Central Question

Given that we don't know the MVP voting results of the NBA 23-24 Season, can we accurately predict which players could be top 5 MVP candidates for the 2023-2024 season?

Strictly predicted based on stats from the past seasons: 2019-2020, 2020-2021, 2021-2022, and 2022-2023 and its correlation to corresponding season's MVP voting results

The player stats for the current 2023-2024 season will be utilized to predict the current season's MVP candidates using the trained data with historic stats.

Data Collection

- 1) Downloaded 2019, 2020, 2021, and 2022 seasons player statistics
  - Focused primarily on PPG, APG, RPG, BPG, SPG for machine learning purposes
  - Merged 4 season's into one cohesive dataset.
    - Added season numbers after name = every cell is different data
      - eg) Steven Adams "19 20"
- 2) Created a new dataset from scratch to include NBA MVP voting results from the NBA website
  - 2019, 2020, 2021, and 2022 season's MVP voting results (Lists top players from each season)
  - Merged 4 season's into one cohesive dataset
    - Added season numbers after name = every cell is different data
      - eg) Joel Embiid "20 21"
- 3) Downloaded 2023-2024 seasons player statistics
  - To predict the 2023-2024 season's MVP voting results using machine learning trained with historic data
- 4) Dropped "1st Place Votes", "2nd Place Votes", "3rd Place Votes", "Total Points" in the when merging the datasets
  - Will consider **correlate rank** of each player

Total_Player_Stats.csv

Player	PPG	APG	RPG	SPG	BPG	G
Steven Adams 19 20	10.9	2.3	9.3	0.8	1.1	63
Bam Adebayo 19 20	15.9	5.1	10.2	1.1	1.3	72
LaMarcus Aldridge 1	18.9	2.4	7.4	0.7	1.6	53
Kyle Alexander 19 2	1	0	1.5	0	0	2
Nickeil Alexander-Wa	5.7	1.9	1.8	0.4	0.2	47
Grayson Allen 19 20	8.7	1.4	2.2	0.3	0.1	38
Jarrett Allen 19 20	11.1	1.6	9.6	0.6	1.3	70
Kadeem Allen 19 20	5	2.1	0.9	0.5	0.2	10
Al-Farouq Aminu 19	4.3	1.2	4.8	1	0.4	18
Justin Anderson 19 2	2.8	0.8	2.1	0	0.6	10
Kyle Anderson 19 20	5.8	2.4	4.3	0.8	0.6	67
Ryan Anderson 19 2(	2.5	1	3.5	0.5	0	2
Giannis Antetokounn	29.5	5.6	13.6	1	1	63
Kostas Antetokounmį	1.4	0.4	0.6	0	0	5

Total_MVP_Voting_Results.csv

Player	1st Place Votes	2nd Place Votes	3rd Place Votes	Total Points	Rank
Giannis Antetokounn	85	16	0	962	1
LeBron James 19 20	16	84	1	753	2
James Harden 19 20	0	1	64	367	3
Luka Doncic 19 20	0	0	14	200	4
Kawhi Leonard 19 2(	0	0	9	168	5
Anthony Davis 19 20	0	0	5	82	6
Chris Paul 19 20	0	0	3	26	7
Damian Lillard 19 2(	0	0	1	23	8
Nikola Jokic 19 20	0	0	2	18	9
Pascal Siakam 19 2(	0	0	2	17	10
Jimmy Butler 19 20	0	0	0	9	11
Jayson Tatum 19 20	0	0	0	1	12
Nikola Jokic 20 21	91	8	1	971	1
Joel Embiid 20 21	1	62	23	586	2

"Method" for Decision Tree Classifier

- **Created Decision Tree Classifier:** for 23-24 MVP candidate prediction on 23-24 dataset of player stats
  - X: training features, Y: target labels
    - Training data: PPG, APG, RPG, BPG, SPG
    - Uses MVP voting results for labels
    - Trained using the historic 4 years of player statistics
    - Tested using 23-24 (current) season's player data
- Prediction: Return players and their corresponding stats (full row) where the predicted MVP Voting Rank is < 6

```python
stats_newseason = pd.read_csv('Player_Stats_23_24.csv', header = 0)
'Player Stats for 23–24 season'
```

```python
X_new_season = []
# extract X
for index, row in stats_newseason.iterrows():
    X_new_season.append(row['PPG':].tolist())
```

```python
from sklearn.tree import DecisionTreeClassifier
# from sklearn.model_selection import cross_val_score

dtree = DecisionTreeClassifier(criterion="entropy", random_state=110)
dtree.fit(X, Y)
prediction = dtree.predict(X_new_season)
```

```python
import numpy as np
np.where(prediction < 6)
stats_newseason.iloc[list(np.where(prediction < 6))[0]]
```

"Result" for Decision Tree Classifier

- **Results of Decision tree Classifier :**
  - 9 players predicted to be classified as Top Players within Rank 1-5 for current season (2023-2024 NBA)
    - Rank refers to Top 1 - 5 predicted MVP voting players
- **Comparison with current ongoing vote:** Top 4 voted MVP candidates for 23-24 MVP
  - Result shows 1 corresponding players out of the top 4 candidates
    - "Giannis Antetokounmpo"

	Player	PPG	APG	RPG	SPG	BPG	G
11	Giannis Antetokounmpo 23 24	30.9	5.7	11.3	1.3	1.2	32
115	Anthony Davis 23 24	25.0	3.3	12.3	1.2	2.6	32
131	Kevin Durant 23 24	29.9	6.0	6.3	0.9	1.1	28
137	Joel Embiid 23 24	35.0	6.0	11.7	1.2	2.0	25
200	Tyler Herro 23 24	23.4	4.5	5.5	1.3	0.1	15
231	LeBron James 23 24	25.4	7.4	7.4	1.4	0.7	31
275	Damian Lillard 23 24	25.5	6.8	4.5	1.0	0.2	31
438	Anfernee Simons 23 24	27.1	5.3	3.1	0.9	0.1	11
534	Trae Young 23 24	28.3	11.3	3.0	1.4	0.2	30

Player	1st place	2nd place	3rd place	Points
1. Nikola Jokic	6	2	0	36
2. Luka Doncic	2	3	2	21
3. Shai Gilgeous-Alexander	0	3	5	14
4. Giannis Antetokounmpo	0	0	1	1

Source:
https://www.sportingnews.com/us/nba/news/sn-nba-awards-2024-voting-results-best-player/122e61e3599c2b7923491d87#MVP

"Method" for K-Nearest Neighbors

- **Performed KNN: Predictions for each value of K to show players that correspond with each K values condition**
  - Initialized accs = empty list to store results
    - Iterates over values 1 - 10 using 'range'
  - X_new_season data: Predictions using trained classifier and stores the predictions in prediction_i
    - It finds the predictions that are **less than 6** and stores them
  - Less than 6 executes the MVP candidate players (Ranked 1-5)
    - Retrieves player names that correspond to these predictions from the DF and stores them

```python
from sklearn.neighbors import KNeighborsClassifier

accs = []
#changing the variables
for i in range(1,11):
  classifier = KNeighborsClassifier(n_neighbors=i).fit(X, Y)
  prediction_i = classifier.predict(X_new_season)

  indices = np.where(prediction_i < 6)[0]

  accs.append(f"KNN with k = {i}: ")
  players = stats_newseason.iloc[indices]['Player'].tolist()
  accs.extend(players)
```

"Result" for K-Nearest Neighbors

- **Result of KNN:** The list contains names of 2023-2024 players predicted at K values 1-10 with its rank less than 6
  - Rank: Players predicted to be 23-24's MVP Rank 1-5
  - KNN indicates different 2023-2024 players that meet the criteria of being predicted as top MVP candidates
    - Top 1-5 voted MVPs
  - Noticeable Player (Listed in many K values):
    - "Joel Embiid"
    - "Giannis Antetokounmpo"
- **Comparison with current ongoing vote:** Top 4 voted MVP candidates for 23-24 MVP
  - Result shows 1 corresponding players out of the top 4 candidates
    - "Giannis Antetokounmpo"

```
['KNN with k = 1: ',
'Joel Embiid 23 24',
'James Harden 23 24',
'Fred VanVleet 23 24',
'KNN with k = 2: ',
'Giannis Antetokounmpo 23 24',
'Luka Doncic 23 24',
'Kevin Durant 23 24',
'Joel Embiid 23 24',
'Shai Gilgeous-Alexander 23 24',
'James Harden 23 24',
'Fred VanVleet 23 24',
'KNN with k = 3: ',
'Giannis Antetokounmpo 23 24',
'Joel Embiid 23 24',
'KNN with k = 4: ',
'Giannis Antetokounmpo 23 24',
'Joel Embiid 23 24',
'KNN with k = 5: ',
'Giannis Antetokounmpo 23 24',
'Joel Embiid 23 24',
'KNN with k = 6: ',
'Giannis Antetokounmpo 23 24',
'Joel Embiid 23 24',
'KNN with k = 7: ',
'Giannis Antetokounmpo 23 24',
'KNN with k = 8: ',
'KNN with k = 9: ',
'KNN with k = 10: ']
```

Player	1st place	2nd place	3rd place	Points
1. Nikola Jokic	6	2	0	36
2. Luka Doncic	2	3	2	21
3. Shai Gilgeous-Alexander	0	3	5	14
4. Giannis Antetokounmpo	0	0	1	1

Source: https://www.sportingnews.com/us/nba/news/sn-nba-awards-2024-voting-results-best-player/122e61e3599c2b7923491d87#MVP

Overall "Conclusion" for the Two ML methods (KNN, Decision Tree Classifier)

- **Both ML methods are accurate** at identifying top MVP candidates
  - All players listed in the predicted ML method had **great individual stats** compared to the remaining players (reasonable prediction)
  - There are corresponding players between the predicted ML model and the actual 23-24 MVP voting result
    - Only based on **individual statistics** and the **correlated rank** from two merged datasets
- But there are **limitations:**
  - 1) Historical data (from 19-20, 20-21, 21-22, 22-23) from different players used to train the ML methods
  - 2) Players are classifices only based on **correlated "rank"** (order of MVP Voting Result)
    - Previously dropped: "1st Place Votes", "2nd Place Votes", "3rd Place Votes", "Total Points" in the when merging the datasets
  - 3) Individual statistics aren't the only standards when selecting the Season MVP
    - Team's statistics, player's popularity, and individual's storyline that impacted the NBA's overall popularity are also considered

Fun Fact

The bird on the previous Twitter logo is named after Larry Bird of the Boston Celtics

It was chosen to honor him because the Co-Founder of Twitter, Biz Stone, is a beloved fan of the Boston Celtics

Larry Bird was a 3 time NBA MVP