



# **NBA Win Prediction**

## **With Machine Learning Models**

7 May 2021



# Table Of Contents

**01**

Introduction

**02**

Methodology

**03**

Process Flow

**04**

Results

**05**

Conclusion





**01**

# Introduction

Problem Statement & Definition



## Problem Statement

As the manager of the basketball team in NBA, I want to understand and know how my team players are performing and predict if the team have the chance of winning the next game at our familiar home grounds.



# Problem Definition



## Goal

Ability to predict if the team win or lose a game.



## Classification Problem

Predict if the team falls in the winning or losing category.



## Target Audience

Anyone who manage the basketball team



## Outcome

To see whether to adjust training routines of players to better perform in the games.



“Talent win games, but teamwork and intelligence wins championships.”

**–Michael Jordan**



# 02

# Methodology

Datasets, Models, Metrics, Tools



# Datasets

EDA done for all csv, but will focus on players.csv and games\_details.csv  
Machine Learning will be using data from games.csv to predict win/lose.

**24,146**

Games

**7,228**

Players

**30**

Teams

**602,767**

Games Details

**191,112**

Ranking Records

Source: (KAGGLE) <https://www.kaggle.com/nathanlauga/nba-games>





# Machine Learning For NBA Games



## ML Models

K-Nearest Neighbor (KNN)  
Naïve Bayes  
Logistic Regression



## Metrics

Precision  
Recall  
F1 Score



## Tools

Google Colab  
Numpy  
Pandas  
Matplotlib  
Seaborn  
Sklearn



**03**

# Process Workflow

EDA, Data preparation, Data analysis,  
ML model training/evaluation



# Data Preparation & Transformation

## Data Cleaning

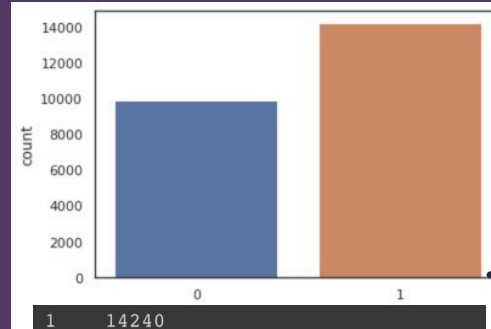
- Checking for NULL values and dropping rows
- Dropping of unused columns

```
GAME_ID      0
HOME_TEAM_ID  0
VISITOR_TEAM_ID  0
SEASON      0
HOME_TEAM_WINS  0
TEAM_ID_homeTeam  0
SEASON_ID    0
TEAM_homeTeam  0
G_homeTeam    600
W_homeTeam    600
L_homeTeam    600
W_PCT_homeTeam  600
TEAM_ID_awayTeam  0
SEASON_ID_awayTeam  0
TEAM_awayTeam  0
G_awayTeam    564
W_awayTeam    564
L_awayTeam    564
W_PCT_awayTeam  564
dtype: int64
df without nans size: 23479
```

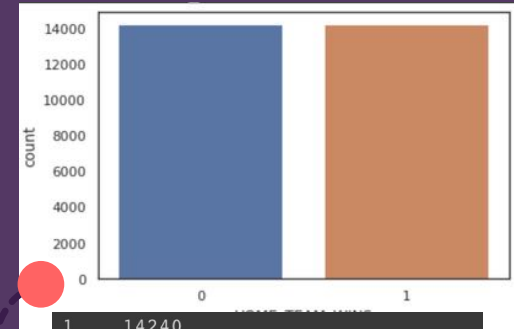
```
GAME_ID      0
HOME_TEAM_ID  0
VISITOR_TEAM_ID  0
SEASON      0
HOME_TEAM_WINS  0
dtype: int64
```

## Balancing Target Variable

- Balancing the records ensure more accurate prediction.



```
1    14240
0     9906
Name: HOME_TEAM_WINS, dtype: int64
```



```
1    14240
0    14240
Name: HOME_TEAM_WINS, dtype: int64
```



# Exploratory Data Analysis

Performed initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.





# Did You Know This?



```
GAME_ID          24146
HOME_TEAM_ID      30
VISITOR_TEAM_ID   30
SEASON            18
HOME_TEAM_WINS     2
dtype: int64
```



18 seasons of NBA Games.



Total of 24146 Games completed so far.



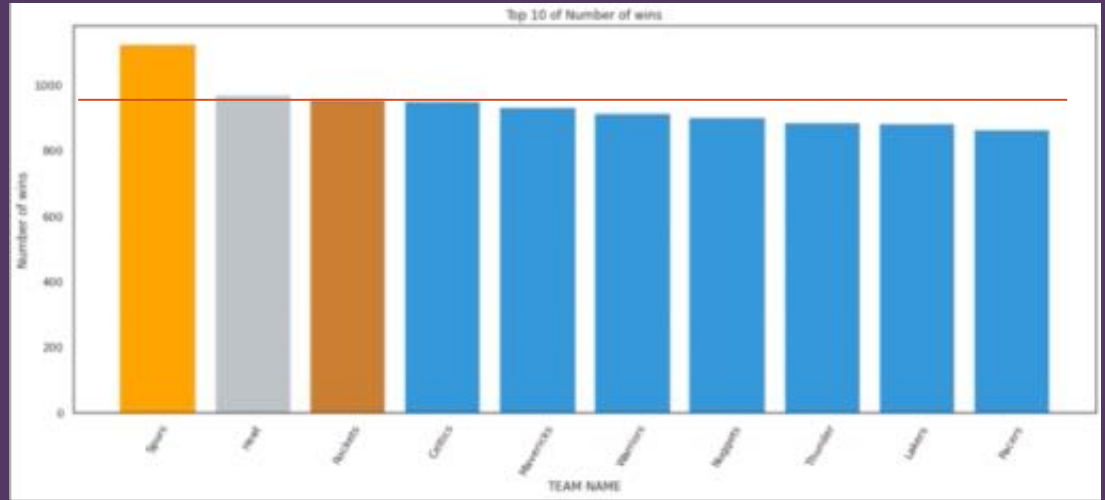
LeBron James played in 1689 games, most in history!

# Team That Won The Most Games Since 2004.



**1126 Games WON!**

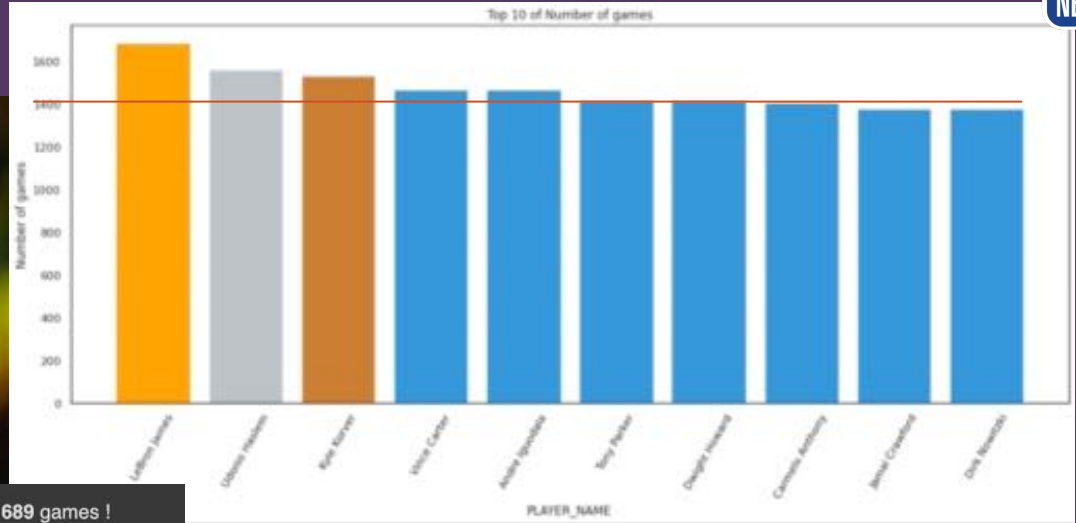
Spurs won the most games in history!



Warriors in the 6<sup>th</sup>, Lakers in the 9<sup>th</sup>.



# Most Games Played By Who?



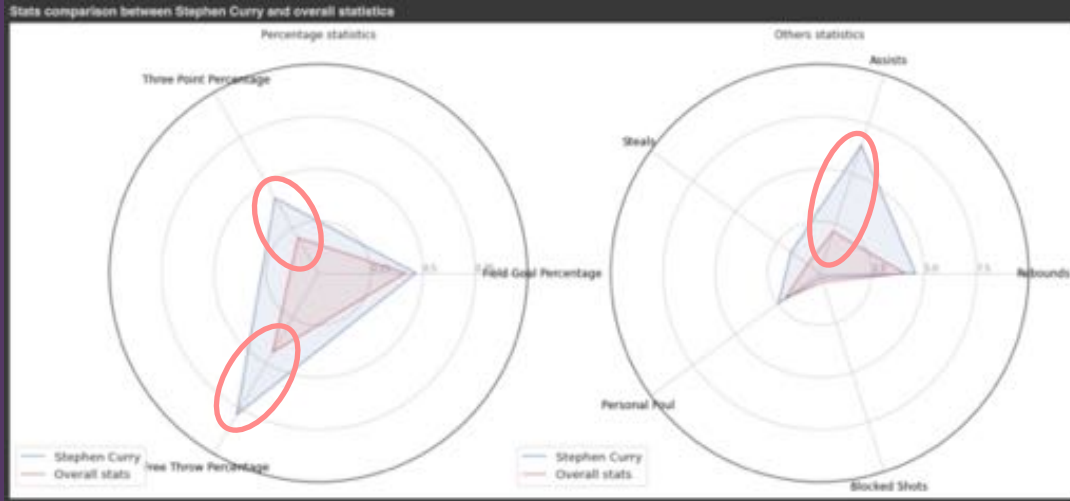
LeBron James played 1689 games !

	PLAYER_NAME	Number of games
0	LeBron James	1689
1	Udonis Haslem	1561
2	Kyle Korver	1534
3	Vince Carter	1470
4	Andre Iguodala	1468

LeBron James played **1689** games!



# Stats Of Stephen Curry



Curry is quite strong in most areas compared to the average stats of all others players

Curry played **947** games for **Warriors** which won **914** games to date.



# Legend Kobe Bryant Vs Stephen Curry



Curry is a better 3-pointer!

Curry is better at assisting fellow team-mates

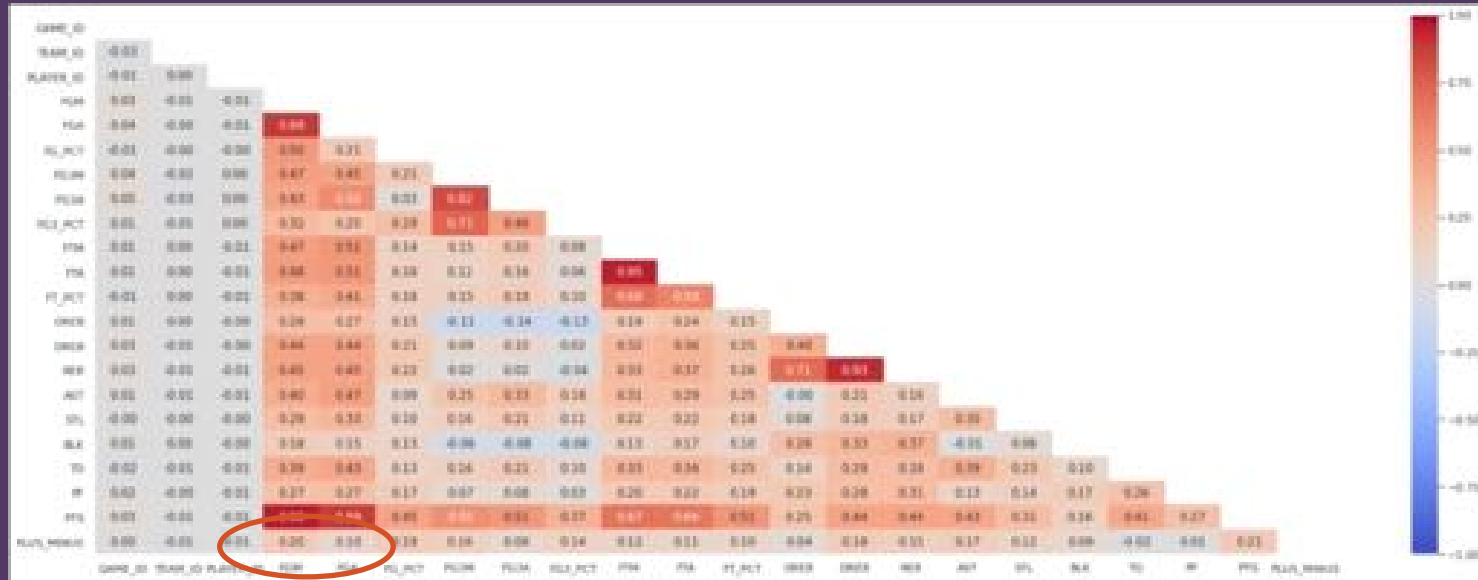
Kobe is better at rebounds



Kobe played **1102** games for **Lakers** which won **883** games to date.



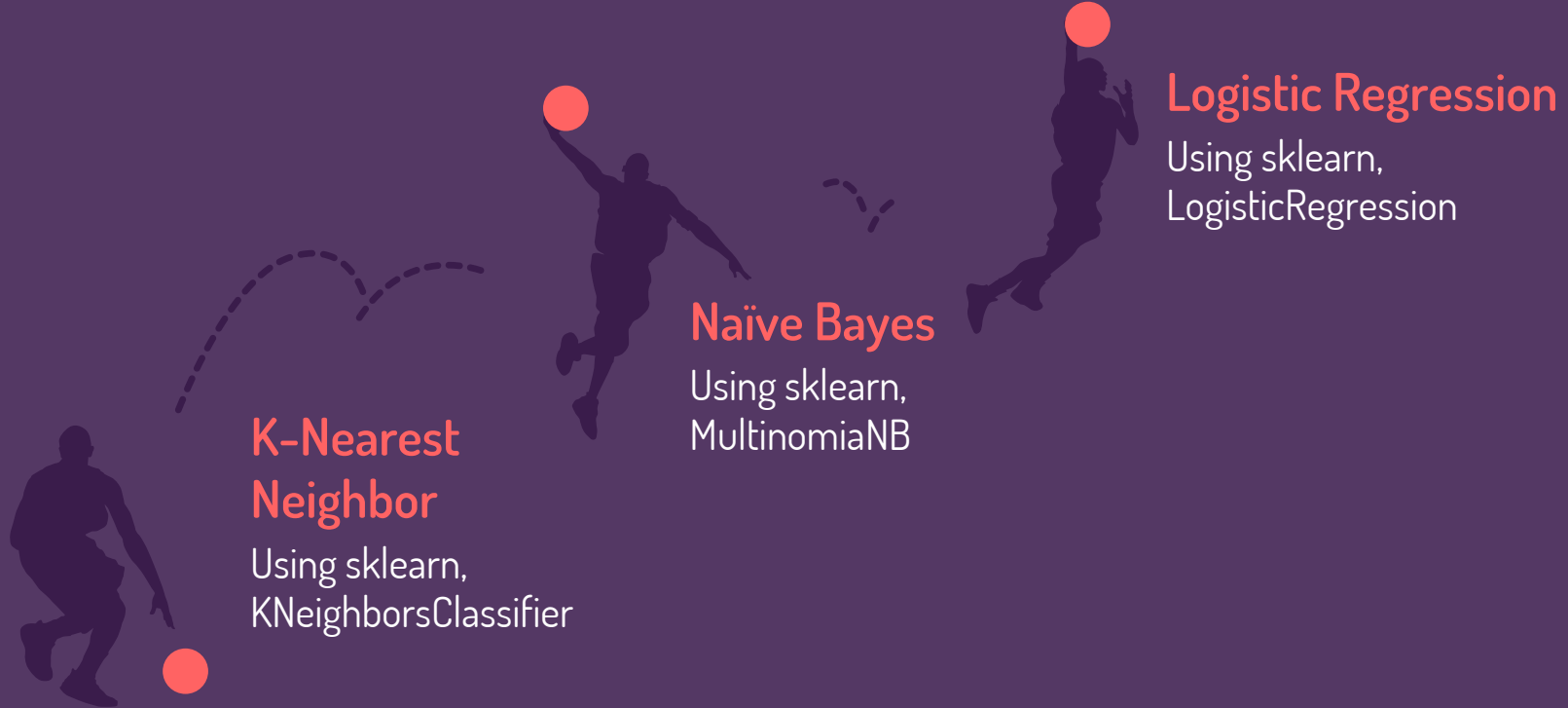
# Correlation Of Player Stats



The highest correlation for player to score points(PTS) with Field Goal Made (FGM: 0.96) and Field Goal Attempted(FGA: 0.88) . This means that for the team with total points scoring of FGA and FGM per player will influence the win prediction.



# ML Model Training/Evaluation





# Hyperparameter Optimization

Model	Validation Method	Scores
K-Nearest Neighbor(KNN)	Best/Optimal K = 1	Accuracy = 62.43%' with 'K = 1' F1-score = 62.31%' with 'K = 1'
Naïve Bayes	K-fold cross validation CV = 5	F1-score = 47.45% Alpha = 1
Naïve Bayes	K-fold cross validation CV = 10	F1-score = 47.45% Alpha = 1
Naïve Bayes	Randomized K-fold cross validation.	F1-score at: 47.45% Alpha = 0.1837
Logistic Regression	K-fold cross validation	Mean F1-score= 47.0% Standard deviation = 0.01



# 04

## Results+

Which Model Produce Better Prediction?



# ML Using K-Nearest Neighbor (KNN)

## Best Estimated Accuracy and F1 Score

Results for accuracy in F1 score is about 55%

Classification report:

	precision	recall	f1-score	support
0	0.54	0.70	0.61	2848
1	0.57	0.40	0.47	2848
accuracy			0.55	5696
macro avg	0.56	0.55	0.54	5696
weighted avg	0.56	0.55	0.54	5696





# ML Using Naïve Bayes



## Best Estimated Accuracy and F1 Score

Results for accuracy in F1 score is less than 50%

Classification report:

	precision	recall	f1-score	support
0	0.50	0.71	0.59	2848
1	0.49	0.28	0.35	2848
accuracy			0.49	5696
macro avg	0.49	0.49	0.47	5696
weighted avg	0.49	0.49	0.47	5696

# ML Using Logistic Regression



## Best Estimated Accuracy and F1 Score

Results for accuracy in F1 score is less than 50%

Somehow the classification report for Logistic Regression is similar to that for Naïve Bayes.

### Classification report:

	precision	recall	f1-score	support
0	0.50	0.71	0.59	2848
1	0.49	0.28	0.35	2848
accuracy			0.49	5696
macro avg	0.49	0.49	0.47	5696
weighted avg	0.49	0.49	0.47	5696





# Which ML Model Better?

**55%**

Accuracy

## K-Nearest Neighbor(KNN)

Is a better model to use for training and prediction as it has a higher accuracy compare to the other model.

Desired Output (Actuals)	Predicted Output	
3939	1	1
1212	1	1
9611	1	0
21570	1	0
3964	1	1
25897	0	0
24554	0	1
28080	0	0
15661	1	1
5980	1	1

**54%**

Macro Average F1-Score

**55%**

Macro Average Precision

**56%**

Micro Average Recall



**05**

# Conclusion





# Conclusion

Yes, although the KNN training model do help to predict if the game is a win or lose, however, we cannot really depend on the predicted results to help with the team performance as the accuracy is actually quite low, at 55%.

	Desired Output (Actuals)	Predicted Output
9611	1	0
21570	1	0
24554	0	1
20461	1	0
20115	1	0
16842	0	1
28458	0	1
1353	0	1
9405	1	0
11368	1	0

There is a 45%  
chance that the  
prediction is wrong!

```
Prediction: 0  
My prediction is a LOSE.
```

```
This was the input data:  
GAME_ID          20400657  
HOME_TEAM_ID     1610612760  
VISITOR_TEAM_ID  1610612759  
HOME_TEAM_WINS   0  
Name: 2085, dtype: int64
```

```
Prediction: 1  
My prediction is a WIN.
```

```
This was the input data:  
GAME_ID          21700386  
HOME_TEAM_ID     1610612765  
VISITOR_TEAM_ID  1610612738  
HOME_TEAM_WINS   0  
Name: 19928, dtype: int64
```

Error Prediction???



# Future Opportunities



- Machine learning training with other models, etc. Random Forest or Decision Tree
- Fine tune feature selection: Predict potential attribute of players that helps with the classification performance for win/lose.
- Improve fine tuning and optimization with hyperparameter may improve the accuracy and F1-score.



**Questions?**