



# NBA - Season 2021/2022



- Analisi di dati statistici di cestisti della Regular Season del campionato di NBA 2021/2022.
- Dataset: “2021-2022 NBA Player Stats” scaricato da Kaggle
- Dimensioni dataset: 611 righe x 30 colonne
- Lavoro suddiviso in Data Cleaning, Data Manipulation, Data Visualization e Machine Learning



# Data Cleaning

- Processo di pulizia di record e colonne.
- Importazione di dataset da file csv in un Data Frame.
- Miglioramento leggibilità colonne.
- Rimozione dei cestisti che non hanno giocato nessuna partita.
- Dimensione dataset ridotta a 590 righe x 30 colonne



# Data Manipulation

- Processo di manipolazione dei dati
- Estrazione nuove colonne:
  - Totale tentativi di canestro per partita
  - Totale punti per minuto
  - Totale punti segnati
- Estrazione dei migliori/peggiori cestisti in base ai punti al minuto ed in base ai punti totali.
- Estrazione dei team con media dei punti al minuto
- Analisi delle Performance: UFuncs vs For Loop



# Data Manipulation

Migliori/peggiori cestisti in base ai punti totali

	player	age	team	total_points
586	Trae Young	23	ATL	1143.9
11	Giannis Antetokounmpo	27	MIL	1115.4
122	Stephen Curry	33	GSW	1104.6
513	Jayson Tatum	23	BOS	1083.6
130	DeMar DeRozan	32	CHI	1066.0
...	...	...	...	...
144	PJ Dozier	25	DEN	97.2
511	Keifer Sykes	28	IND	85.8
279	Stanley Johnson	25	LAL	79.2
402	Nerlens Noel	27	NYK	63.0
16	Trevor Ariza	36	LAL	44.4



# Data Manipulation

Migliori/peggiori cestisti in base ai punti al minuto

	player	age	team	points_by_minutes
11	Giannis Antetokounmpo	27	MIL	0.874618
156	Joel Embiid	27	PHI	0.863222
149	Kevin Durant	33	BRK	0.802740
586	Trae Young	23	ATL	0.801724
265	LeBron James	37	LAL	0.789617
...	...	...	...	...
211	Maurice Harkless	28	SAC	0.217391
254	Andre Iguodala	38	GSW	0.204878
182	Taj Gibson	36	NYK	0.193717
16	Trevor Ariza	36	LAL	0.192708
402	Nerlens Noel	27	NYK	0.152174



# Data Manipulation

Estrazione dei team con media dei punti al minuto.

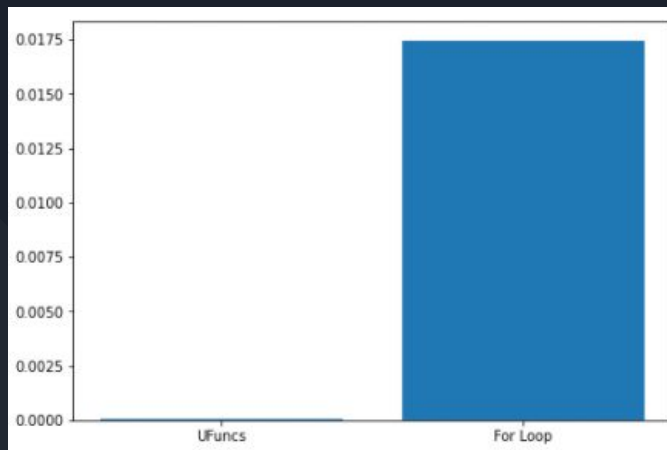
Points per minutes			
team			
MEM	0.491776	TOR	0.451626
UTA	0.488111	NOP	0.450838
MIL	0.486429	DAL	0.449961
ATL	0.481344	BRK	0.445857
SAS	0.478731	LAC	0.443000
POR	0.477954	IND	0.442142
CHO	0.475532	DET	0.441898
MIA	0.471582	PHI	0.438017
PHO	0.470775	SAC	0.435941
MIN	0.469328	ORL	0.432193
CLE	0.466845	GSW	0.431988
CHI	0.464206	LAL	0.430389
WAS	0.458304	DEN	0.426061
HOU	0.456910	OKC	0.409204
BOS	0.452864	NYK	0.407321

- Memphis Grizzlies è il miglior team per punti al minuto.
- New York Knicks risulta il peggior team per punti al minuto.



# Data Manipulation

## Analisi delle Performance: UFuncs vs For Loop

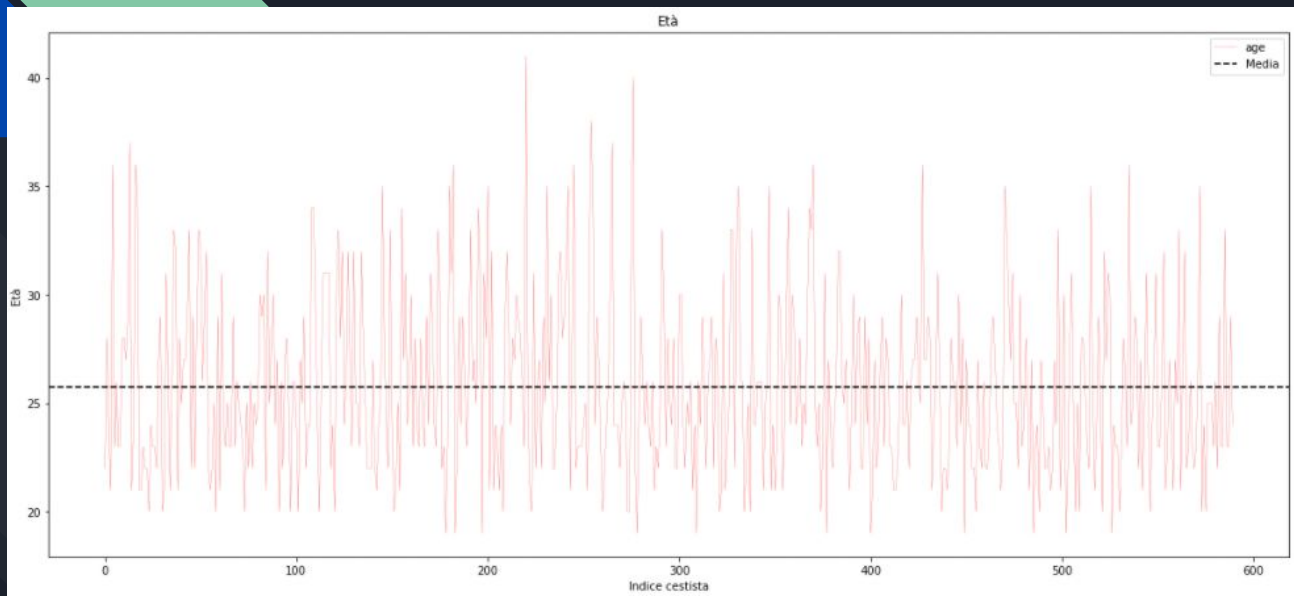


- L'analisi delle performance è effettuato tramite il calcolo dei punti al minuto (punti per partita / minuti giocati per partita).
- UFuncs: 54  $\mu$ s  $\pm$  385 ns per loop
- For Loop: 17.6 ms  $\pm$  54.7  $\mu$ s per loop



# Data Visualization

Grafico età cestisti in base al rank

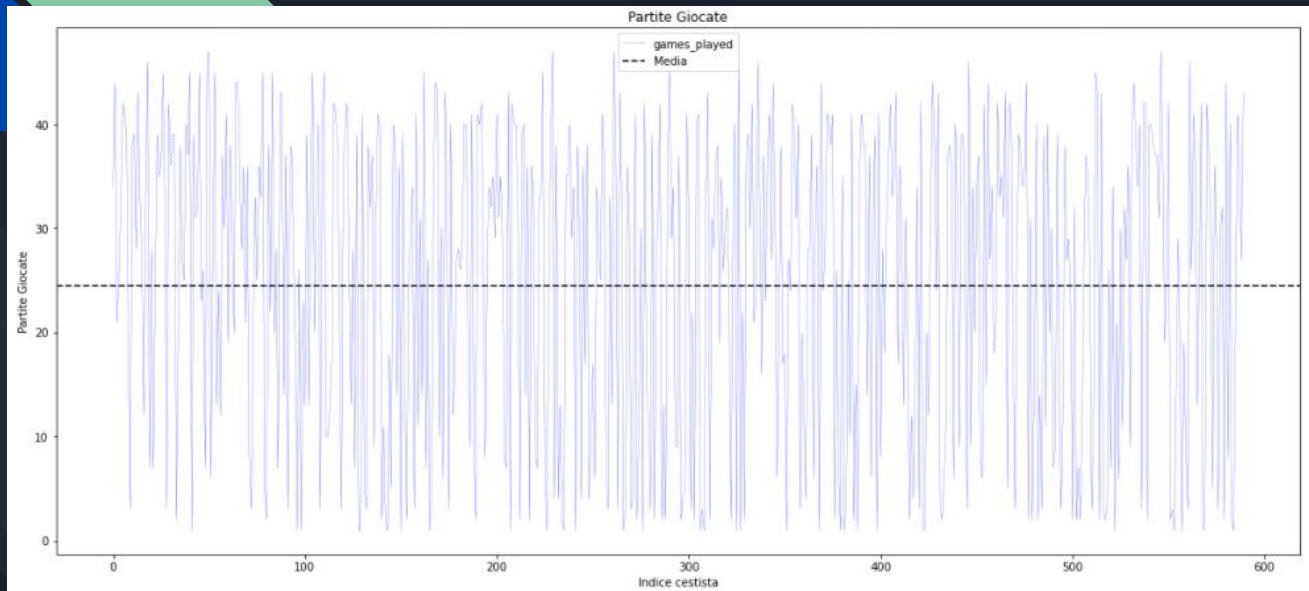






# Data Visualization

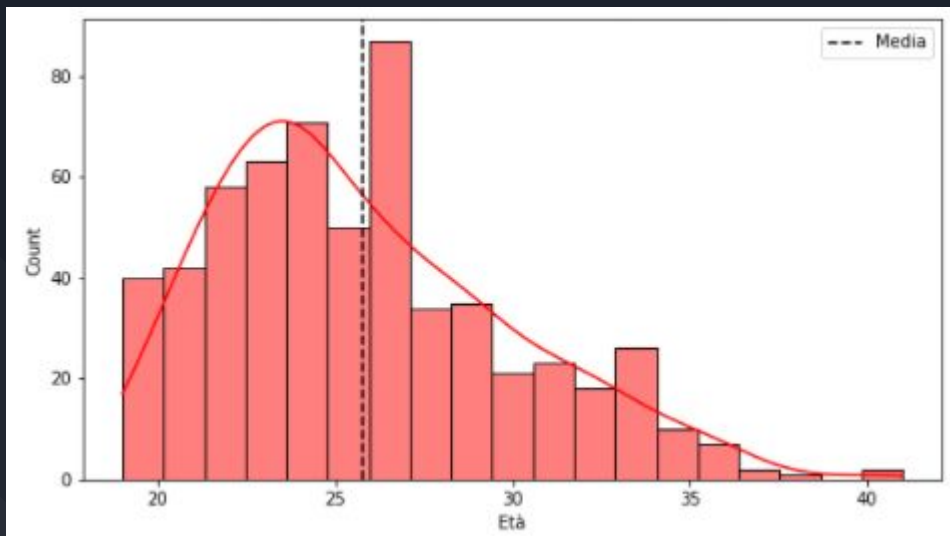
Grafico delle partite giocate dai cestisti in base al rank





# Data Visualization

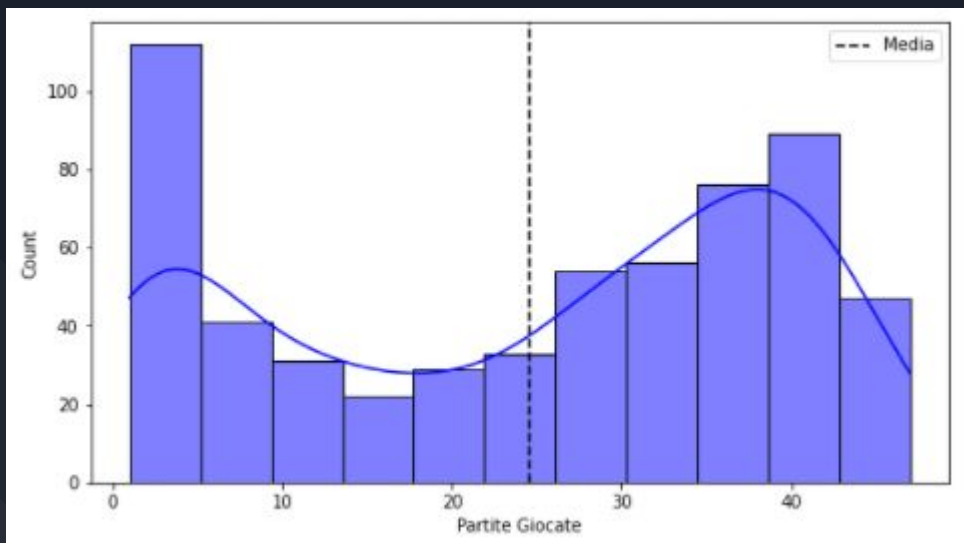
Distribuzione età dei cestisti: la media si attesta intorno ai 26 anni





# Data Visualization

Distribuzione delle partite giocate dai cestisti: la media si attesta intorno alle 25 partite

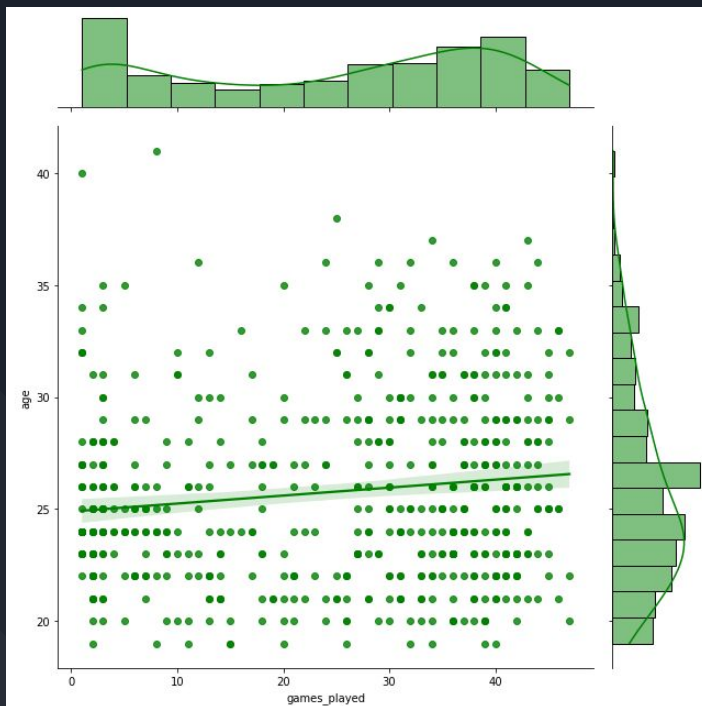




# Data Visualization



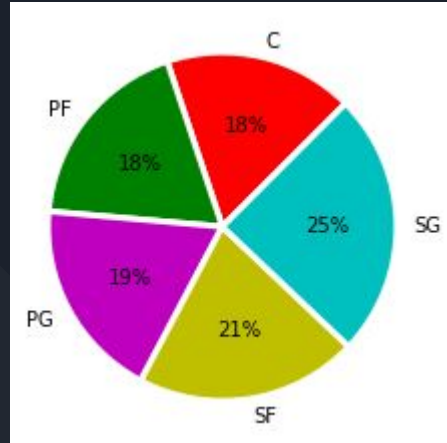
Grafico di densità età - partite giocate





# Data Visualization

## Confronto delle posizioni

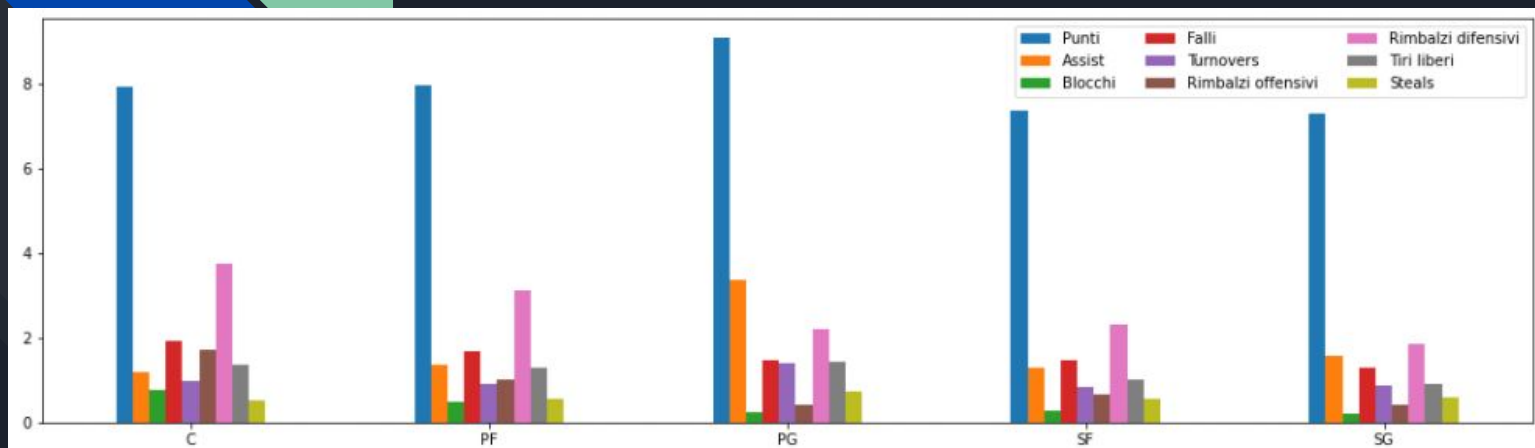


- C: Centro → 18%
- PF: Ala grande → 18%
- SF: Ala piccola → 21%
- SG: Guardia Tiratrice → 25%
- PG: Playmaker → 19%



# Data Visualization

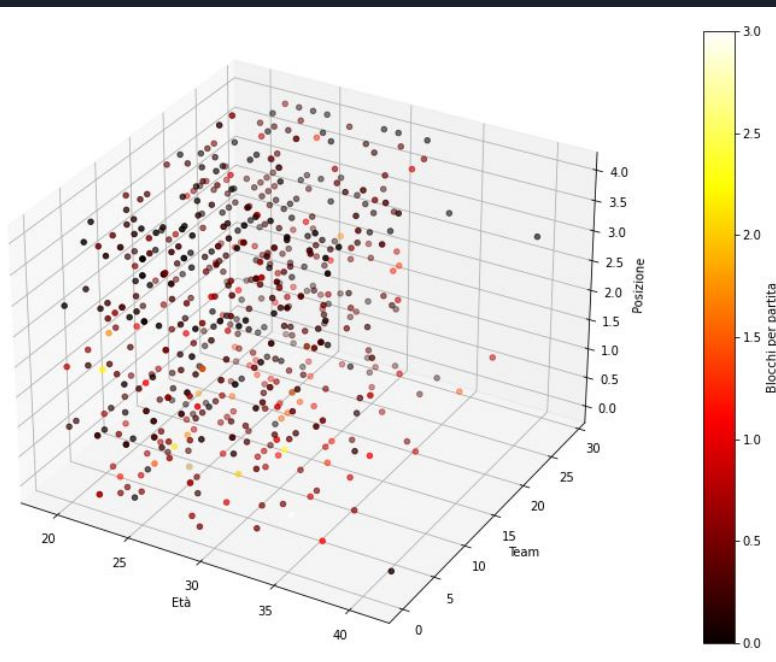
Visualizzazione divisa per ruolo di: Punti, Assist, Blocchi, Falli, Turnovers, Rimbalzi offensivi e difensivi, Tiri liberi e Steals. Tutti i valori sono riferiti per partita giocata.





# Data Visualization

## Visualizzazione in 4d dei Blocchi per Partita



### Valori posizioni:

- 0 → Centro
- 1 → Ala piccola
- 2 → Ala grande
- 3 → Guardia Tiratrice
- 4 → Playmaker

Dal grafico risulta che il ruolo che effettua maggiori blocchi è il **Centro**.



# Machine Learning

- Analisi delle correlazioni delle funzionalità e apprendimento automatico sui cestisti che hanno realizzato almeno un punto.
- Per l'analisi delle correlazioni tutti i valori non numerici sono rimpiazzati con numeri da 0 fino al massimo del valore univoco della colonna.
  - Posizione: stringhe trasformate in 5 valori numerici unique
  - Team: stringhe trasformate in 30 valori numerici unique
- Principali correlazioni:

tot_goal_attempt	field_goal_attempts_per_game	0.999989
points_per_game	field_goals_per_game	0.991264
free_throws_per_game	free_throw_attempts_per_game	0.985976
tot_goal_attempt	points_per_game	0.977573
points_per_game	field_goal_attempts_per_game	0.977566
2-point_field_goal_attempts_per_game	2-point_field_goals_per_game	0.976163
3-point_field_goals_per_game	3-point_field_goal_attempts_per_game	0.972948
defensive_rebounds_per_game	total_rebounds_per_game	0.972350
tot_goal_attempt	field_goals_per_game	0.971526
field_goal_attempts_per_game	field_goals_per_game	0.971504
fiel_goal_percentage	effective_field_goal_percentage	0.960202
2-point_field_goal_attempts_per_game	field_goals_per_game	0.944854
field_goals_per_game	2-point_field_goals_per_game	0.937739
points_per_game	total_points	0.933787





# Machine Learning

- Machine learning sul totale dei punti segnati dai cestisti con il ruolo di Playmaker
- Il dataset è filtrato sui Playmakers che hanno fatto almeno un punto: Shape (106, 8)
- Utilizzo del modello KNeighborsRegressor
- Input:
  - Età
  - Partite giocate
  - Minuti giocati per partita
  - Punti per partita
  - Totale tentativi di canestro
  - Percentuale effettiva di canestri dal campo
  - Percentuale tiri liberi
- Target:
  - Totale punti segnati



# Machine Learning

## Configurazione

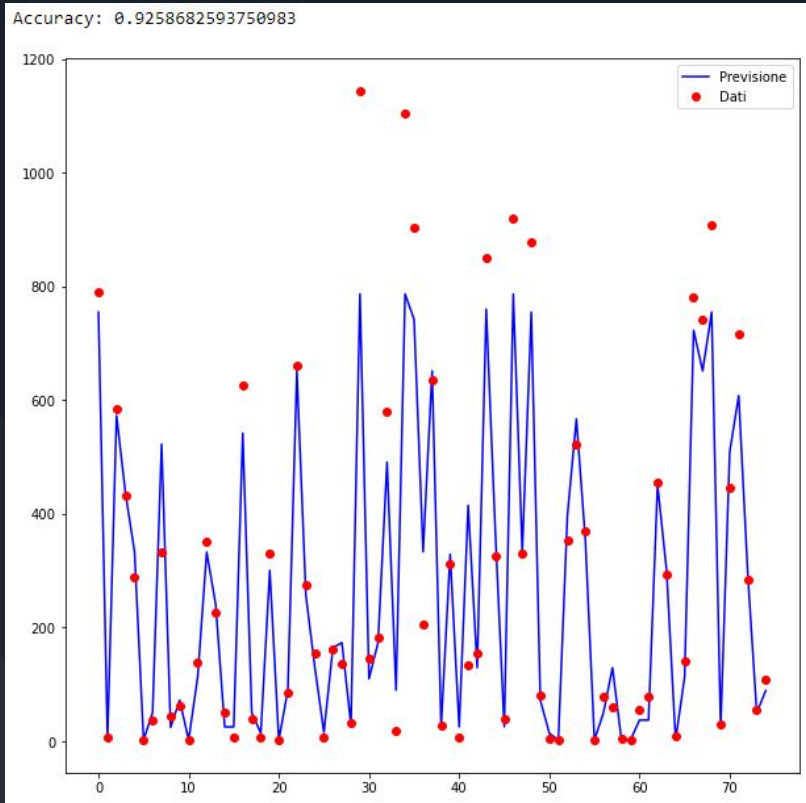
- n° neighbors: (1-5)
- train size: 0.3
- random state: 1

## Risultati Accuracy Score:

- n\_neighbors = 1  $\rightarrow$  0.8888182776662534  $\sim$  0.89
- n\_neighbors = 2  $\rightarrow$  0.903420954874883  $\sim$  0.90
- n\_neighbors = 3  $\rightarrow$  0.9258682593750983  $\sim$  0.93
- n\_neighbors = 4  $\rightarrow$  0.9024695355808362  $\sim$  0.90
- n\_neighbors = 5  $\rightarrow$  0.8834166473932522  $\sim$  0.88



# Machine Learning



Miglior Configurazione

- n° neighbors: 3
- train size: 0.3
- random state: 1

Accuracy

- 0.93



NBA - Season 2021/2022



Grazie  
per l'attenzione!