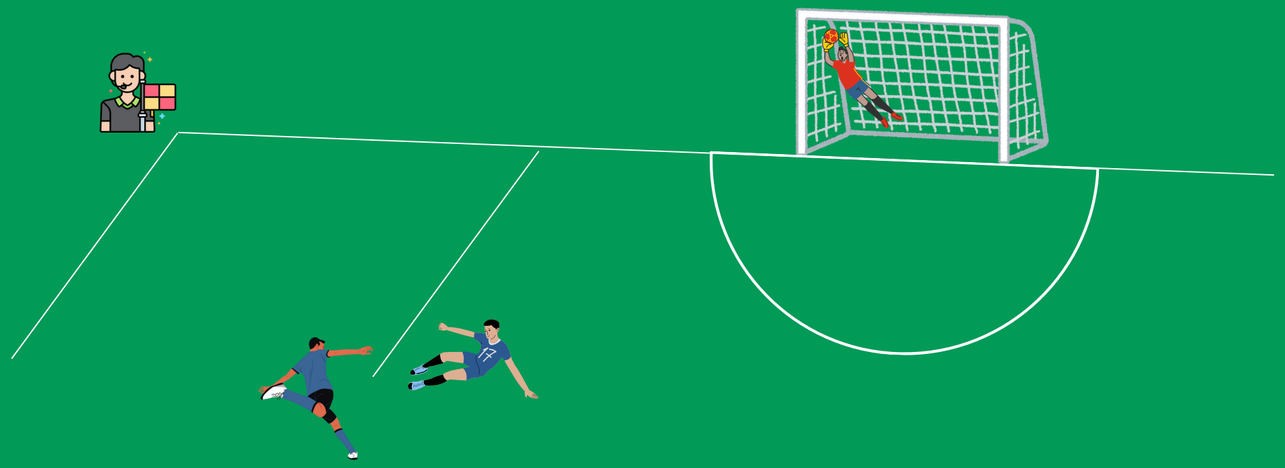
# Exploratory Data Analysis (EDA) on Premier League Data From 2017/18 Using Python




**Group 13**

**22224516 Gurung Yash**

**22224014 Khatri Priya**

**22224040 Busal Sagar**

# Table of Contents

# Libraries used in this Data Analysis

Pandas

Matplotlib



Scipy

# Data Source


***FotMob is a popular mobile app for football (soccer) fans, providing live scores, match stats, news, and updates for teams and leagues worldwide. It offers in-depth match analyses, and real-time commentary, making it a go-to app for following football events.***

```
[2]  import requests
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns

     test_url = 'https://www.fotmob.com/api/leagues?id=47&ccode3=JPN&season=2017%2F2018'

     data_1 = requests.get(url = test_url).json()
```

# The data we extracted


*API JSON Data*

*Pandas Dataframe*

| | Name | PL | W | D | L | GD | PTS | Goals_Scored | Goals_Conceded | Draws_Home | Goals_Scored_Home | Goals_Conceded_Home | Draws_Away | Goals_Scored_Away | Goals_Conceded_Away |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Man City | 38 | 32 | 4 | 2 | 79 | 100 | 106 | 27 | 2 | 61 | 14 | 2 | 45 | 13 |
| 1 | Man United | 38 | 25 | 6 | 7 | 40 | 81 | 68 | 28 | 2 | 38 | 9 | 4 | 30 | 19 |
| 2 | Tottenham | 38 | 23 | 8 | 7 | 38 | 77 | 74 | 36 | 4 | 40 | 16 | 4 | 34 | 20 |
| 3 | Liverpool | 38 | 21 | 12 | 5 | 46 | 75 | 84 | 38 | 7 | 45 | 10 | 5 | 39 | 28 |
| 4 | Chelsea | 38 | 21 | 7 | 10 | 24 | 70 | 62 | 38 | 4 | 30 | 16 | 3 | 32 | 22 |
| 5 | Arsenal | 38 | 19 | 6 | 13 | 23 | 63 | 74 | 51 | 2 | 54 | 20 | 4 | 20 | 31 |
| 6 | Burnley | 38 | 14 | 12 | 12 | -3 | 54 | 36 | 39 | 5 | 16 | 17 | 7 | 20 | 22 |
| 7 | Everton | 38 | 13 | 10 | 15 | -14 | 49 | 44 | 58 | 4 | 28 | 22 | 6 | 16 | 36 |
| 8 | Leicester | 38 | 12 | 11 | 15 | -4 | 47 | 56 | 60 | 6 | 25 | 22 | 5 | 31 | 38 |
| 9 | Newcastle | 38 | 12 | 8 | 18 | -8 | 44 | 39 | 47 | 4 | 21 | 17 | 4 | 18 | 30 |
| 10 | Crystal Palace | 38 | 11 | 11 | 16 | -10 | 44 | 45 | 55 | 5 | 29 | 27 | 6 | 16 | 28 |
| 11 | Bournemouth | 38 | 11 | 11 | 16 | -16 | 44 | 45 | 61 | 5 | 26 | 30 | 6 | 19 | 31 |
| 12 | West Ham | 38 | 10 | 12 | 16 | -20 | 42 | 48 | 68 | 6 | 24 | 26 | 6 | 24 | 42 |
| 13 | Watford | 38 | 11 | 8 | 19 | -20 | 41 | 44 | 64 | 6 | 27 | 31 | 2 | 17 | 33 |
| 14 | Brighton | 38 | 9 | 13 | 16 | -20 | 40 | 34 | 54 | 8 | 24 | 25 | 5 | 10 | 29 |
| 15 | Huddersfield | 38 | 9 | 10 | 19 | -30 | 37 | 28 | 58 | 5 | 16 | 25 | 5 | 12 | 33 |
| 16 | Southampton | 38 | 7 | 15 | 16 | -19 | 36 | 37 | 56 | 7 | 20 | 26 | 8 | 17 | 30 |
| 17 | Swansea | 38 | 8 | 9 | 21 | -28 | 33 | 28 | 56 | 3 | 17 | 24 | 6 | 11 | 32 |
| 18 | Stoke | 38 | 7 | 12 | 19 | -33 | 33 | 35 | 68 | 5 | 20 | 30 | 7 | 15 | 38 |
| 19 | West Brom | 38 | 6 | 13 | 19 | -25 | 31 | 31 | 56 | 9 | 21 | 29 | 4 | 10 | 27 |

# Research Questions & Hypothesis

## Question 1:

**Do top five teams have better attacking and defence than the bottom five teams?**

**Hypothesis 1: Top teams have better attacking and defending than bottom teams.**

## Question 2:

**Do teams perform better at home than away?**

**Hypothesis 2: Teams tend to perform better at home than away.**

## Question 3:

**Does the location of the game (home vs. away) influence the likelihood of a draw?**

**Hypothesis 3: The location of the game does not significantly affect the likelihood of a draw.**

# Data Preparation

| | name | shortName | id | pageUrl | deduction | ongoing | played | wins | draws | losses | scoresStr | goalConDiff | pts | idx | qualColor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Manchester City | Man City | 8456 | /teams/8456/overview/manchester-city | None | None | 38 | 32 | 4 | 2 | 106-27 | 79 | 100 | 1 | #2AD572 |
| 1 | Manchester United | Man United | 10260 | /teams/10260/overview/manchester-united | None | None | 38 | 25 | 6 | 7 | 68-28 | 40 | 81 | 2 | #2AD572 |
| 2 | Tottenham Hotspur | Tottenham | 8586 | /teams/8586/overview/tottenham-hotspur | None | None | 38 | 23 | 8 | 7 | 74-36 | 38 | 77 | 3 | #2AD572 |
| 3 | Liverpool | Liverpool | 8650 | /teams/8650/overview/liverpool | None | None | 38 | 21 | 12 | 5 | 84-38 | 46 | 75 | 4 | #2AD572 |
| 4 | Chelsea | Chelsea | 8455 | /teams/8455/overview/chelsea | None | None | 38 | 21 | 7 | 10 | 62-38 | 24 | 70 | 5 | #0046A7 |

```
df_2 = pd.DataFrame(data_1['table'][0]['data']['table']['home'])
df_2.head()
```

| | name | shortName | id | pageUrl | deduction | ongoing | played | wins | draws | losses | scoresStr | goalConDiff | pts | idx | qualColor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Manchester City | Man City | 8456 | /teams/8456/overview/manchester-city | None | None | 19 | 16 | 2 | 1 | 61-14 | 47 | 50 | 1 | #2AD572 |
| 1 | Arsenal | Arsenal | 9825 | /teams/9825/overview/arsenal | None | None | 19 | 15 | 2 | 2 | 54-20 | 34 | 47 | 2 | #2AD572 |
| 2 | Manchester United | Man United | 10260 | /teams/10260/overview/manchester-united | None | None | 19 | 15 | 2 | 2 | 38-9 | 29 | 47 | 3 | #2AD572 |
| 3 | Liverpool | Liverpool | 8650 | /teams/8650/overview/liverpool | None | None | 19 | 12 | 7 | 0 | 45-10 | 35 | 43 | 4 | #2AD572 |
| 4 | Tottenham Hotspur | Tottenham | 8586 | /teams/8586/overview/tottenham-hotspur | None | None | 19 | 13 | 4 | 2 | 40-16 | 24 | 43 | 5 | #0046A7 |

```
df_3 = pd.DataFrame(data_1['table'][0]['data']['table']['away'])
df_3.head()
```

| | name | shortName | id | pageUrl | deduction | ongoing | played | wins | draws | losses | scoresStr | goalConDiff | pts | idx | qualColor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Manchester City | Man City | 8456 | /teams/8456/overview/manchester-city | None | None | 19 | 16 | 2 | 1 | 45-13 | 32 | 50 | 1 | #2AD572 |
| 1 | Tottenham Hotspur | Tottenham | 8586 | /teams/8586/overview/tottenham-hotspur | None | None | 19 | 10 | 4 | 5 | 34-20 | 14 | 34 | 2 | #2AD572 |
| 2 | Manchester United | Man United | 10260 | /teams/10260/overview/manchester-united | None | None | 19 | 10 | 4 | 5 | 30-19 | 11 | 34 | 3 | #2AD572 |
| 3 | Chelsea | Chelsea | 8455 | /teams/8455/overview/chelsea | None | None | 19 | 10 | 3 | 6 | 32-22 | 10 | 33 | 4 | #2AD572 |
| 4 | Liverpool | Liverpool | 8650 | /teams/8650/overview/liverpool | None | None | 19 | 9 | 5 | 5 | 39-28 | 11 | 32 | 5 | #0046A7 |

```
df_1 = pd.DataFrame(data_1['table'][0]['data']['table']['all'])
df_2 = pd.DataFrame(data_1['table'][0]['data']['table']['home'])
df_3 = pd.DataFrame(data_1['table'][0]['data']['table']['away'])

df_1 = df_1.drop(['name', 'id', 'pageUrl', 'deduction', 'ongoing', 'idx', 'qualColor'], axis = 1)
df_2 = df_2.drop(['name', 'id', 'pageUrl', 'deduction', 'ongoing', 'played', 'wins', 'losses', 'goalConDiff', 'pts', 'idx', 'qualColor'], axis = 1)
df_3 = df_3.drop(['name', 'id', 'pageUrl', 'deduction', 'ongoing', 'played', 'wins', 'losses', 'goalConDiff', 'pts', 'idx', 'qualColor'], axis = 1)


df_1.columns = ["Name", "PL", "W", "D", "L", "+/-", "GD", "PTS"]
df_2.columns = ["Name", "Draws_Home", "+/-"]
df_3.columns = ["Name", "Draws_Away", "+/-"]

df_1[['Goals_Scored', 'Goals_Conceded']] = df_1['+/-'].str.split('-', expand=True)

df_1['Goals_Scored'] = df_1['Goals_Scored'].astype(int)
df_1['Goals_Conceded'] = df_1['Goals_Conceded'].astype(int)

df_1 = df_1.drop(columns = ['+/-'])

df_2[['Goals_Scored_Home', 'Goals_Conceded_Home']] = df_2['+/-'].str.split('-', expand=True)

df_2['Goals_Scored_Home'] = df_2['Goals_Scored_Home'].astype(int)
df_2['Goals_Conceded_Home'] = df_2['Goals_Conceded_Home'].astype(int)

df_2 = df_2.drop(columns = ['+/-'])

df_3[['Goals_Scored_Away', 'Goals_Conceded_Away']] = df_3['+/-'].str.split('-', expand=True)

df_3['Goals_Scored_Away'] = df_3['Goals_Scored_Away'].astype(int)
df_3['Goals_Conceded_Away'] = df_3['Goals_Conceded_Away'].astype(int)

df_3 = df_3.drop(columns = ['+/-'])
```

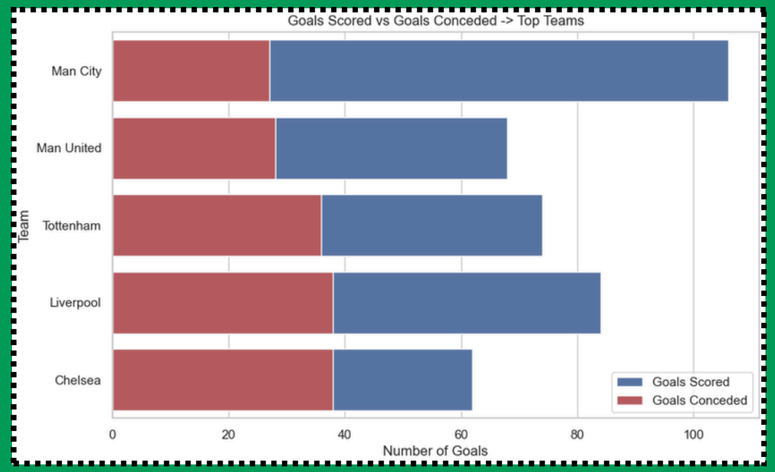
```
merged_data = pd.merge(df_1, df_2, on=['Name'])
merged_data = pd.merge(merged_data, df_3, on=['Name'])
merged_data.head()
```

| | Name | PL | W | D | L | GD | PTS | Goals_Scored | Goals_Conceded | Draws_Home | Goals_Scored_Home | Goals_Conceded_Home | Draws_Away | Goals_Scored_Away | Goals_Conceded_Away |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Man City | 38 | 32 | 4 | 2 | 79 | 100 | 106 | 27 | 2 | 61 | 14 | 2 | 45 | 13 |
| 1 | Man United | 38 | 25 | 6 | 7 | 40 | 81 | 68 | 28 | 2 | 38 | 9 | 4 | 30 | 19 |
| 2 | Tottenham | 38 | 23 | 8 | 7 | 38 | 77 | 74 | 36 | 4 | 40 | 16 | 4 | 34 | 20 |
| 3 | Liverpool | 38 | 21 | 12 | 5 | 46 | 75 | 84 | 38 | 7 | 45 | 10 | 5 | 39 | 28 |
| 4 | Chelsea | 38 | 21 | 7 | 10 | 24 | 70 | 62 | 38 | 4 | 30 | 16 | 3 | 32 | 22 |

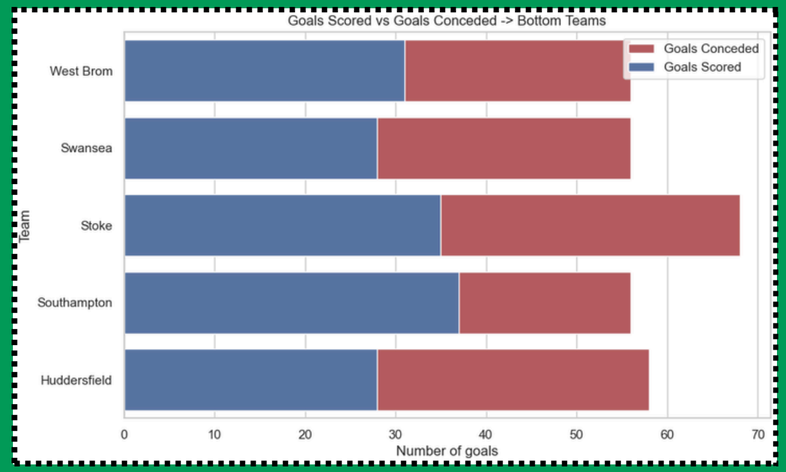
**Final Data**

# Hypothesis 1:

# The top five teams exhibit superior attacking and defensive capabilities compared to the bottom five teams.




Top five teams



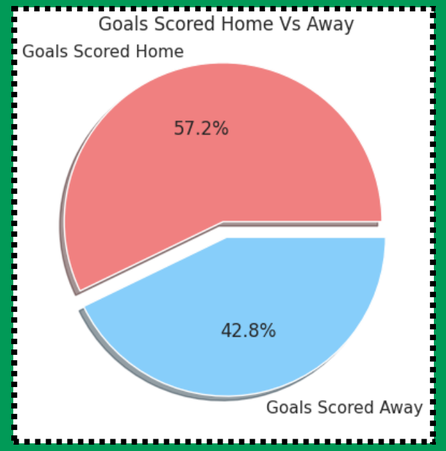

Bottom five teams

## Statistical Testing:

1. ***T-test correlation coefficient (Goals Scored, Top vs. Bottom)*: 5.931809596103181**
2. ***P-value*: 0.00034902852424575 35**

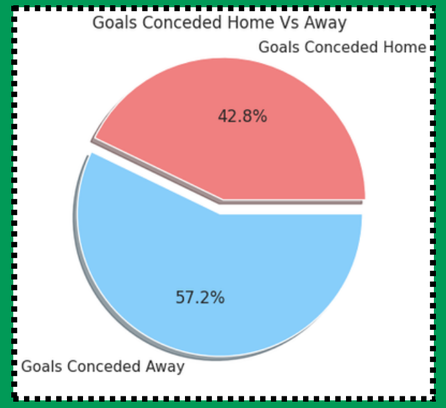
1. **T-test correlation coefficient (Goals Conceded, Top vs. Bottom): -7.522830747540472**
2. **P-value: 6.77976432014433e-05**

# Hypothesis 2:
# Teams tend to achieve higher scoring outcomes in home games compared to away games.




**Statistical Testing:**

1. ***T-test correlation coefficient (Home vs. Away Goals Scored)*: 3.7867493091016193**
2. ***P-value*: 0.0012469170419907249**

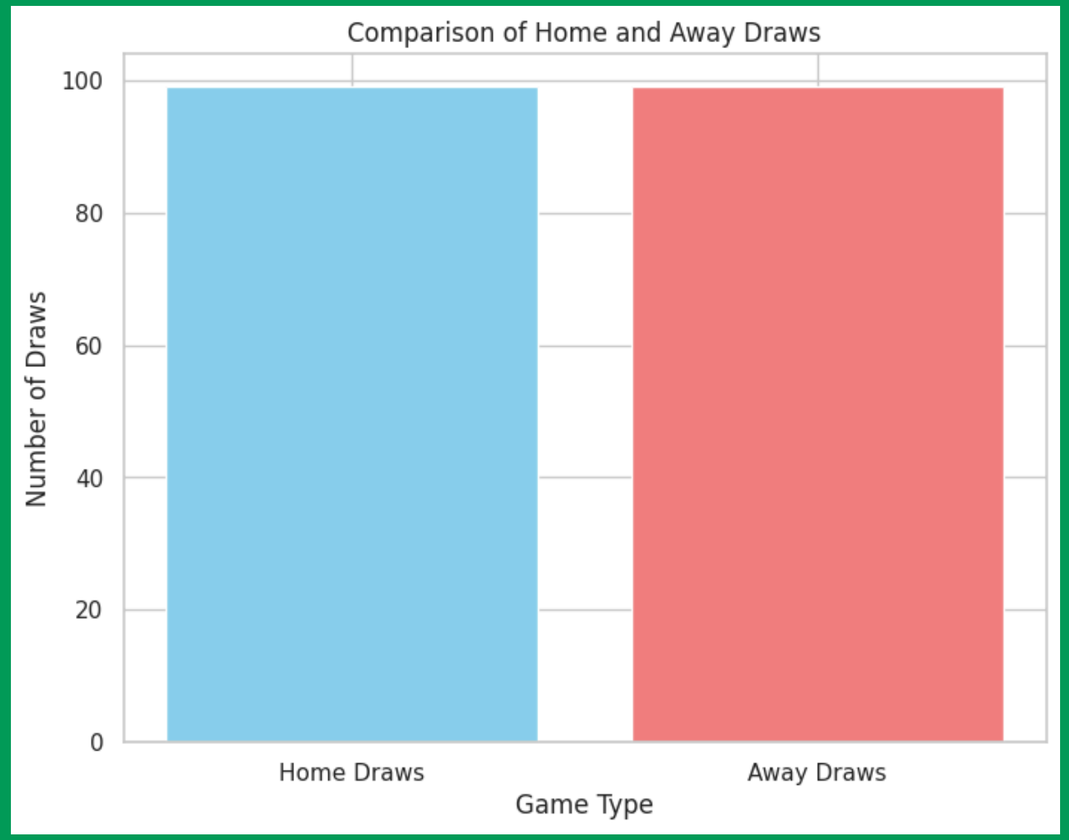



**Statistical Testing:**

1. ***T-test correlation coefficient (Home vs. Away Goals Conceded)*:5.505051964024391**
2. ***P-value*: 2.606115058055847e-05**

**Both tests confirm a statistically significant difference, indicating that teams tend to score more and concede fewer goals at home compared to away games.**

# Hypothesis 3:
# The location of the game does not significantly affect the likelihood of a draw.




**Statistical Testing:**

1. ***T-test correlation coefficient (Home vs. Away Goals Scored)*: 0.0**
2. ***P-value*: 1.0**

**This result indicates no significant difference between the number of draws in home and away games, as evidenced by the high p-value of 1.0.**

# Summary and Findings

- Hypothesis 1: Top teams score more goals than bottom teams, supported by a statistically significant p-value.
- Hypothesis 2: Teams score more goals at home and concede less, showing a home advantage in terms of offensive and defensive performance.
- Hypothesis 3: There is no significant difference in draw frequency between home and away games, suggesting that draws are equally likely regardless of location.

# Conclusion

The analysis highlights significant patterns in scoring and draws, offering insights into team performance dynamics. While offensive strength varies between top and bottom teams and is influenced by game location, draws are equally likely regardless of whether teams play at home or away.

# References


- Premier League table 2017/2018, form and next opponent. (n.d.). FotMob. https://www.fotmob.com/api/leagues?id=47&ccode3=JPN&season=2017%2F2018
- Level Up: Mastering statistics with Python - Stack Overflow. (2021, February 16). https://stackoverflow.blog/2021/02/16/level-up-mastering-statistics-with-python/
- Dataquest. (2022, January 20). Web Scraping NBA stats with Python: Data Project [Part 1 of 3] [Video]. YouTube. https://www.youtube.com/watch?v=JGQGd-oa0l4

**Thank you for listening!**
**We will be happy to answer any questions you may have about the analysis.**

# Any Questions?