DATA ANALYSIS – PORTFOLIO 1

With a diverse background spanning from borehole operations technician to elementary classroom teacher, cashier, front-desk banking officer, and now a data scientist, I have embarked on a journey to continually prove my worth and quantify my relevance in various fields.

My tenure as an intern at Lufemos Consult Nig Ltd has equipped me with valuable skills in data analytics using Microsoft Excel spreadsheets, SQL Server Management Studio for database management, Power BI for visualization, and Python programming along with data science libraries like NumPy, Pandas, and Matplotlib.

The acquisition of this set of skills for data science tasks and a future of endless learning has driven a sense of career fulfilment within me and I am passionate and optimistic of a long and successful career in data science.

PROFESSIONAL BACKGROUND

Armed with a bachelor's degree in Applied Geology, I possess practical knowledge in water borehole construction techniques and have experience as a sales representative, providing me with a well-rounded understanding of diverse operational environments, including the banking sector. Motivated by a desire for a sustainable and dynamic career that aligns with my objectives, I transitioned to data science, accumulating two years of experience in utilizing tools such as Microsoft Excel, SSMS Workstation, Power BI, and Python for data analysis, visualization, and reporting.

My proficiency in Python programming language for data science tasks with libraries such as Pandas, and NumPy, drives me to contribute to clearer data-driven decision-making processes through insightful reports. With dedication and consistency, I am committed to leveraging my skills and experiences to make meaningful contributions in the field of data science and beyond.

TABLE OF CONTENT

INTRODUCTION

PROJECT BACKGROUND

The English Premier League stands as one of the most esteemed football competitions globally, featuring 380 matches hosted at various venues throughout the country. The officiating of each match holds significant sway over match outcomes and ultimately impacts the league standings. The insights derived from this report hold relevance for analysts, league viewers, the English Football Association, and all associated stakeholders.

PROBLEM

The primary objective of this analysis and report is to examine the officiating patterns of individual referees across all matches of the EPL 2021/2022 season. Utilizing available data on key incidents in each game, the aim is to provide insights into the unique contributions of referees in shaping the outcomes of matches.

DESIGN

The data extraction process involved sourcing data from the official Premier League website, followed by data cleaning procedures conducted using Google Sheets and Data World. Aggregate functions like average and round, alongside conditional logic statements such as "case when," were employed to derive insights from the dataset. Furthermore, filtering functions such as "group by" and "order by" were utilized to organize and analyze the dataset comprehensively. The dataset comprises various metrics related to teams, referees, and match statistics for both home and away sides, including fouls, shots, and cards. To extract and present the required information from the dataset, SQL Server Management Studio (SSMS) queries were executed, and the results were tabulated. Additionally, Microsoft Excel was leveraged for data visualization purposes.

ANALYSIS AND VISUALISATION OF INSIGHTS FROM DATASET

Games officiated by individual referees

The dataset was queried to determine the number of matches officiated by each referee during the 21/22 EPL season

```
select Referee,
    count (*) as Game_Officiated
from [EPLsoccer21-22]
group by Referee
order by Game_Officiated desc;
```

Figure 1: SSMS code for columns in table 1

The chart depicted in Figure 2 indicates that referees A Taylor and P Tierney officiated in the top 10% of matches during the season, suggesting a heightened level of recognition for their officiating capabilities from players, fans, and regulatory bodies. Referees who oversaw twenty or more games demonstrate a higher level of value and overall acceptance compared to referees such as M Salisbury and T Harrington, who were involved in just three games each and may be perceived as having less influence or significance in officiating roles.

Referee	Game_Officiated
A Taylor	28
P Tierney	27
C Pawson	26
M Atkinson	26
M Dean	26
M Oliver	26
J Moss	25
D Coote	20
S Attwell	20
A Marriner	19
C Kavanagh	18
K Friend	18
A Madley	16
D England	15
S Hooper	15
R Jones	12
G Scott	12
P Bankes	12
J Gillett	9
J Brooks	4
M Salisbury	3
T Harrington	3

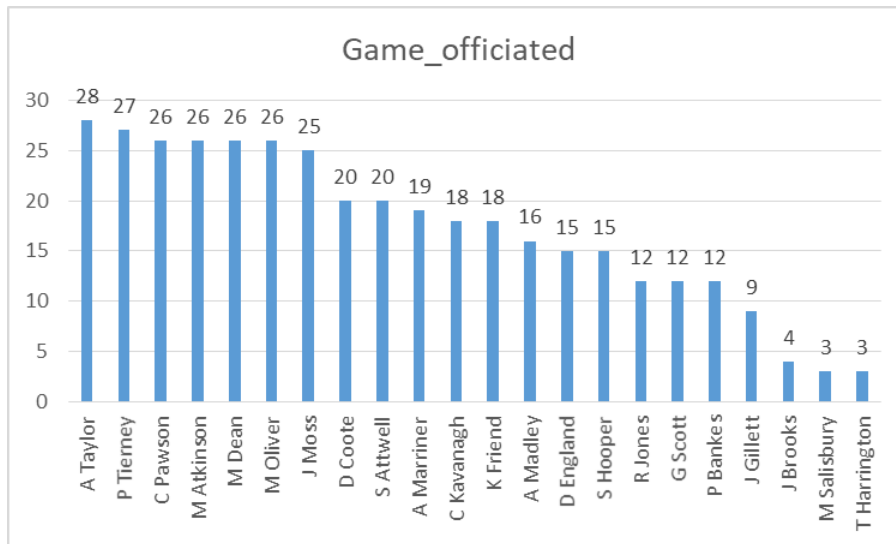Table 1: Number of games officiated by each referee (cells in red indicate top 10%)

Figure 2: Chart showing number of games officiated by individual referees

Average number of goals, fouls, yellow cards, and red cards distribution per game for each referee

The goals, fouls committed, yellow cards and red cards been issued during the game are key statistical indicators for any soccer match. The columns below shows how these indicators are influenced by individual referees per match officiated.


```sql
--SELECT *
--from [EPLsoccer21-22]

select Referee,
        round(avg(FTHG + FTAG), 0) as avg_goals_per_game,
        round(avg(HF + AF), 0) as avg_fouls_per_game,
        round(avg(HY + AY), 0) as avg_yellow_cards_per_game,
        round(avg(HR + AR), 2) as avg_red_cards_per_game
    from [EPLsoccer21-22]
    group by Referee;
```

Figure 3: SSMS code query for the columns in table 2

Referee	avg_goals_per_game	avg_fouls_per_game	avg_yellow_cards_per_game	avg_red_cards_per_game
A Madley	3	23	2	0.13
A Marriner	3	19	3	0.11
A Taylor	3	20	4	0.18
C Kavanagh	3	20	4	0.17
C Pawson	3	22	4	0.08
D Coote	2	24	5	0.1
D England	3	21	4	0.07
G Scott	3	22	3	0.08
J Brooks	2	20	5	0
J Gillett	4	19	3	0.11
J Moss	3	18	2	0.16
K Friend	3	20	4	0.11
M Atkinson	3	16	2	0.08
M Dean	3	19	3	0.08
M Oliver	3	21	3	0.27
M Salisbury	3	23	4	0
P Bankes	2	21	5	0.08
P Tierney	3	22	4	0.19
R Jones	3	21	3	0
S Attwell	3	18	4	0.05
S Hooper	2	21	3	0
T Harrington	1	21	2	0

Table 2: Showing average goals, fouls, yellow cards and red cards per game for each referee (cells in red indicate top 10%).

Considering the discrepancy in the number of matches officiated by individual referees, it's notable from Table 2 that the average number of goals scored per game across all referees was 3. Referees with an average goals per game exceeding this overall average indicate a higher likelihood of their matches ending with three or more goals (on J Gillet exceeds this average), while those with lower averages suggest the opposite tendency. This analysis holds particular significance for both punters and analysts, offering valuable insights into potential goal-scoring patterns based on individual referee officiating tendencies.
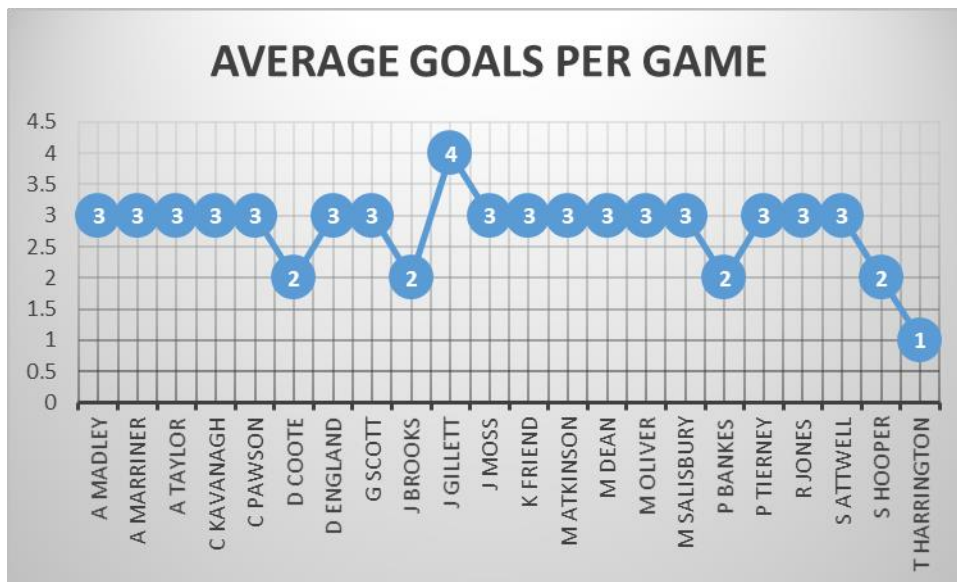
Figure 4: Chart showing the average goals scored in each game for each referee

The average from the "average_fouls_per_game" column in Table 2 is computed to be 21. The accompanying chart illustrates the average fouls per game for individual referee. Referees with an average fouls per game exceeding 21 suggest a propensity to intervene more frequently in matches, potentially disrupting the flow of play and increasing the likelihood of issuing cards. This information holds particular relevance for teams in devising tactical strategies ahead of matches, allowing them to adapt their gameplay accordingly based on the officiating tendencies of the referee assigned to oversee the game.
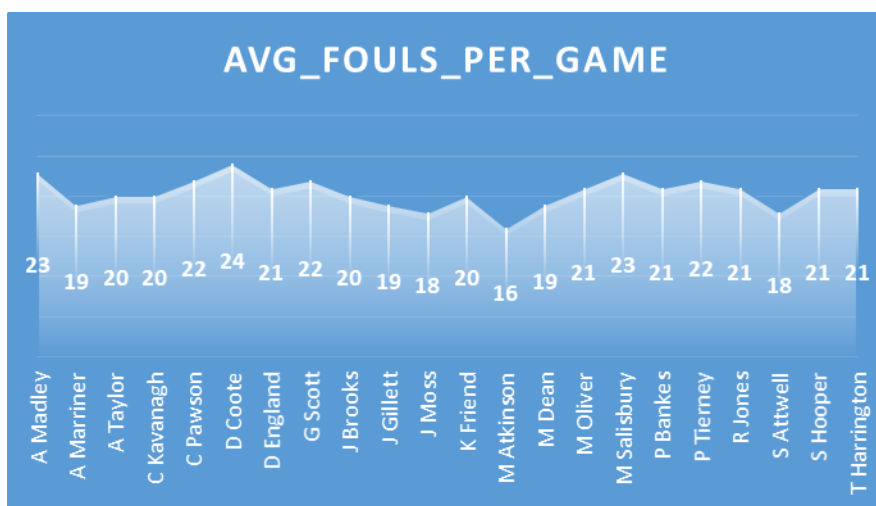

Figure 5: Chart showing the average fouls committed in each game for individual referees

Based on the data presented in Table 2, the average number of yellow cards per game is calculated to be 4 and also the chart below shows the trendline, suggesting that referees with individual averages above this threshold are more inclined to issue cards compared to those below the overall

average. D Coote, as indicated in Table 1, officiated in 20 games, a figure surpassed by only seven out of a total of twenty two other referees. Despite this, the chart below reveals that he maintained an average yellow card per game of 5, which is the highest average from the "avg_yellow_cards_per_game" column. This observation indicates a significant likelihood for D Coote to issue yellow cards during the matches he officiates. Such insights are particularly valuable to a range of stakeholders, including teams, players, analysts and punters alike.
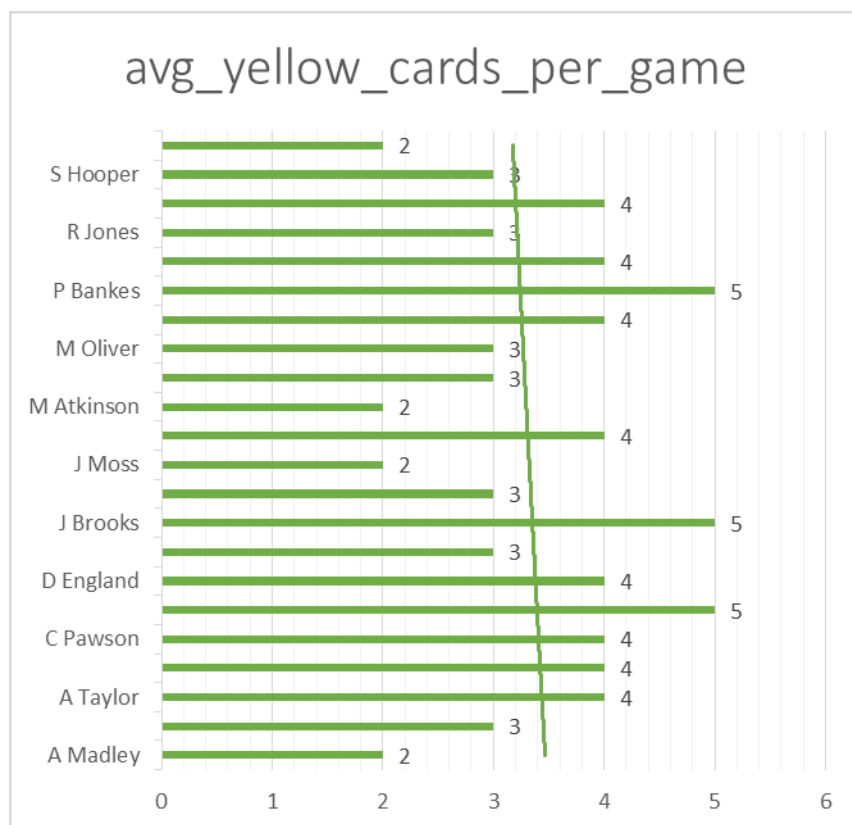

Figure 6: Chart showing the average yellow card issued in each game for each referee

The "avg_red_cards_per_game" pie chart below shows the top four referees with the highest red card tendencies per game from table 2 above. M Oliver stood the highest chance of issuing a red card in all the games he officiated
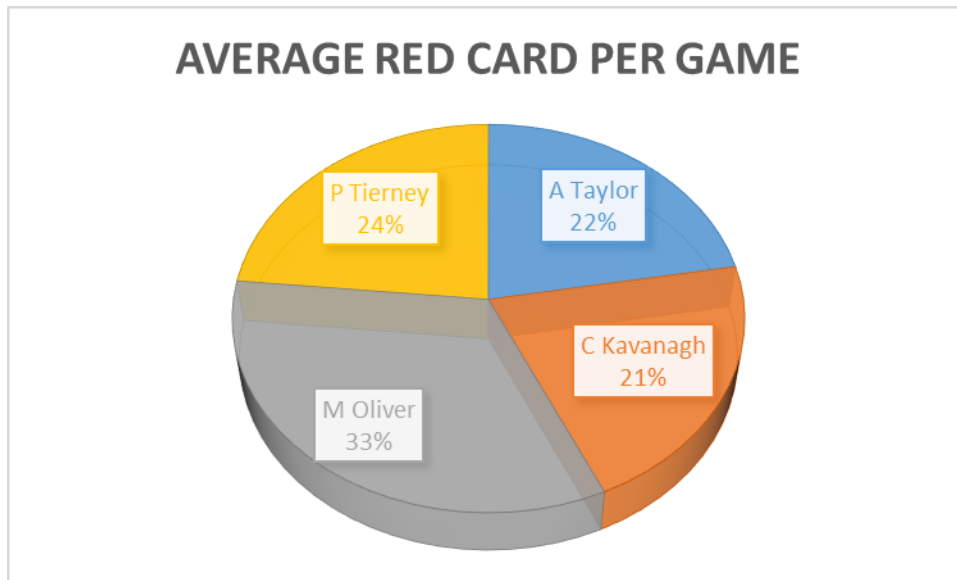
Figure 7: Chart showing average number of red cards shown in each game for individual referees

Average home shots on target, away shots on target, home possession and away possession per game for individual referees.

The columns below were queried from the EPL table to determine the average home/away shots and average home/away ball possessions for each referee per match.

```
--SELECT *
--from [EPLsoccer21-22]

select Referee,
        round(avg(HST), 0) as avg_home_shots_on_goal,
        round(avg(AST), 0) as avg_away_shots_on_goal,
        round(avg(HST / (HST + AST) * 100), 0) as avg_possession_home,
        round(avg(AST / (HST + AST) * 100), 0) as avg_possession_away
    from [EPLsoccer21-22]
    group by Referee;
```

Figure 8: SSMS code querying the average home/away shots in each game for individual referees and average ball possession in each game for individual referees

Referee	avg_home_shots_on_goal	avg_away_shots_on_goal	avg_possession_home	avg_possession_away
A Madley	6	4	57	43
A Marriner	4	5	41	59
A Taylor	5	4	55	45
C Kavanagh	5	5	52	48
C Pawson	4	5	47	53
D Coote	4	4	54	46
D England	5	4	54	46
G Scott	5	3	60	40
J Brooks	4	3	52	48
J Gillett	5	5	51	49
J Moss	5	3	61	39
K Friend	5	5	53	47
M Atkinson	4	4	54	46
M Dean	5	5	50	50
M Oliver	5	3	59	41
M Salisbury	4	6	46	54
P Bankes	5	3	58	42
P Tierney	5	5	48	52
R Jones	4	5	44	56
S Attwell	5	3	62	38
S Hooper	5	5	52	48
T Harrington	3	3	54	46

Table 3: Showing average home/away shots and possessions in each match for individual referees.

From table 3 above the average shots on goal for both the home and away side can be seen as compared for individual referees with A Madley and M Oliver making the top 10% for "avg_home_shots_on_goal" and on the other hand, M Salisbury and J Gillett made up the top 10% for "avg_away_shots_on_goal", the visual analysis in the charts below show the average level of possibilities to get a goal both the home or away side stands against individual referees per game officiated. This is especially important for regulators to analyze and detect anomalies in referee's biases.
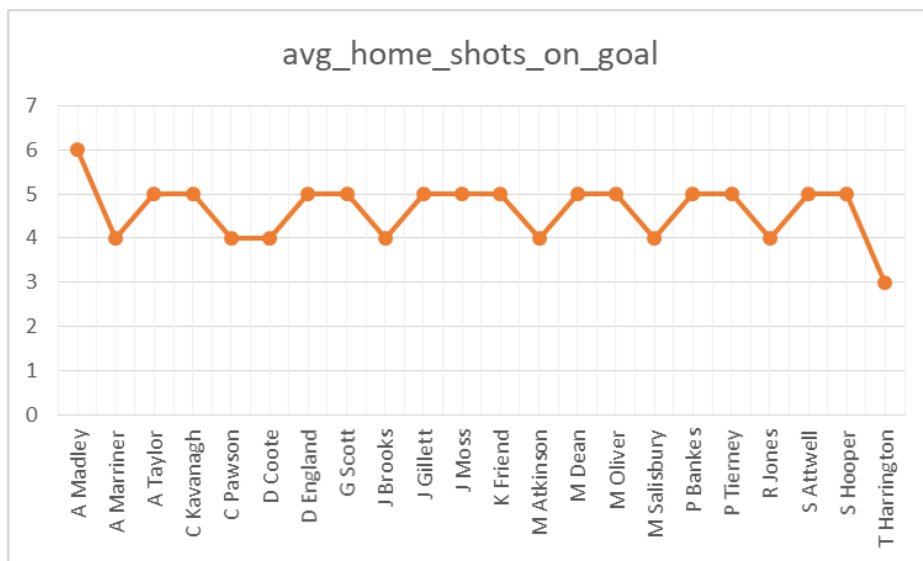
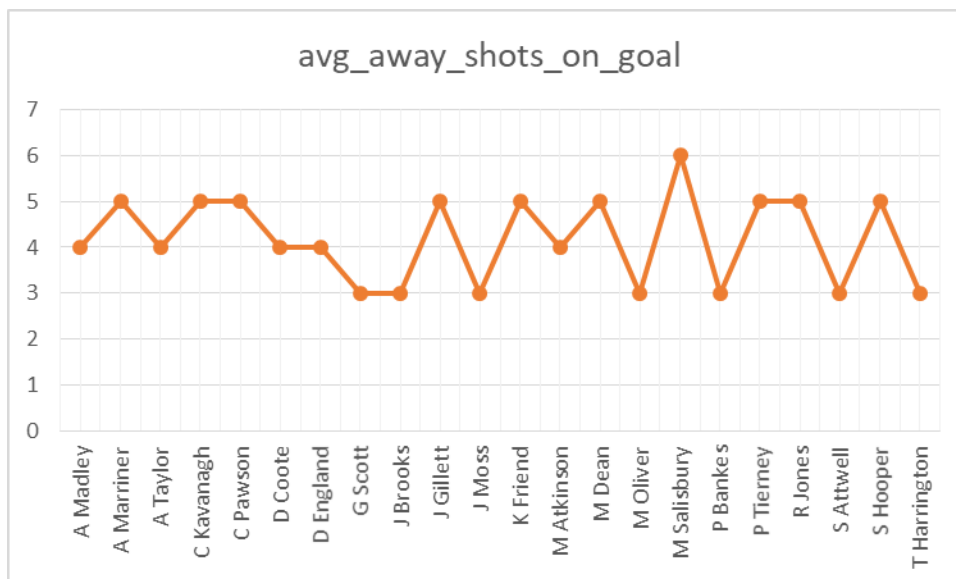Figure 9: Chart showing the average home shots on target per game for individual referees

Figure 10: Chart showing the average away shots on target per game for individual referees.

The chat below shows the average possession for both the home and away team for individual referees, this can also serve as an indicator to spot anomalies in individual referee biases.
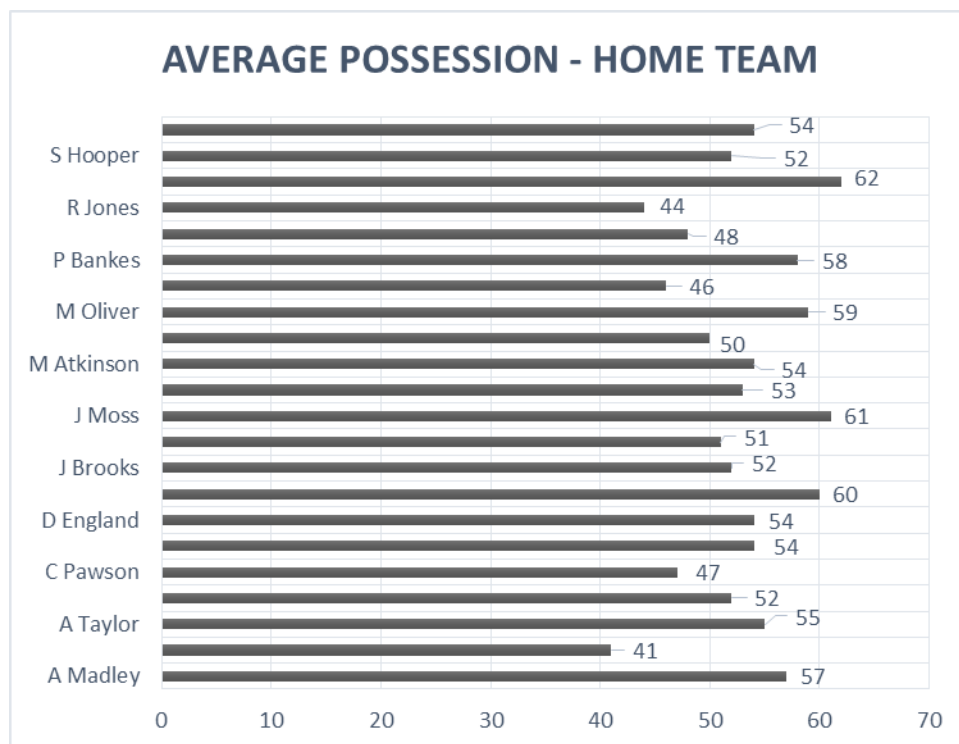

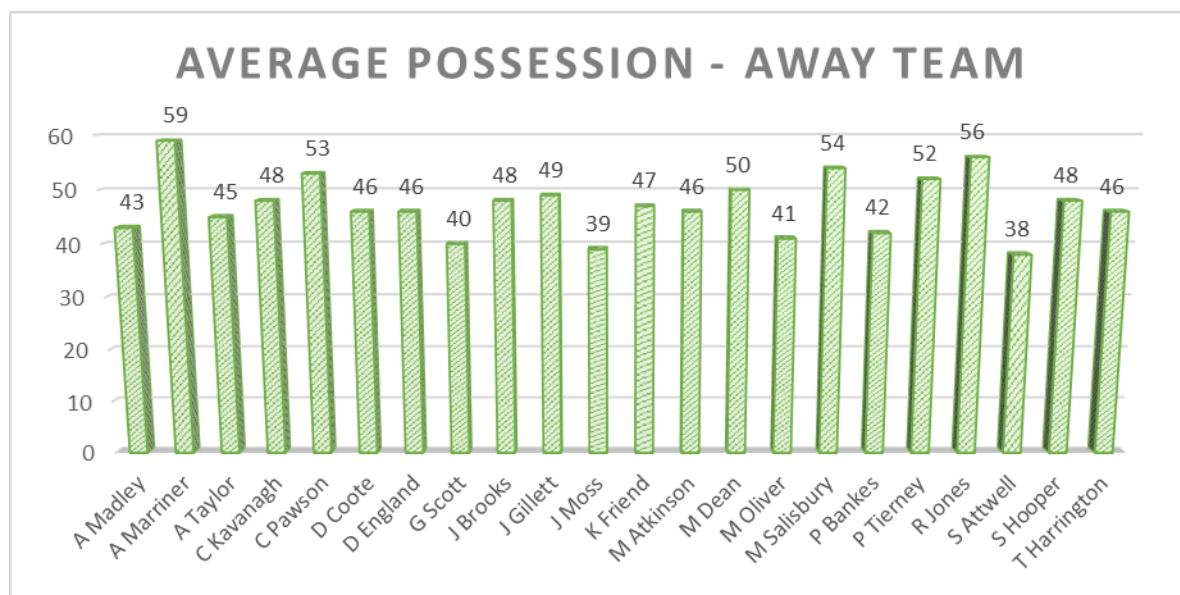Figure 11: Chart showing the average ball possession for the home team for individual referees

Figure 12: Chart showing the average ball possession for the away team for individual referees

Average home/away team winning percentage per game for individual referees

Table 4 below was queried in SQL Server Management Studio (SSMS) to assess the likelihood of the home and away teams winning each match, as influenced by individual referees, expressed as percentages, during the 2021/2022 English Premier League (EPL) soccer season.
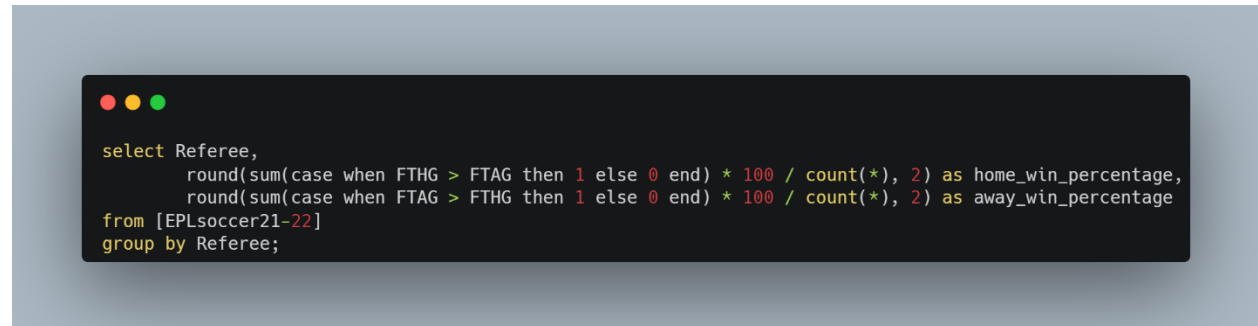
```
select Referee,
        round(sum(case when FTHG > FTAG then 1 else 0 end) * 100 / count(*), 2) as home_win_percentage,
        round(sum(case when FTAG > FTHG then 1 else 0 end) * 100 / count(*), 2) as away_win_percentage
from [EPLsoccer21-22]
group by Referee;
```

Figure 13: SSMS code querying the rate of home/away win percentage for individual referees per game

Referee	home_win_percentage (%)	away_win_percentage (%)
A Madley	62	25
A Marriner	26	57
A Taylor	42	25
C Kavanagh	33	33
C Pawson	50	34
D Coote	25	40
D England	46	26
G Scott	50	41
J Brooks	25	50
J Gillett	44	33
J Moss	64	24
K Friend	33	50
M Atkinson	53	30
M Dean	34	34
M Oliver	53	26
M Salisbury	66	33
P Bankes	41	33
P Tierney	37	37
R Jones	8	50
S Attwell	60	20
S Hooper	26	33
T Harrington	33	33

Table 4: Showing the home/away win percentage for individual referees per game.

From table 4 above the percentage of home and away wins when individual referees were in charge can be seen from the charts below with M Salisbury and J Moss making the top 10% indicating a higher chance the home team would win whenever they were in charge, and A Mariner having the highest tendencies from the away chart.
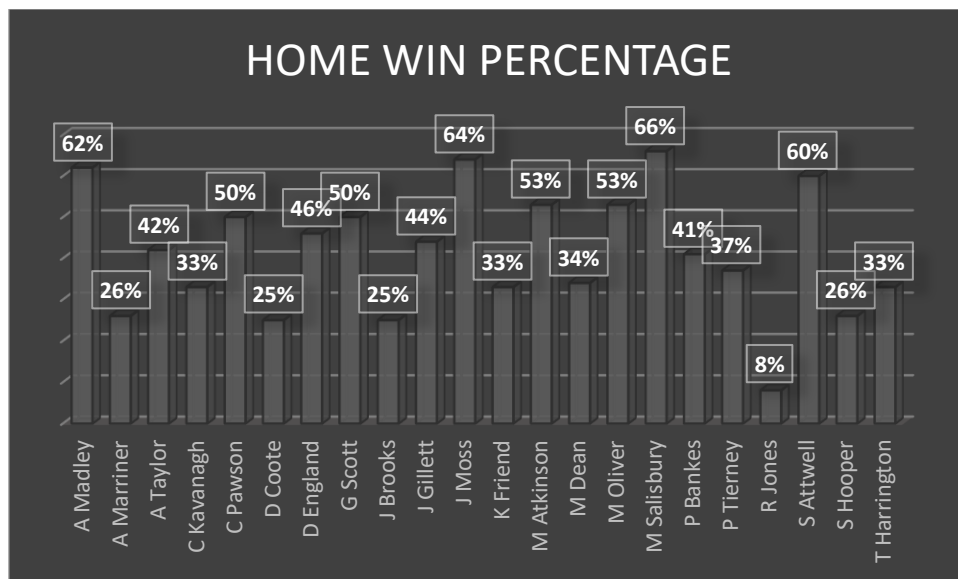

Figure 14: Chart showing the home winning percentage for individual referees per match
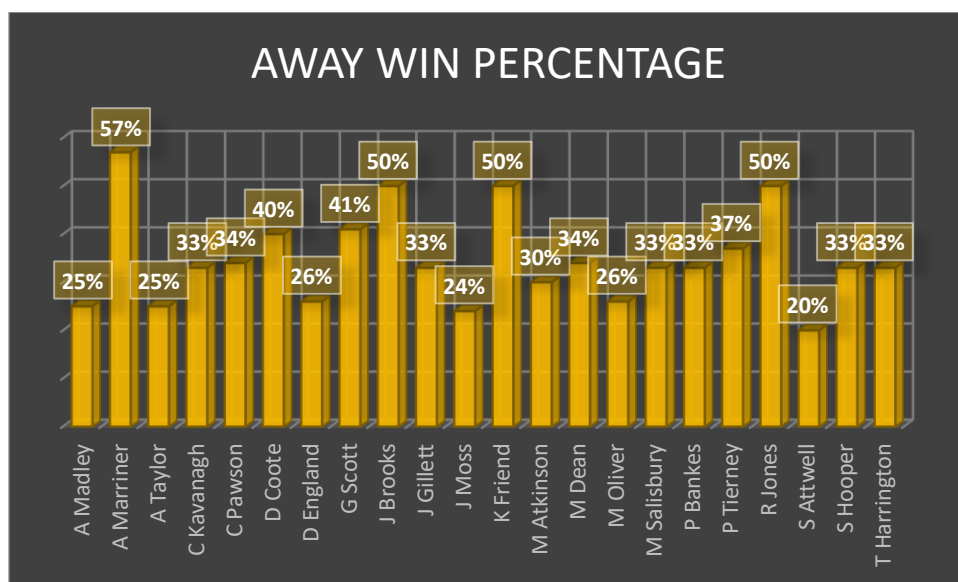

Figure 14: Chart showing the away winning percentage for individual referees per match

IMPORTANT FINDINGS

- The top/most used EPL referees for the EPL's 21/22 season were A Taylor, P Tierney, C Pawson, M Atkinson, M Dean, M Oliver and J Moss which shows the level of importance and trust the regulators have in them.
- M Salisbury and T Harrington more-or-less looked like hired-help from the lower divisions with just three games each.
- J Gillett with nine games officiated had the highest average goals scored per game of 4 goals per game.
- A Madley D Coote and M Salisbury had the highest tendencies to make a foul call as compared to other referees.
- With fouls being the direct cause of a warning card (yellow card) and subsequent sending off (red card), D Coote had the highest fouls average per game amongst all the referees and subsequently made the top podium for yellow cards issued per game, this is a direct indication that the higher the fouls average, the higher to cards tendency.
- M Oliver with twenty six games officiated (third highest) has a 13% chance of issuing a red card in each of the games he officiated in.
- Home teams had the highest amount of shots average when M Oliver and A Madley were in charge. On the other hand the away team had highest average shots when J Gillett, M Salisbury and A Marriner were in charge.
- The home team ball percentage was highest on an average when S Attwell was in charge with 62% of the possession going to the home side in his games on an average. Conversely, A Marriner and R Jones had the highest average from the away column for average possession per game for referee.
- The home side stood a better chance of winning most whenever A Madley, J Moss, M Salisbury, S Attwell due to the higher winning percentage in games officiated by them as seen from table 3. For the away column Andre Marriner had a 57% away winning average making him the top man for away win tendencies.

CONCLUSION

The English premier league is the biggest soccer league in the world and analysis on referee performances and general study of biases as well as anomalies are relevant for all stakeholders involved in preparation for future seasons.

REFERENCE

Badreddine, S., & Spranger, M. (2021). Extending Real Logic with Aggregate Functions. In NeSy (pp. 115-125).

Data source http://data.world