Analyzing Manchester City Soccer Games from 2018 to May 2023.

Abstract:

This project, which seeks to analyze soccer games of Manchester City from 2018 to May 2023, has 2 aims. This research project examines different soccer games of Manchester City and seeks to answer the question if there is a relation between the formation played by Manchester City and the outcome of the game. The study also addresses if player statistics such as shots, shots on target, possession percentage, fouls, corners, and pass completion rate relates with the total goals scored by Manchester City. Utilizing a sample of 200 randomly selected games, the analysis employs statistical tests and linear modeling techniques to answer these questions. Results showed that Manchester City generally win regardless of any formation they play in and "shots on target" is the most related factor with the total goals scored for Manchester City. We can predict total goals scored by Manchester City only by using Shots on Target, Yellow Cards, and Corners.

Written By:
Gaurab Baral
gaurabusa@gmail.com

Introduction (Background/Motivation):

Before we start looking at our research paper, let's understand why we're doing this. People who love soccer, analysts, and those who care about the game are curious about what makes a team like Manchester City keep winning. Manchester City is undoubtedly the best team, winning the most trophies in the past decade. By studying how they play, we want to figure out not only how they win but also share useful ideas about soccer strategies and performance.

In this research paper, we extracted and analyzed the soccer games played by Manchester City from 2018 to May 2023. We extracted our data from Google's database (Google Database) (see more at References) which stood as the exclusive repository for soccer game information and was highly reputable. Considering Manchester City's consistent success in winning various trophies in recent years, our objective was to scrutinize their gameplay to identify the factors such as passes, goals, and fouls that significantly impacted their performance. We were interested in two questions which led us to delve into various aspects and draw conclusions for each at the conclusion of the research paper. The questions were:

1. Was there any influence of playing formations (e.g., 4-4-2 or 4-4-3) on Win/Not Win outcomes of Manchester City soccer matches?
2. How did player statistics such as shots, shots on target, possession percentage, fouls, corners, and pass completion rate influence the total goals scored in a soccer match played by Manchester City?

Methods:

- <u>Variables:</u>
    - **Outcome**: Categorical (Win/Lose/Draw) => Result of the Soccer Game.
    - **Formation**: Categorical => The specific arrangement in which Manchester City played during the game.
    - **Shots**: Quantitative => A shot occurs when the ball is directed towards the opposition's goal, typically when the ball is near the goal post.
    - **Shots on Target:** Quantitative => Instances where the ball strikes the goal post, regardless of whether the goalkeeper saves it or not.
    - **Possession**: Quantitative => The percentage of time Manchester City had control of the ball during the game.
    - **Passes**: Quantitative => The total number of passes successfully completed by Manchester City.
    - **Fouls**: Quantitative => The overall count of fouls committed by Manchester City during the match.
    - **Yellow Cards**: Quantitative => The total number of yellow cards received for dangerous fouls committed by Manchester City players. Yellow cards are received when the fouls are extreme and dangerous.
    - **Corners**: Quantitative => The total number of corners awarded to Manchester City during the game. (Look at Soccer terms glossary in references for more details)
- <u>Data Collection:</u>
    - Initially, we gathered data from the Google Database of Manchester City. This data had 331 rows and 15 columns.
    - We sampled 200 random rows from our initial population dataset. (Appendix 1)
- <u>Data Pre-Processing:</u>
    - Out of the initial 200 data rows, we chose 182 rows by excluding games where the formation used was unique to the sample.

- We modified our data, categorizing "win" as "win", "lost" and "draw" as "no win".
- We visualized our data to look for any outliers but did not find any for the 182 rows.

- First Research Question:
    - For the first research question, we conducted a Test for Independence to determine if there was a relationship between the outcome of the game and the formation, they played in.
- Second Research Question:
    - For the second research question, we examined the correlation between all the explanatory variables and then removed one non-significant variable at a time, building appropriate linear models on each removal while keeping track of the adjusted R^2 value.
    - We also investigated the relationship between "Shots on Target (Quantitative)", "Outcome (Categorical)" and the Total goals (Quantitative and response variable) by building a regression model eliminating one variable at a time. We also predicted a few values (total scores) in different scenarios using this model.

Results:

Table1: Outcome based on Formation

Formation⬇	Not Win	Win	Total
3.2.4.1	3(27.3%)	8(72.7%)	100%
4.2.3.1	9(30%)	21(70%)	100%
4.3.3	29(20.6%)	112(79.4%)	

Figure1: Outcome based on Formation.

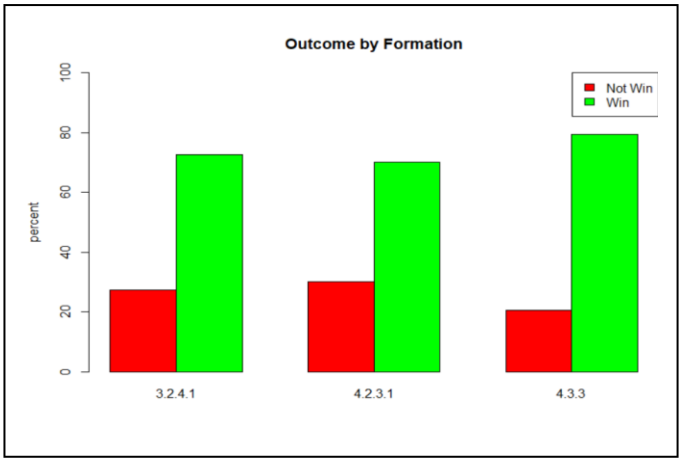


Figure 2: Wins distributed to Formations.

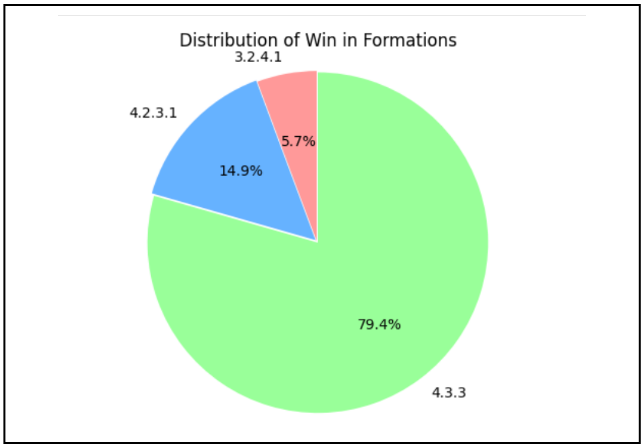
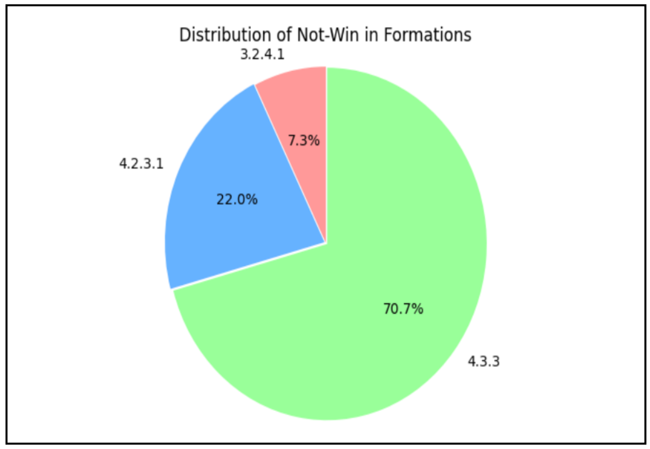


Figure 3: Not-Wins distributed to Formations.



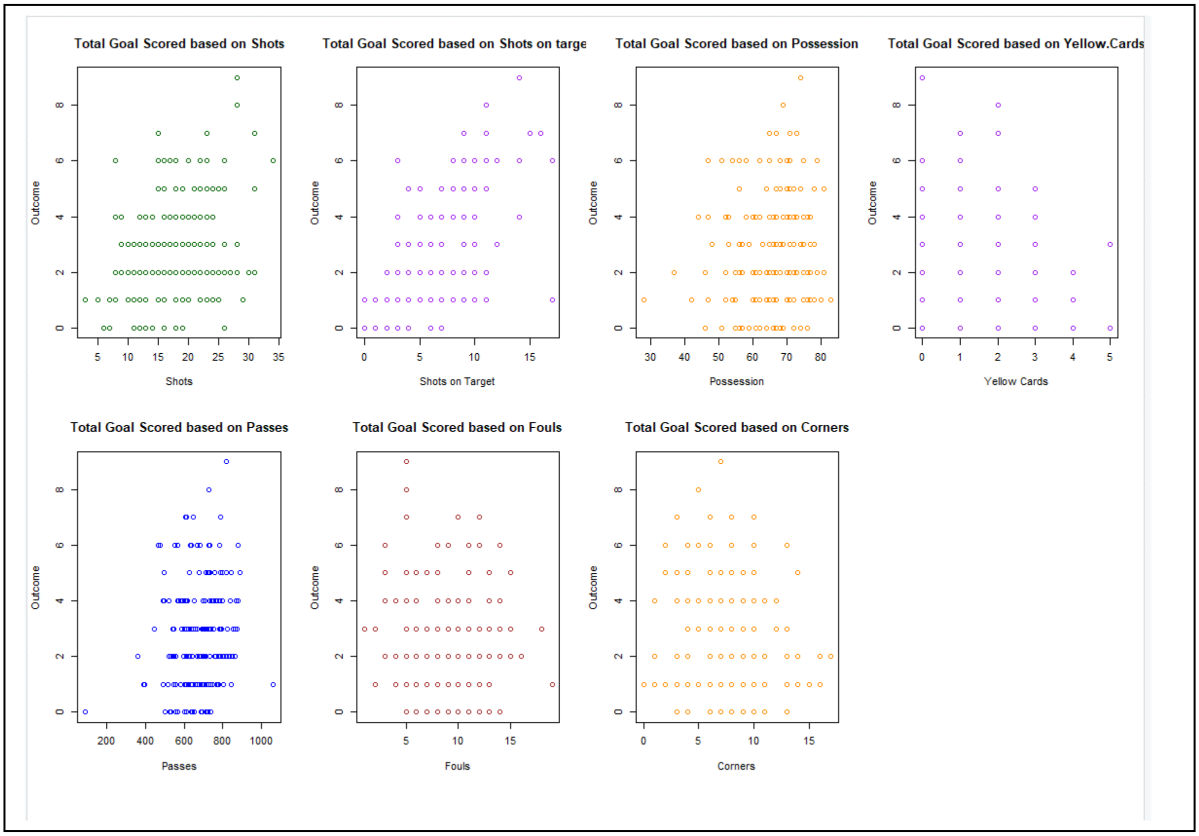
Looking at the numerical and graphical summaries *(Table 1 and Figure 1,2 and 3)*, we can say that the likelihood of Manchester City winning in their games is greater than the likelihood of not winning. We found that the chi-squared tests yielded p-values of 0.5277(Appendix 2A) indicating that there is no significant relationship between the outcome of the game alone and the formation played by Manchester City.
At the 95% significance level, the proportion of games that Manchester City wins is estimated to be around 70.58% to 83.20% and the proportion of games that Manchester City did not win is estimated to be around 16.82% to 29.42%. (Appendix 2B)

To answer research question #2, we calculate the correlation coefficient between the y (Total Goal Scored) and x (explanatory variables). We also built a scatter plot for the data which is displayed below.

Table2: Correlation Coefficient of Goal Scored with different factors

Goal Scored vs Shots	Goal Scored vs Shot on Target	Goal Scored vs Possession	Goal Scored vs Passes	Goal Scored vs Fouls	Goal Scored vs Yellow Cards	Goal Scored vs Corners
0.3331	0.6109	0.0964	0.1645	-0.0557	-0.1798	-0.1224

Figure 4: Scatter Plots of Total Goal Scored with explanatory variables



Looking at the graphical and numerical summaries (Table 2 and Figure 4) of:

- Total goals vs shots: relation is linear and positive. (Correlation: 0.3331)
- Total goals vs shot on target: relation is linear and positive. (Correlation: 0.6109)
- Total goals vs possession: relation is linear and neither positive nor negative. (Correlation: 0.0964)
- Total goals vs Passes: relation is linear and neither positive nor negative (Correlation: 0.1645)
- Total goals vs Fouls: relation is linear and neither positive nor negative (Correlation: -0.0557)
- Total goals vs yellow cards: relation is linear and neither positive nor negative. (Correlation: -0.1798)
- Total goals vs Corners: relation is linear and neither positive nor negative (Correlation: -0.1224)

The linear model with all 7 variables (Shots, Shots on Target, Possession, Passes, Fouls, Yellow: Cards and Corners) of Manchester City had a p-value of 2.2e-16. (Appendix 3). This signifies that there is a relation between the explanatory variables. We removed one insignificant variable at a time and the details of backward elimination model is presented in the table below: (Appendix 4)

	Adjusted R^2:	Remove:
All 7 Variables:	0.4554	Fouls (p-value: 0.9451)
All 7 Variables excluding Fouls	0.4585	Possession (p-value: 0.6417)
All 7 Variables excluding Fouls, Possession	0.4609	Passes (p-value: 0.3694)
All 7 Variables excluding Fouls, Possession, Passes	0.4615	Shots (p-value: 0.2824)
All 7 Variables excluding Fouls, Possession, Passes and Shots	0.461	Nothing

For the final model, 46.99% of the variation in total goal scored by Manchester City can be explained by Shots on Target, Yellow Cards and Corners received by Manchester City. The final equation for this model is:

$$\hat{y} = 1.58224 + 0.39679(Shots\ on\ Target) - 0.21107(Yellow\ Cards) - 0.16494(Corners)$$

where $\hat{y}$ is the total goal scored by Manchester city.

From the provided equation, the following observations can be made:

- Holding Yellow Cards and Corners constant, an increase in Shots on Target is associated with an expected increase of 0.39679 in the total goals scored by Manchester City.
- With Yellow Cards and Shots on Target held constant, obtaining an additional corner is associated with an expected decrease of 0.16494 in the total goals scored by Manchester City.
- Keeping Shots on Target and Corners constant, each yellow card received by Manchester City is associated with an expected decrease of 0.21107 in the total goals scored.

Five different scenarios were created to predict the total goals scored by Manchester City. (Appendix 5):

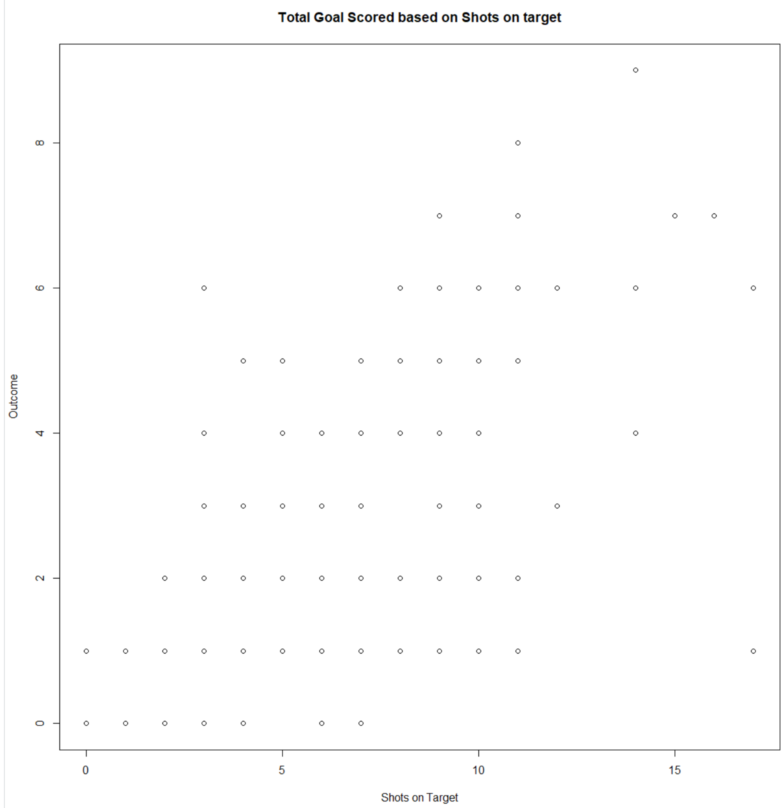
Scenario	Shots on Target	Yellow Cards	Corners	Total Goal Scored ($\hat{y}$)	Lower Prediction Interval	Higher Prediction Interval
1	5	1	2	3.025469	0.2805848	5.770354
2	6	2	3	3.046282	0.3064310	5.786092
3	8	2	2	4.004786	1.2530952	6.756478
4	10	1	6	4.349693	1.6116563	7.087730
5	15	3	4	6.241407	3.4198916	9.062922

Scenario 1: In a game where Manchester City had 5 shots on target, 1 yellow card and 2 corner kicks, it can be predicted that the total goal scored will be between 0 and 6. (95% confidence, Appendix 5)

Scenario 2: In a game where Manchester City had 6 shots on target, 2 yellow cards and 3 corner kicks, it can be predicted that the total goal scored will be between 0 and 6. (95% confidence, Appendix 5)

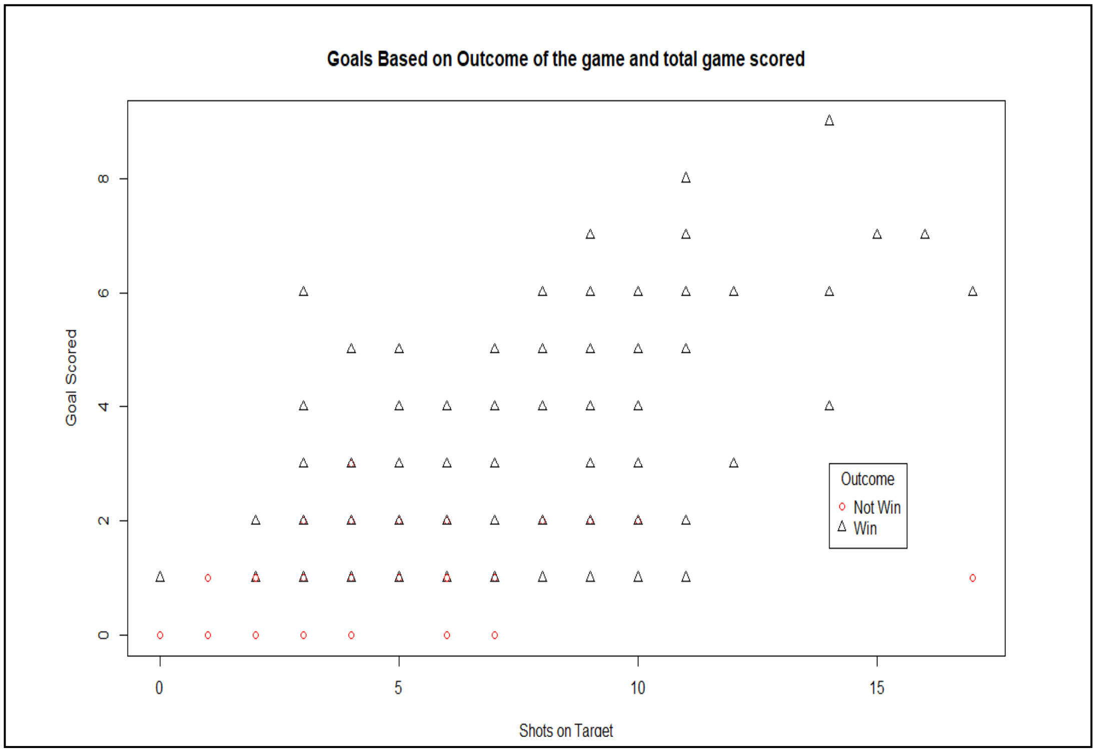
Scenario 3: In a game where Manchester City had 8 shots on target, 2 yellow cards and 2 corner kicks, it can be predicted that the total goal scored will be between 1 and 7. (95% confidence, Appendix 5)
Scenario 4: In a game where Manchester City had 10 shots on target, 1 yellow card and 6 corner kicks, it can be predicted that the total goal scored will be between 1 and 8. (95% confidence, Appendix 5)
Scenario 5: In a game where Manchester City had 15 shots on target, 3 yellow cards and 4 corner kicks, it can be predicted that the total goal scored will be between 3 and 10. (95% confidence, Appendix 5)

Figure 5: Scatter Plot of Total Goal Scored with the most significant explanatory variable.



The most significant explanatory variable in our case is Shots on Target from the correlation table (Table 2), the figure (Figure 5) and the linear model (Appendix 3). Hence, we look to see if total goals scored by Manchester city can be predicted by the outcome of the game and the total shots on target by Manchester City.

Figure 6: Scatter Plot of Total Goal Scored with outcome and shots on target.



Upon examining Figure 6, the dataset exhibits a linear trend with minimal outliers, leaning towards sort of a positive correlation. The instances marked by the red circle, indicating 'Not Win,' generally appear slightly lower compared to those marked by the black triangle, denoting 'Win.' Specifically, games won by Manchester City tend to have a higher number of goals scored compared to those lost or drawn by the team.

When looking at all variables to predict the goals scored by the team, it was determined that there is statistically significant evidence to support that both shots on target and outcome of the game are useful. (p-value=2.2e-16) (Appendix 6). For this model, 49.42% of the variation in goals scored can be explained by shots on target and the outcome of the game.

Our final equation that uses the best model to predict the total goals scored by Manchester City is:

$$\hat{y} = -0.41739 + 0.27841(Shots\ on\ Target) + 1.67882(Outcome\ of\ the\ game)$$

where $\hat{y}$ is the total goals scored by Manchester City.

From this equation we can say that:

- For every unit increase in Shots on Target, we expect the total goals to rise by 0.27841, assuming the outcome of the game remains constant. (Appendix 6)
- When Shots on Target remains constant, if Manchester City is winning the game, the goals scored will rise by 1.67882. (Appendix 6).

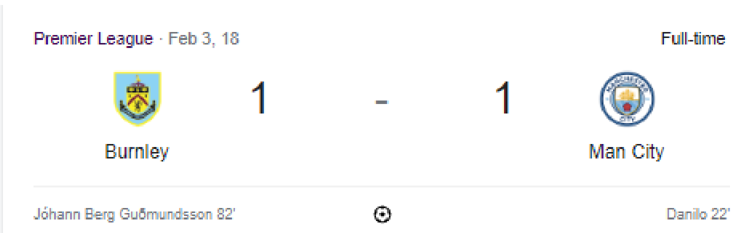
Two different game scenarios were created to predict the goal scored for Manchester City. (Appendix 7)

Scenario	Outcome of the Game	Shots on Target	Total Goal Scored	Lower Prediction Interval	Higher Prediction Interval
1	0(Lose)	8	1.81	-0.88	4.50
2	1(Win)	8	3.49	0.83	6.15

Scenario 1: For a game lost by Manchester City with 8 shots on target, it can be predicted that the total goal scored will be between -1 and 5 where -1 implies that they concede a goal.
Scenario 2: For a game won by Manchester City with 8 shots on target, it can be predicted that the total goal scored will be between 1 and 7.

Navigating Outliers:

Looking at Figure 6, we see a red dot near shots on target = 16. In this case, Manchester City lost the game. Why so? How did they lose when they had 16 shots in target which is higher than usual. Looking into the dataset, we can see that the game was played between Burnley and Manchester City.



	Year	Goals.SCORED	Goals.Conceded	WIN.LOSE.DRAW	Formation	Shots	Shot.on.Target	Possession	Passes	Fouls	Yellow.Cards	Red.Cards	OffSides	Corners	RedCardsFlag
1	2018	1	1	Draw	4.3.3	20	17	71	695	6	1	0	0	13	No

City were chasing their 23rd win in 26 league games this season and trying to become only the third team in the Premier League era (after Manchester United in 2000 and Chelsea in 2006) to hold an 18-point lead at the top of the table.
Some of the main reasons why I think City lost this game are:

- Manchester City tried to be defensive after the first goal they scored and when Burnley scored in the last quarter of the game, they could not switch from defensive to offensive gameplay, especially as they had already subbed out one of the main strikers.
- Burnley showcased their top-notch performance, executing a robust game plan with a relentless focus on tackles, leaving little room for the opposition to mount any threats on their goal.
- City did not play this game at their home and did not have the support of the home crowd.
- City did learn from this and proceeded to win 5-1 the next game and even won the season with 19 points higher than the runner up team.

Discussion/Conclusion:

This study aimed to assess the impact of formation on game outcomes for Manchester City and understand the influence of key performance indicators (shots, shot on target, possession, yellow cards, corners) on total goals scored. Moving forward, we can do the following:

- **Strategic Adjustments**: Given that formation did not show a significant association with outcomes, future strategies for Manchester City might consider focusing on other aspects such as player dynamics, tactics, or opponent analysis to enhance performance.
- **Key Performance Indicators (KPIs):** The identified KPIs (shots on target, corners, yellow cards) can guide coaching staff and players in training and match preparations. Emphasizing these areas could potentially improve goal-scoring opportunities.
- **Continuous Monitoring:** Regularly monitoring and analyzing these KPIs during the season can provide real-time insights. Adjustments to game strategies and player performance can be made based on ongoing data analysis.

Some limitations and suggestions for future research is mentioned below:

- **Data Quality:** Future studies should strive for more extensive and accurate datasets, possibly incorporating more granular details such as player-specific statistics.
- **Additional Variables:** While the study focused on specific variables, exploring additional factors such as weather conditions, player injuries, or match context could provide a more nuanced perspective on game outcomes.
- **Machine Learning Approaches:** Integrating machine learning models can offer more sophisticated insights, allowing for nonlinear relationships and predictive analyses. This could enhance the accuracy of outcome predictions and provide more actionable recommendations.
- **Qualitative Analysis:** Supplement quantitative findings with qualitative analysis, including player interviews or coach assessments, to gain a better understanding of the factors influencing team performance.
- **External Validity:** We can think about whether the findings we found out in our study can apply to all football teams or if they only relate to one team. We can check if there are special things about each team that affect the results.

In summary, our study revealed that while formation may not strongly influence game outcomes for Manchester City, key performance indicators like shots on target, corners, and yellow cards play a crucial role in goal-scoring. These insights can inform strategic decisions and training focus, emphasizing the importance of specific actions during matches for achieving success.

References:

manchester city games - Google Search. (n.d.). https://www.google.com/search?q=manchester+city+games&rlz=1C1UEAD_enUS931US931&oq=manch&gs_lcrp=EgZjaHJvbWUqCAgAEEUYOBg7MggIABBFGDgYOzIKCAEQLhixAxiABDIGCAIQRRg5MgYIAxAjGCcyDwgEEC4YFBiHAhixAxiABDIGCAUQRRg9MgYIBhBFGDwyBggHEEUYPdIBBzkyMmoxajeoAgCwAgA&sourceid=chrome&ie=UTF-8#sie=t;/m/01634x;2;/m/02_tc;mt;fp;1;;;

Soccer terms glossary. (n.d.). https://www.socceramerica.com/glossary/

Appendix 1:
Analysis Choosen: Sample Random Rows
Justification: We sampled out 200 games randomly from the original dataset.

```
# Load necessary package
install.packages("dplyr")
library(dplyr)
# Read the dataset
data <- read.csv("C:\\Users\\OSAKWEJ1\\Downloads\\STA305.csv")
# Randomly sample 200 rows
sampled_data <- data %>% sample_n(200)
```

Appendix 2A:
Analysis Choosen: Chi-Square Test for Independence
Justification: Categorical Data with 3 options.

```
RQ1.table <- xtabs(formula = ~Formation + WIN.LOSE.DRAW, data = new_data)
chisq.test(RQ1.table,simulate.p.value = TRUE)
```

```
Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data:  RQ1.table
X-squared = 1.4122, df = NA, p-value = 0.5277
```

Appendix 2B:
Analysis Choosen: One Sample Proportion Test
Justification: RRUS: no known bias, np,nq = (41,141) >=10

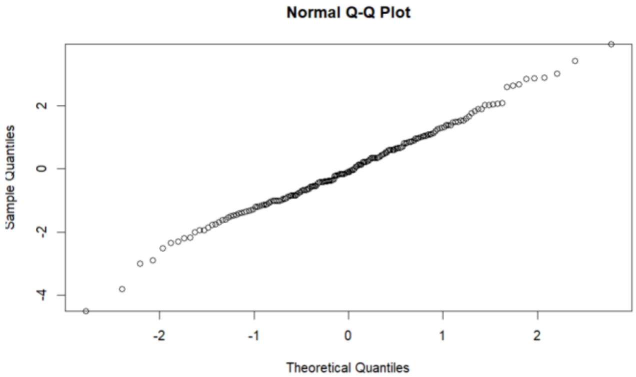
```
prop_test_result <- prop.test(141, 182, conf.level = 0.95)
prop_test_result
prop_test_result2 <- prop.test(41, 182, conf.level = 0.95)
prop_test_result2
```

```
95 percent confidence interval:
 0.7057738 0.8318128
```

```
95 percent confidence interval:
 0.1681872 0.2942262
```

Appendix 3:
Analysis Choosen: Linear Model
Justification:

- Normality:

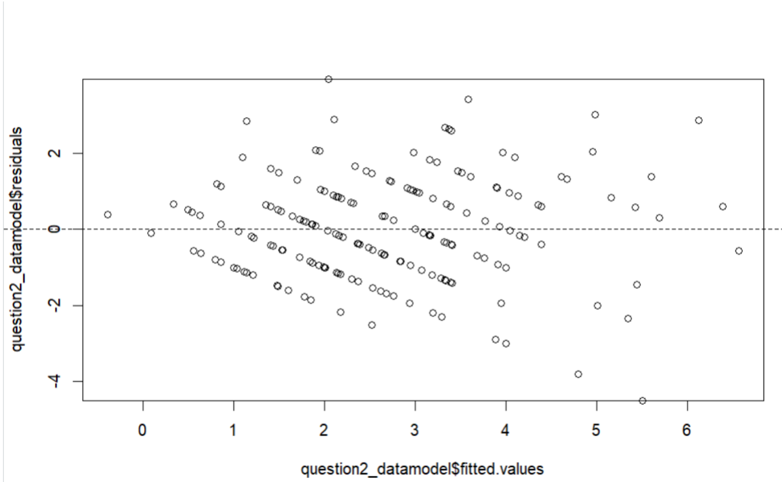



```
          Shapiro-Wilk normality test

data:  question2_datamodel$residuals
W = 0.99385, p-value = 0.6501
```

The QQ plot looks linear and the Shapiro wilk value = 0.6501, fail to reject normality.

- Constant Standard Deviation:



The plot of predicted vs residuals seems to have somewhat linear spread.

- Independence:

There is no known dependence. No reason to match one game to another.

Linear Model:

question2_datamodel<-lm(Goals.SCORED~ Shots + Shot.on.Target + Possession + Passes +Fouls +Yellow.Cards+ Corners, data=question2_data)
> summary(question2_datamodel) #adf.r^2 = 0.4554

```
Call:
lm(formula = Goals.SCORED ~ Shots + Shot.on.Target + Possession +
    Passes + Fouls + Yellow.Cards + Corners, data = question2_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5059 -0.9326 -0.0874  0.8679  3.9534

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.012374   0.924609   1.095   0.2751
```

```
Shots           0.034277   0.027986   1.225   0.2223
Shot.on.Target  0.357038   0.045180   7.903 2.99e-13 ***
Possession     -0.008476   0.018915  -0.448   0.6546
Passes          0.001288   0.001369   0.941   0.3481
Fouls           0.002528   0.036641   0.069   0.9451
Yellow.Cards   -0.181057   0.104879  -1.726   0.0861 .
Corners        -0.183779   0.039025  -4.709 5.06e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.385 on 174 degrees of freedom
Multiple R-squared:  0.4764,	Adjusted R-squared:  0.4554
F-statistic: 22.62 on 7 and 174 DF,  p-value: < 2.2e-16
```

Appendix 4:

Analysis Choosen: Linear Model

Justification: Removing one significant explanatory variable at a time.

```
Call:
lm(formula = Goals.SCORED ~ Shots + Shot.on.Target + Possession +
    Passes + Yellow.Cards + Corners, data = question2_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5133 -0.9373 -0.0917  0.8652  3.9483

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.045096   0.791408   1.321   0.1884
Shots           0.034287   0.027906   1.229   0.2209
Shot.on.Target  0.357198   0.044992   7.939 2.36e-13 ***
Possession     -0.008682   0.018624  -0.466   0.6417
Passes          0.001286   0.001365   0.942   0.3475
Yellow.Cards   -0.178250   0.096389  -1.849   0.0661 .
Corners        -0.183743   0.038910  -4.722 4.76e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.381 on 175 degrees of freedom
Multiple R-squared:  0.4764,	Adjusted R-squared:  0.4585
F-statistic: 26.54 on 6 and 175 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Goals.SCORED ~ Shots + Shot.on.Target + Passes +
    Yellow.Cards + Corners, data = question2_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5295 -0.9622 -0.1026  0.8551  3.9677

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.8790368  0.7051387   1.247   0.2142
```

```
Shots          0.0310058  0.0269440   1.151   0.2514
Shot.on.Target  0.3594183  0.0446394   8.052 1.18e-13 ***
Passes          0.0008066  0.0008964   0.900   0.3694
Yellow.Cards   -0.1801092  0.0960925  -1.874   0.0625 .
Corners        -0.1892484  0.0369927  -5.116 8.11e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.378 on 176 degrees of freedom
Multiple R-squared:  0.4758,	Adjusted R-squared:  0.4609
F-statistic: 31.95 on 5 and 176 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Goals.SCORED ~ Shots + Shot.on.Target + Yellow.Cards +
    Corners, data = question2_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5958 -0.9363 -0.0449  0.8997  3.8198

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.42424    0.36052   3.950 0.000113 ***
Shots           0.02893    0.02683   1.078 0.282368
Shot.on.Target  0.36596    0.04402   8.313 2.39e-14 ***
Yellow.Cards   -0.19975    0.09353  -2.136 0.034083 *
Corners        -0.18682    0.03687  -5.066 1.01e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.377 on 177 degrees of freedom
Multiple R-squared:  0.4734,	Adjusted R-squared:  0.4615
F-statistic: 39.77 on 4 and 177 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Goals.SCORED ~ Shot.on.Target + Yellow.Cards + Corners,
    data = question2_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4.9727 -0.9170 -0.0416  0.9511  3.7681

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.58244    0.32947   4.803 3.30e-06 ***
Shot.on.Target  0.39679    0.03348  11.852  < 2e-16 ***
Yellow.Cards   -0.21107    0.09298  -2.270   0.0244 *
Corners        -0.16494    0.03080  -5.355 2.62e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 1.377 on 178 degrees of freedom
Multiple R-squared: 0.4699, Adjusted R-squared: 0.461
F-statistic: 52.6 on 3 and 178 DF, p-value: < 2.2e-16

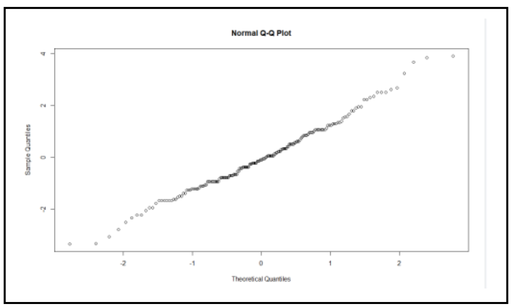
Appendix 5:
Analysis Choosen: Prediction Interval
Justification: Prediction Interval at 95% confidence interval

```
new_data <- data.frame(
  Shot.on.Target = c(5, 6, 8, 10, 15),
  Yellow.Cards = c(1, 2, 2, 1, 3),
  Corners = c(2, 3, 2, 6, 4))
predict(question2_datamodel5, newdata
= new_data, level = 0.95, interval =
"predict")
```

```
       fit       lwr      upr
1 3.025469 0.2805848 5.770354
2 3.046262 0.3064310 5.786092
3 4.004786 1.2530952 6.756478
4 4.349693 1.6116563 7.087730
5 6.241407 3.4198916 9.062922
```

Appendix 6:
Analysis Choosen: Linear Regression with Indicators.
Justification: The linear regression to predict total goal scored using shots on target and outcome of the game as indicator.

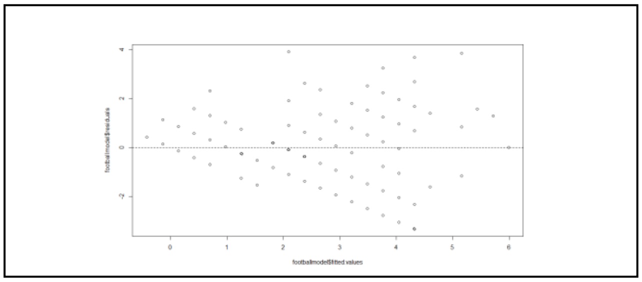
- Normality: The QQ plot looks linear and the Shapito Wilk p-value: 0.1677. Thus, it fails to reject normality.



```
        Shapiro-Wilk normality
test

data:  footballmodel$residuals
W = 0.98892, p-value = 0.1677
```

- Constant Standard Deviation: The plot of residuals vs fitted values seems to have somewhat of a constant spread.



- Independence: No known Dependence

```
Call:
lm(formula = Goals.SCORED ~ Shot.on.Target + WIN.LOSE.DRAW, data = extra
_credit)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.3239 -0.9319 -0.0925  0.8927  3.9034 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -0.41739    0.25683  -1.625    0.106    
Shot.on.Target  0.27841    0.03383   8.231 3.75e-14 ***
WIN.LOSE.DRAW   1.67882    0.25650   6.545 6.06e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.342 on 179 degrees of freedom
Multiple R-squared:  0.4942,	Adjusted R-squared:  0.4886 
F-statistic: 87.46 on 2 and 179 DF,  p-value: < 2.2e-16
```

Appendix 7:
Analysis Choosen: Prediction Interval
Justification: Prediction Interval for model built in Appendix 6.

Scenario	Outcome of the Game	Shots on Target	Total Goal Scored	Lower Prediction Interval	Higher Prediction Interval
1	0(Lose)	8	1.809857	-0.8807028	4.500416
2	1(Win)	8	3.488674	0.8311277	6.146221

```
new <- data.frame(Shot.on.Target=c(8, 8), WIN.LOSE.DRAW=c(0, 1))
predict(footballmodel, newdata=new, level = 0.95, interval = "predict")
```