

oooo

FOOTBALL DATA ANALYTICS



oooo

ฟุตบอลเป็นกีฬานิยมหนึ่งชนิดที่ได้รับความนิยมมากที่สุดในปัจจุบัน ซึ่งปัจจัยสำคัญที่ส่งผลต่อการแข่งขันนั้นมีหลายประการ ไม่ว่าจะเป็นจำนวนการส่งบอลการสกัดการดวลกันหรือแม้แต่สกิติรับบอลหรือบอลที่เข้าประตู

ทางกลุ่มของเรารอว่างได้เลิ่งเห็นถึงความสำคัญและความน่าสนใจของชุดข้อมูลอิ่งได้ทำการนำมาวิเคราะห์เชิงลึกเพื่อพิสูจน์ข้อเท็จจริงต่างๆจากชุดข้อมูลที่มี

ข้อมูล ณ วันที่ 9 มีนาคม 2566



DATA CLEANING





```
#players_22_23
df_standard23 = pd.read_csv('players_stat/standard(22-23).csv')
df_passing23 = pd.read_csv('players_stat/passing(22-23).csv')
df_creation23 = pd.read_csv('players_stat/creation(22-23).csv')
df_shooting23 = pd.read_csv('players_stat/shooting(22-23).csv')
df_defensive23 = pd.read_csv('players_stat/defensive(22-23).csv')
df_possession23 = pd.read_csv('players_stat/possession(22-23).csv')

# add all the dataframes to a list
df_list = [df_standard23, df_passing23, df_creation23, df_shooting23, df_defensive23,
           df_possession23]
#loop to concat all df
for df in df_list:
    new_cols = [col for col in df.columns if col not in df_standard23.columns]
    df_standard23 = pd.concat([df_standard23, df[new_cols]], axis=1)
```

```
#Set the directory where the CSV files are located
directory = 'players_stat/keeper/Standard/'

#Initialize an empty list to store the dataframes
dfs = []

#Loop through each file in the directory and load it into a dataframe
for filename in os.listdir(directory):
    if filename.endswith('.csv'):
        # Load the CSV file into a dataframe
        df = pd.read_csv(os.path.join(directory, filename))
        # Add a column to indicate the season (based on the filename)
        df['Season'] = filename[7:12] # ex. 20-21, 21-22
        # Add the dataframe to the list
        dfs.append(df)

#Concatenate all the dataframes into a single dataframe
std_gk = pd.concat(dfs, ignore_index=True)
.

.

.

#Add all the std_gk, adv_gk to a list
new_cols = [col for col in adv_gk.columns if col not in std_gk.columns]
gk = pd.concat([std_gk, adv_gk[new_cols]], axis=1)
```



```
● ● ●

#cut string
df_standard23.Nation = df_standard23.Nation.str.split(n=1).str.get(1)
df_standard23.Comp = df_standard23.Comp.str.split(n=1).str.get(1)
df_standard23.Age = df_standard23.Age.str.split('-').str.get(0)
df_standard23 = df_standard23.drop(['Matches', '-9999'], axis=1)

#set index
df_standard23 = df_standard23.set_index('Rk')

#export file
df_standard23.to_csv('players_22-23.csv', encoding='utf-8')
```

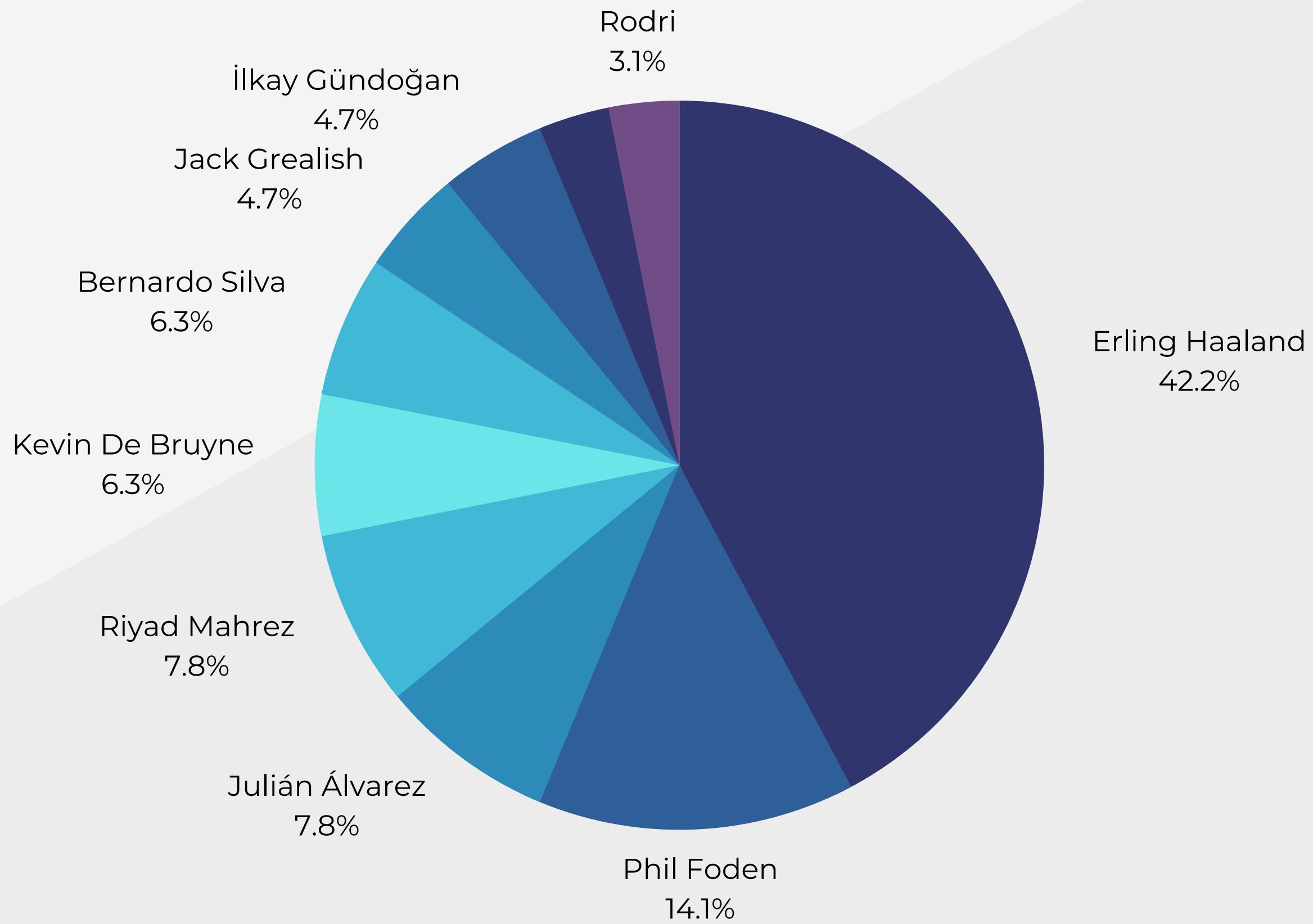
○○○○

ວະໄໄທກຳໃຫ້ HAALAND ເປັນນັກ ເຕະກີກຳປະຕູໄດ້ສູງ ກີສຸດໃນ PREMIER LEAGUE



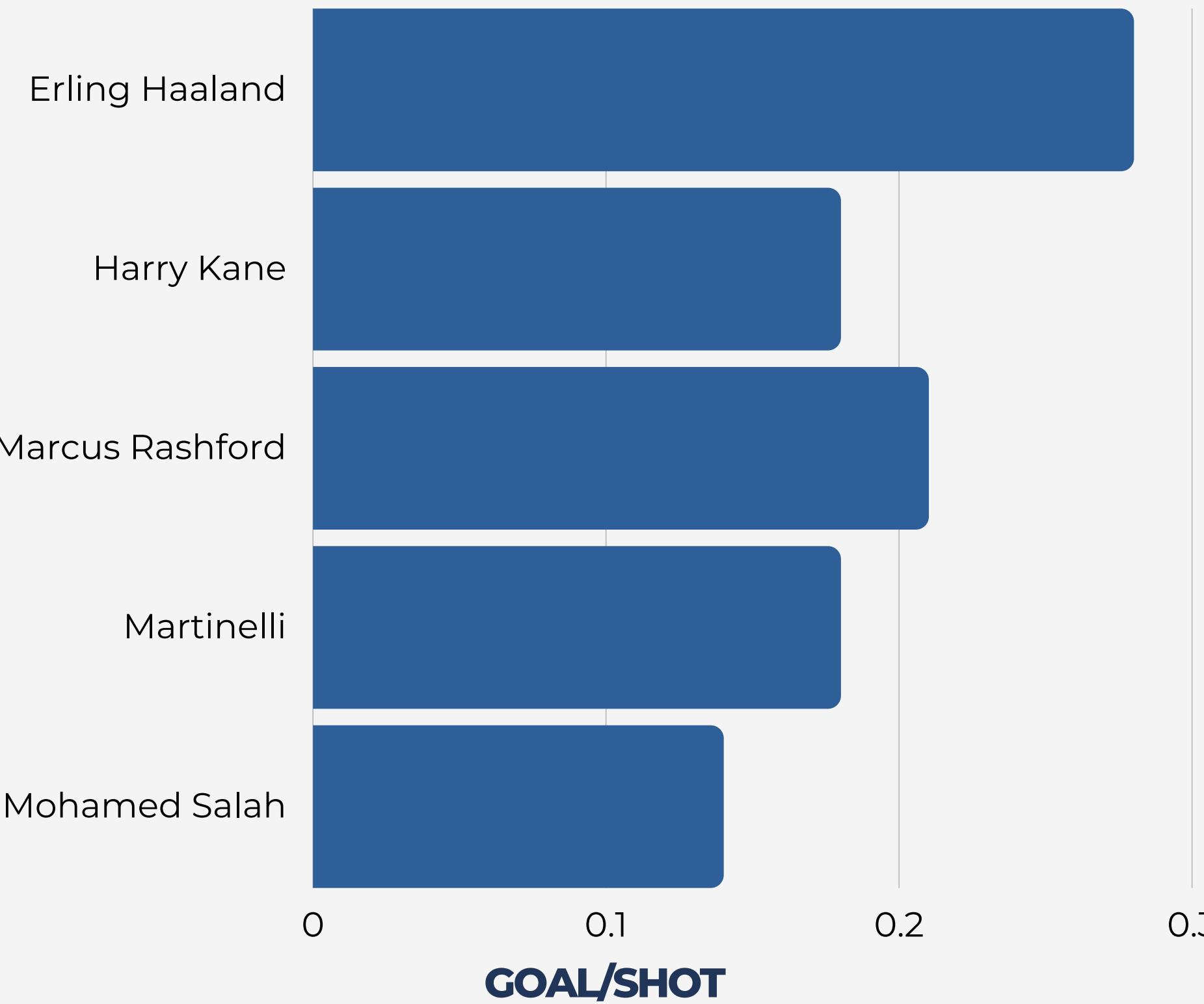
○○○○

GOALS FROM MANCHESTER CITY



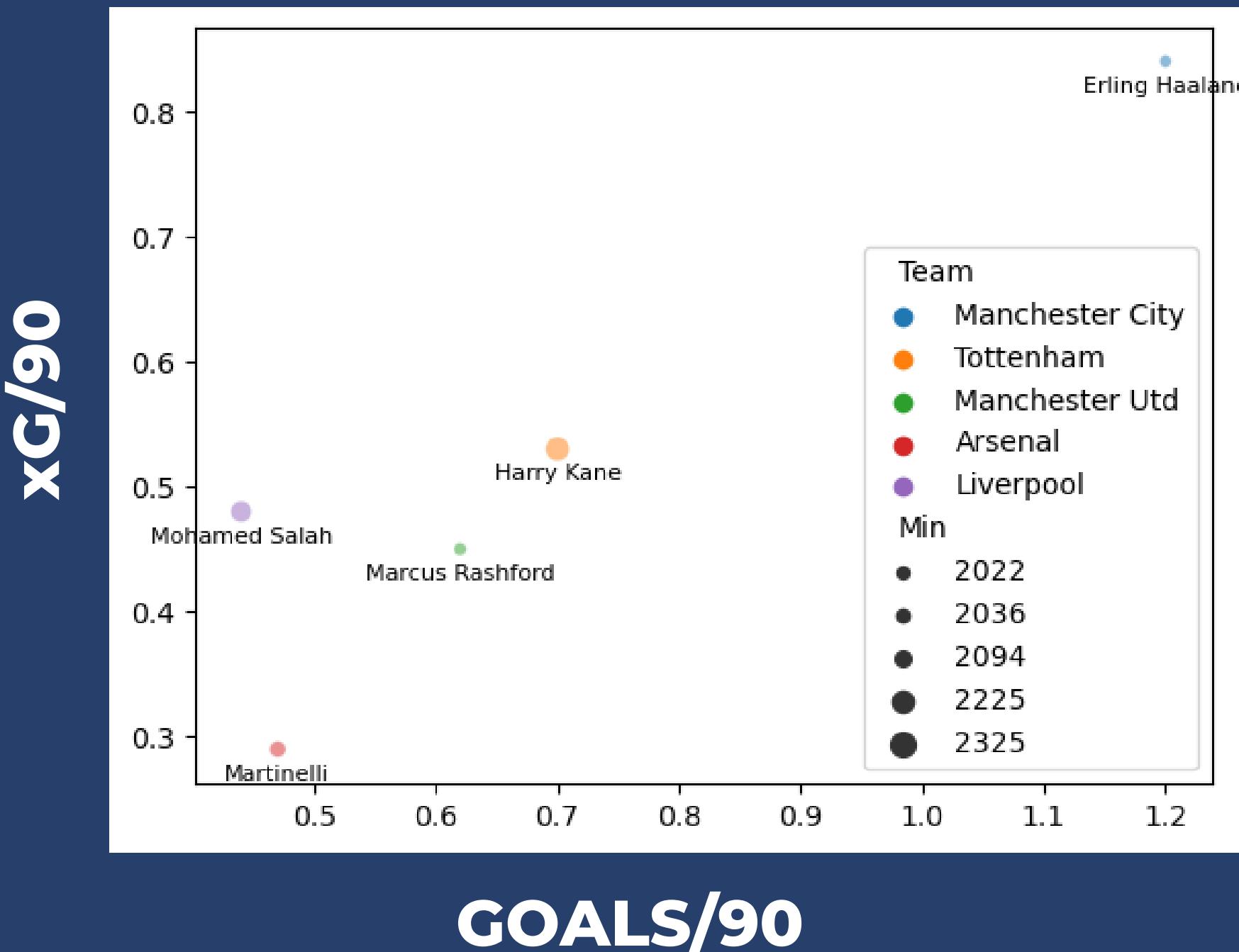
CONVERSION RATES

Top 5 goal scorers in EPL who have started a minimum of 10 games.



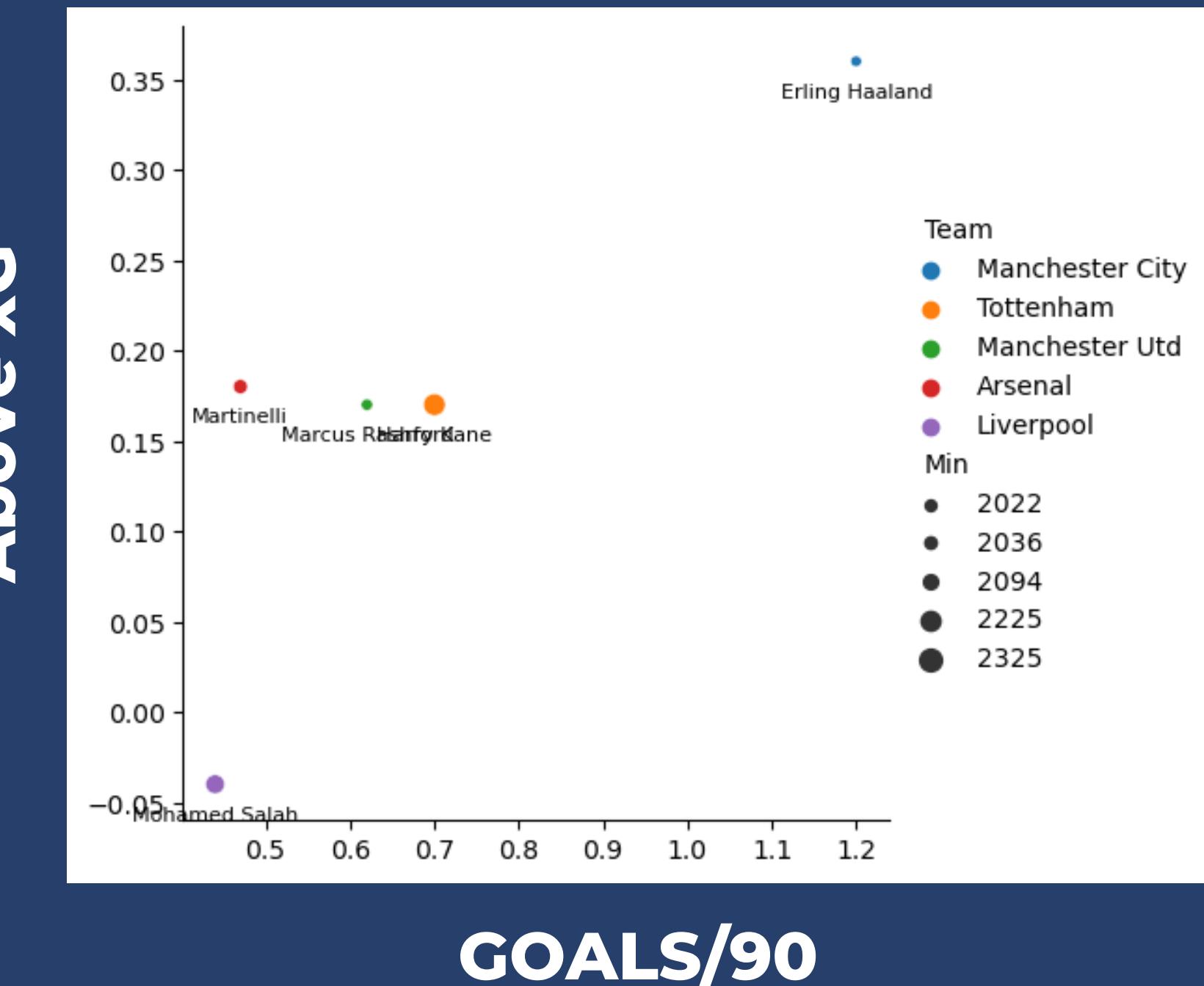
PLAYER GOALS VS XG

Top 5 goal scorers in EPL who have started a minimum of 10 games.



PLAYER GOALS VS ABOVE XG

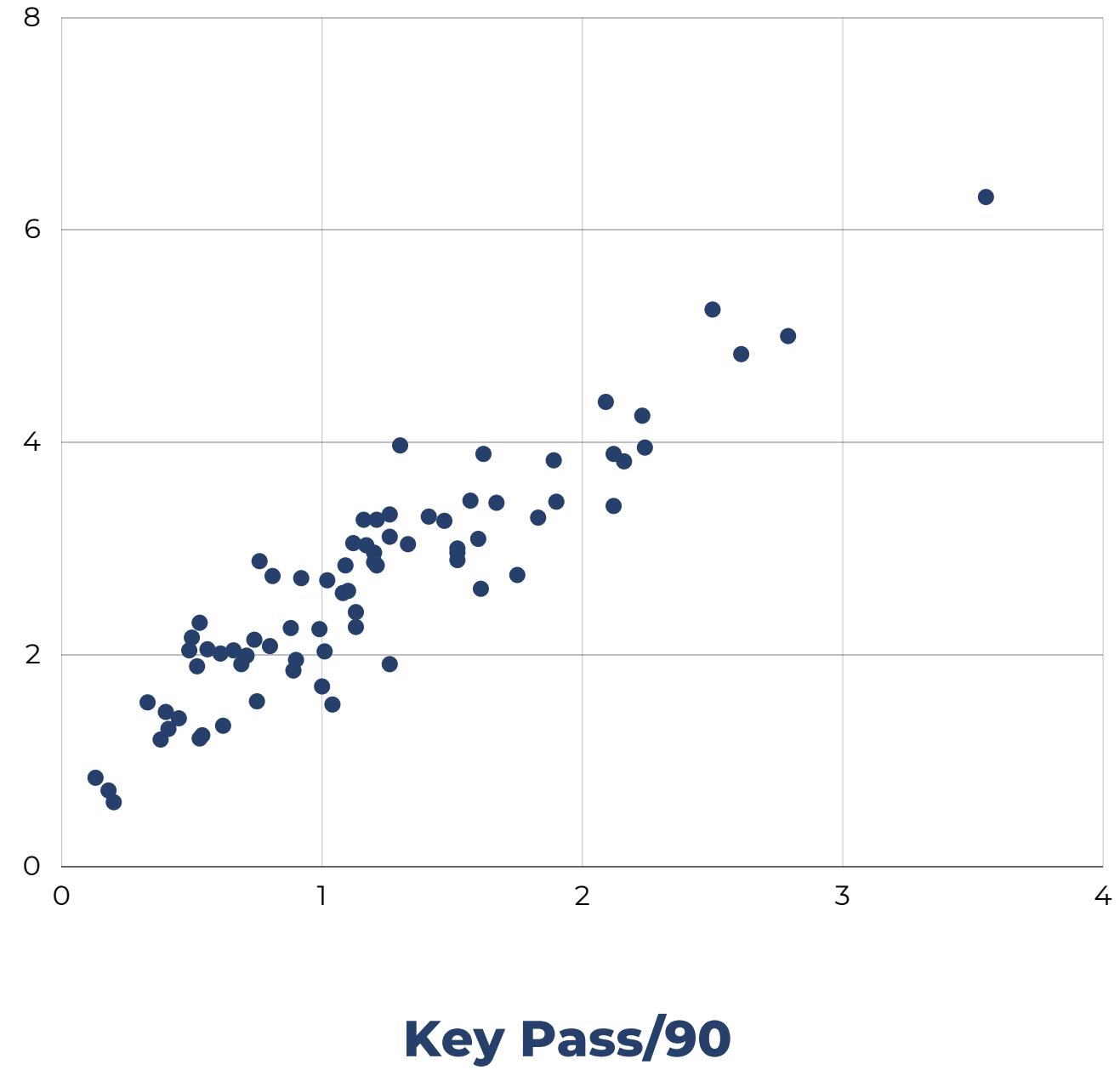
Top 5 goal scorers in EPL who have started a minimum of 10 games.



Key Pass - Shot Creating Action

All midfield from EPL who have started a minimum of 10 games.

Shot Creation Action/90



■ Shot Creation Action/90 ■ Key Pass/90

Manchester City

Manchester Utd

Liverpool

Crystal Palace

Newcastle Utd

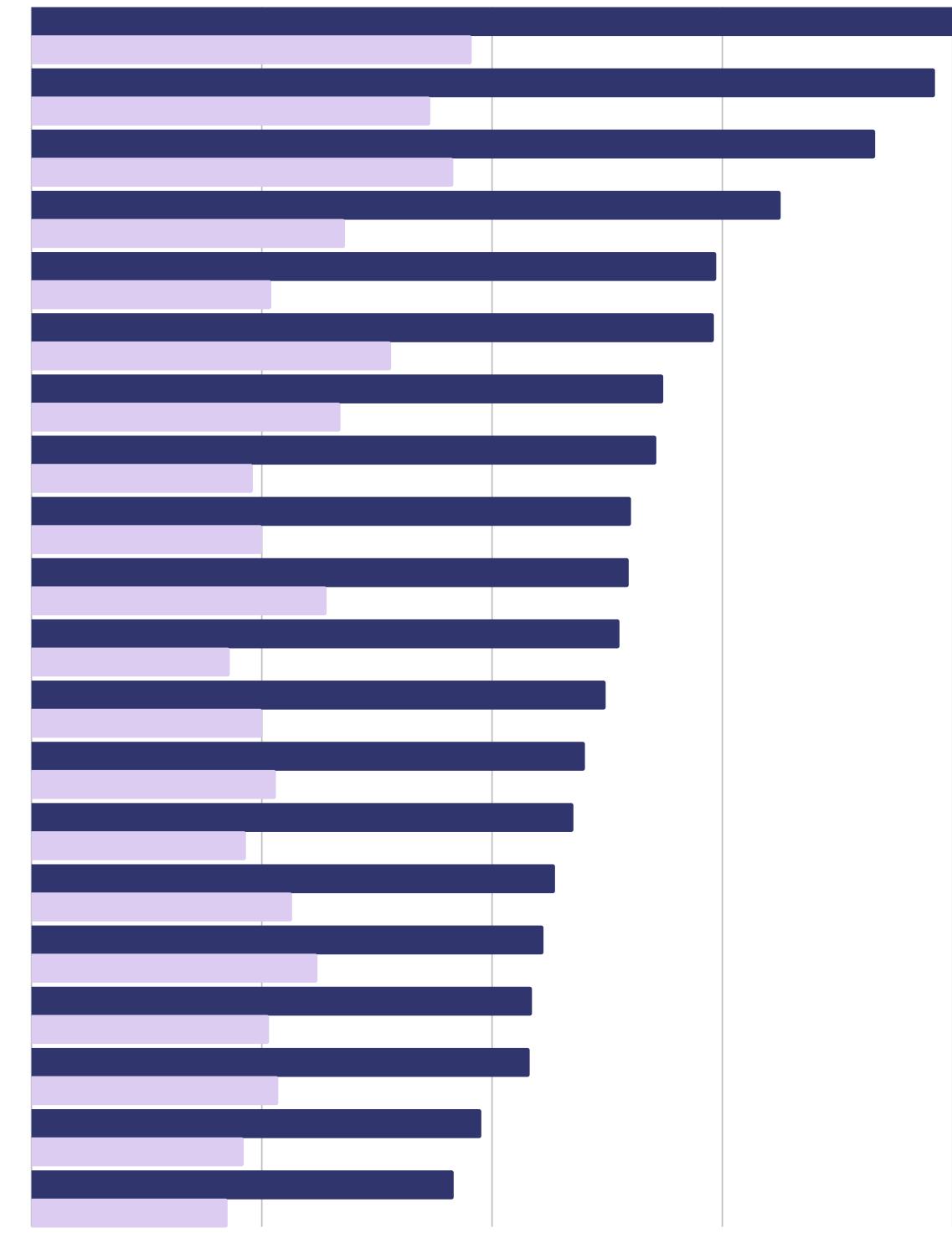
Brentford

Leicester City

Fulham

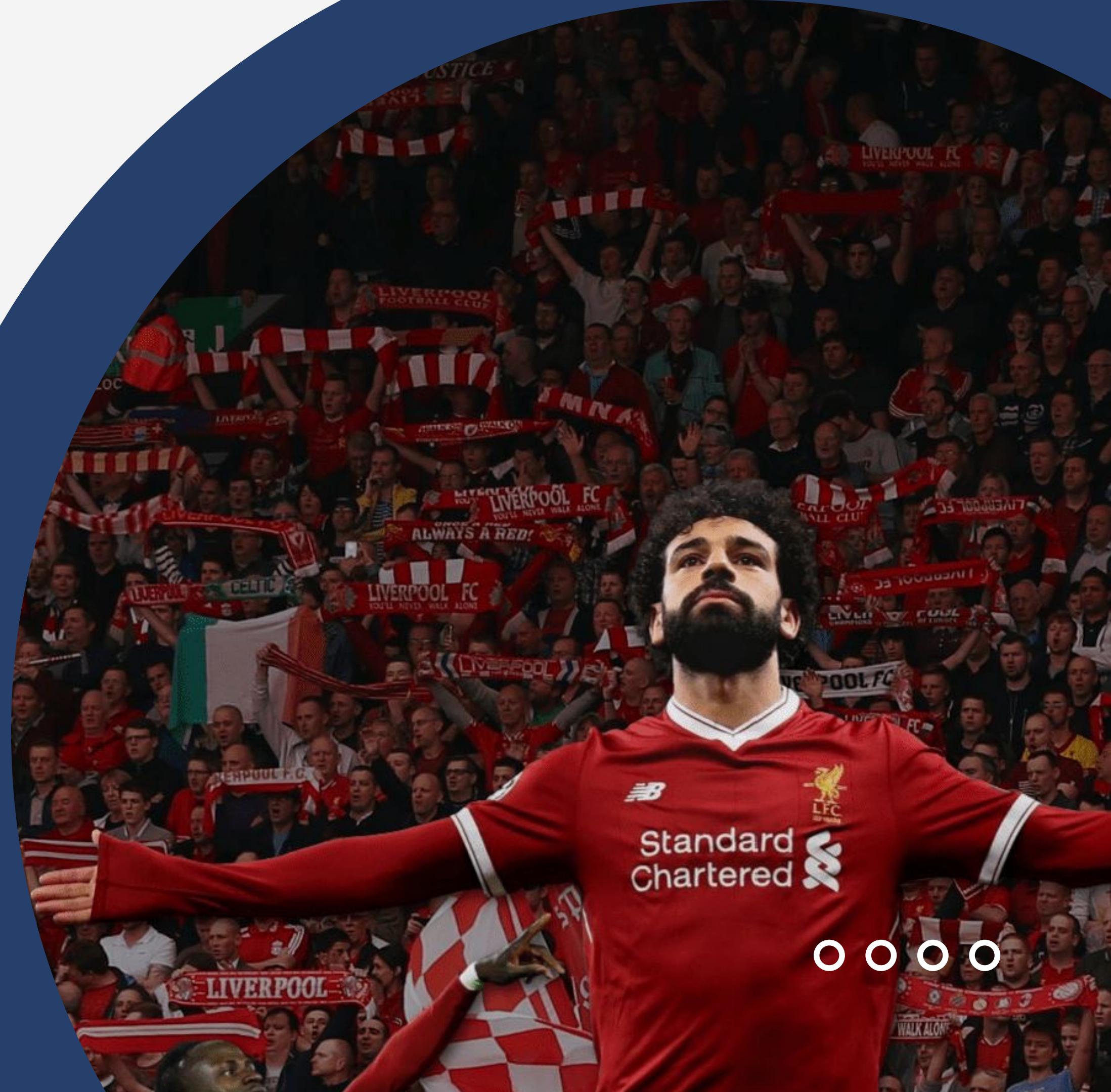
Nott'ham Forest

Bournemouth



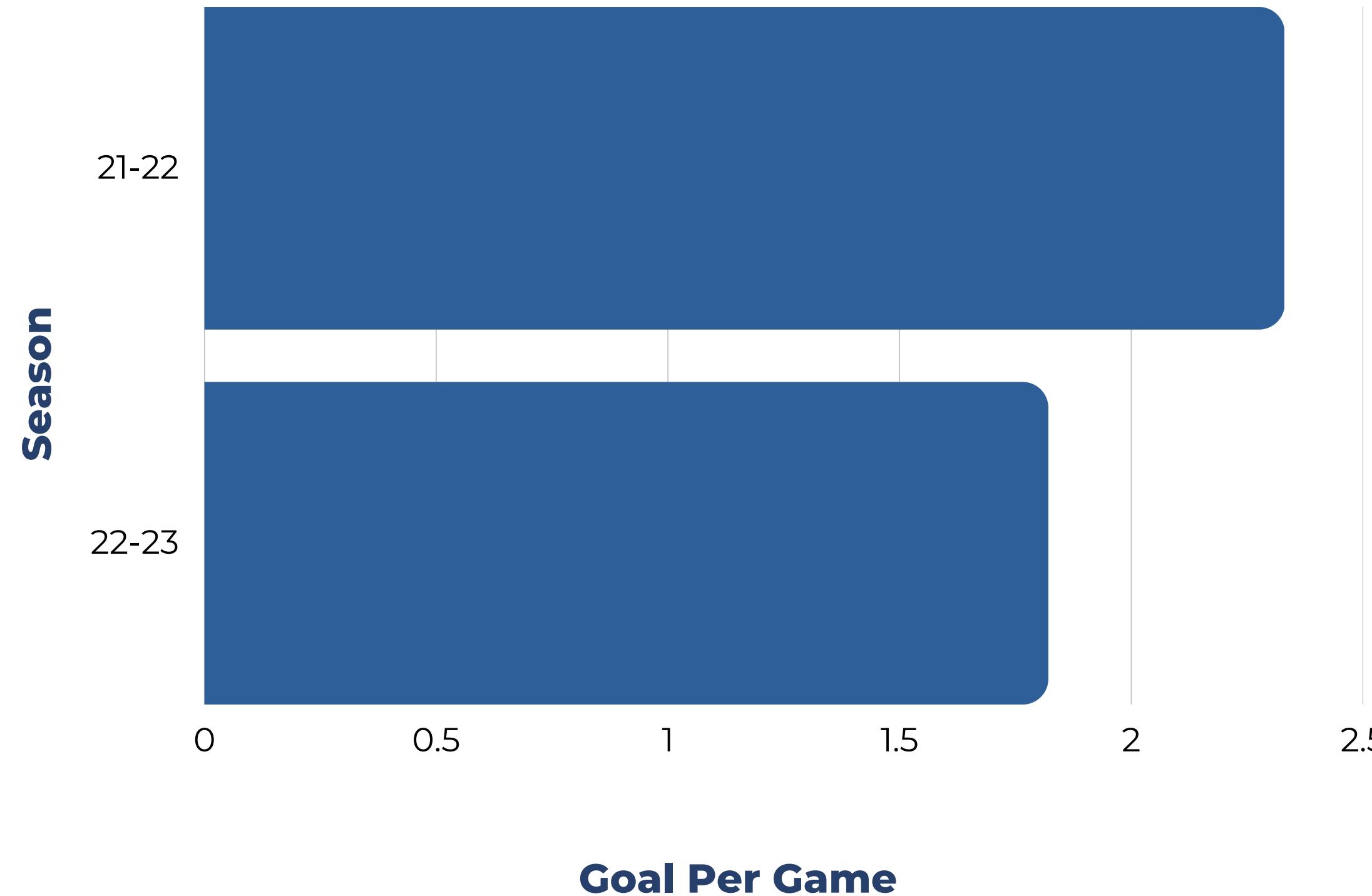
oooo

นักเตะคนใดในทีม
LIVERPOOL
ที่สมควรได้รับ
เชล่าในการเล่น
เพิ่ม

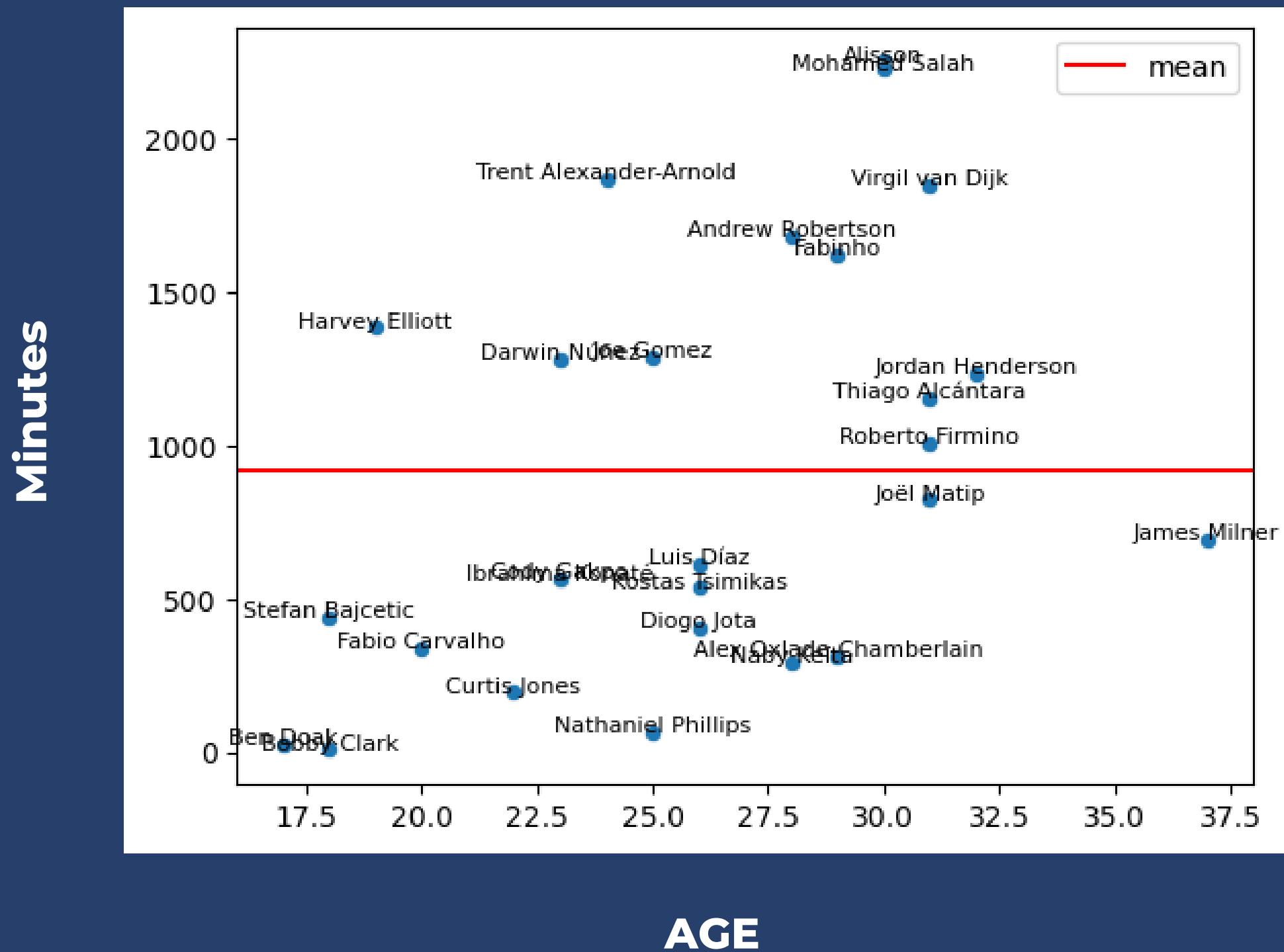


oooo

LIVERPOOLS' GOALS PER GAME

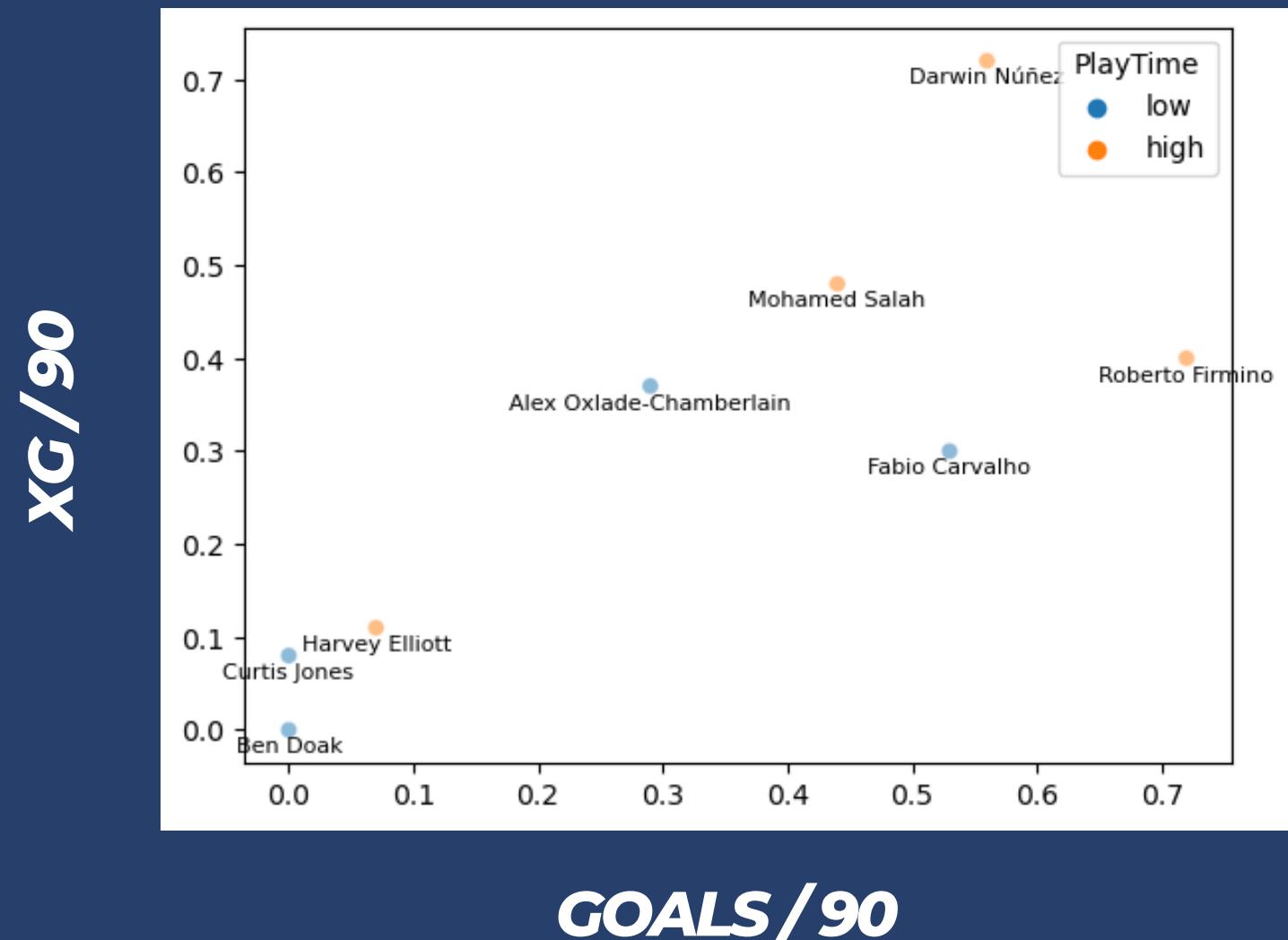


PLAYER'S PLAYTIME

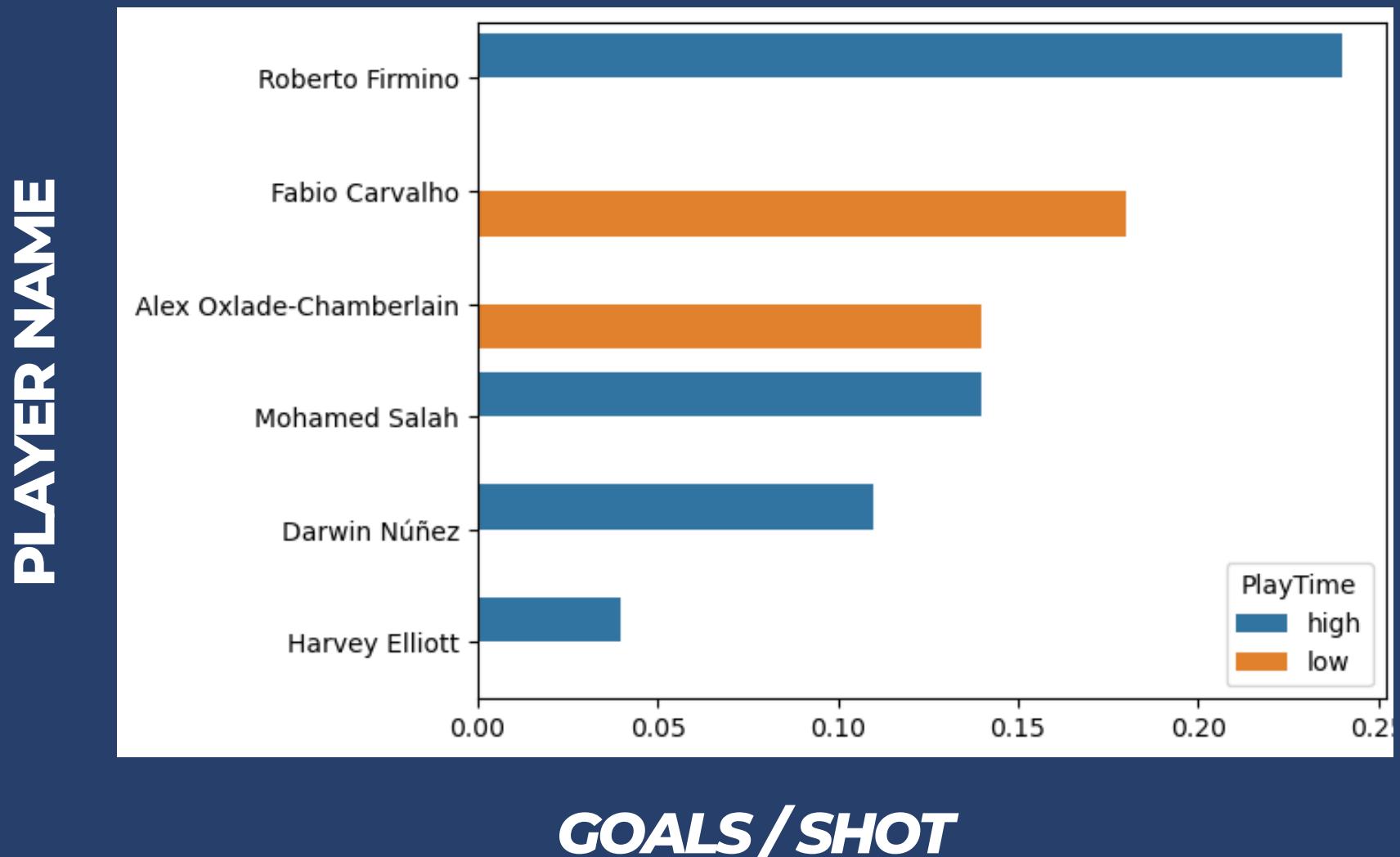


PLAYER GOALS VS XG

All attacker from LFC (injured player not included)



CONVERSION RATES OF DIFFERENT PLAYERS



〇〇〇

ជុំដឹកសុំបុង

ជូនការប្រាក់

ឈរកែវាយ

27-31 មេ



អង់យុវតានាក់ទៅនៅក្នុង^៩ ការវិគរាងប៉ុម្មតិន្នន័យ

(STANDARD METRIC FIT TO ANALYZE THE POPULATION)

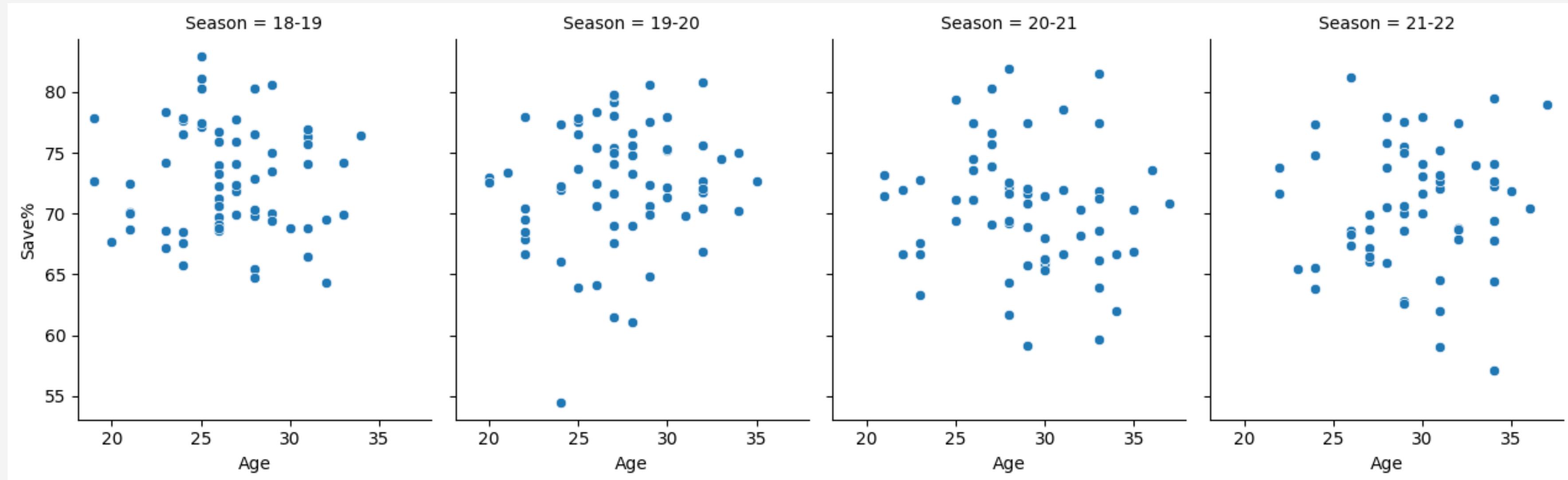
- Save Percentage (Save%)
- Cleansheet Percentage (CS%)
- Penalty Kicks Save Percentage (PSave%)
- Defensive action outside of penalty area per 90 minutes
(OPA/90)
- Post-Shot Expected Goals minus Goals Allowed per 90 minutes
(PSxG-GA/90)

ขั้นตอนการวิเคราะห์

- plot Relation plot ดูภาพรวมของข้อมูล metric นั้นๆเทียบกับอายุ
- List top player ของ metric
- หา percentile ที่ 95 ของ metric
- List player
 - player ที่มีค่า metric เหนือกว่า 95% ของประชากรกับ (top_player)
 - player ที่มีค่า metric เหนือกว่า 95% ของประชากรและมีอายุอยู่ในช่วง 27 -31 (top_player_age_27_31)
- นำจำนวนของกลุ่มตัวอย่างทั้งสองกลุ่มมาหาสัดส่วน ดังนี้
 - เปอร์เซ็นต์ของผู้เล่นที่อายุอยู่ในช่วง 27 -31 เทียบกับ top player ทั้งหมด (ANS)
 - $ANS = (top_player_age_27_31 / top_player) * 100$

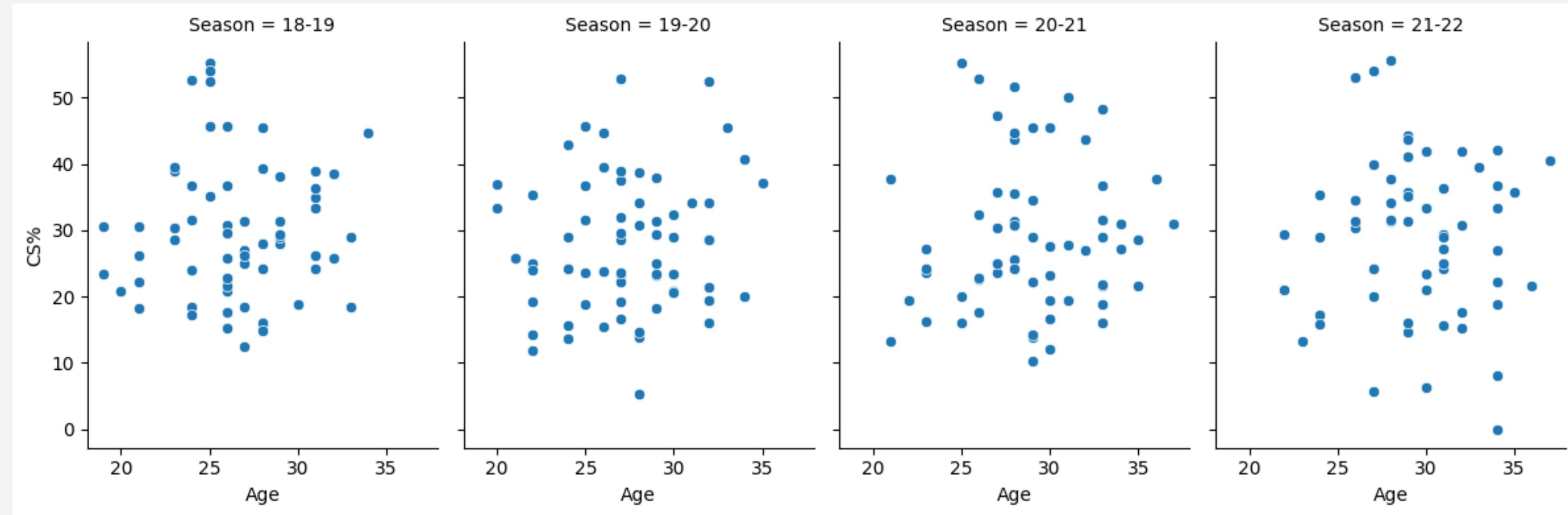
Save% : Save Percentage

(Shots on target against - Goals against) / Shots on target against



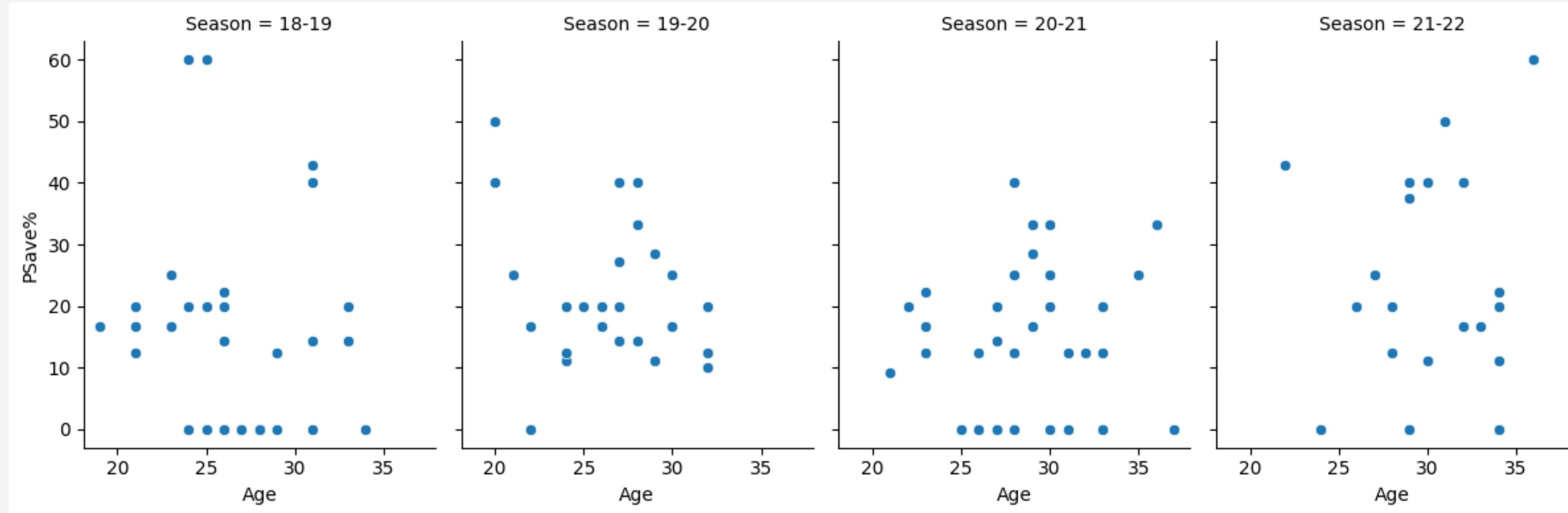
ผู้รักษาประตูที่เป็นผู้รักษาประตูแนวหน้าในเรื่องของ **Save%** และมีช่วงอายุอยู่ระหว่าง **27-31** ปี อยู่ที่ประมาณร้อยละ **53.85** ของกั้งหมด

Cleansheet Percentage (CS%)



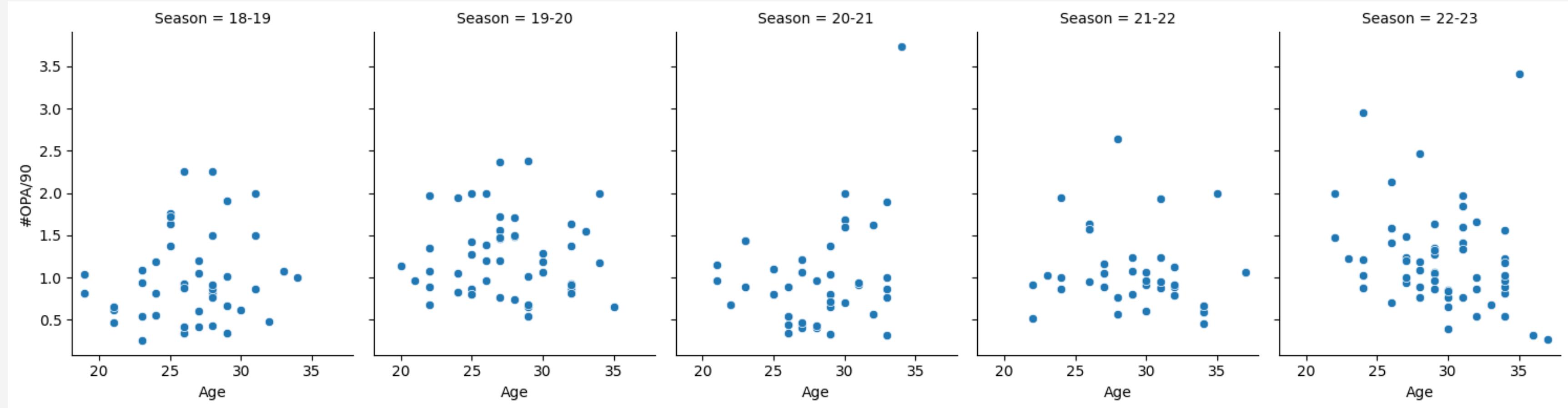
เมื่อเราทำการวิเคราะห์เบื้องต้นการเก็บ **cleansheet**, เราจะพบว่าผู้รักษาประตูที่มี **CS%** เป็นอันดับต้นน้ำและอายุอยู่ในช่วง **27-31** ปี มีมากถึง **38.46%** ของทั้งหมด

Penalty Kicks Save Percentage (PSave%)



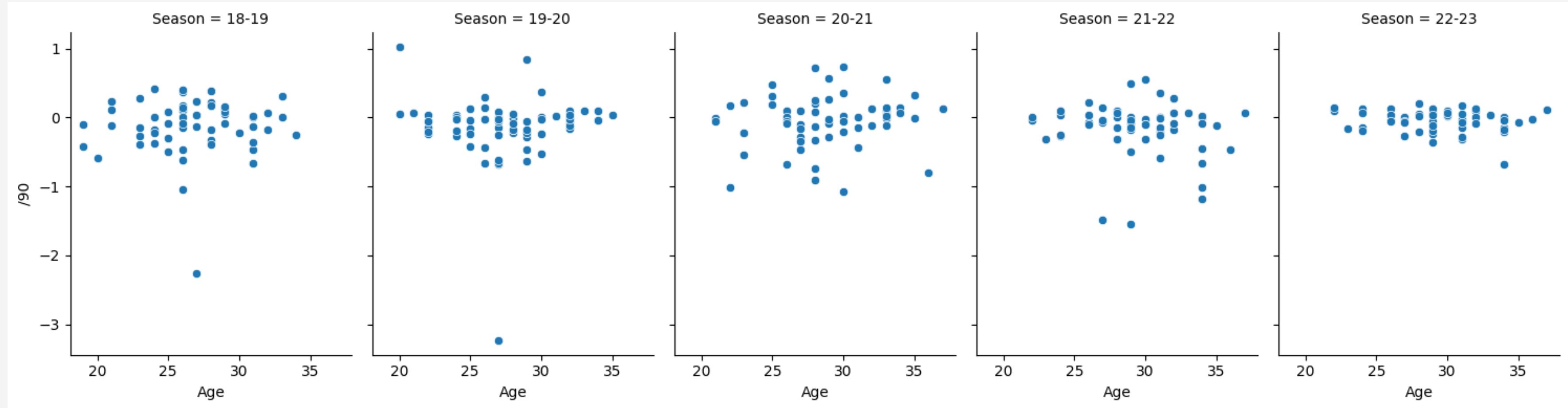
เมื่อเรามากำหนดว่าเปอร์เซ็นต์ของการเชฟลูกโทษแล้ว, เราจะพบว่าผู้รักษาประตูที่มี **PSave%** เป็นอันดับต้นๆ และอายุอยู่ในช่วง **27-31** ปี มีมากถึง **42.86%** ของทั้งหมด

Defensive action outside of penalty area per 90 minutes (#OPA/90)



เมื่อเรามากำหนดค่า **#OPA/90** แล้ว, เราจะพบว่าผู้รักษาประตูที่มี **#OPA/90** เป็นอันดับต้นน้ำและอายุอยู่ในช่วง **27-31** ปี มีมากถึง **35.29%** ของทั้งหมด

Post-Shot Expected Goals minus Goals Allowed per 90 minutes (PSxG-GA/90)



เมื่อเราทำการวิเคราะห์ **PSxG-GA/90** แล้ว (หน่วยัดหรือค่าสถิติก็มีความเสถียรที่สุดในการที่เราจะระบุได้ว่าผู้รักษาประตูนั้นเป็นผู้รักษาประตูที่ดีหรือแย่ ตั้งแต่เริ่มแรก), เรา ก็จะพบว่าผู้รักษาประตูที่มี **PSxG-GA/90** เป็นอันดับต้นนี้และอายุอยู่ในช่วง **27-31** ปี มีมากถึง **57.14%** ของทั้งหมด

SUMMARIZATION

- **Save%** : **53.85%**
- **CS%** : **38.46%**
- **PSave%** : **42.86**
- **#OPA/90** : **35.29%**
- **PSxG-GA/90** : **57.14%**

จากสถิติที่เราได้สรุปไว้บี้น
ทุกรายการมีค่าเกิน **30%** ของประชากร
ทั้งหมด และสำหรับค่าสกัดทั้งหมดเราใช้
เปอร์เซ็นไทล์ **95** เป็นเกณฑ์การวัดว่าใคร
เป็นผู้รักษาประถุแวงหน้า (เหนือกว่า **95%**
ของประชากร) ซึ่งหมายความว่า "อายุที่เป็น^{จุดพิเศษ}ของผู้รักษาประถุอยู่ที่ระหว่าง
27 และ **31** ปี" เป็นความจริงตามที่ได้
วิเคราะห์มา

○ ○ ○ ○

THANK
YOU

นายชนิตพล ໂລະມັກລ 6410406533
นายชยกร เจียรสุวิกานต์ 6410450117
นางสาวแพรวรุ่ง พุดชะวา 6410451253
นายอุรวิศ เจียรจินดา 6410406932

○ ○ ○ ○