims 0.7525
Jaden Springer 0.7500
tableau

DASHBOARD 3


Player Stats & Prediction
Leaderboard
Player Comparison
NBA
Player Comparison based on position of the player
Player Position
Power Forward
Player1
Admiral Schofield
Player2
Chuma Okeke
Age
25
Best Rank
433 / 605
Average Points (2021-2022 & 2022-2023)
4.200
Average Assists (2021-2022 & 2022-2023)
0.8000
Power Forward
Age
23
Best Rank
420 / 605
Average Points (2021-2022 & 2022-2023)
8.600
Average Assists (2021-2022 & 2022-2023)
1.700
Points Prediction
3.800
4.200
6.735
2022
2023
2024
Stats (2021-2022 & 2022-2023)
32.40%
64.60%
45.10%
54.90%
0.2000
0.1000
Average Three Point %
Average Two Point %
Average Field Goal %
Average Effective Field Goal %
Average Steals per Game
Average Blocks per Game
Stats (2021-2022 & 2022-2023)
31.80%
48.00%
37.60%
47.80%
1.400
0.600
Points Prediction
8.600
4.700
5.899
2022
2023
2024
+ableau

CHALLENGES

•Bad Encoding: The original data had encoding issues that we struggled to handle. We had to apply encoding techniques to ensure proper handling and interpretation of the data.

• External Factors: While player statistics provide valuable insights, it's important to note that other factors can influence a player's performance on the court. Factors such as injuries, team dynamics, coaching strategies, and external circumstances were not included in our analysis. Considering these external factors could further enhance the accuracy and predictive power of the model.

•Outliers: To handle these outliers, we implemented a post-processing step where we replaced any negative predicted values with zeros. This approach allowed us to address the outliers and ensure that the predicted statistics remain within a valid range. By zeroing out the negative values, we mitigated the impact of outliers on the model's performance and ensured that the predicted player statistics align with the expectations of NBA player performance.

RESULT

The prediction system achieved **R-squared value of 0.9998**, indicating a high level of accuracy in predicting player statistics based on the historical data.

The NBA Player Statistics Analysis and Prediction System leverages historical player data, applies machine learning techniques, and provides valuable insights and predictions on player performance. The system can assist with team selection, player scouting, and forecasting player statistics for the upcoming season.

THANK YOU

QUESTIONS ARE WELCOME