Middlesex University Dubai

Business Intelligence
Course Work 2

English Premier League Data

(2014-2020)

Muhammad Ehtesham Bahoo
M00920464
Miss Chinnu George

Table of Content:

Introduction:

Introduction:

Welcome to the dataset containing match statistics from 2014 to 2020. This dataset provides a comprehensive collection of information related to football matches, allowing enthusiasts and researchers to delve into the metrics that impact the game. The dataset encompasses various aspects of matches, including possession, passing success, shot statistics, clearances, fouls, cards, team information, goals scored, and the corresponding years.

By exploring this dataset, one can gain insights into the performance of teams, analyze their possession percentages, evaluate their pass success rates, assess their shooting abilities, and track their defensive actions. Additionally, team ratings are included, allowing for comparisons and correlations between team performance and match outcomes. Furthermore, the dataset enables the examination of foul counts, yellow and red cards issued to teams, and the goals scored by both the home and away teams.

Data Analysis and Visualization:

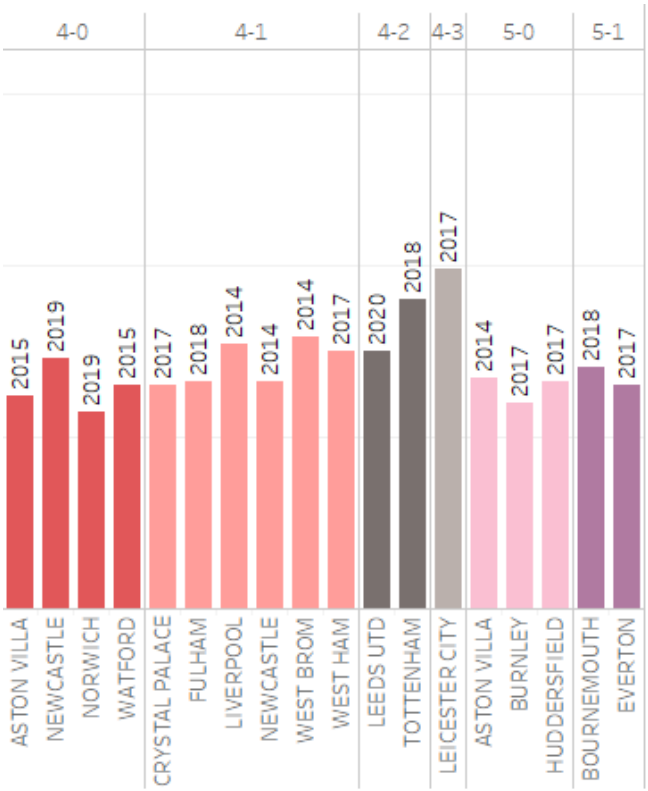
This section presents Tableau figures to visually represent the dataset, enhancing understanding and providing valuable insights. The interactive and user-friendly visualizations offer a comprehensive exploration of match statistics, shots, fouls and cards, team information, and years. These Tableau figures unlock valuable insights and make the data more accessible and engaging for football enthusiasts, researchers, and sports analytics enthusiasts.

Question 1: How does the number of goals affect Match excitement.

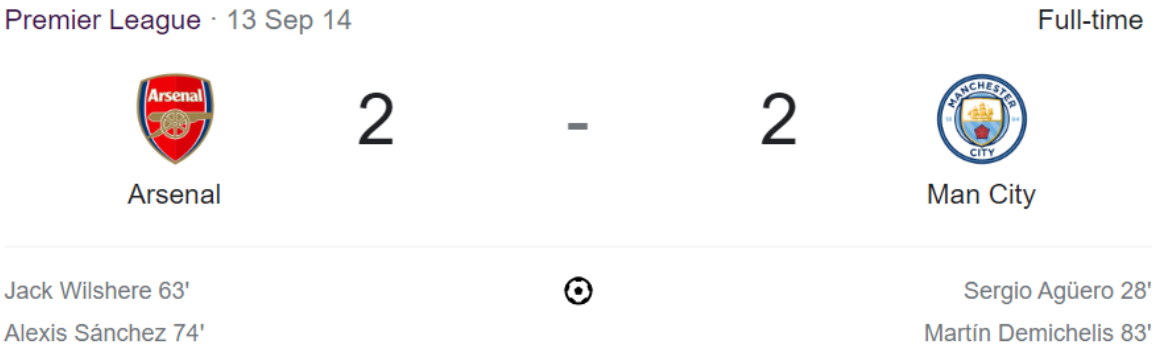
This Part of the report aims to explore the relationship between the number of goals scored and the level of match excitement in Arsenal football matches. By analyzing the dataset, we can gain insights into how goal scoring impacts the excitement level of the matches. Specifically, we will focus on Arsenal as the home team and examine various scenarios to understand the factors influencing match excitement.

Among the matches featuring Arsenal as the home team, the most common game score was **2-0**, where Arsenal scored two goals. These matches had a normal level of excitement. This can be attributed to the fact that most opposing teams had lower ratings than Arsenal. However, it is noteworthy that games against Manchester United generated higher levels of excitement. This could be due to the historical rivalry between the two teams or their close positions in the league table, both fighting for qualification in the **UEFA Champions League**.

Contrary to expectations, high-scoring matches did not necessarily guarantee elevated levels of match excitement. For instance, victories against Everton (5-1), Burnley (5-0), and Huddersfield (5-0) did not generate significant excitement among viewers. This could be attributed to the relative positions of the teams in the Premier League table. Chelsea finished 5th, Burnley finished 7th, Everton finished 8th, and Huddersfield finished 16th, leading to their relegation. The



matches against these teams, despite the high goal count, did not involve teams of similar ratings or intense competition, resulting in a lack of excitement.



The Most Exciting Match: Arsenal vs. Manchester City (2014/2015):

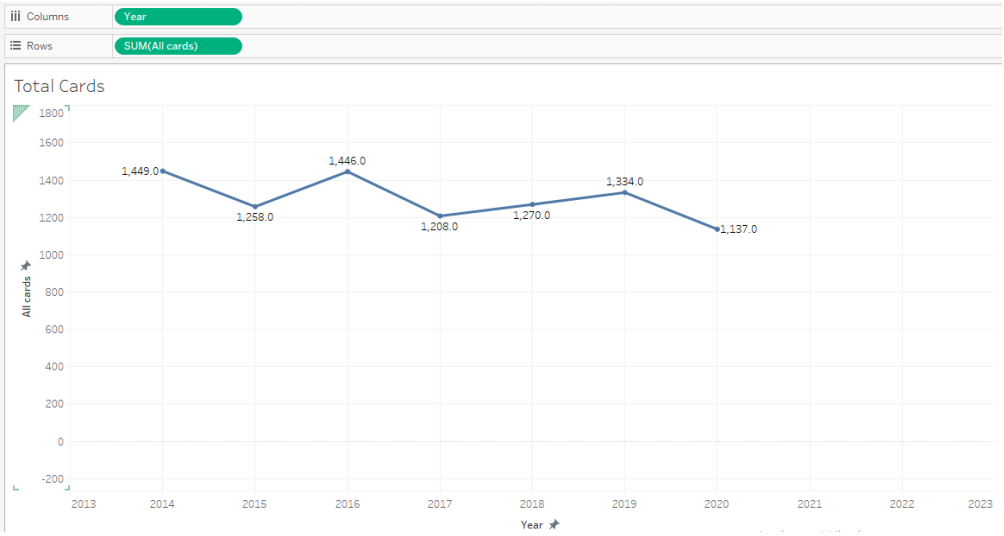
One particularly thrilling match was Arsenal's encounter with Manchester City during the 2014/2015 season. The match ended in a draw, but it was filled with excitement. Sergio Aguero scored the opening goal for Manchester City in the first half. However, Jack Wilshere equalized for Arsenal in the 63rd minute, followed by Alexis Sanchez putting Arsenal in the lead in the 74th minute. The excitement continued as Martin Demichelis scored an equalizer, resulting in a draw. This match stood out due to the similar team ratings and the intense back-and-forth nature of the game.

Based on the analysis of Arsenal's matches, it is evident that the number of goals scored does not singularly determine the excitement level of a match. Factors such as historical rivalries, team ratings, and the competitive nature of the game play significant roles in shaping match

excitement. Matches against top-rated teams or those with closely positioned teams tend to be more thrilling, irrespective of the final score. The dataset provides valuable insights into the relationship between goal scoring and match excitement, highlighting the multifaceted nature of football matches.

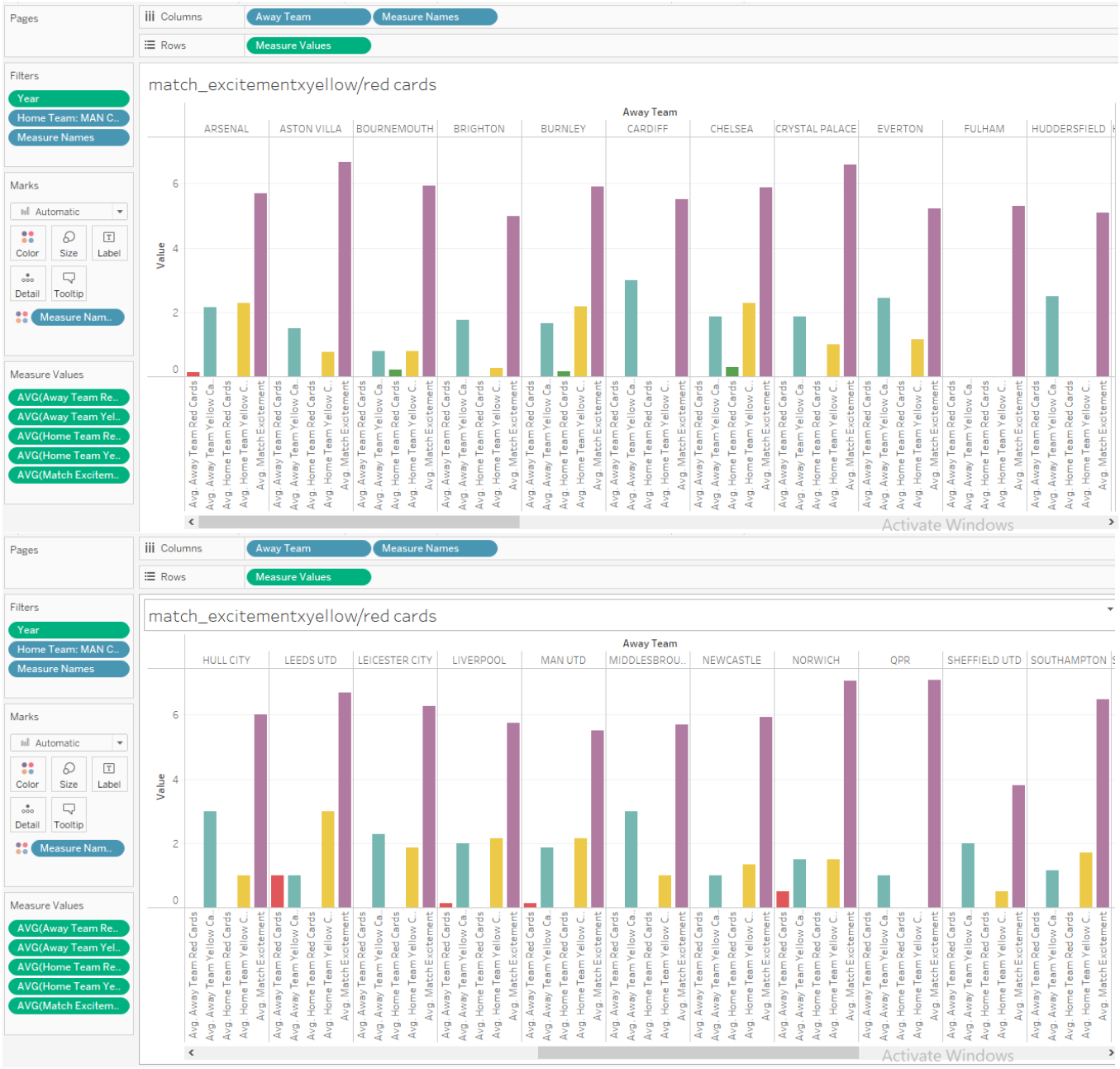
Question 2: How does the red cards and yellow cards affect Match excitement?

This analysis's objective is to explore the influence of red and yellow cards on match excitement. By examining the dataset and visualizing the card distribution, we can gain insights into how different card counts affect team performance and the overall quality of the match. The assumption is that an increased number of cards issued during a game should result in more thrilling matches, as it introduces elements of suspense and potential disadvantage for the penalized team.



Here we can use this visualization to check which year had the most red/yellow/2nd yellow cards so we are choosing 2014 as our calendar year to check how cards affect match excitement.

When considering the measures considered for visualization, we aimed to gain insights into how varying numbers of cards could impact a team's performance and the overall quality of the match. Initially, there was an assumption that a higher number of cards issued to players would lead to more exciting matches, as it creates a thrilling effect in the game. The reasoning behind this assumption is that if a player is sent off due to a red card, it would put their team at a disadvantage. Additionally, if a defender receives a yellow card, their tackling performance may be affected due to the fear of receiving a subsequent red card.



While analyzing the visualization, it becomes evident that yellow and red cards do have an impact on the quality of the game, although the effect is not significant. One interesting finding is that, on average, away teams tend to receive more cards compared to home teams.

The reason behind this pattern could be attributed to the pressure faced by the away team when playing in a hostile environment. Away teams are often confronted with a crowd of home supporters, which can create a more challenging atmosphere. To overcome these challenges, away teams may adopt a more aggressive playing style, leading to an increased likelihood of committing fouls and subsequently receiving cards.

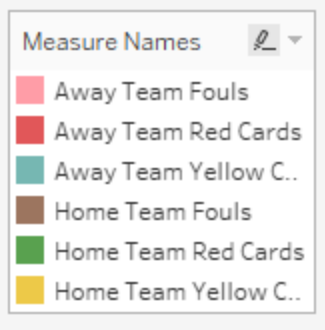
However, it is important to note that the impact of yellow and red cards on match excitement goes beyond the card count alone. Team ratings and the fame of clubs also play a significant role in determining the level of excitement in a match. As discussed in Question 4, the reputation and historical rivalries of the teams involved, as well as their overall performance ratings, contribute to the anticipation and excitement surrounding the game.

A specific example that demonstrates this point is the match between West Ham and Burnley, where West Ham received a red card and three yellow cards. Despite the presence of cards, the match was not particularly exciting. This can be attributed to both teams having lower ratings, which may have influenced the overall quality and intensity of the gameplay.

Therefore, while yellow and red cards do impact the game, their influence on match excitement is not solely determined by the card count. Factors such as team ratings, club fame, historical rivalries, and the game's competitive nature contribute to the overall excitement level of a match.

Question 3: How does the number of fouls result in yellow/red cards?

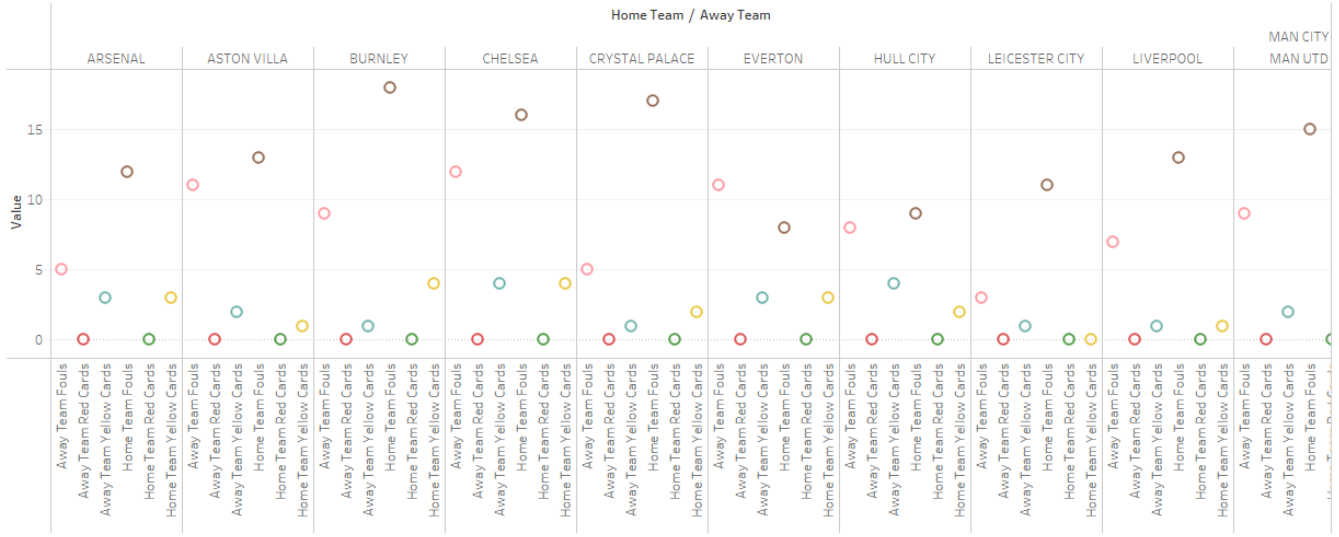
In analyzing the relationship between the number of fouls committed and the issuance of yellow and red cards, we focused on the performance of Manchester City as the home team



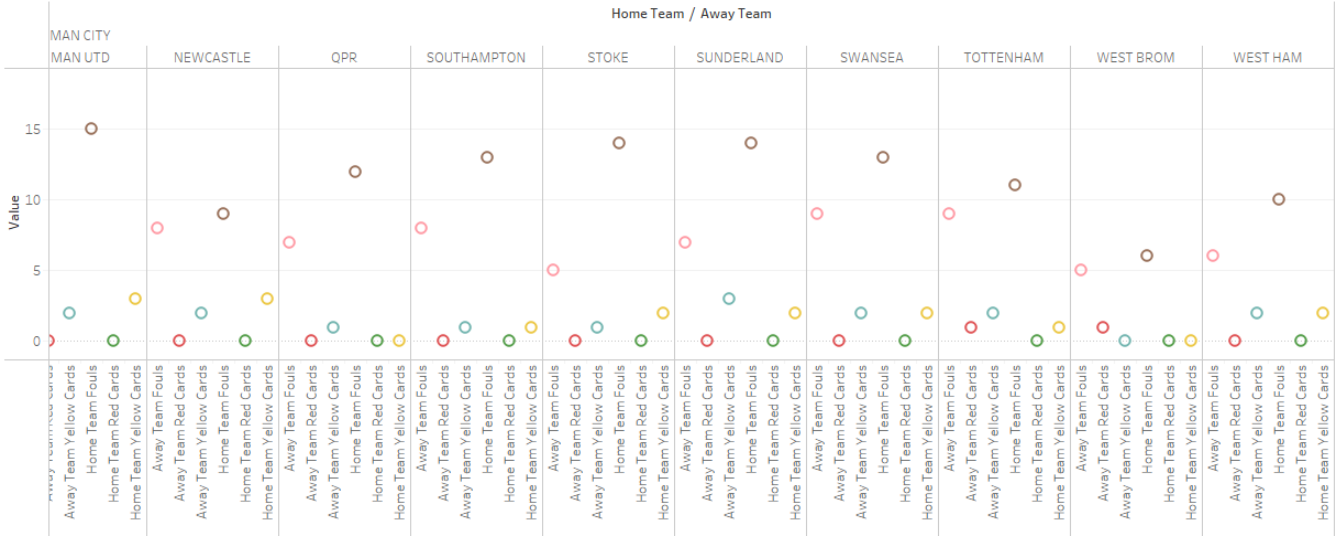
during the 2014/2015 season. This season was chosen as it had the highest number of yellow and red cards issued, as shown in Figure 2.1.

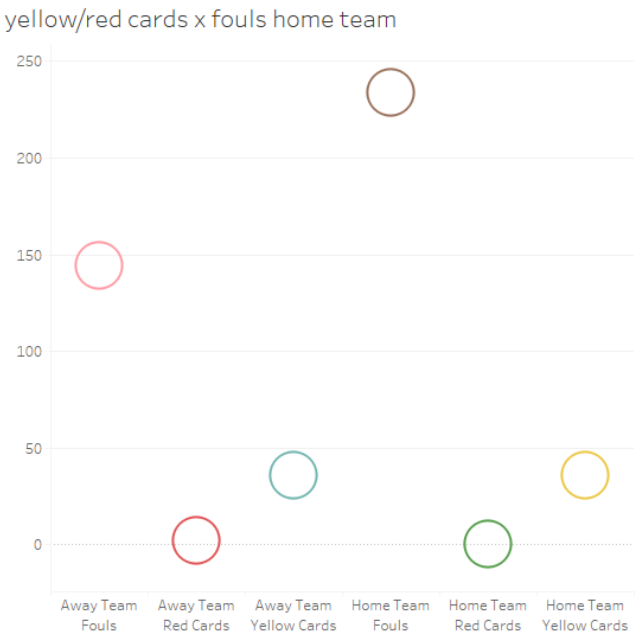
Upon examining the data, it became clear that Manchester City committed many fouls throughout the season. This suggests that their playing style may have been more aggressive or prone to committing fouls compared to other teams. Notably, Manchester City had a higher number of fouls against teams such as Newcastle, Manchester United, Chelsea, and Burnley. This observation indicates that the frequency of fouls is not necessarily linked to club rivalries but rather reflects the gameplay dynamics and strategies employed by different teams.

yellow/red cards x fouls home team



yellow/red cards x fouls home team





When analyzing the consequences of fouls in terms of yellow and red cards, we found that there is a 15.3% chance, on average, of a home team receiving a yellow card for each foul committed. On the other hand, when the foul is committed by the away team, there is a higher likelihood of receiving a yellow card, with an average probability of 25%. This discrepancy might be attributed to factors such as the referee's interpretation of the foul, the intensity of the foul, or the potential bias towards the home team.

In addition to yellow cards, there is a 1.38% chance of a red card being issued for dangerous fouls. This suggests that the severity and nature of the foul play a role in determining whether a red card is given. Note that red cards are rare compared to yellow cards, indicating referees reserve this disciplinary action for more egregious or dangerous fouls.

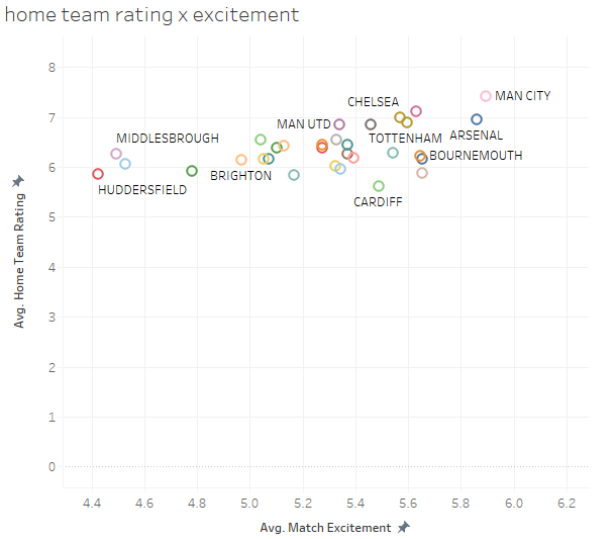
Overall, the data suggests that the number of fouls committed by a team can influence the likelihood of receiving yellow or red cards. However, other factors such as the context of the foul, the referee's judgment, and the severity of the foul also play a significant role in determining the disciplinary actions taken.

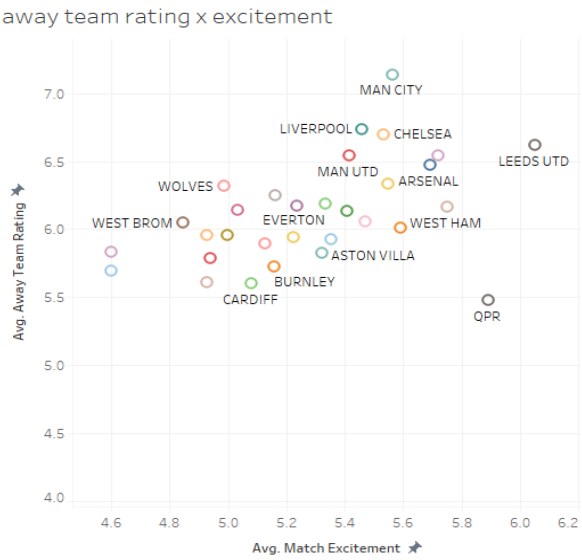
Question 4: How do Team Ratings affect Match Excitement?

Based on the provided data, higher-rated teams tend to have higher match excitement when playing as the home team. This could be attributed to factors such as larger stadiums and a larger fan base, which can contribute to a more vibrant and enthusiastic atmosphere during home matches.

However, the data also reveals an interesting anomaly where Leeds, despite not having the highest rating, shows the highest match excitement when playing as an away team. The exact reason for this observation cannot be determined solely based on the provided data. There could be a range of factors influencing the match excitement for away teams, such as the team's playing style, the competitiveness of the match, or the passionate support from the away fans.

To gain a more comprehensive understanding of the relationship between team ratings, home/away status, and match excitement, further analysis and additional variables may be necessary.





Question 5: Goals scored compared to different types of shots.

When comparing the number of goals scored to the shots taken by teams, we examined both the home team and the away team. The rationale behind focusing on the away team is that they often face more pressure in an unfamiliar environment, which could potentially result in more shots being taken.

Without the dataset visualizations, one might expect a positive correlation between the number of shots taken and the goals scored. The reasoning behind this expectation is that the more shots a team takes, the greater the likelihood of scoring goals. However, there are several factors that can influence this relationship.

Firstly, the quality of shots matters. Taking many shots does not guarantee a higher number of goals if they are not accurate or well-placed. Therefore, the accuracy and precision of the shots taken by a team play a crucial role in converting those shots into goals.

Secondly, the defensive capabilities of the opposing team can also impact the number of goals scored. If the opposing team has a strong defense, they may be successful in blocking or deflecting a sizable number of shots, resulting in a lower goal conversion rate.

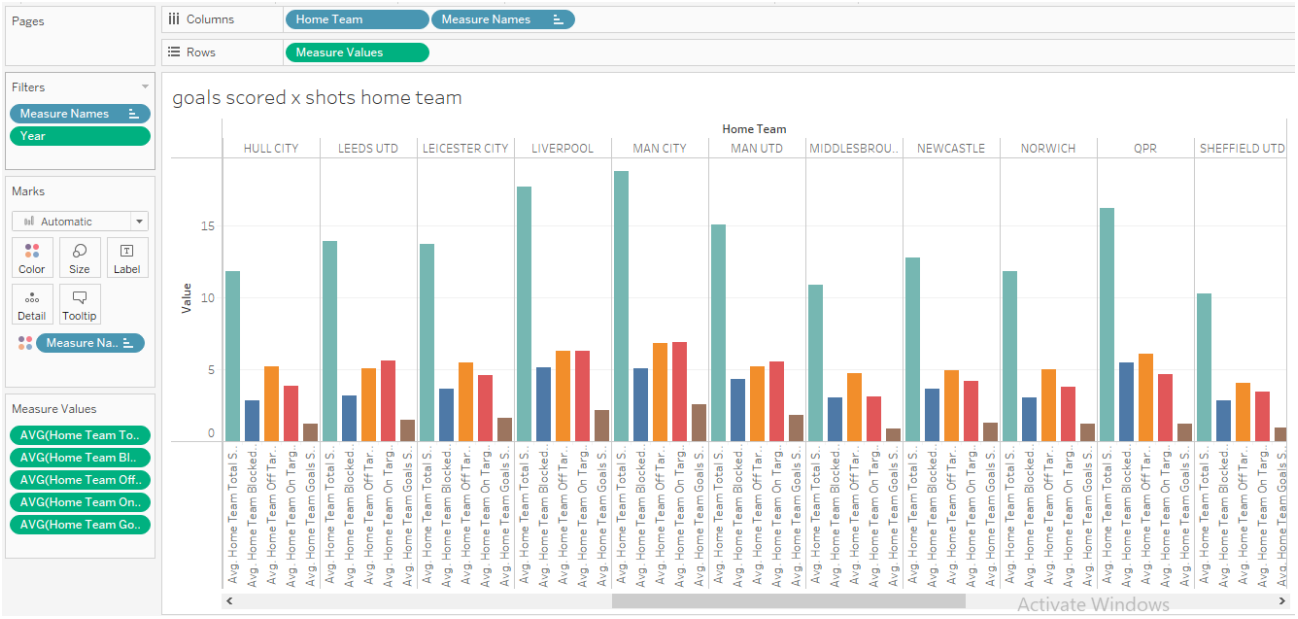
Lastly, the skill level and performance of the teams involved also contribute to the goals scored. A team with highly skilled players and effective attacking strategies may have a higher goal-scoring rate, even with a lower number of shots taken.

Therefore, while it is reasonable to expect a positive relationship between shots taken and goals scored, it is important to consider the quality of shots, the defensive capabilities of the opposing team, and the overall skill level of the teams involved to fully understand the impact

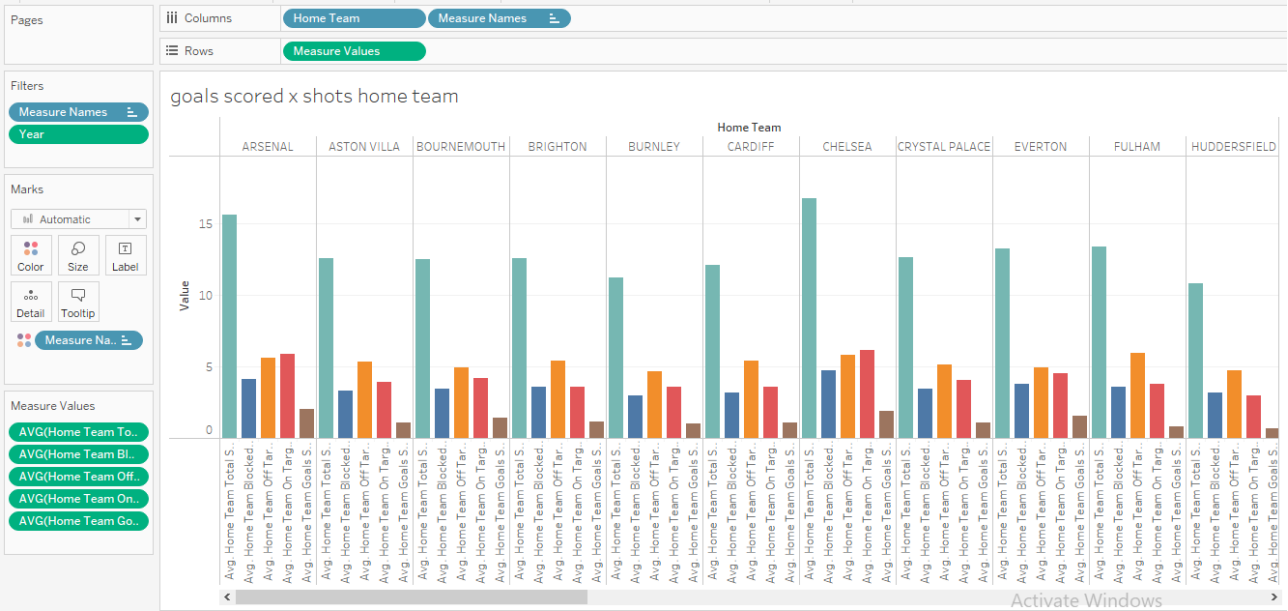
of shots on goal-scoring. The dataset visualizations can provide more concrete insights into these relationships and help validate or refine these expectations.

Home Teams:

When analyzing the performance of home teams, an interesting trend emerges. Teams with higher overall ratings, such as Manchester City and Liverpool, tend to take more shots, particularly on target. For instance, from 2014 to 2020, Manchester City averaged 18.81 total shots per match. Out of these shots, 5.08 were blocked by defenders or the goalkeeper, 6.84 were off target, and 6.89 were on target. On average, Manchester City scored 2.60 goals as a home team, indicating a conversion rate of approximately 13.8% for each shot taken. This highlights the correlation between a team's shooting volume and their ability to generate goals, suggesting that teams with higher ratings tend to be more aggressive and precise in their attacking play.



In contrast to higher-rated teams, lower-rated teams such as Burnley FC and Brighton FC tend to have fewer shots in their home games. Taking Burnley FC as an example, between 2014 and 2020, they had an average of 11.21 shots per home game. Out of these shots, 2.98 were blocked by defenders or the goalkeeper, 4.64 were off target, and approximately 3.61 were on target. On average, Burnley FC scored 1.04 goals per game at home, indicating a conversion rate of around 9.27% for each shot taken. This suggests that lower-rated teams might focus more on defensive strategies and have a more conservative approach in their attacking play, resulting in fewer shots and a lower goal-scoring rate compared to higher-rated teams.



Away teams:

When considering away games, higher-rated teams such as Manchester City, Chelsea, and Liverpool tend to take fewer shots compared to their home games. Taking Manchester City as an example, in an away game, their players took an average of 8.42 shots. Out of these shots, 2.03 were blocked, 3.34 were off target, and 3.08 were on target, resulting in an average of 0.95 goals per game. This indicates a 44.67% decrease in the number of shots compared to their home games.

There could be several reasons for this phenomenon. One probable reason is the motivation of the players. Playing on away turf can be challenging as teams have fewer fans to support them, which might affect their confidence and performance. Additionally, the strategy implemented by the team's manager could play a role. Managers may adopt a more cautious approach in away games, prioritizing defensive stability and counter-attacking opportunities rather than taking a high volume of shots.

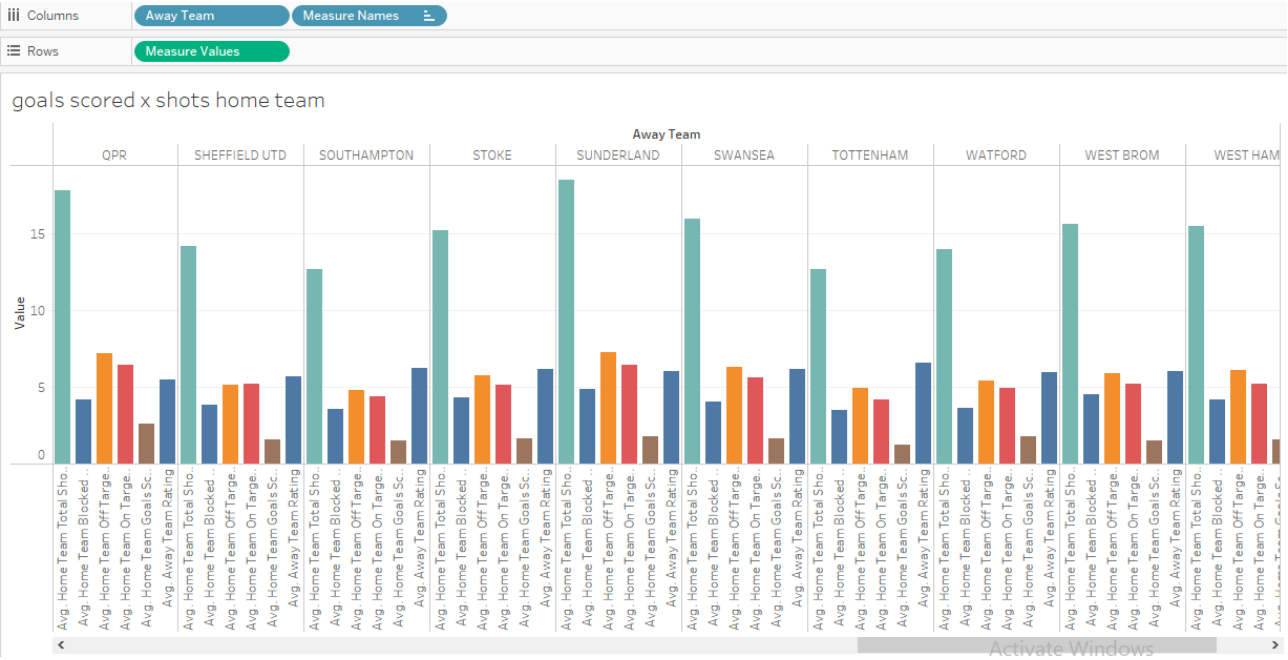
Other factors that could contribute to this trend include the style of play of the home team, the quality of the opposition's defense, and the overall tactics employed by the away team. To gain a deeper understanding and confirm these theories, it would be beneficial to analyze additional metrics such as game lineups, game timelines, and player statistics. These factors can provide more insights into the specific strategies and dynamics at play during away games for higher-rated teams.

Contrary to higher-rated teams, lower-rated teams such as Burnley FC, QPR, Sunderland, and Swansea tend to take more shots when playing away games. Taking Burnley FC as an example, their players, on average, take 17.66 shots per game in away matches. Out of these shots, 5.71 are blocked, 6.46 are off target, and 5.48 are on target, resulting in an average of 1.58 goals per game. This indicates a 36.53% increase in the number of shots compared to their home games.

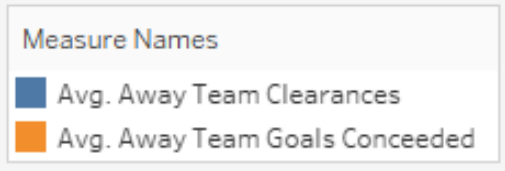
Several factors could contribute to this trend for lower-rated teams. One explanation is that these teams may adopt a more proactive and attacking style of play when playing away. As

underdogs, they may seize the opportunity to take more shots and be more adventurous in their approach, aiming to score goals and secure points against stronger opposition. Additionally, the defensive vulnerabilities of higher-rated home teams might create more opportunities for lower-rated away teams to create chances and take shots.

It is important to note that these observations are based on the data provided, and further analysis would be required to confirm these trends and explore the underlying factors in more detail. Additional metrics such as possession statistics, game situations, and player performance can provide further insights into the strategies and dynamics of lower-rated teams in away games.



Question 6: How does Clearances affect Goals conceded?

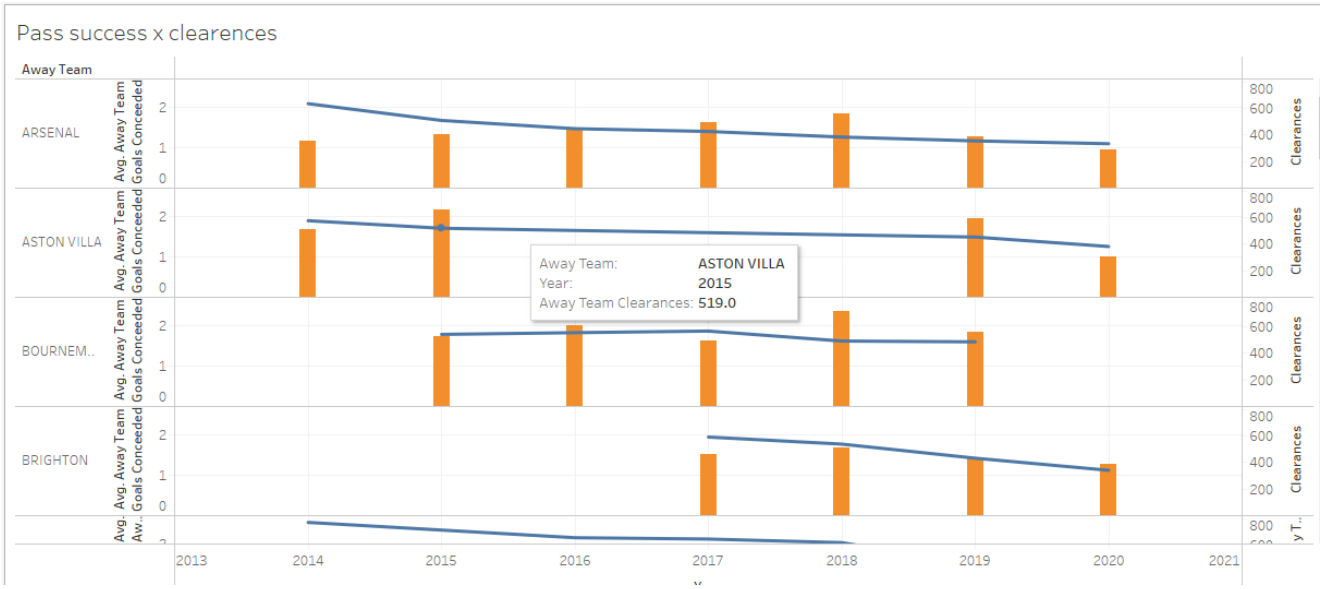


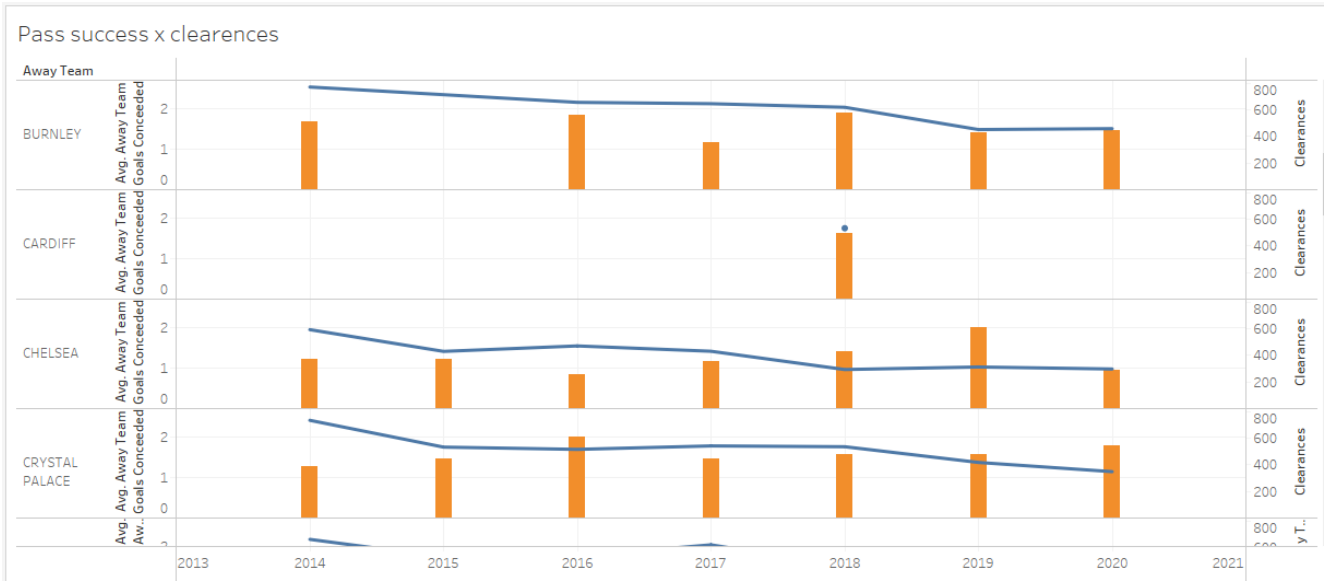
To analyze the relationship between goals conceded and clearances made by defenders, we can examine the data and explore any potential connections. Clearances are an important defensive action taken by players to prevent the opposing team from scoring, so it is reasonable to assume that a higher number of clearances could be associated with fewer goals conceded.

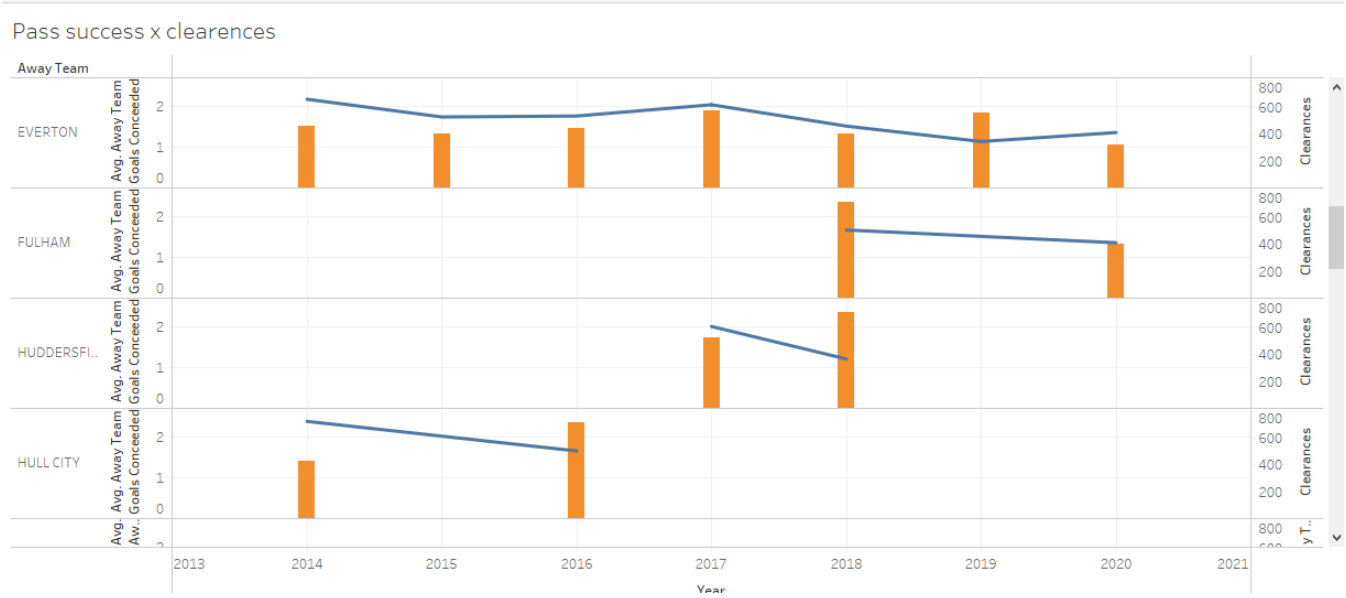
By comparing the statistics of various teams, we can observe whether there is a correlation between the number of clearances and goals conceded. If teams with

more clearances tend to concede fewer goals, it would suggest a strong defensive performance.

connection to clearances made by defenders,





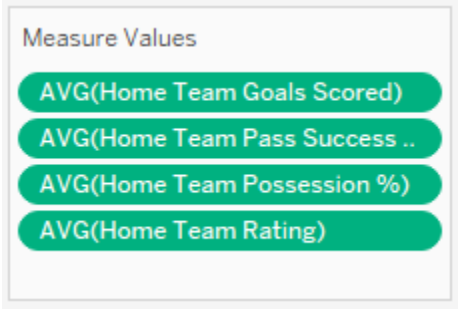


Upon analyzing the data, it is evident that there is an inverse relationship between clearances and goals conceded. Teams that make a higher number of clearances tend to concede fewer goals, indicating a connection between these two factors. Clearances serve as a defensive measure to block or remove the ball from dangerous areas, minimizing the opponent's chances of scoring.

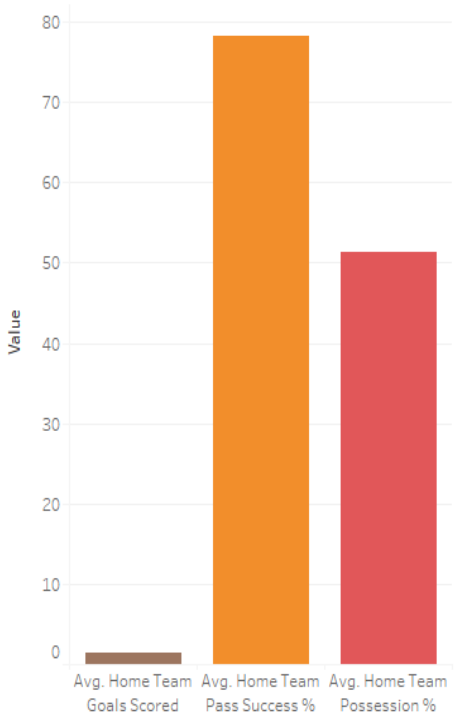
However, it is important to consider that there can be exceptions to this trend. Factors such as the quality of opposition attacks, individual defensive errors, and overall defensive organization can also influence the number of goals conceded. Additionally, simply relying on clearances alone may not guarantee a dwindling number of goals conceded, as other defensive aspects, such as interceptions, tackles, and positioning, also play significant roles in preventing goals.

In conclusion, while there is a correlation between clearances and goals conceded, it is essential to consider other defensive factors and the overall defensive strategy employed by a team. Clearances contribute to defensive stability and can help reduce the number of goals conceded, but they are just one aspect of a comprehensive defensive approach.

Question 7: AVG Pass success/Possession result on goals scored.



When examining the relationship between average pass success rate, possession, and goals scored, it becomes evident that these factors play a significant role in a team's offensive performance. Pass success rate reflects the ability of a team to maintain control of the ball and execute accurate passes, while possession measures the amount of time a team spends with the ball in their control.

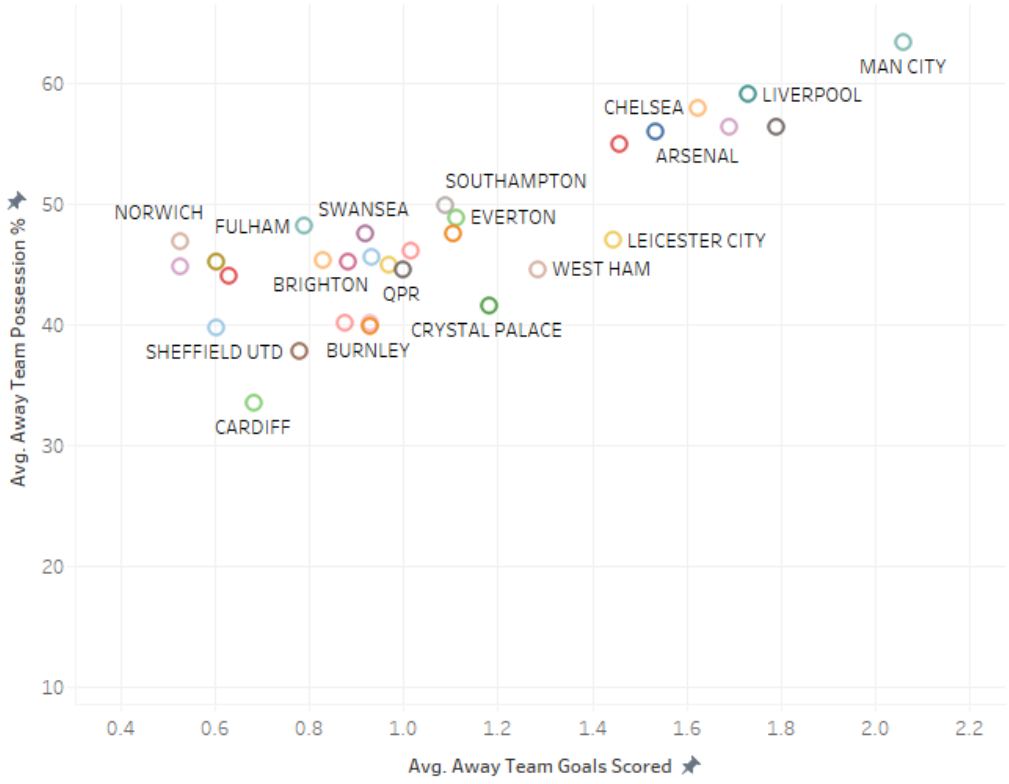


On average, teams have a pass success rate of approximately 78.32% and a possession rate of 51.32%, which leads to an average of 1.50 goals scored per game. However, there is variation among teams, with some demonstrating higher pass success rates and possession, resulting in more goals scored.
Figure 7.2 illustrates that Manchester City stands out with an impressive pass success rate of 85.62%, leading to an average of 2.06 goals per game. This highlights the importance of effective passing in creating scoring opportunities and converting them into goals. Conversely, teams like Cardiff FC, with a pass success rate of 63.32%, score an average of 0.684 goals per game, indicating a lower efficiency in their offensive play. Similarly, possession plays a crucial role in goal-scoring success. Figure 7.3 displays Manchester City's dominance in possession, averaging 63.43% per game, resulting in 2.06 goals. Higher possession allows teams to dictate the flow of the game, create more opportunities, and maintain offensive pressure. In contrast, lower-rated teams like Cardiff FC, with a possession rate of 33.53%, struggle to maintain control and consequently score fewer goals (0.684 per game).

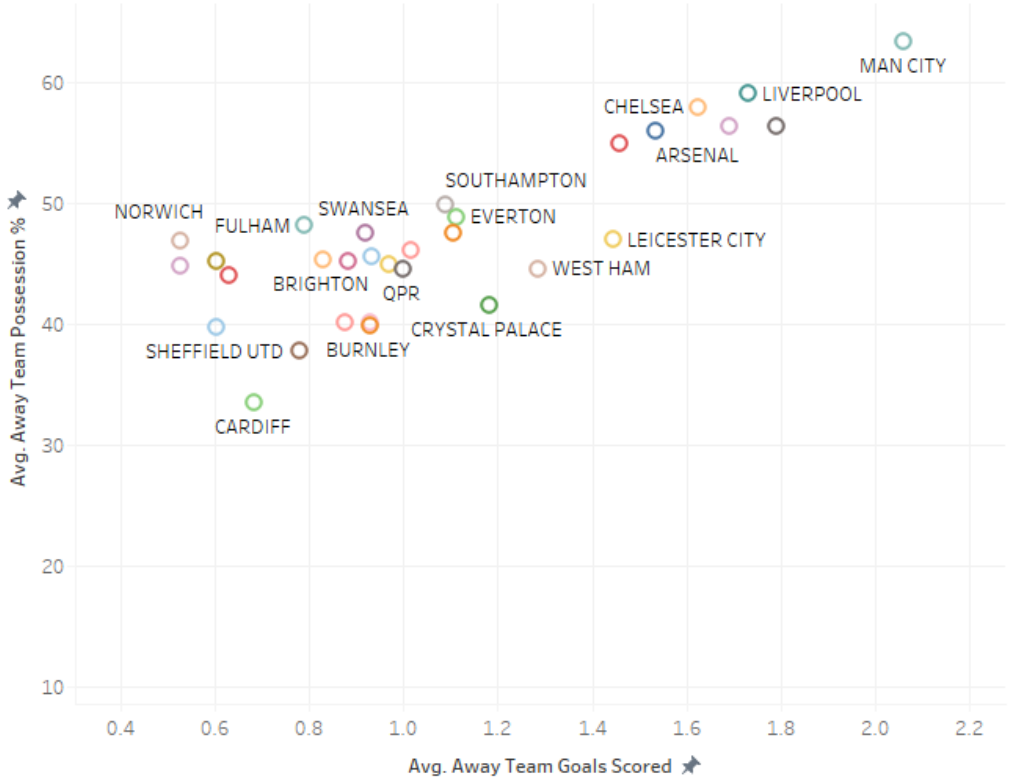
These findings suggest that both a high pass success rate and possession positively influence a team's ability to score goals. Effective passing facilitates ball movement and penetration into the opposition's defensive areas, while higher possession provides more opportunities to create goal-scoring chances.

However, it is essential to consider other factors that contribute to goal-scoring success, such as the quality of players, attacking strategies, and individual player skills. Additionally, a team's defensive abilities and the strength of the opposing team also impact the number of goals scored.

Possession x Goals Scored



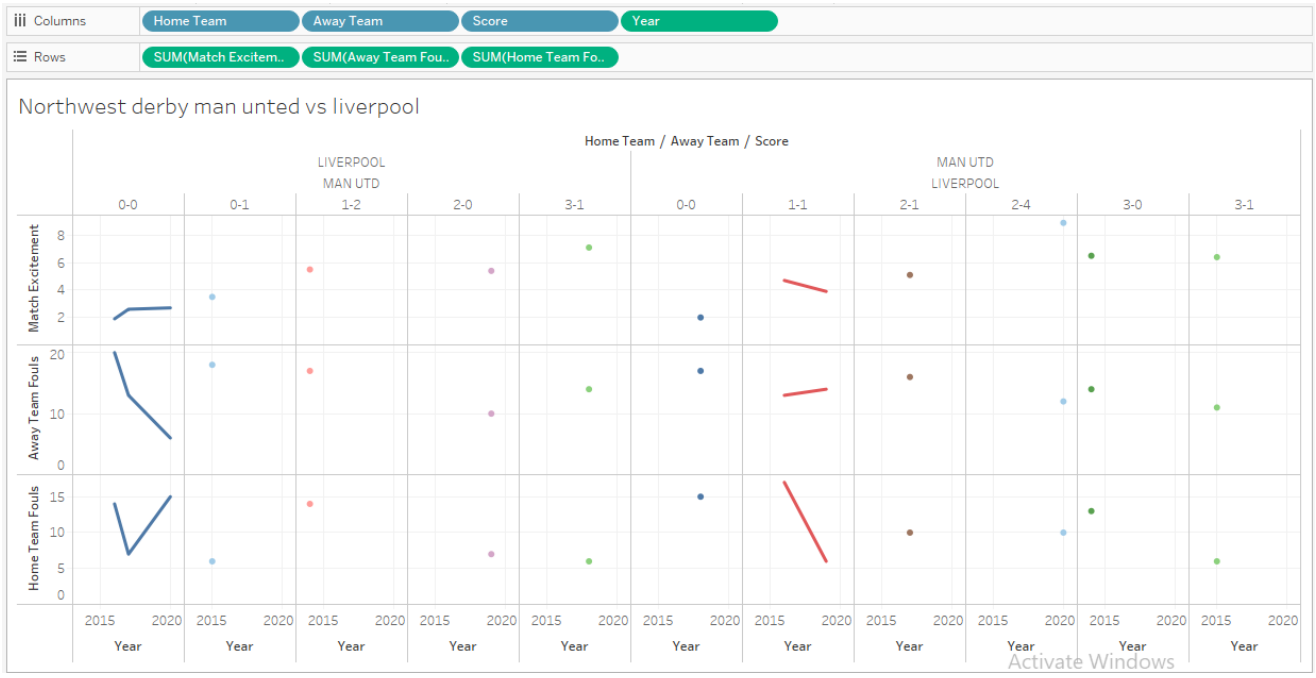
Possession x Goals Scored



In conclusion, a higher pass success rate and possession are associated with an increased likelihood of scoring goals. Teams like Manchester City demonstrate the positive impact of effective passing and controlling possession on their offensive performance. Nonetheless, it is crucial to consider these factors alongside other aspects of the game to gain a comprehensive understanding of a team's goal-scoring capabilities.

Question 8: Northwest Derby Manchester United vs Liverpool home vs away games stats analysis

The quality of the Northwest Derby between Manchester United and Liverpool, the biggest rivalry in the English Premier League, depends on numerous factors such as goals scored, pass success, possession, shots taken, number of passes, yellow cards, and fouls. These factors contribute to the overall excitement and entertainment value of the match. Analyzing specific statistics and trends would provide a deeper understanding of the home and away games in this rivalry.



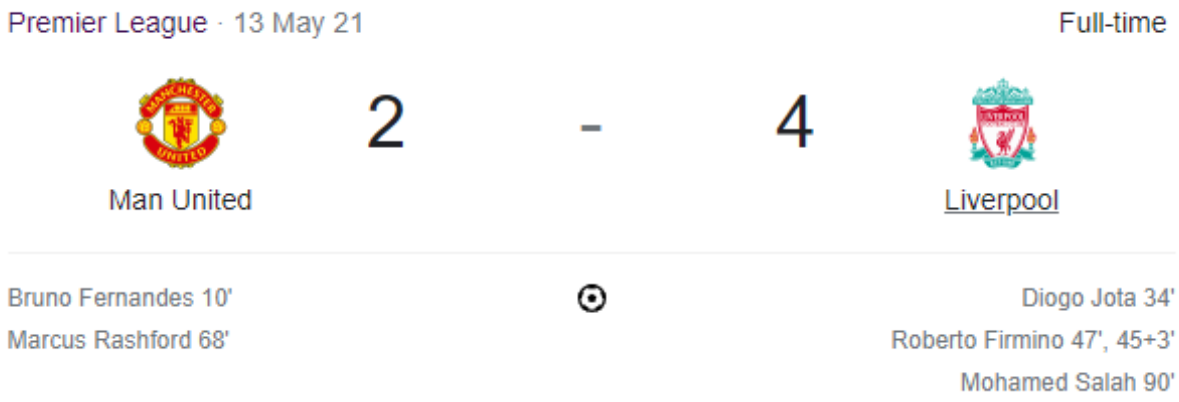
In a thrilling home game at Old Trafford during the 2020-2021 season, Liverpool defeated Manchester United with a score of 4-2. The match highlighted some interesting aspects that align with our previous analysis.

The first half ended with Liverpool leading by one goal, courtesy of Roberto Firmino's last-second goal. In the second half, Firmino scored again just two minutes after the game started, increasing Liverpool's lead. Marcus Rashford's goal in the 68th minute gave Manchester United hope, but Mohamed Salah sealed the deal for Liverpool with a goal in the 90th minute.

This game provides a notable example that supports our analysis. Despite Manchester United having more shots (18) compared to Liverpool's 17, the number of shots on target was crucial. Manchester United had only 3 shots on target, while Liverpool had 8, resulting in 4 goals. This demonstrates that the number of shots alone does not determine the outcome or excitement of a game.

Additionally, Liverpool's higher number of fouls in the game may indicate their effective defensive strategy, leading to fewer goals conceded. On the other hand, Manchester United receiving three yellow cards, all in the second half, suggests that players, particularly defenders, might resort to making poor decisions and playing aggressively when under pressure, especially when losing in a home game.

This game exemplifies the complex dynamics that contribute to the quality and excitement of a match, highlighting the importance of shots on target, defensive strategies, and player behavior under different circumstances.



	TEAM STATS	
18	Shots	17
3	Shots on target	8
54%	Possession	46%
511	Passes	454
78%	Pass accuracy	78%
10	Fouls	12
3	Yellow cards	0
0	Red cards	0
1	Offsides	1
6	Corners	4

Overall Trends & Patterns:

In the analysis of the relationship between the number of goals and match excitement, it was found that the number of goals scored does not singularly determine the excitement level of a match. Factors such as historical rivalries, team ratings, and the game's competitive nature also play significant roles. Matches against top-rated teams or those with closely positioned teams tend to be more thrilling, irrespective of the final score.

When examining the influence of red and yellow cards on match excitement, it was observed that while yellow and red cards do impact the game, their influence on match excitement is not solely determined by the card count. Factors such as team ratings, club fame, historical rivalries, and the game's competitive nature contribute to the overall excitement level of a match.

In the analysis of the relationship between the number of fouls and yellow/red cards, it was found that there is a correlation between the number of fouls committed and the likelihood of receiving yellow or red cards. However, other factors such as the context of the foul, the

referee's judgment, and the severity of the foul also play a significant role in determining the disciplinary actions taken.

When comparing the number of goals scored to the shots taken by teams, it was observed that the quality of shots, the defensive capabilities of the opposing team, and the skill level of the teams involved all influence the relationship between shots and goals. While a positive correlation between shots and goals is expected, these factors need to be considered to fully understand the impact of shots on goal-scoring.

In analyzing team ratings and match excitement, it was observed that higher-rated teams tend to have more shots and a higher goal-scoring rate compared to lower-rated teams. This suggests that higher-rated teams are more aggressive and precise in their attacking play, leading to more excitement in their matches.

In the analysis of goals conceded and clearances made by defenders, an inverse relationship was observed. Teams that make a higher number of clearances tend to concede fewer goals, indicating the importance of clearances as a defensive measure to minimize the opponent's scoring chances.

Overall, these analyses demonstrate the multifaceted nature of football matches and how numerous factors, such as historical rivalries, team ratings, disciplinary actions, and defensive measures, contribute to the excitement and outcome of the games.

Data Mining Algorithms:

In this Project I am analyzing football match excitement, goals scored, and fouls and variables that affect them, as my dataset includes more numerical values, I will be using M5P algorithm to figure patterns out, in this section we will be removing irrelevant attributes creating categories to make it easier for Weka to apply the M5P algorithm.

Data Processing:

Removing Irrelevant Attributes

Attributes Removed:

- "Half Time Score"
- "Home Team Corners" and "Away Team Corners"
- "Home Team Throw Ins" and "Away Team Throw Ins"
- "Home Team Aerials Won" and "Away Team Aerials Won"
- "Home Team Goals Conceded" and "Away Team Goals Conceded"

The decision to remove the "Home Team Goals Conceded" and "Away Team Goals Conceded" columns was based on the availability of alternative variables that can provide similar information. Instead of focusing on goals conceded by each team, we can consider the goals

scored by the away team in home games and the goals scored by the home team in away games as indicators of defensive performance. This substitution allows us to maintain the analysis of defensive strength without explicitly including the goals conceded columns.

Regarding the "Home Team Aerials Won" and "Away Team Aerials Won" columns, while aerial battles can impact the game, their inclusion in the analysis did not significantly contribute to understanding match excitement compared to other variables. Therefore, the decision was made to remove these columns, as they were not crucial for the specific focus of our analysis.

The "Home Team Corners," "Away Team Corners," "Home Team Throw Ins," and "Away Team Throw Ins" columns were removed as they had limited relevance to understanding match excitement. Factors like possession, shots, and goals scored provide more direct indicators of gameplay dynamics and excitement. Removing these columns streamlined the analysis to focus on more influential variables.

Creating Categories

Since Weka works best with categorial data all numerical values were modified into meaningful categories.

Match Statistics:

- Possession:
    - Home Team Possession %
    - Away Team Possession %
- Pass Success:
    - Home Team Pass Success %
    - Away Team Pass Success %
- Match Excitement

Shots:

- Home Team Off Target Shots
- Home Team On Target Shots
- Home Team Total Shots
- Home Team Blocked Shots
- Away Team Off Target Shots
- Away Team on Target Shots
- Away Team Total Shots
- Away Team Blocked Shots

Fouls and Cards:

- Home Team Fouls
- Away Team Fouls

- Home Team Yellow Cards
- Home Team Red Cards
- Away Team Yellow Cards
- Away Team Red Cards

Team Information:

- Home Team
- Away Team
- Home Team Rating
- Away Team Rating

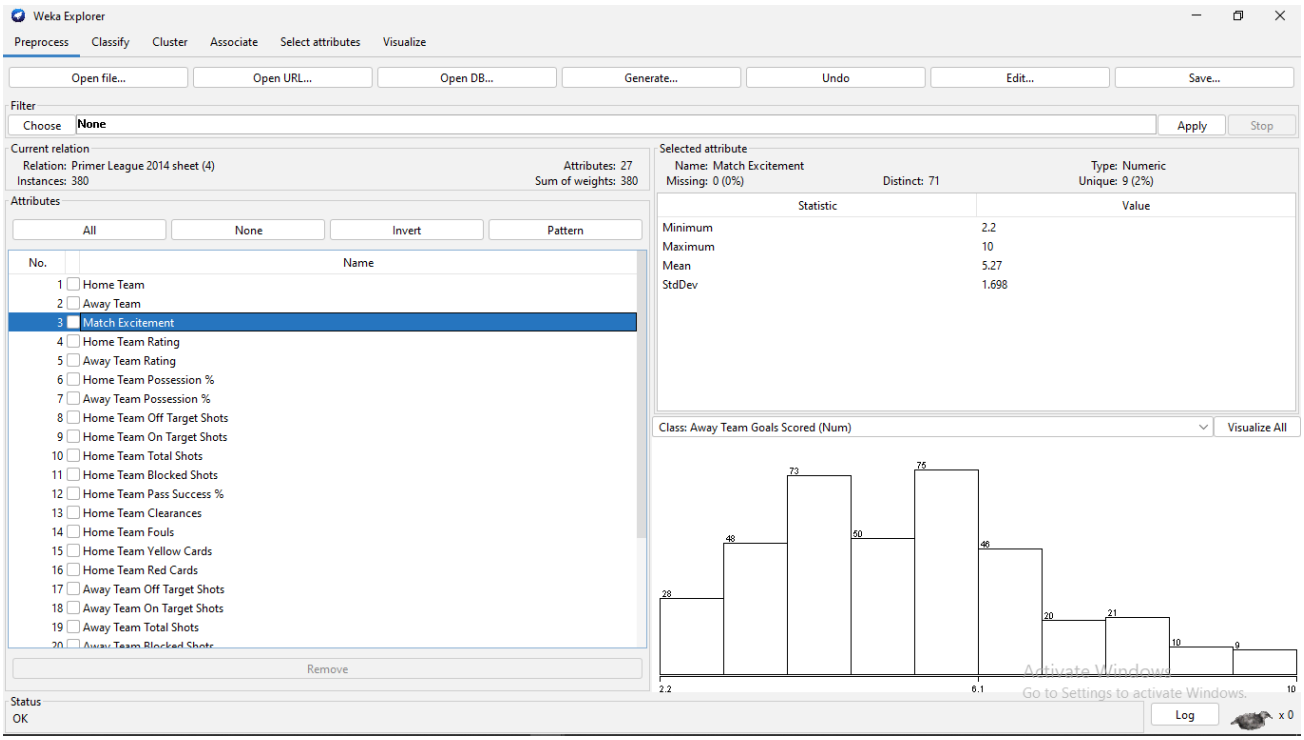
Goals:

- Home Team Goals Scored
- Away Team Goals Scored

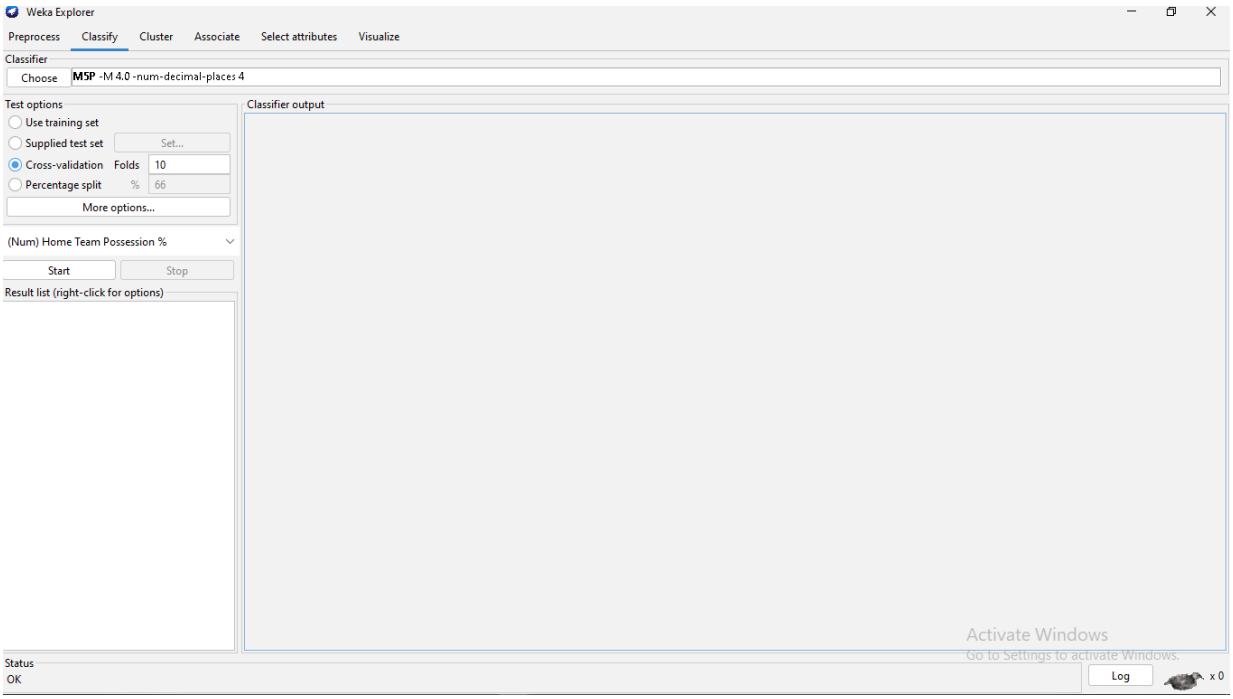
Year

Data Mining:

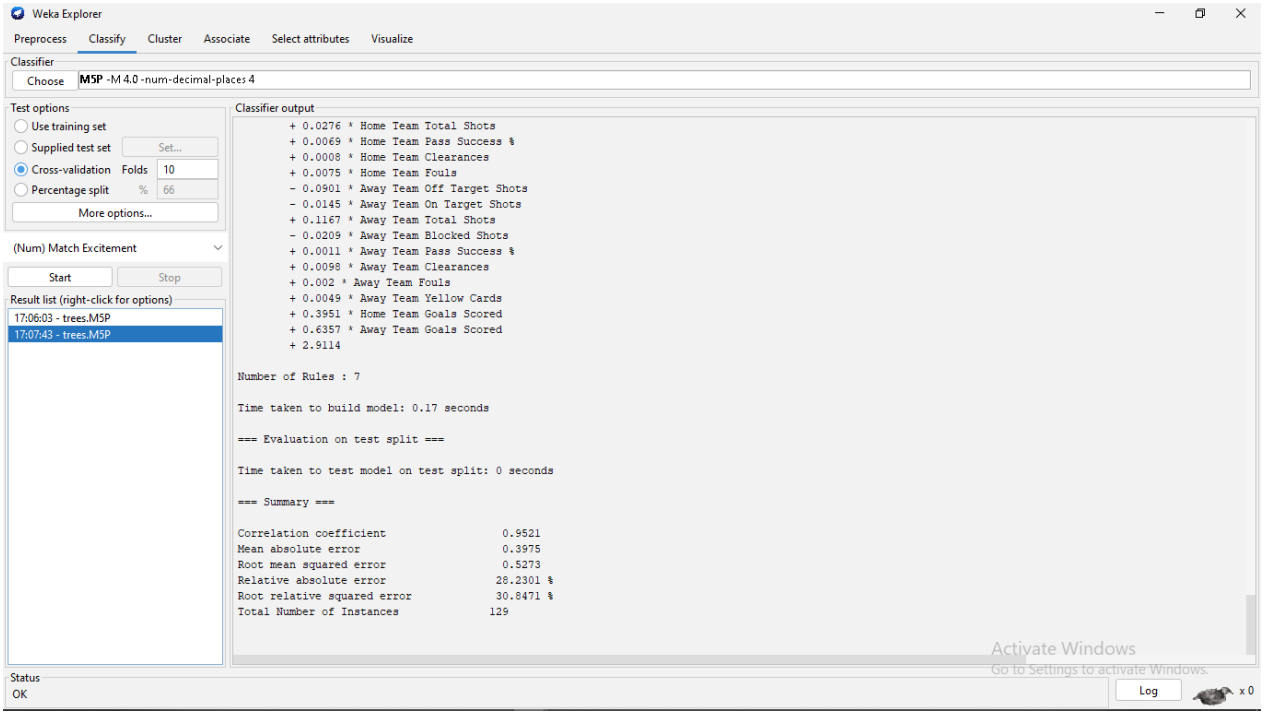
➢ Step1: Importing the file: Weka Explorer imports CSV file containing categorial data



➢ Step2: Configuring Classifier/choosing the algorithm

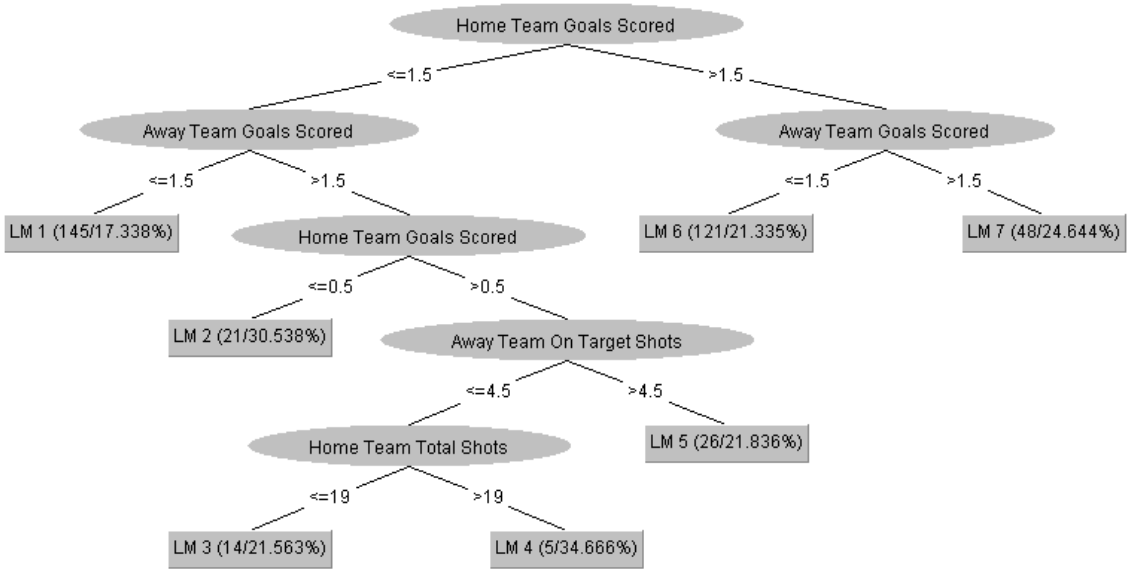


- Step3: running the algorithm showing the tree



Our analysis of the visualizations has uncovered a significant factor that influences match excitement: the number of goals scored by the home team. As we previously observed in our visualizations, the quality of shots plays a crucial role in determining the goal count. However,

we also discovered that the performance of the opposing team, particularly in terms of on-target shots or goals, can affect the home team's performance. This can be attributed to the pressure faced by the home team, making it more challenging for them to create scoring opportunities, especially when the opposing team is leading or dominating the game.



Data Ethics:

In conducting this research, it is important to acknowledge that all the information regarding match statistics used in this dataset is public and provided by FIFA. The dataset has been compiled in a CSV format with the intention of offering football enthusiasts an opportunity to analyze and explore the various metrics that influence the game. It is crucial to emphasize that no classified information was utilized, no policies were violated, and no individuals were impacted by the dissemination of this information. The dataset's sole purpose is to promote research and enhance understanding of the factors that shape the game of football.

Conclusion:

Visualization:

Our exploration of visualizations has revealed several key insights that shed light on various aspects of football matches. Some of the notable findings include:

- **Home Team Advantage:** The analysis of the home team's performance compared to the away team consistently shows a higher number of goals scored by the home team. This suggests a potential home team advantage in football matches.
- **Team Ratings:** The distribution of team ratings indicates that the teams in the Primer League have varying levels of skill and performance. This difference in ratings could be a contributing factor to the home team advantage observed.
- **Goal Distribution:** The histogram of goals scored by teams demonstrates a normal distribution centered around 1.5 goals. This suggests that most matches tend to have a dwindling number of goals, with few matches having a significantly higher goal count.
- **Possession Percentage and Shots:** The higher the number of accurate shots towards the goal, the more goals scored. This observation aligns with the famous saying in football that "Attack is the best Defense." It suggests that teams with a greater focus on offensive gameplay, as reflected by their ability to generate accurate shots towards the goal, are more likely to score goals and potentially create more excitement in matches.

Data mining:

Using the M5P tree classifier, we gained valuable insights into one of the key metrics in this dataset:

Match Excitement: The M5P tree classifier model analyzed numerous factors, including home goals, away goals, and the difference in goals between teams, to predict match excitement. By focusing on goals scored by teams, the model provided valuable information on the level of excitement in a match. However, it is important to note that while this analysis offers significant insights, there may be additional variables that could further enhance our understanding of match excitement dynamics. Further exploration and incorporation of other factors could provide a more comprehensive perspective.

Business Intelligence:

Based on the dataset analysis, the following key insights can be highlighted for the Business Intelligence section:

- **Goal Distribution:** Matches tend to have a dwindling number of goals, following a normal distribution. This helps businesses in sports analytics, betting, and fantasy football set appropriate expectations.
- **Match Excitement Prediction:** Machine learning models can predict match excitement based on factors like team ratings, possession percentage, shots on target, fouls, and red cards. This aids businesses in marketing, ticket sales, and broadcasting decisions.

- **Offensive Gameplay:** There is a positive correlation between possession percentage and shots, emphasizing the importance of offensive strategies for scoring. Businesses in team performance analysis and sports equipment can leverage this insight.
- **Impact of Opposing Team Performance:** The opposing team's performance, especially on-target shots, or goals, affects the home team's performance. Businesses in team analysis and sports psychology can help teams overcome challenges posed by dominating opponents.
- **Shot Quality:** Accurate shots towards the goal significantly contribute to goal scoring. This highlights the importance of improving shot accuracy for better outcomes. Businesses in player training and sports technology can focus on enhancing shot quality.
- **Home Crowd Influence:** The dataset implies that home crowd support may impact match outcomes, contributing to the observed home team advantage. Businesses in fan engagement and stadium operations can leverage this factor to enhance the overall fan experience.

These insights empower businesses in the sports industry to make data-driven decisions, optimize strategies, and create an engaging fan experience.

Reference:

Sanjeet Singh Naik. (n.d.). (2022) Football Data - Top 5 Leagues. [Data file]. Retrieved from https://www.kaggle.com/datasets/sanjeetsinghnaik/football-data-top-5-leagues?select=english+premier+league+data.csv

(Accessed June 3rd, 2023)