Special Topics in Data Analytics-Soccer

In this project, as the data chief of an English Premier League (EPL) sports organization, we selected the PLSquadPassingStats 2022-2023 dataset. Using this dataset and Tableau, our goal is to derive actionable insights for better performance in the EPL. By leveraging data analytics and building visualizations, we aim to present to stakeholders the optimal playing style the team should adopt to be more successful next season. Success is defined as making it to the Champions League (top 4) or qualifying for European competitions (top 7). We are focused on developing a more effective playing style to win more games and achieve a top 7 finish.

From this dataset we focus in the following 26 features and 20 rows (squads) of data:

1. **Squad**: Name of team

2. **#Pl**: Number of Players used in Games

3. **#Age**: Average Age

4. **Poss**: Possession- calculated as the percentage of passes attempted

5. **MP**: Matches played

6. **Gls**: Goals scored or allowed

7. **Ast**: Assists expected

8. **Goa+Ass**: Goals and Assists

9. **G-PK**: Non penalty goals

**10.** **Penalty kicks made**

11. **PKatt**: Penalty kicks attempted

12. **CrdY**: Yellow cards

13. **CrdR**: Red Cards

14. **xG**: Expected Goals

15. **npxG**: non penalty expected goals

16. **xAG**: expected assisted goals

17. **npxG+xAG**: non penalty expected goals plus assisted goals

18. **PrgCarr**: Progressive carrier

19. **PrgPass**: Progressive passes

We removed the columns related to the 90-minute statistics. (we didn’t use these features):

**Gls**: Goals scored or allowed

ii. **Ast**: Assists expected

iii. **G+A**: Goals and Assists

iv. **G-PK**: Non penalty goals

v. **PK**: Penalty kicks made

vi. **PKatt**: Penalty kicks attempted

vii. **CrdY**: Yellow cards

viii. **CrdR**: Red Cards

ix. **xG**: Expected Goals

x. **npxG**: non penalty expected goals

xi. **xAG**: expected assisted goals

xii. **npxG+xAG**: non penalty expected goals plus assisted goals

xiii. **PrgC**: Progressive carrier

xiv. **PrgP**: Progressive passes

Instead, we identified the need for specific statistics to achieve our goal of developing a more effective playing style to win more games and achieve a top 7 finish. Therefore, we added the following::

1. **Rank:** Squad finish in competition (1 is the best)

2. **Wins**

3. **Draws**

4. **Losses**

5. **GF:** Goals for

6. **GA:** Goals against

7. **GD**: Goal Difference

8. Non-Penalty Goals

9. **Cmp%** Passes Completion Percentage.

10. **Cmp:** Passes completed

11. **Att**: Passes Attempted

12. **TotDist** Total Distance

13. **PrgDist:** Progress distance

14. **KP**: Key Passes

15. **xA**: Expected Assists.

16. **PPA**: Passes into Penalty Area

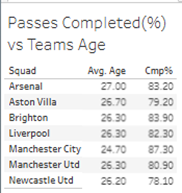
17. **CrsPA**: Crosses into Penalty Area Completed passes that move the ball towards the opponent's goal line at least 18 yards not included set pieces

We use Tableau to visualize key statistics for each squad.

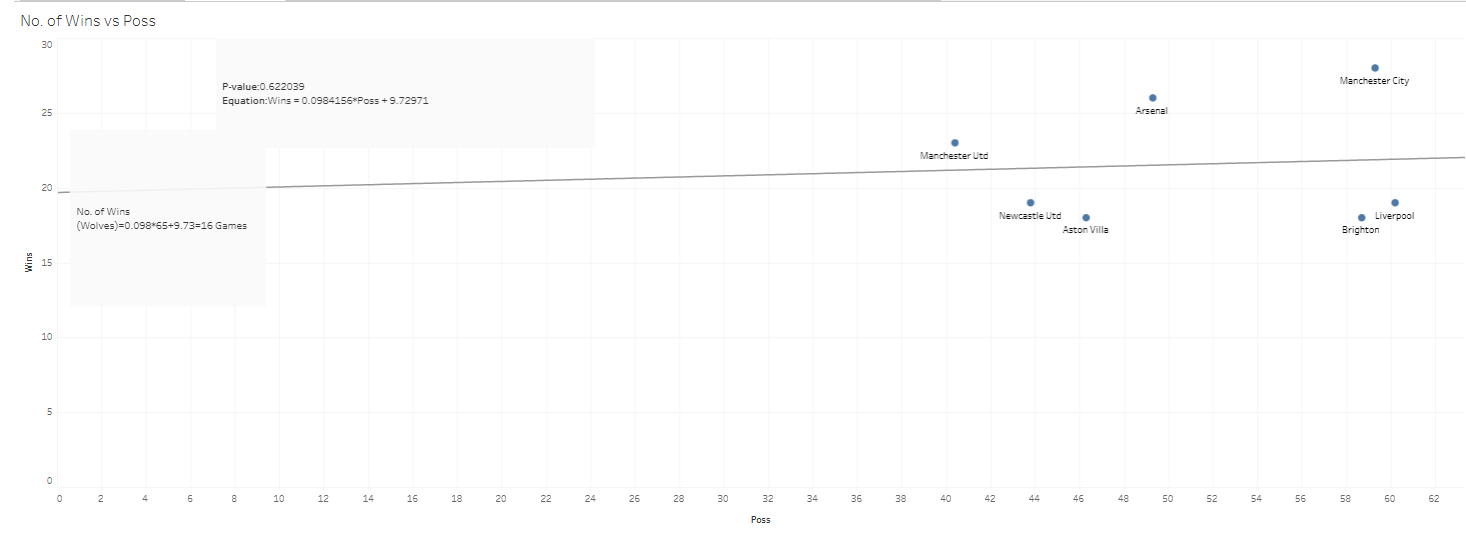
We did the following visualizations:

1.Passes completed % vs Teams Age:

To examine the relationship between passes completed percentage and team age for the top 7 Premier League teams, we created a table summarizing these metrics. This table facilitates a straightforward comparison of the average age of the teams against their passes completed percentage. However, no discernible trends were observed, as the average age across the teams is quite similar.

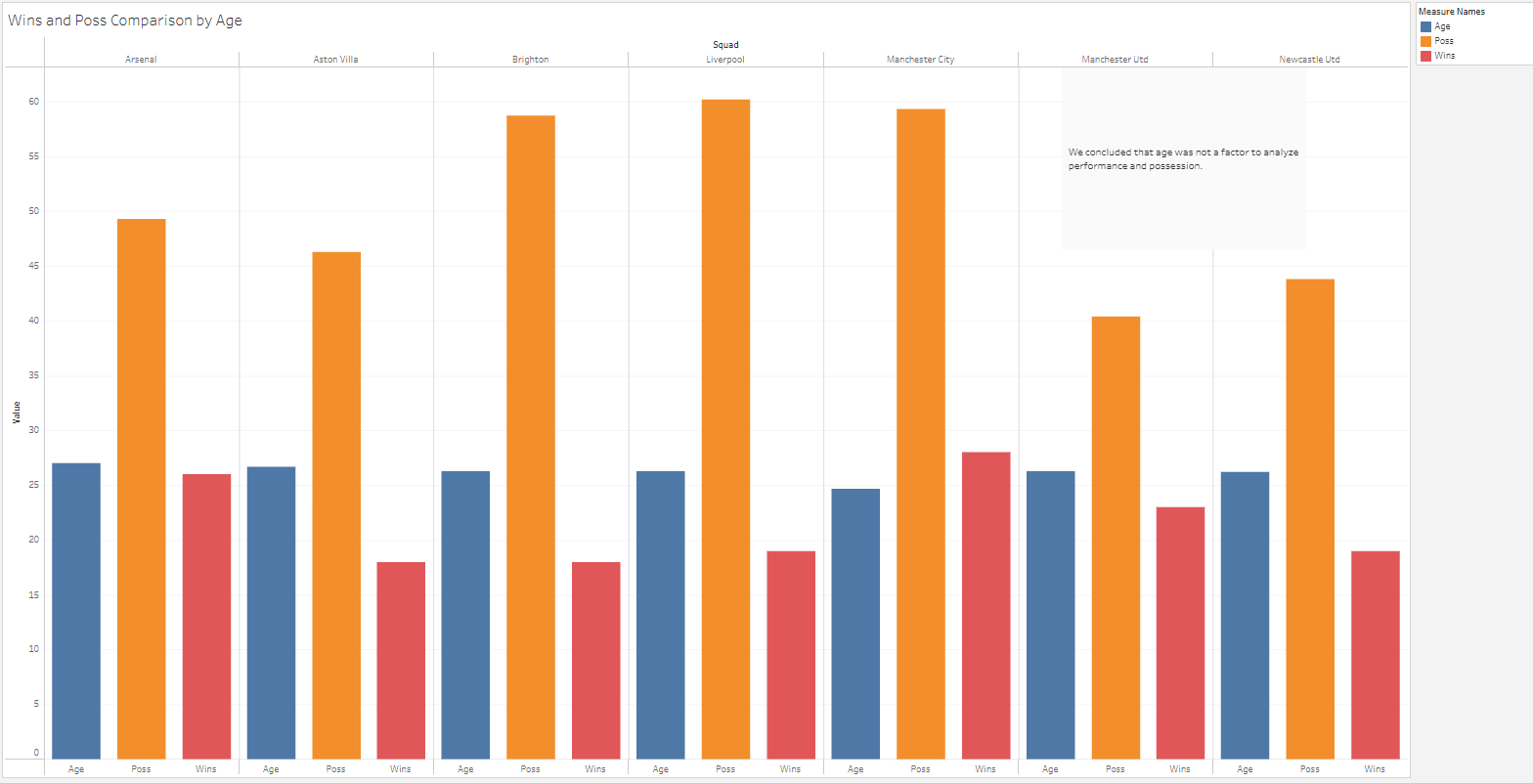


2. No. of wins vs possession:



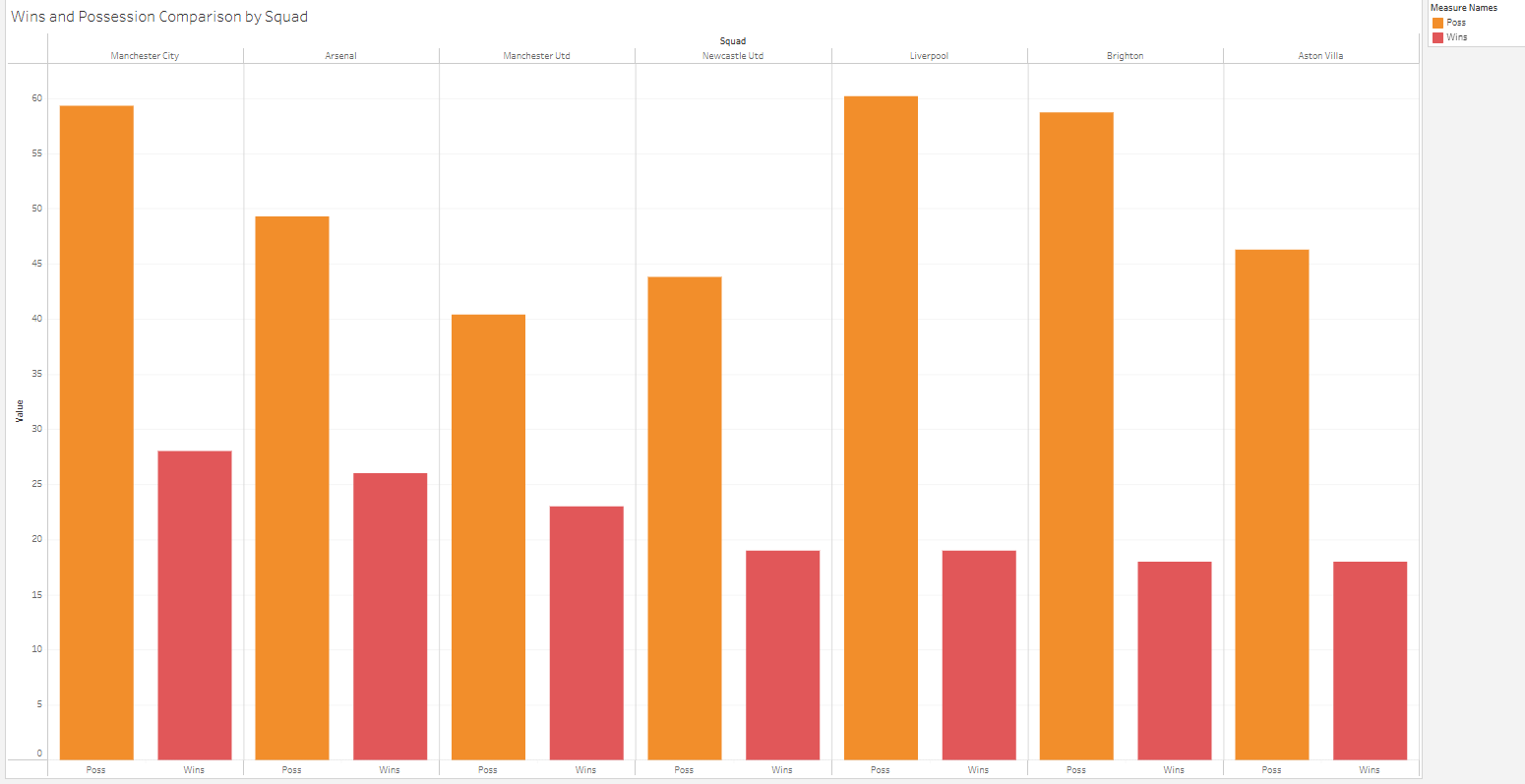
To analyze the relationship between wins and possession percentage for the top all the Premier League teams, we created a scatter plot. This visualization helps identify any patterns or correlations between these two metrics. In the scatter plot, the number of wins is plotted on the y-axis and possession percentage on the x-axis. Each point represents one of the top entire teams, allowing us to see if teams with higher possession percentages also tend to have more wins. For example, Manchester City, with the highest number of wins, is plotted alongside its corresponding possession percentage. The results suggest that possession percentage is not a statistically significant predictor of success among the top 7 teams.

3. Wins and possession comparison by age:



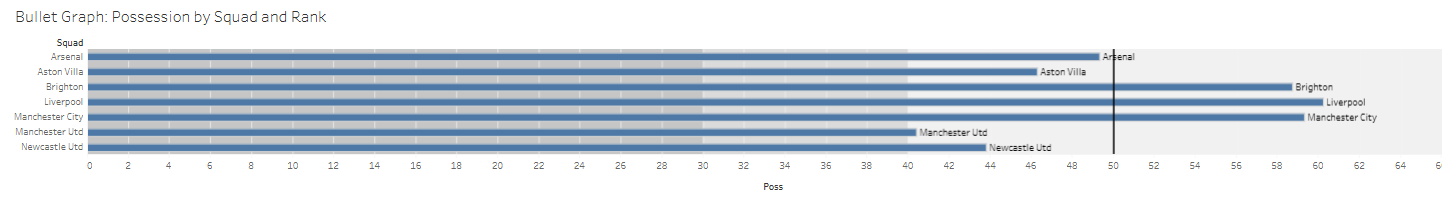
We used a bar graph to explore the relationship between possession, wins, and squad age for the top 7 Premier League teams. The graph allows us to compare these metrics and assess if there is a correlation between a squad’s average age and their possession percentage or number of wins. Although we aimed to determine if older squads have higher possession rates, our analysis concluded that squad age did not significantly impact performance or possession, as the average ages across the top teams were quite similar.

4. Wins and possession comparison by squad:



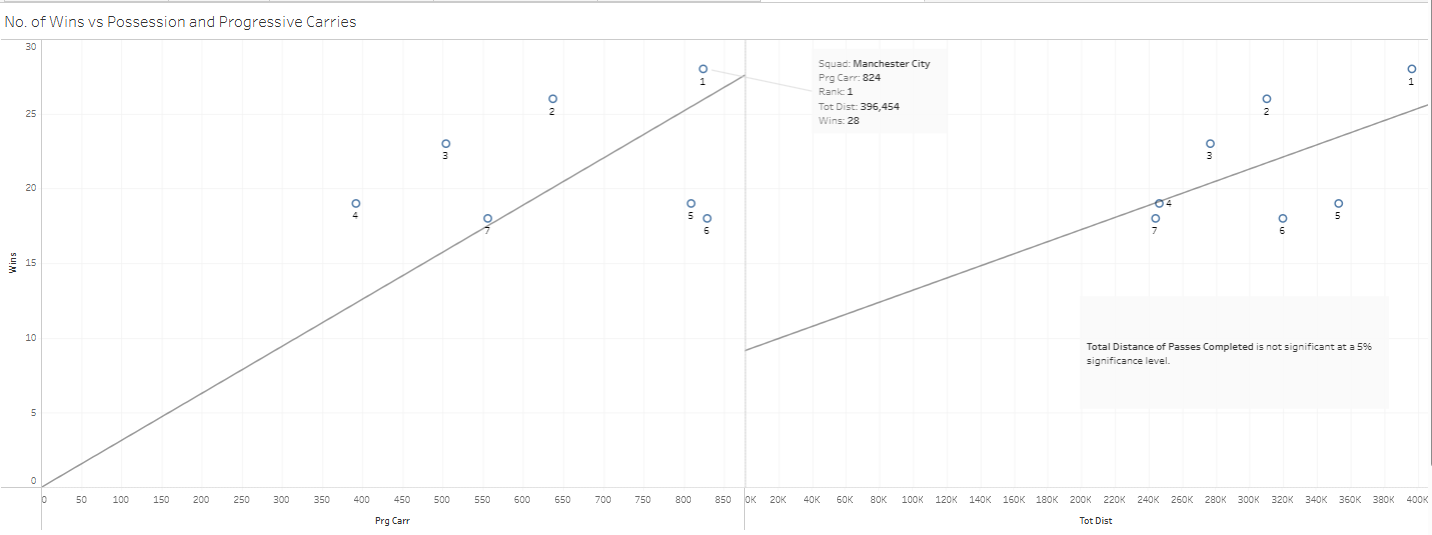
We used a bar graph to identify the relationship between possession percentage and wins for the top 7 Premier League teams. This visualization helps to compare the possession percentages and the number of wins for each team side by side. We can see that Manchester City, with the highest number of wins, also has a high possession percentage. However, the results suggest that possession alone is not a definitive predictor of success, as other teams with high possession percentages do not necessarily have the most wins or even rank in the top 7.

5. Wins vs Possession by Squad



We used a bullet graph to analyze the relationship between possession and wins for the top 7 Premier League squads. The graph displays possession percentages and the number of wins for each squad, allowing us to visually compare these metrics. This analysis reveals whether teams with higher possession percentages tend to have more wins. For example, Manchester City, ranked first, shows both high possession (59.3%) and a high number of wins (28). Additionally, we created a calculated field named "Expected Possession," which represents the baseline at 50% possession.

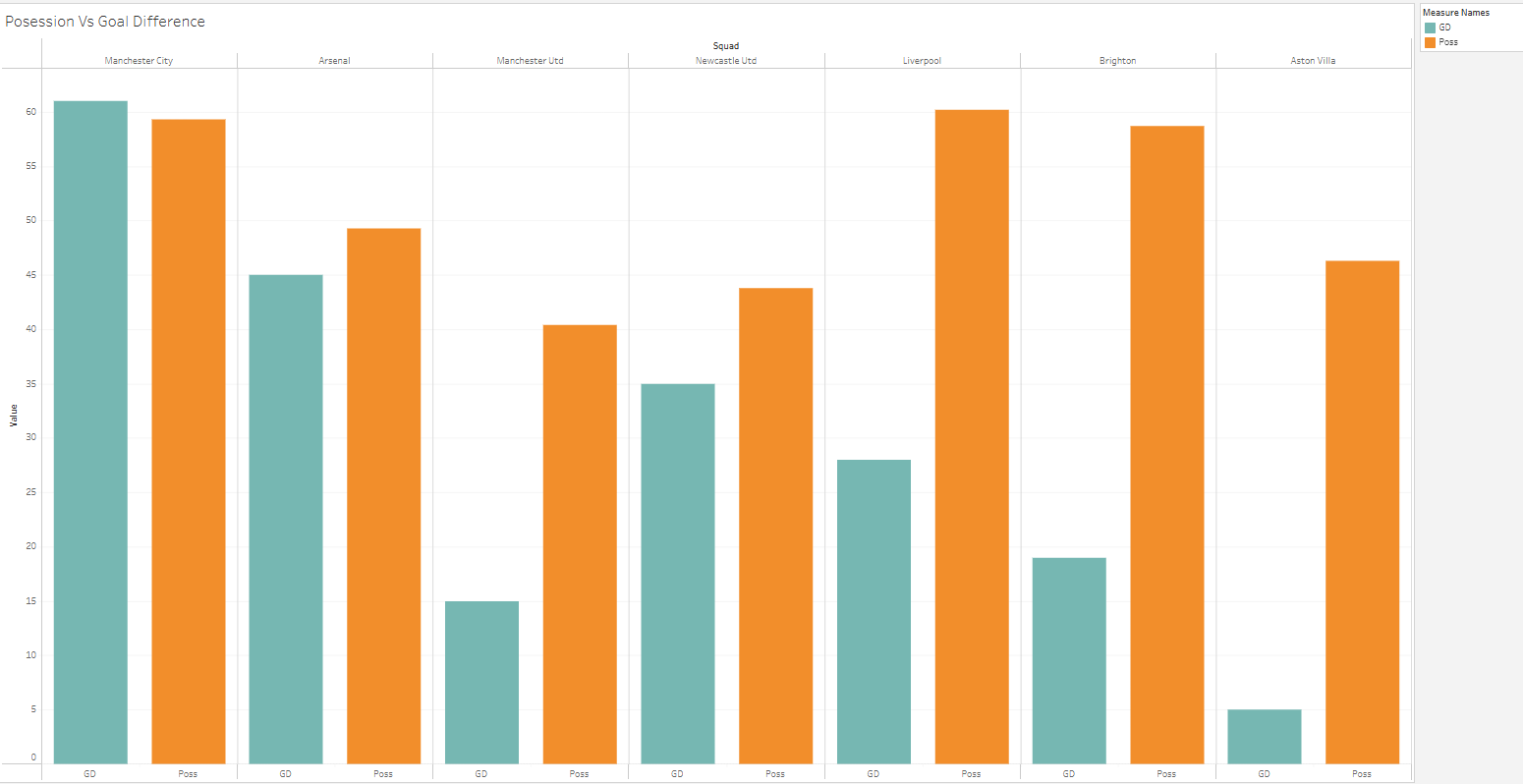
6. No. of Wins vs Possession and Progressive Carries:



· We created a scatter plot with the number of wins on the y-axis and progressive carries on the x-axis to identify visible patterns or trends. For instance, Manchester City, ranked first, is shown with 28 wins and 824 progressive carries. Analysis focused on the relationship between the number of wins and progressive carries by squad. This analysis can provide insights into the effectiveness of a team's playing style. We determined that there is a statistical relationship between the number of wins and progressive carries by calculating the correlation coefficient through linear regression. The analysis revealed a significant correlation, with the regression formula being Wins = 0.03 × Progressive Carries. This indicates that an increase in progressive carries is associated with an increase in the number of wins by 0.03.

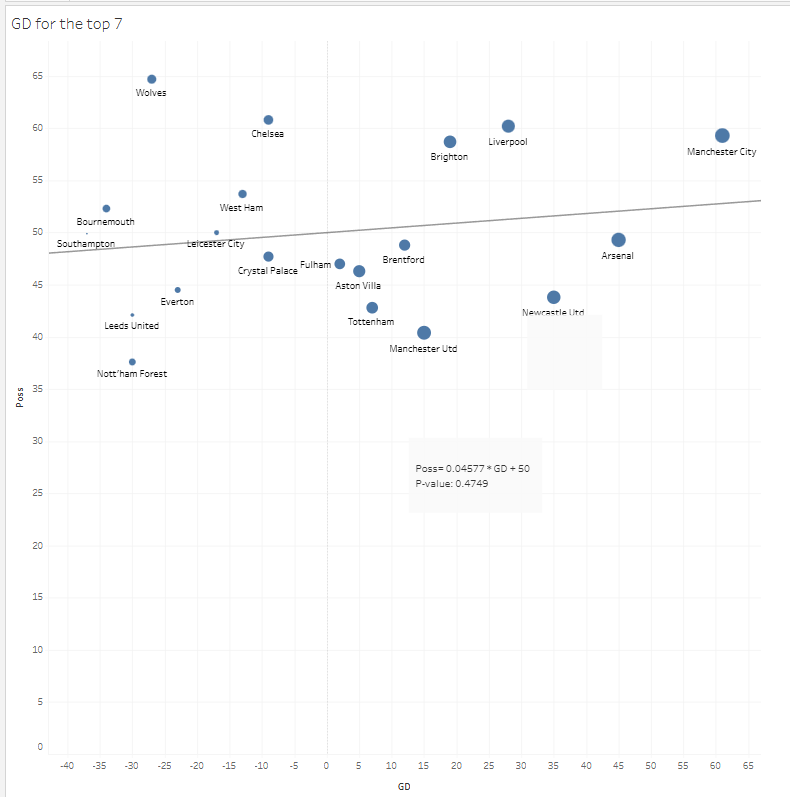
· We used a scatter plot with the number of wins on the y-axis and the total distance of passes completed on the x-axis to identify any visible patterns or trends. For example, Manchester City, ranked first, is displayed with 28 wins and a total distance of 396,454 completed passes. This analysis examined the relationship between the number of wins and the total distance of passes completed by each squad. The results suggest that the total distance of passes completed does not have a statistically significant correlation with the number of wins, indicating that increasing passing distance may not be associated with better performance.

7. Possession vs Goal Difference



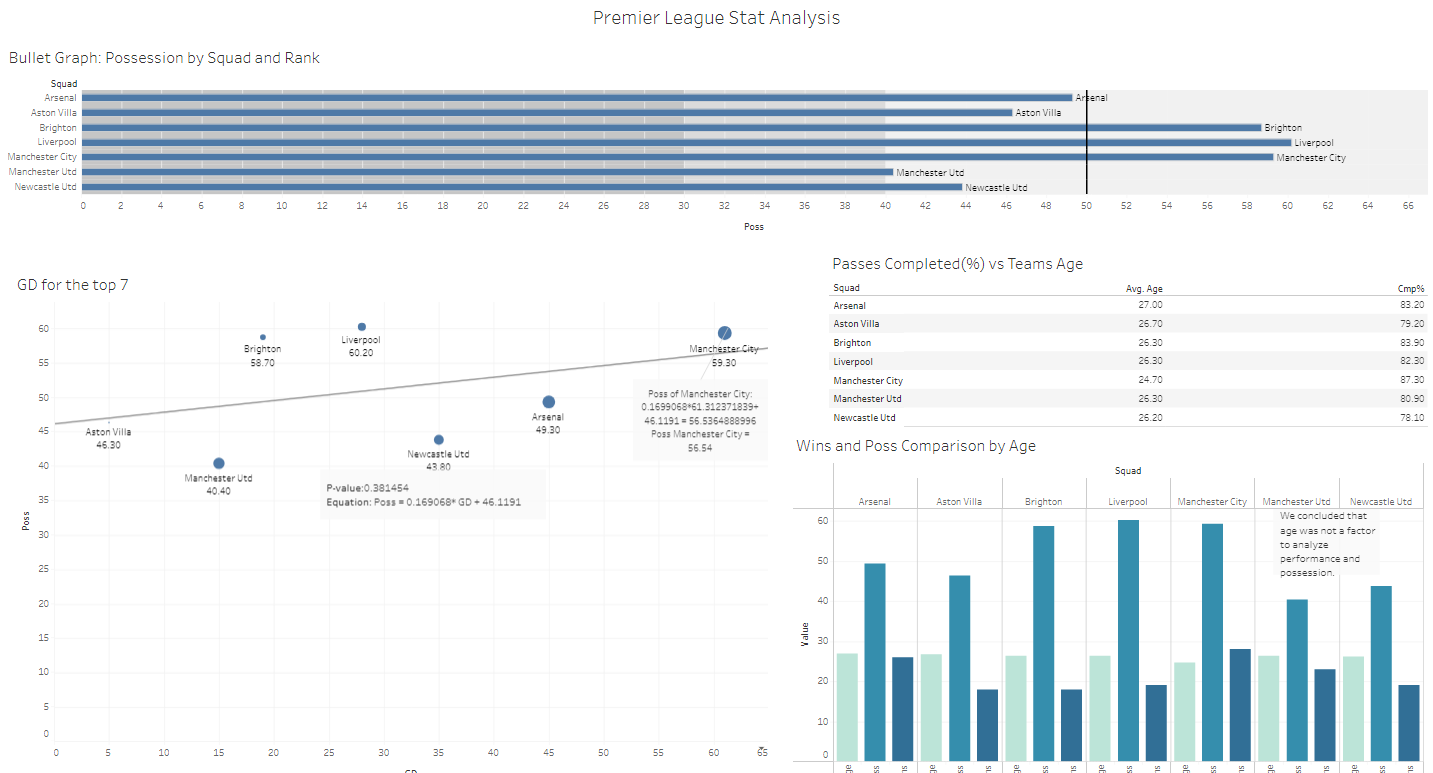
We used a bar graph to explore the relationship between possession percentage and Goal Difference for the top 7 Premier League teams. This visualization allows for a side-by-side comparison of possession percentages and Goal Difference for each team. For example, Manchester City, which has the highest Goal Difference, also boasts a high possession percentage. However, the results indicate that Goal Difference is a key predictor of success. Notably, achieving a Goal Difference of at least 19 appears to correlate with a top 7 finish. Additionally, high possession percentages do not always equate to the most wins or a top 7 ranking.

8.- Goal Difference vs Ranking:



To analyze the relationship between Possession Percentage, Ranking, and Goal Difference, we created a scatter plot. This visualization helps identify patterns or correlations between these metrics. In the scatter plot, Possession Percentage is plotted on the y-axis and Goal Difference on the x-axis, with each point representing a team. For example, Manchester City, which has the highest Goal Difference, is plotted with its corresponding Possession Percentage. The results suggest that Goal Difference is not a statistically significant predictor of Possession Percentage. However, achieving a Goal Difference of at least 19 appears to correlate with a top 7 finish.

9.- Dashboard:



### **Conclusion and Recommendations**

Based on our analysis, we recommend the following optimal playing style of Dynamic Progressive Possession for the team to enhance their chances of success in the Premier League.

This means that we want to implement a strategic approach that combines maintaining possession of the ball with a high tempo, fluid movement, and proactive play. It also emphasizes controlling the game through ball retention while also being aggressive and forward-thinking in creating scoring opportunities. This also encourage players, especially in the midfield and forward areas, to take risks, attempt dribbles, and make incisive passes.

We will also have a 4-3-3 formation and in some teams more known of having a more defensive style a 4-2-3-1 formation.

* + Goalkeeper: Quick distribution and confident in playing out from the back.
  + Defenders: Ball-playing center-backs and adventurous full-backs who can contribute to attacks.
  + Midfielders: A mix of a holding midfielder for balance, and central midfielders capable of progressive passes and dynamic movement.
  + Forwards: Wingers who can cut inside and a central striker who can link up play and finish chances.

This is our results that led to our findings:

1. **Scout Younger Players:**

Focus on recruiting players younger than 28 years old. Younger players tend to have higher stamina and can adapt to a dynamic playing style, contributing to overall team performance.

1. **Prioritize Experienced Defenders with High Possession:**

Ensure the team includes experienced defenders who can maintain high possession. Possession is crucial for controlling the game and reducing the opponent’s chances to score.

1. **Increase Progressive Passes:**

Emphasize progressive passes as a key strategy. Our analysis indicates a significant correlation between progressive passes and the number of wins. Increasing the number of progressive passes can lead to more scoring opportunities and, ultimately, more victories.

1. **Achieve a Goal Difference (GD) of at Least 19:**

Aim for a minimum goal difference of 19. Our findings show that teams with a GD of 19 or more are more likely to finish in the top 7. This means focusing on both scoring more goals and improving defensive strategies to concede fewer goals.

By implementing these strategies, the team can develop a more effective playing style, increasing their chances of finishing in the top 7 and qualifying for European competitions.