

## INM433 Visual Analytics: Group 3 Coursework

### Group members:

- Sahan Chowdhury
- Murad Noor
- Nicholas Tsioras
- Egemen Yeke

### Dataset:

<https://www.kaggle.com/datasets/thamarsekhri/premier-league-stats-2022-2023?rvi=1>

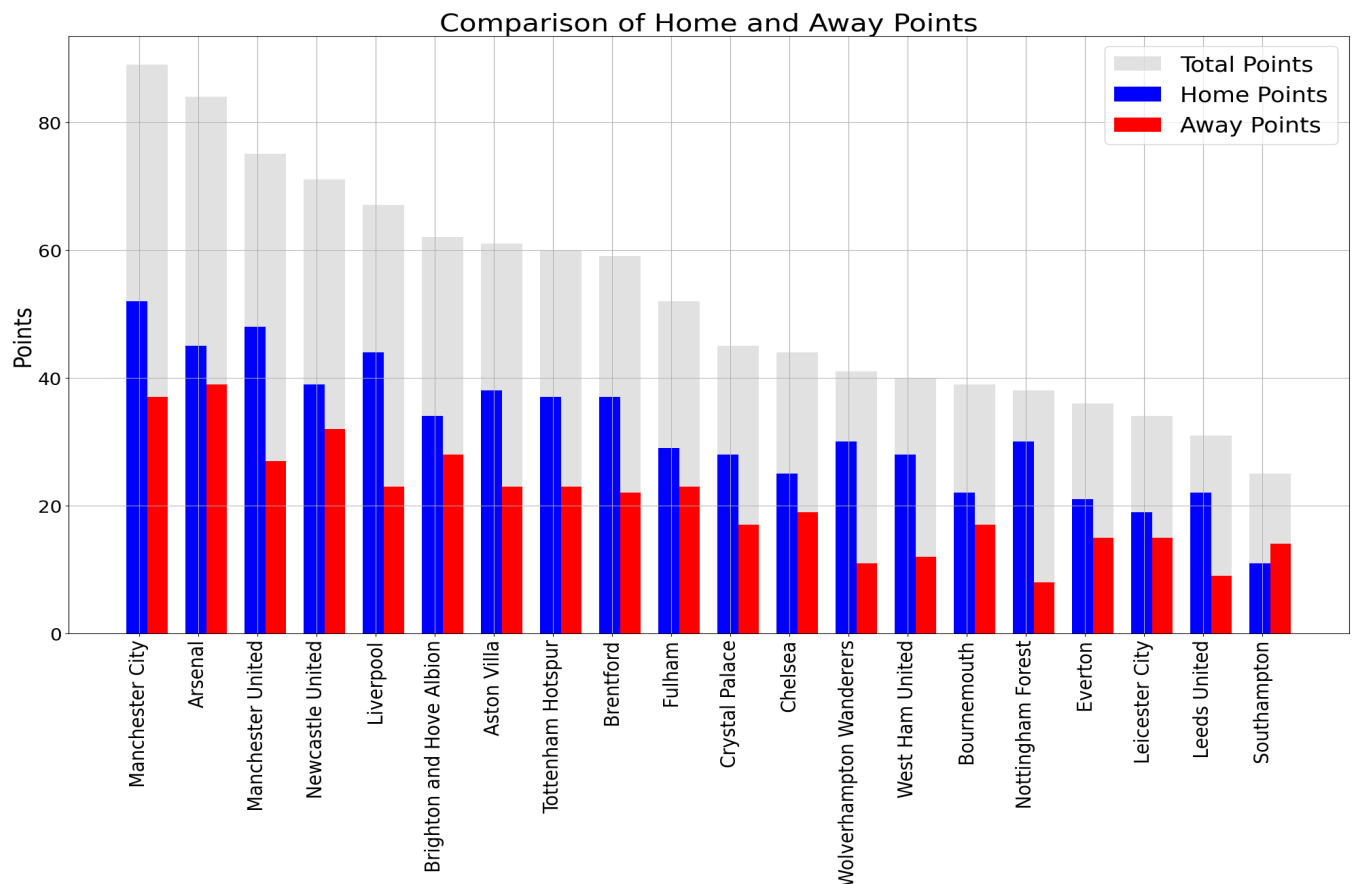
### Research question:

Was there a significant difference in performance between Premier League teams during home and away matches in the 2022/23 season, and can this difference be linked to their final rankings for that season?

### Introduction:

Our dataset displays the 2022–2023 Premier League season by providing comprehensive home and away performance data for every team. Even though the concept of "home-field advantage" is well known, our objective is to dive deeper and examine whether a team's performance differs noticeably between home and away games when they play. We will do this by using visual aids to draw conclusions.. Additionally, we aim to establish if there is a connection between these performance trends and the final rankings achieved by the teams in the 2022/23 season. This analysis uses key metrics, such as possession stats, goals scored and conceded, passing, blocks, saves, attempted shots and fouls, examining how these variables stack up in both home and away fixtures for all 20 teams.

Graph 1: A double bar chart showing comparison of total home/away/total points for all teams

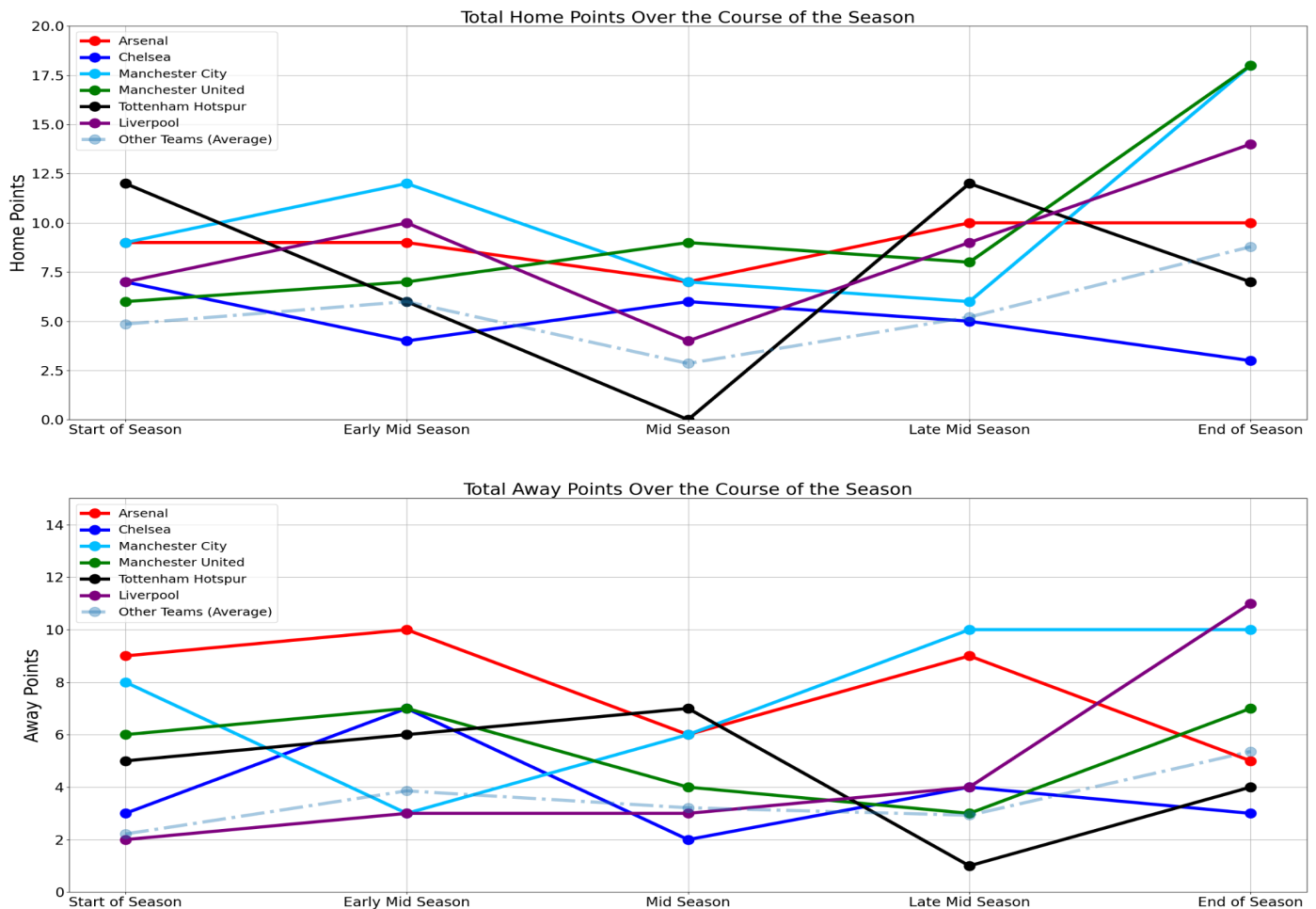


- **Purpose:** Compare the points tally for home and away games in the Premier League for all teams. This will help us see which teams showed more consistency in home and away games and how that impacted their total points tally and their overall league position.
- **Justification:** We use bar charts to compare numerical data and split them to categories. We use 3 different bars for home, away and total points to see how they correlate. We also use different colors for each bar to show the differences between the home and away points.
- **Interpret:** Most of the teams except Southampton got more points at home than away. This shows the home advantage that many teams tend to exploit. Also, there seems to be a positive correlation between the total points at home and away games. This indicates that teams that get more points at home, seem to get more points away from home as well. Liverpool, Tottenham Hotspur, Nottingham Forest and Wolverhampton Wanderers are the teams with the highest point differential from home and away games. This indicates that they are strong at home games but also seem to struggle at games away from home. Arsenal and Manchester City are the 2 teams that got the most points from away games and are also the 2 teams who finished in the league winner and

runner up positions. This shows how crucial it is for teams to be able to pick up points away from home and how strongly it correlates with finishing high up the table. Also, home advantage is crucial for teams, especially for those who are fighting to avoid relegation.

- **Critique:** There is an absence of any distribution line for both home and away games which could help us identify any overall trends that represent the data. Maybe if we added a separate distribution line for home and away games we would be able to highlight differences and make clearer comparisons between the home and away performances of the teams

Graph 2: Line graph comparing the 'Big 6' teams points across different periods in the season to the average of all other teams.

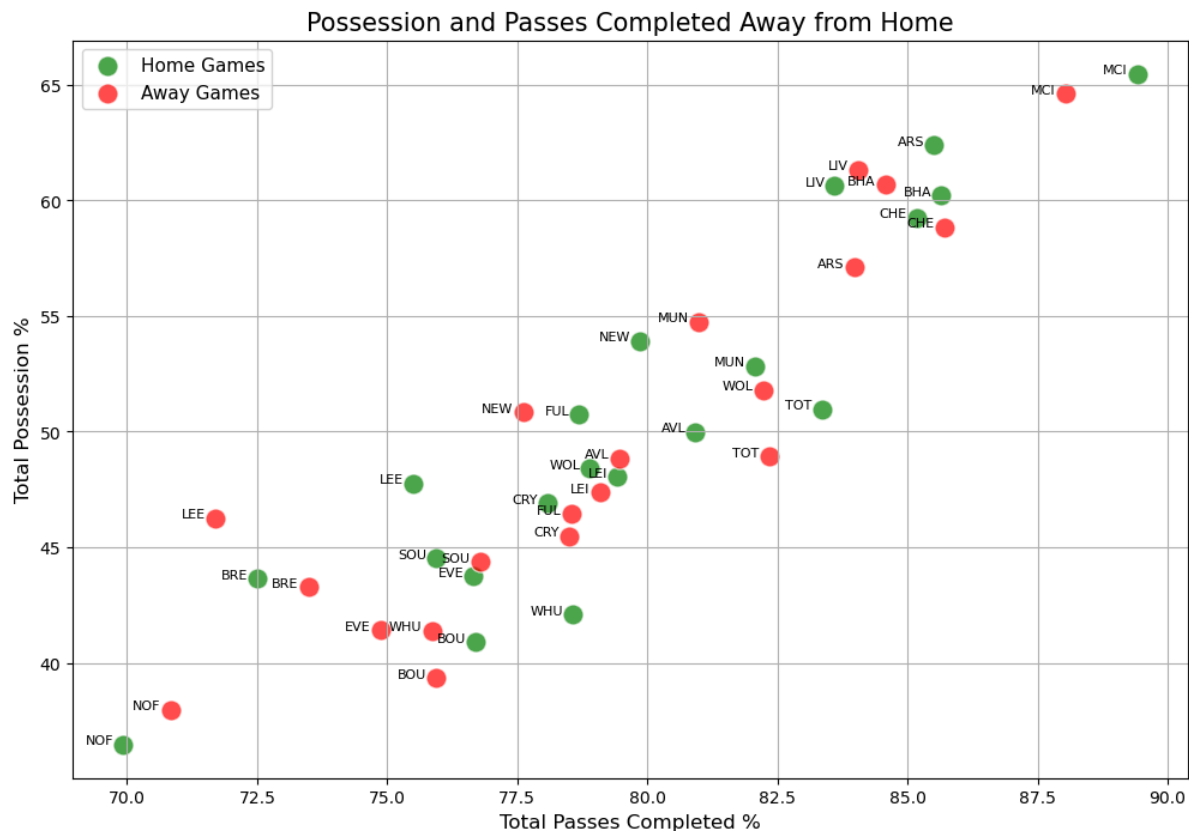


- **Purpose:** We created 2 subplots for home and away games to evaluate the 'big 6' team's performance throughout the season compared to the average performance of the other teams of the premier league. This will show us if any of the teams were consistently leading or trailing in the league in terms of points and we can look for patterns and trends in a team's performances over the course of the season.
- **Justification:** Line plots are particularly useful and well-suited for analyzing time series data. In our case we are analyzing the points throughout the season. Line plots allow us to spot trends, patterns and differences between the data. The reason we only have one line that represents the average of the rest of the teams in the Premier League is that we want to avoid overcrowding. We want the viewers to distinguish between the lines. We chose to show the data for the 'big 6' teams (Arsenal, Chelsea, Liverpool, Manchester United, Manchester City and Tottenham Hotspur) as they are the most popular teams in the Premier League and draw more attention.
- **Interpret:** It seems that most teams in the 'big 6' follow a similar trend in the games at home but not so much in the away games. It seems that most team's performances

declined during the mid season and particularly between early mid season and mid season (October-January) at both home and away games but they had a strong end to the season (April-May). The main reason behind this is fixture congestion as there are more premier league games in midseason in a short space of time which can also lead to many injuries. Tottenham Hotspur seems to be a team that stands out from the rest and follows a distinct trend at both home and away games. All of the 'big 6' teams perform better than the rest of the teams, especially at home except Chelsea who finished 12th in the league. Arsenal and Manchester City are the most consistent teams at home and away games. This is the reason that they finished in first and second place in the Premier League.

- **Critique:** There are a few improvements that could be made in the graph. Maybe it would be better to show the points over the course of the season for all the teams instead of counting the average for the teams outside the 'big 6'. The reason though we chose to do that is so the graph is more simple and easier for the viewers to interpret. So we have a separate line for the teams in the 'big 6' and one extra line that represents the rest of the teams. This does not help us visualize the data for all the teams separately and makes it more difficult for us to do comparison. Also maybe it would be better to show the total points per month so the graph can give us a more detailed and granular view of the team's performances.

Graph 3: Scatter Plot demonstrating the correlation between average total passes completed to the average total possession for all teams in both home and away games

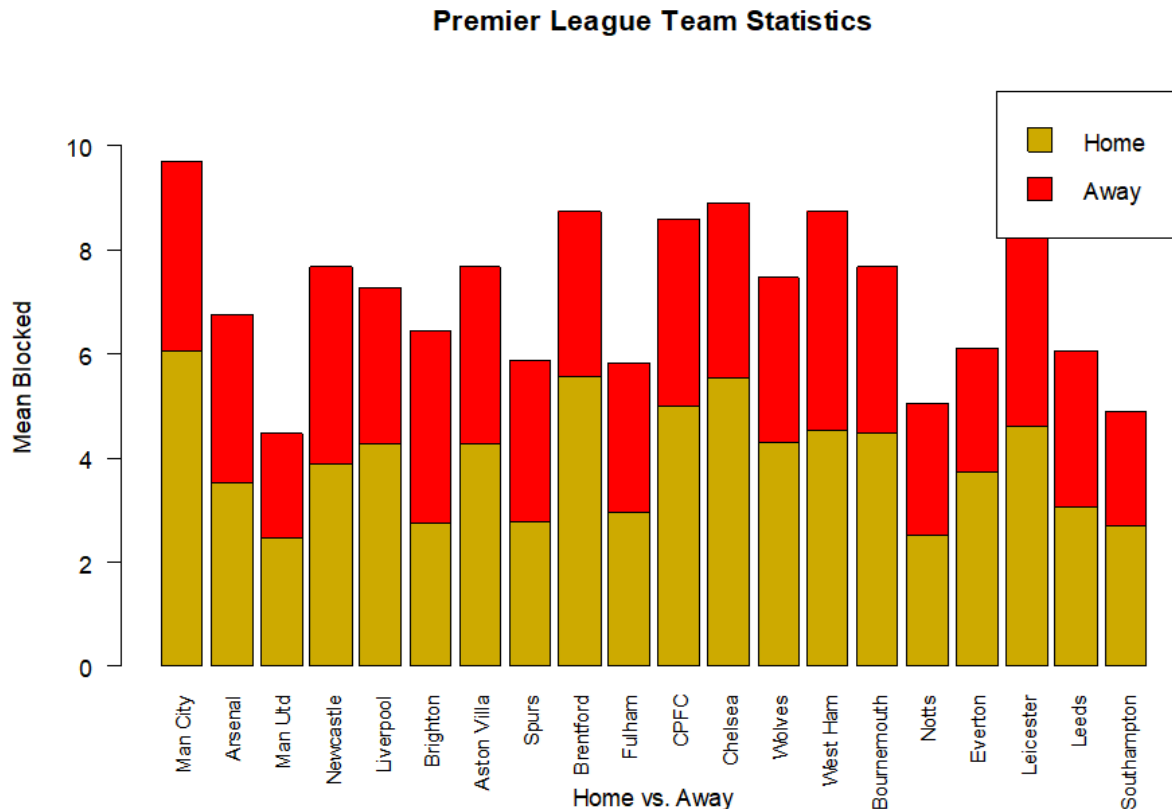


- **Purpose:** In this plot we analyze the possession and pass completion of teams in home and away games. This can provide us valuable insights about the team's strengths and weaknesses and also their playing style. Our goal is to see which teams have a more consistent playing style at home and away games.
- **Justification:** Scatterplots are useful for identifying relationships between numerical values. We use different colors for home and away statistics for better visual clarity so the viewer can understand better the context of the data. We also added the abbreviations of the teams in the plot to distinguish the statistics of each team. We chose to show the average values for both home and away games to avoid overcrowding and identify patterns more easily.
- **Interpret:** There is a high correlation between possession and pass completion which indicates that successful passes are important for maintaining possession. It seems that most teams tend to have more of the ball at home games than at away. This indicates that most of the teams have a more dominant playing style at home games but are more passive in the away games. Nottingham Forest and Man City are the teams that stand out the most, but for different reasons. Manchester City has far more ball possession at both home and away games than any other team in the Premier League. Nottingham Forest on the other hand, seems to be the team with the least amount of possession at both home and away games. Also teams like Arsenal, Liverpool and Brighton are very

dominant in possession at both home and away games, while the majority of the teams have possession between 45-55% and 77.5-82.5% total passess completed.

- **Critique:** It would be better if there was another legend which showed all the team abbreviations. This will be very useful for viewers who do not know the team's abbreviations so they can interpret the graph better. Also maybe it would be better if each data point of the graph was a separate game, instead of counting the averages for each team. This will help us identify variability or consistency in a team's performance and can help us do a more detailed analysis.

Graph 4: Stacked Bar chart comparing the average blocks per game for all teams in both home and away setting

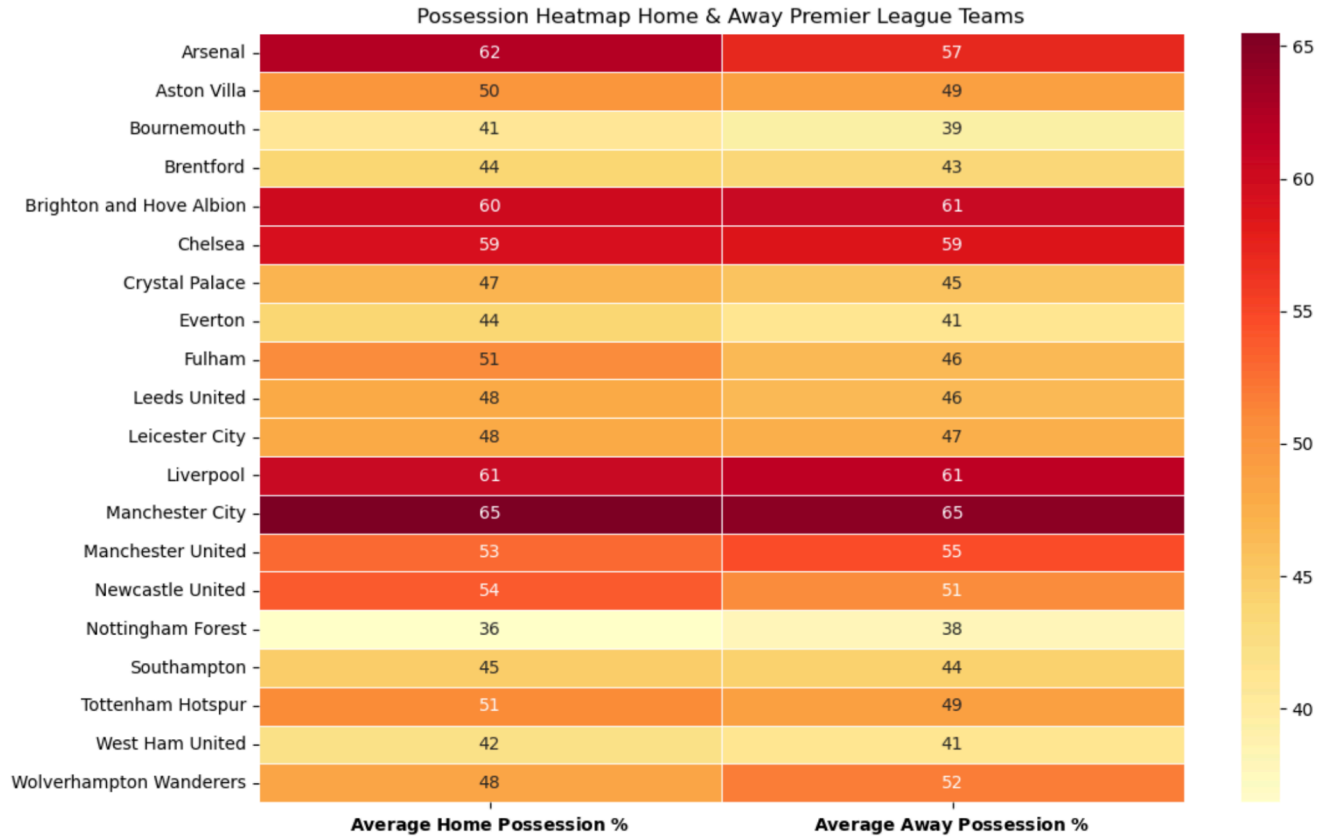


- Purpose:** The purpose of plotting this graph is to study the average blocks by all 20 English premier league teams in both home and away games. We want to see whether the stadium crowds in both scenarios make a difference to the teams defensive performance and also potentially see if their proportion of home and away blocks is linked to their performance across the entire season.
- Justification:** Our justification for using a stacked bar chart for this purpose was to ensure that all the data was in one place on the visualisation for us to analyse. The stacked bar plot was an appropriate option because the data for the x axis of our graph is qualitative nominal, making it easy for us to represent each group of the data set with bars. Also because the data of the y axis is quantitatively discrete, we can very easily divide it into groups. Here we are interested in seeing the ratios between the data of the two groups. I had chosen really dark distinct colours here to make the difference between the two groups clearer to see. I also adjusted the y-axis length so that it is suitable as Manchester City's sum of average home and away blocks was much larger than those of the other teams so the stacked bar plot for that team went above the axis. This was chosen over other graphs like a histogram to allow for wider representation.



- **Interpret:** My interpretation of the presented statistical graphic would be that overall the distribution of blocks at home and away are the same across all 20 premier league teams. Also, In general all teams perform more blocks at home than they do away, showing that a home ground with a supporting crowd and stadium gives the players in the team a big advantage. Looking at the 6 most popular big premier league teams, it is visible that their distribution of home blocks and away blocks is the same as of the rest of the teams in the league playing that season. Chelsea, Manchester City, Tottenham Hotspur, Liverpool and Manchester United have significantly more home blocks than away blocks. Arsenal has approximately equal amount of blocks in both home and away games. Leicester on the other hand have more away blocks than home blocks. These two cases are unique when we look at the rest of the data visualisation. There are no clear outliers in this data which is good because there is less likely to be skew in the visualisation meaning that it is accurate and representative of the trends seen in performance by all of the teams across the season.
- **Critique:** A problem with this graph is that it is a bit too simple, as it condenses the information a lot, not allowing us to see the entire spread of home and away blocks in all the matches, which prevents us from interpreting possible variations in our dataset such as number of blocks across matches. The choice to use averages is a tradeoff for concise representation over inclusion of detail. Additionally there is an absence of any distribution line in this graph to show overall trends that represent the data. This forces us to look at the positioning of objects on the graph ourselves to determine patterns within the data shown to us graphically in the graph. The top of the bar for Leicester City is not visible because of the graph's colour key. Also because the graph only includes averages we cannot see the central tendency or statistical dispersion.

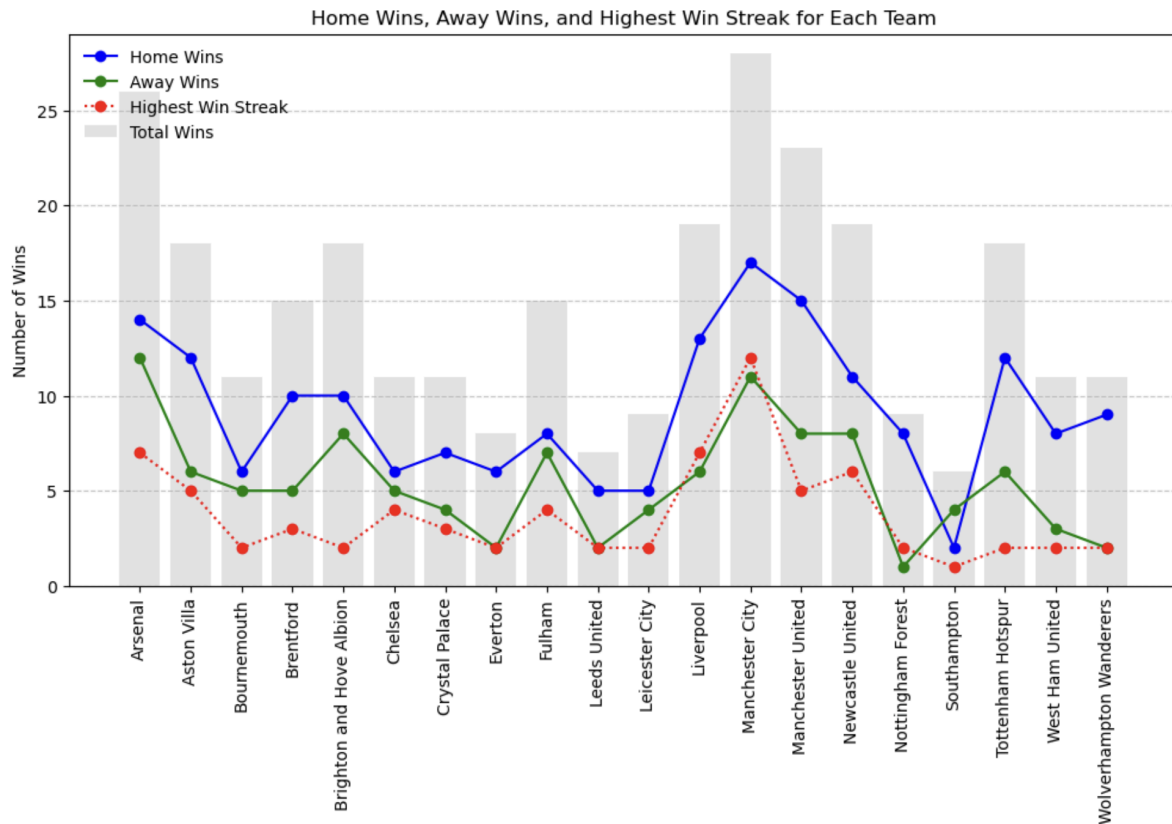
Graph 5: Heatmap of the average possession per game for all teams for home and away games



- Purpose:** The purpose of this heatmap is to visually represent the average possession per game for all Premier League teams during both their home and away fixtures in the 2022/23 season. It allows us to visually see and compare the possession statistics across different teams. Darker cells indicate a higher percentage of ball possession per game and the lighter cells suggest a lower possession.
- Justification:** Heatmaps are effective for visually representing average possession percentages using color intensity, displaying the comparison of possession variations. They are well-suited for swift data analysis, and are very simple to understand.
- Interpret:** Examining the heatmap, we notice an abundance of orange cells, indicating that the majority of teams fall within the 45%-50% range for average possession. The similar color intensity for both home and away games suggests a consistency in possession statistics amongst all teams. This also suggests that in most games teams had closed matched possession statistics. Notably, teams such as Manchester City, Manchester United, Liverpool, Brighton, Chelsea, and Arsenal display higher color intensity, implying their superior possession percentages. This aligns with their rankings for the 2022/23 season, except for Chelsea (finished 12th). Therefore, it is reasonable to conclude that teams that displayed higher possession stats for both home and away matches tend to secure higher positions in the league.

- **Critique:** Despite the fact heatmaps provide a high-level summary, it does not offer detailed possession figures for each team, and where on the pitch this possession is prominent for every individual team, as this would give greater depth and understanding as to where teams dominate on the pitch when in possession. It's worth noting that heatmaps may not accurately depict the extent of differences in possession statistics between teams, especially when colors/shades are similar. Additionally, they do not account for outliers or extreme possession values, which can potentially distort the overall perception.

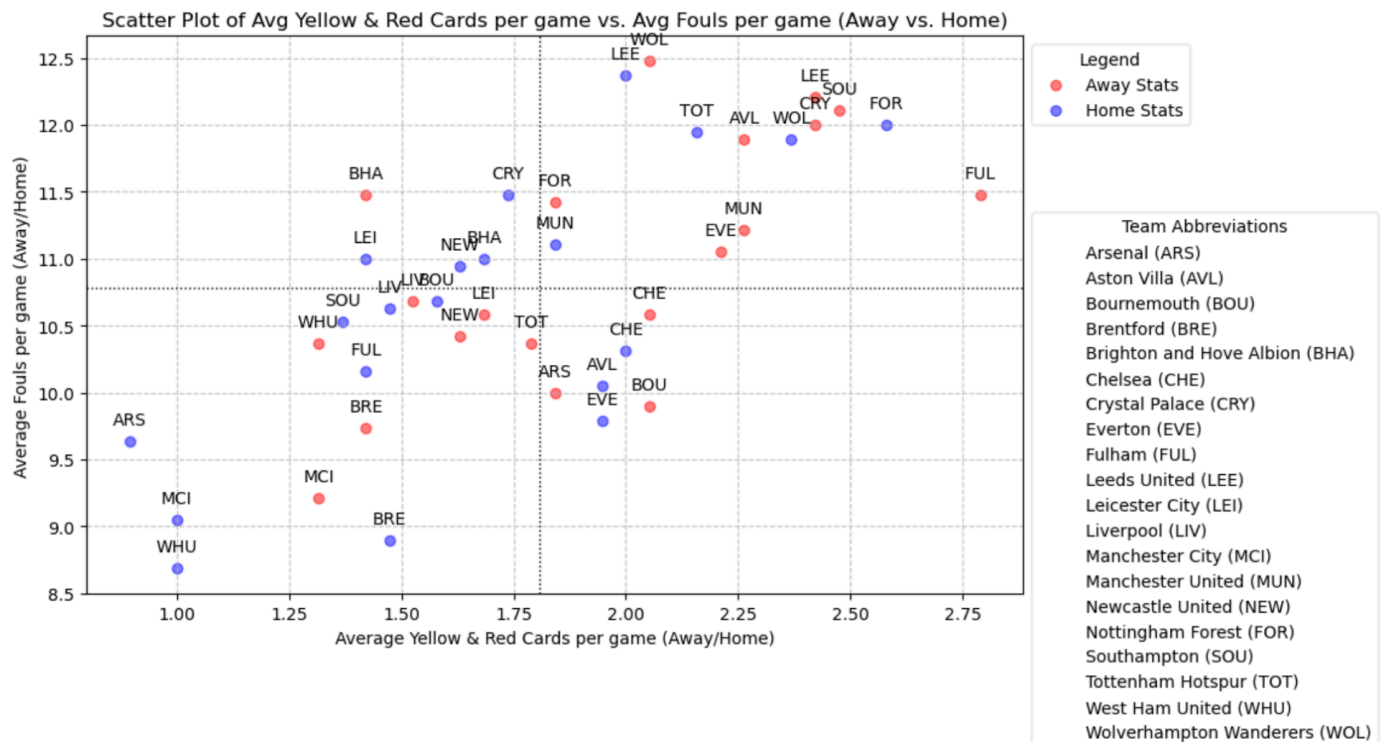
Graph 6: Line graph & bar chart combination. Solid Line charts showing number of wins per team. Dotted line showing each teams highest win streak (highest number of consecutive wins), and faded bar chart showing every teams total wins.



- Purpose:** The purpose of this Line graph & bar chart combination is to provide a comprehensive overview of Premier League teams' performance by visualizing their number of wins. This visualization allows for a more in depth comparison of teams based on the total wins in away and home games and identify the teams with the highest win streaks (consecutive wins), and understand how these factors contribute to each team's performance in the league. The line graph is effective in the manner we quickly spot trends, and the faded bar chart is there to provide an overall summary.
- Justification:** The Line graph & bar chart combination is an effective choice for this task as it allows for a simultaneous representation of two key performance indicators - total wins and win streaks, and difference in away and home wins - for each Premier League team. The solid line graph provides a clear visualization of the number of wins, enabling straightforward trend analysis, while the dotted line highlighting win streaks offers insights into the teams' consistent success. The faded bar chart complements the line graph, offering an overview of total wins for all teams. This combination provides a clear view of team performance and enables easy comparisons and trend identification.

- **Interpretation:** From the graph we can say that there is a positive correlation between home wins and away wins. Teams with more home wins also tend to have more away wins, and overall teams win more games at home, signifying the home field advantage. The only exception is Southampton whose number of away wins exceed home wins. Southampton also finished last, and had the lowest number of home wins (2). This indicates the importance of home wins. There is also a positive correlation between home wins and win streak. The win streak line graph tends to mirror the trend of the home win line graph. On average the majority of teams had a win streak of 5 or less, except Arsenal, Manchester City, Newcastle United, Liverpool, who all finished within top 5 during the 2022/23 season. All teams that had more than 10 home wins have finished within the top 6 of the premier league table, amongst them Manchester city had the highest number of total wins, home wins and win streak, and only narrowly beaten by Arsenal in away wins. Manchester city alongside Nottingham Forest were the only 2 teams whose win streak exceeded the number of away wins. Again justifying Manchester City's prominent league standing.
- **Critique:** The faded bar chart, while useful for providing an overall view of the data, may not be as effective as the line graph when it comes to making precise and detailed comparisons. The line graph offers better visual clarity for examining win streaks because it provides a continuous visual representation of the data. In contrast, the faded bar chart presents discrete bars, which might make it harder for viewers to accurately distinguish and compare specific data points. Furthermore the inclusion of multiple variables in a single graph slightly clutters the visual presentation. Maybe in order to get a better understanding we could do a line graph comparison of wins vs losses, as looking at just wins on a line graph does not depict the full story.

Graph 7: Scatter plot to show correlation between average yellow and red card combined per game vs average fouls per game.



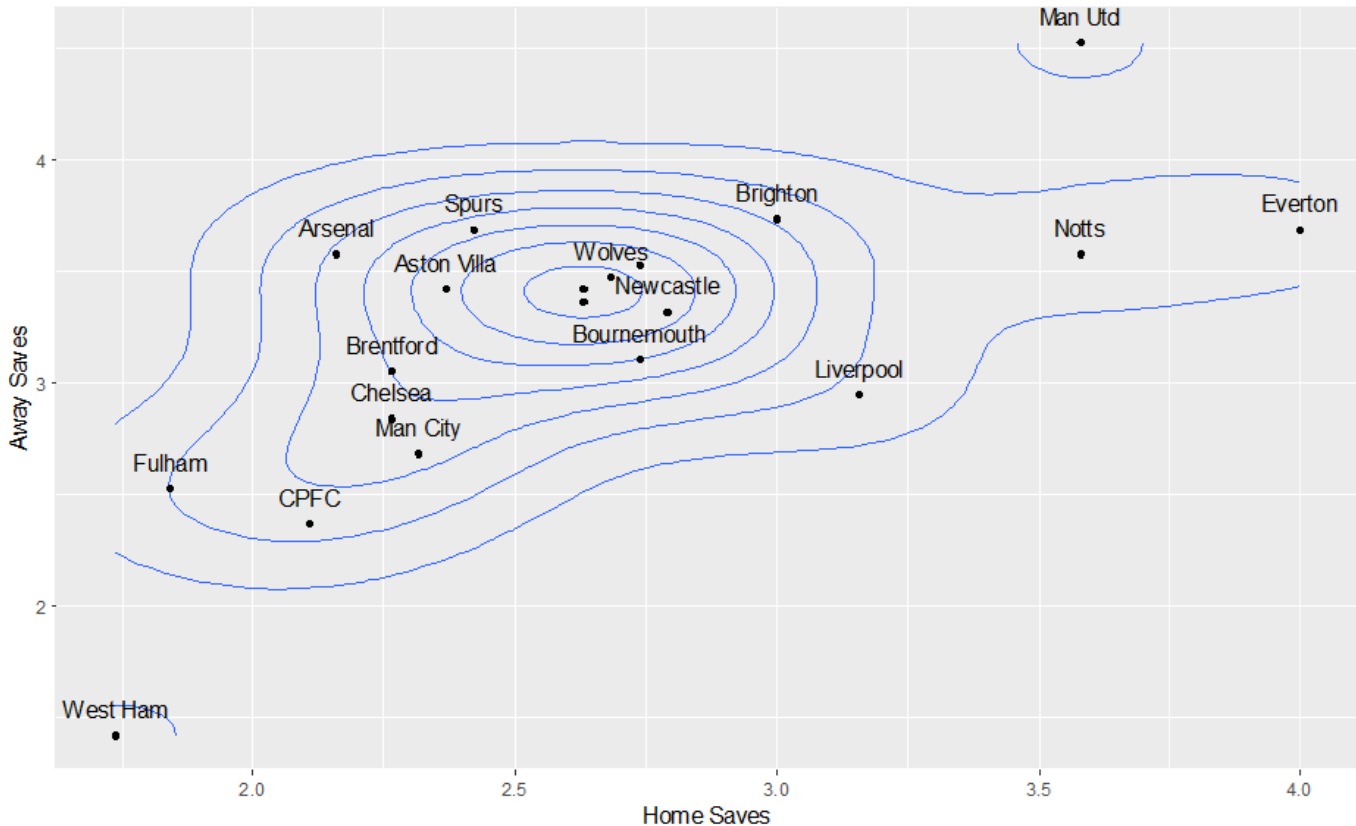
- Purpose:** The purpose of creating a scatter plot is to visually represent and analyze the correlation between the average number of yellow and red cards combined per game and the average number of fouls per game for Premier League teams during the 2022/23 season. This visualisation aims to showcase insights on the relationship between a team's disciplinary record, shown by cards, and their playing style, which is reflected in the number of fouls committed.
- Justification:** A scatter plot is the ideal selection for this analysis as it effectively illustrates the correlation between the average number of fouls and the average number of cards per team. Fouls represent unfair offenses committed against opponents, and based on their severity and recurrence, they result in either a yellow (caution) or a red (dismissal) card. Utilizing a scatter plot offers the advantage of enhanced precision in assessing this relationship, and allows evaluation of multiple variables within a single graph. Additionally, we have distinguished the plot by using colors to differentiate between home and away data points, with team abbreviations accompanying each data point.
- Interpretation:** One of the first things we notice is that the majority of blue data points are on the bottom left quadrant, this indicates teams obtain low cards and low fouls when playing at home. This suggests that playing styles at home are more disciplined and less aggressive. On the other hand, away games (red data points) are more scattered around the centre. The central line is approximately 1.85 cards per game and 10.8 fouls per

game. However, Fulham (FUL) in away games is an outlier. It has the highest number of cards per game for away games of approximately 2.78 cards per game and 11.49 fouls per game. Another trend we notice is all teams either have a higher average card or average foul per game in away games, except, Tottenham (TOT), Nottingham Forest (FOR), Newcastle (NEW) who defy this condition. There doesn't seem to be a correlation between teams home and away cards vs fouls and their final standings as every team has its own playing style. Wolverhampton Wanderers (WOL) and Leeds United (LEE) both have a very high cards per game vs fouls per game for both home and away, whereas Manchester city (MCI) has the lowest card vs fouls per game for both home ground and away ground.

- **Critique:** It's important to note that the scatter plot may be susceptible to overcrowding, potentially affecting the ease of data interpretation and leading to some data points overlapping, which in turn could reduce the effectiveness of the scatter plot. In addition to this it's much more difficult to quickly spot trends from glance.

Graph 8: A 2D (Two Dimensional) Density Plot to show the correlation between the average saves in away games vs the average saves in home games for all 20 teams in the premier league.

2D Density Plot of Home and Away Saves by Team with Labels

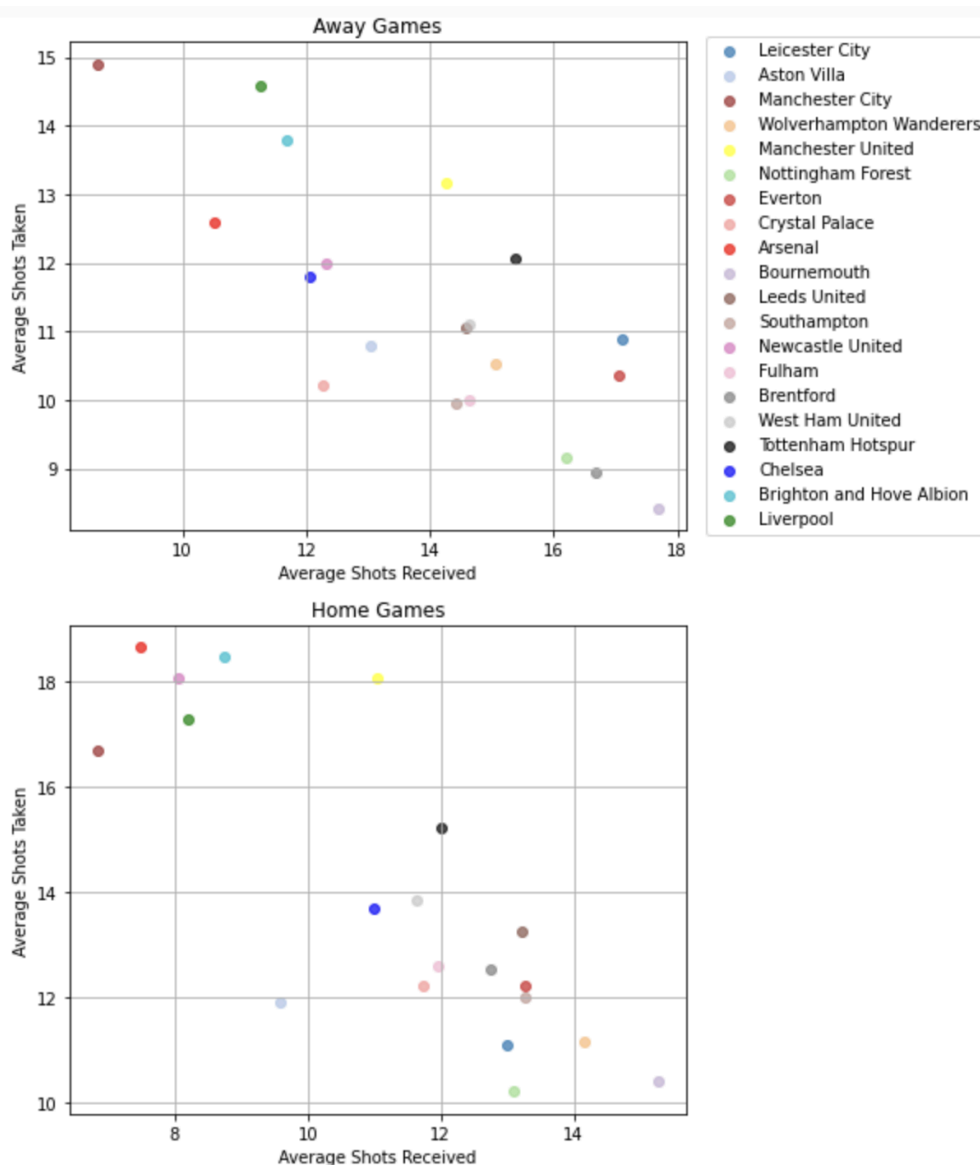


- **Purpose:** The purpose of using a 2d density plot was to see where each one of the 20 premier league teams stand in terms of their goalkeepers ability to save goals at their home stadium compared to their ability to save goals at away stadiums. Also additionally, see how all of the teams compare with each other in this performance metric and what general trends there are.
- **Justification:** A two-dimensional density plot is the most appropriate in this scenario because it shows the distribution of home and away saves across the set of teams, allowing us to analyse the correlation between the two variables to see if there is any relationship between the two categories. It can even make it easier for us to classify groups of teams in the entire set to see which teams are similar to each other by looking at the level contour lines on the graph. The 2d density plot with contours is suitable for our situation because we have very few data points and thus there is no risk of overfitting present. This is much better than for example a Hexbin because of the size of data involved in plotting the graphical representation, also it is generally more accessible due to ease of reading as viewers can more accurately judge point positions.



- Interpret:** Firstly, what we can very clearly see from the graph plot is that two teams are outliers from the overall trend but are very different from each other at the same time. These two teams are Manchester United and West Ham. Manchester United's goalkeeper performs a lot of saves both home and away though slightly more away. Whilst at the same time, West Ham's goalkeeper does not perform a lot of saves either home or away. This is expected since two of these three teams are part of the top 6 in the league for that season. We can see that groups that on the same level curve have roughly the same ratio of home and away saves even if one has slightly more than the other one. There is an expected overall positive correlation with the data where more home saves means that there are more away saves resultantly. Some teams deviate from this trend slightly however not in any way that is unusual. The spacing in between the contour lines slowly increases as the points go further from the center of the plot. This shows that there is an increasing difference between the teams consistency of performance as the gaps between home and away ratios get bigger. This is reflected in the points table of the premier league, where there is no linear trend between the points difference of teams on consecutive rankings.
- Critique:** There are some improvements that can be made to this graph. One of them being that there could be a colour scale for the different concentric circles on the graph similar to a temperature gradient representing the increase in home and away goal ratio across the graph plot. Also, the x and y axis could have been extended to make the graph plot big enough to include all of the contour lines to show all of the trends of home and away save ratios more accurately, and so that an analysis of possible outliers could have been included in the description. Also because only the averages of the data are plotted here, data modality is not visible.

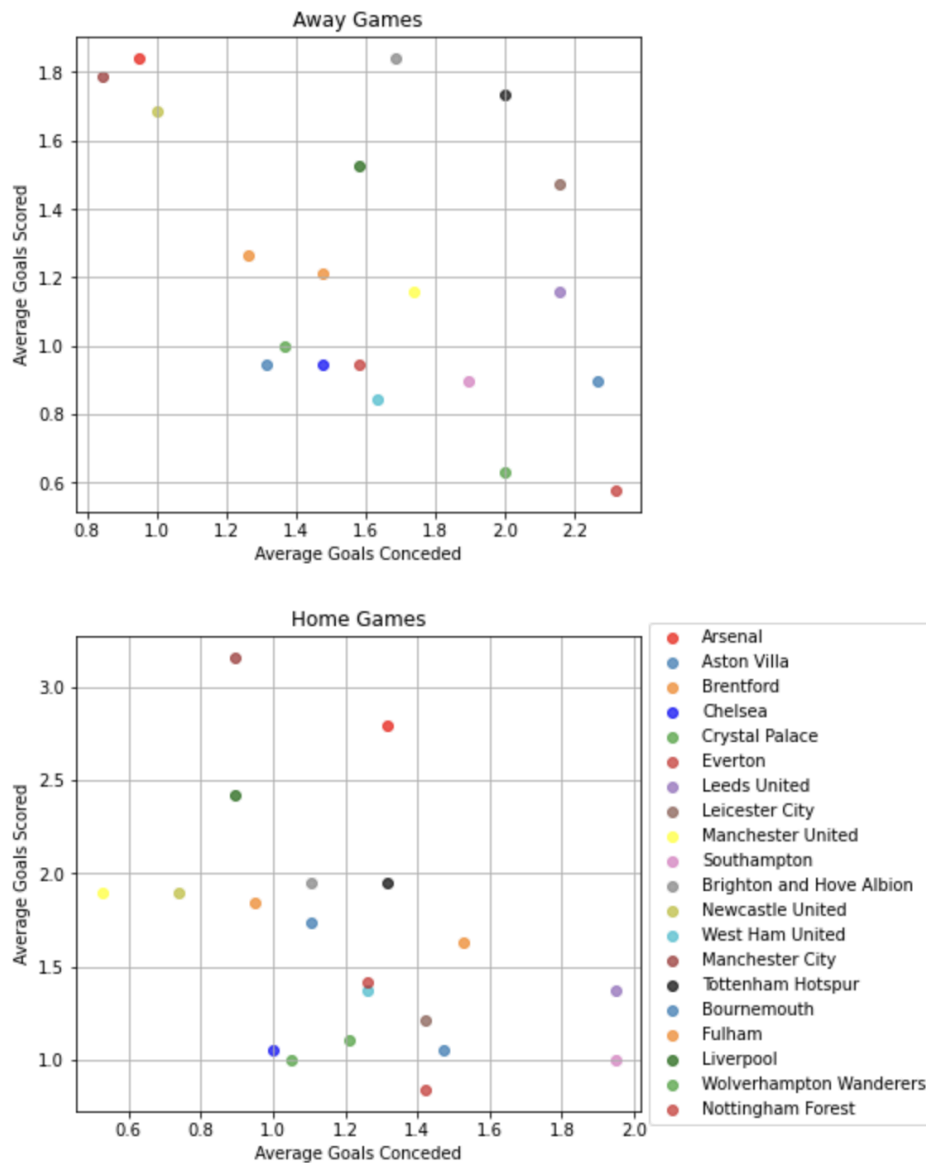
Graph 9: Scatter plot showing the average shots taken and received for both home and away games for all teams in the Premier League



- **Purpose:** The purpose of the scatter plot is to compare the average shots taken and received by all teams for home and away games for the 2022-23 Premier League season.
- **Justification:** A scatter plot is very useful for comparing two numerical variables. In this case, average shots received and average shots taken are the two metrics we want to compare for both home and away games. Different colors are used to distinguish between data points and a legend is used to identify the teams. Subplots are used to visualize the home and away games together. Also the simplicity aspect provides an easy understanding of the visualization.

- **Interpret:** The scatter plot displays that in general more shots are taken by teams at their home games as opposed to away games. Moreover, teams generally receive less shots at their home field. These observations indicate that teams tend to create more shots and limit the opposition's shots more on their own turf. Another insight is that there is a bigger gap between the top and other teams in their home games compared to away games. The top 6 finishers (Manchester City, Arsenal, Manchester United, Newcastle United, Liverpool, and Brighton) are separated from the other teams indicating that home game shot taking and receiving is an important factor for final rankings. Tottenham and Chelsea on the other hand, are both near the middle of the graphs, which correlates with their final rankings of 8<sup>th</sup> and 12<sup>th</sup>.
- **Critique:** The utilization of subplots were very beneficial for visualizing home and away games together. However, a problem is that they occupy unnecessary space on the page.

Graph 10: Scatter plot showing the average goals scored and conceded for both home and away games for all teams in the Premier League



- **Purpose:** The purpose of the scatter plot is to compare the average goals scored and conceded by all teams for home and away games for the 2022-23 Premier League season.
- **Justification:** A scatter plot is a valuable tool for analyzing two numerical variables. Here, we aim to assess the relationship between the average goals conceded and average goals scored for both home and away games. The use of various colors

distinguishes data points, while a legend helps identify the teams. Employing subplots enables a clear comparison between home and away games.

- **Interpret:** The scatter plot indicates that teams tend to score more goals and concede less at their home games compared to their away games. We can also observe that the top 2 league finishers, Manchester City and Arsenal scored more than the rest of the teams in home games. This indicates that having a strong performance at home is beneficial for a top placing. Most teams conceded between 1 to 2 goals on average and the only teams to concede less at are Manchester City, Newcastle United, Brentford, and Leeds United. Moreover, most teams scored below 2 goals on average for home and away games, except Manchester City, Arsenal and Liverpool. This suggests that teams at the higher end of the table are more clinical at home games.
- **Critique:** A downside of this graph is that plotting averages of scored and conceded goals prevents us from observing the dispersion of the data.

## Summary/Conclusion

In conclusion, our analysis proved that teams performed better when they played at home as opposed to away. The concept of home-field advantage in football is evident through our analysis. A major factor in Manchester City, Arsenal, and Liverpool's success in the 2022–2023 season was their ability to win more games, accumulate more points, and establish dominance in the majority of them at home.

From our observation it was evident that home performance was a major deciding factor on every team's success. Teams such as Manchester City and Arsenal, dominated most home matches but also dominated the most away matches and as a result justified their final ranking in the 2022/23 league table. However there were exceptions such as Fulham who dominated mainly at home games and still finished mid-table (10th place). This highlights the complexity of the premier league.

Our analysis shows that teams generally perform better when they're playing at home, and as a result, home-field advantage is a big part of their success. However, doing well in away games is also equally important and often links to a good league position. In the future, we can make our analysis better by looking at other stats, injuries, and how teams are coached.