UFC Analysis - IDS 702 Final Project

Arko Bhattacharya, Eric Ortega Rodriguez, Mu Niu, Nruta Choudhari

2024-12-15

Abstract

Introduction

The Ultimate Fighting Championship (UFC) is the world's leading mixed martial arts (MMA) promotion, known for bringing together elite fighters from diverse combat sports backgrounds. Founded in 1993, the UFC has grown into a global phenomenon, hosting events worldwide that showcase athletes competing in disciplines such as boxing, wrestling, Muay Thai, Brazilian Jiujitsu, and judo. UFC fights take place in a distinct eight-sided cage, known as the Octagon, where fighters test their skills in striking, grappling, and overall strategy under a unified set of rules. The sport has evolved significantly over the years, introducing standardized weight classes, safety regulations, and scoring systems to ensure competitive fairness and fighter safety.

This project examines UFC performances using data on UFC fights from 2010 to the present (last updated in November, 2024). The data, sourced from Kaggle, includes key fighter metrics, fight outcomes, betting odds, and performance indicators such as strikes landed and submission attempts. By leveraging this dataset, we aim to analyze factors influencing fight outcomes and performances.

Our research questions are:

1. How does the reach of the fighter relate to the total number of strikes landed during a fight?
2. Is the fight outcome associated with the number of submission attempts made by a fighter?

These questions are worth exploring because they provide a deeper understanding of UFC performance dynamics. For instance, examining the relationship between a fighter's reach and the total number of strikes landed can underscore the tactical performance of physical attributes in effective striking. Similarly, analyzing the association between fight outcomes and submission attempts can shed light on the strategic role of grappling in securing victories.

The findings from this analysis offer valuable insights for fighters, coaches, and analysts, helping optimize training strategies, improve fight preparation, and enhance understanding of opponents' strengths and weaknesses.

Methods

Data and Preprocessing

The dataset was obtained from Kaggle, a widely recognized platform for sharing datasets and data science resources. Each row of the dataset refers to an individual bout, which refers to an individual match between two fighters. This includes data on fighter attributes such as **height, weight, reach, stance, and age**, as well as fight statistics like **strikes landed, significant strikes, takedowns, submission attempts, and knockdowns**. Additionally, it documents fight outcomes, including the **winner**, **method of victory** (e.g., knockout, submission, decision), the **round in which the fight ended**, and the **total duration of the fight**.

The dataset contains 6,478 rows across 118 columns, with several variables containing missing values. During preprocessing, columns with over 6,000 missing values were dropped due to their lack of significance and the infeasibility of imputation. Other columns had a smaller proportion of missing values, and rows with missing values in key variables (e.g., strikes landed, reach, and weight class) were removed. This resulted in a final dataset with 4,895 rows. Most of the missing values were concentrated in performance metrics, such as submission attempts or specific strike statistics.

In UFC, fighters are assigned to either the red corner or the blue corner, which indicates their position in the Octagon and helps differentiate between competitors. For the first research question, the dataset was filtered to include the variables related to reach, weight class, height, strikes landed and current win streak, ensuring that key confounding variables were included. The data for fighters in the red and blue corners were combined into a single dataframe to facilitate analysis.

For the second research question, a new binary variable, Outcome was created to indicate the winner. A value of 1 was assigned if the fighter in the red corner won, and a value of 0 if the fighter in the blue corner won. The model included variables such as submission attempts, significant strikes landed, fight duration, and weight class to account for both physical attributes and performance metrics. These variables ensured a more comprehensive analysis of the factors influencing fight outcomes while addressing potential confounders.

Model Fitting and Evaluation

To examine the relationship between a fighter's reach and the total number of strikes landed during a fight, a Multiple Linear Regression (MLR) model was utilized. The model included key predictors such as logarithmic transformation of reach, logarithmic transformation of height,

win streak and weight class, with an interaction term between the win streak and weight class to explore potential moderating effects. Outliers and influential points were identified using Cook's distance, and there were removed to improve model robustness. Diagnostics, including residuals vs. fitted plots, were performed to assess linearity and homoscedasticity, while Variance Inflation Factor (VIF) was used to evaluate multicollinearity. Model performance was measured using the R-squared value.

For the second research question, a logistic regression model was employed to predict fight outcomes (binary: win or loss) using submission attempts, reach, significance strikes, fight duration, and weight class as predictors. The model was refined using stepwise selection to identify the most significant predictors, and diagnostics such as Cook's distance, leverage, and deviance residuals were used to detect and remove influential points. The final logistic regression model included key predictors such as logarithmic transformations of submission attemps, logarithmic transformation of reach, logarithmic transformation of significant strikes landed, and logarithmic transformation of fight duration. Model performance was evaluated using the area under the receiver operating characteristic curve (ROC curve), and diagnostic plots were generated to assess the model's fit.

All the analyses were conducted in R.

Results

Research Question 1: How does the reach of the fighter relate to the total number of strikes landed during a fight?

To understand the data distribution, we first computed summary statistics for the key variables: **Reach, Weight Class, Height, Win Streak, and Average Significant Strikes Landed**. Continuous variables are reported as means with standard deviations, and categorical variables are reported as counts and percentages.

```
# Summarize the data
summary_table <- ufc_q1 %>%
  group_by(WeightClass) %>%
  summarize(
    N = n(),
    Avg_Reach = paste0(round(mean(ReachCms, na.rm = TRUE), 1), " ± ", round(sd(ReachCms, na.r
    Avg_Height = paste0(round(mean(Height, na.rm = TRUE), 1), " ± ", round(sd(Height, na.rm =
    Avg_Strikes = paste0(round(mean(AvgSigStrLanded, na.rm = TRUE), 1), " ± ", round(sd(AvgSi
    Median_Streak = paste0(median(WinStreak, na.rm = TRUE), " [", quantile(WinStreak, 0.25, r
  )

# Render the table
kable(summary_table, caption = "Summary Statistics by Weight Class", align = "c")
```

Table 1: Summary Statistics by Weight Class

WeightClass	N	Avg_Reach	Avg_Height	Avg_Strikes	Median_Streak
Bantamweight	1015	174.7 ± 6	170.8 ± 4.3	21.2 ± 21.1	1 [0-2]
Catch Weight	77	180.8 ± 10.1	176.6 ± 8.6	8.6 ± 11.4	1 [0-2]
Featherweight	1118	179.7 ± 5.7	175.2 ± 5	22.3 ± 21.1	1 [0-2]
Flyweight	523	170.4 ± 5.7	167 ± 4.3	20.3 ± 21.4	1 [0-2]
Heavyweight	715	197.4 ± 7.2	190.7 ± 5.8	18.4 ± 17.4	1 [0-2]
Light Heavyweight	757	194.3 ± 6.5	188.2 ± 4.3	21.4 ± 18.2	1 [0-2]
Lightweight	1665	181.8 ± 5.6	177.2 ± 4.6	23.3 ± 19.2	1 [0-2]
Middleweight	1188	190.6 ± 5.9	185 ± 4.3	19.5 ± 17.4	1 [0-2]
Welterweight	1527	187.1 ± 6	181.8 ± 4.4	23.3 ± 19.3	1 [0-2]
Women's Bantamweight	302	170.6 ± 5.3	169.3 ± 4.4	21.2 ± 24.8	1 [0-1]
Women's Featherweight	35	174.6 ± 5.6	171.1 ± 6.1	6.9 ± 10.3	1 [0-1]
Women's Flyweight	355	168.6 ± 5.8	166.5 ± 4	11.3 ± 19.1	1 [0-2]
Women's Strawweight	458	162.1 ± 5.7	161.5 ± 4.6	22.6 ± 28.1	1 [0-2]

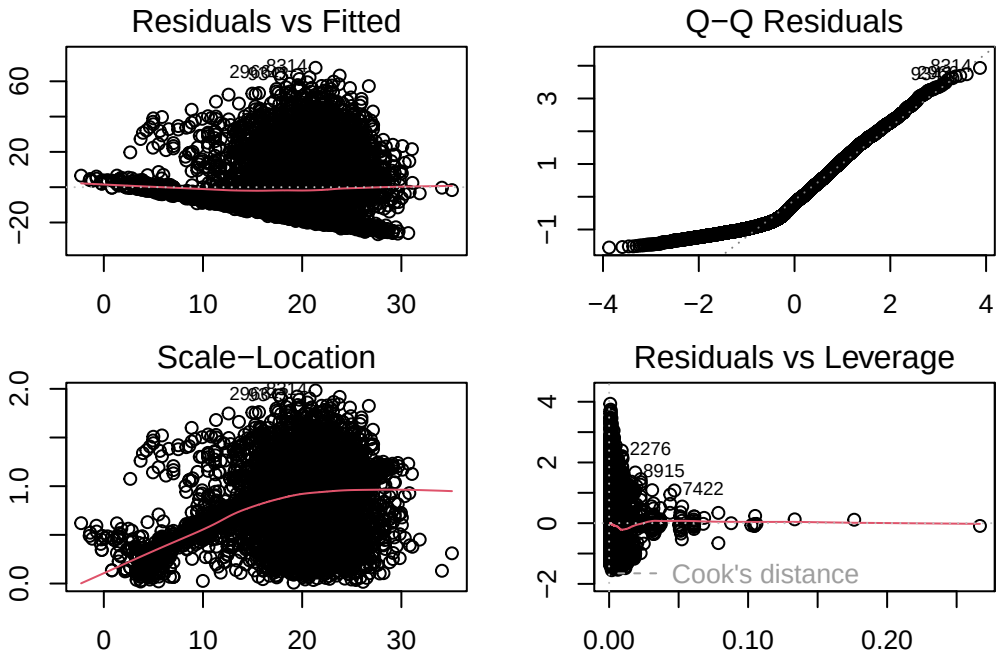
```
model_q1 <- lm(AvgSigStrLanded ~ ReachCms + WeightClass*WinStreak + Height,
               data = ufc_q1)

cooks_d <- cooks.distance(model_q1)
influential <- which(cooks_d > (4 / nrow(ufc_q1)))
ufc_q1_clean <- ufc_q1[-influential, ]

model_q1_clean <- lm(AvgSigStrLanded ~ log(ReachCms) + WinStreak * WeightClass + log(Height)
summary(model_q1_clean)
```

```
Call:
lm(formula = AvgSigStrLanded ~ log(ReachCms) + WinStreak * WeightClass +
    log(Height), data = ufc_q1_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-26.646 -14.706  -3.488  11.234  67.651

Coefficients:
                                          Estimate Std. Error t value
(Intercept)                             370.84260   36.25877  10.228
```

```
log(ReachCms)                                -52.68014    7.20203  -7.315
WinStreak                                      0.88969    0.44797   1.986
WeightClassCatch Weight                      -11.98423    3.11742  -3.844
WeightClassFeatherweight                       3.15345    0.99510   3.169
WeightClassFlyweight                          -5.12434    1.27743  -4.011
WeightClassHeavyweight                         5.67378    1.35218   4.196
WeightClassLight Heavyweight                   8.21926    1.26942   6.475
WeightClassLightweight                         5.88441    0.93235   6.311
WeightClassMiddleweight                        5.28211    1.09455   4.826
WeightClassWelterweight                        7.67438    1.01547   7.557
WeightClassWomen's Bantamweight               -8.38313    1.52870  -5.484
WeightClassWomen's Featherweight             -15.29454    4.30357  -3.554
WeightClassWomen's Flyweight                 -14.94543    1.39726 -10.696
WeightClassWomen's Strawweight               -13.58684    1.45493  -9.339
log(Height)                                  -15.59306    9.11990  -1.710
WinStreak:WeightClassCatch Weight             -0.23074    1.87824  -0.123
WinStreak:WeightClassFeatherweight             0.07549    0.61197   0.123
WinStreak:WeightClassFlyweight                 1.09982    0.85105   1.292
WinStreak:WeightClassHeavyweight              -0.08639    0.65154  -0.133
WinStreak:WeightClassLight Heavyweight        -0.29595    0.63414  -0.467
WinStreak:WeightClassLightweight               0.08217    0.54268   0.151
WinStreak:WeightClassMiddleweight             -0.24912    0.53973  -0.462
WinStreak:WeightClassWelterweight              0.11901    0.55113   0.216
WinStreak:WeightClassWomen's Bantamweight      1.70091    1.02098   1.666
WinStreak:WeightClassWomen's Featherweight    -1.01963    3.74947  -0.272
WinStreak:WeightClassWomen's Flyweight        -1.23959    0.77629  -1.597
WinStreak:WeightClassWomen's Strawweight       3.87668    1.15362   3.360
                                            Pr(>|t|)
(Intercept)                                  < 2e-16 ***
log(ReachCms)                               2.80e-13 ***
WinStreak                                   0.047055 *
WeightClassCatch Weight                     0.000122 ***
WeightClassFeatherweight                    0.001535 **
WeightClassFlyweight                        6.08e-05 ***
WeightClassHeavyweight                      2.74e-05 ***
WeightClassLight Heavyweight                9.98e-11 ***
WeightClassLightweight                      2.89e-10 ***
WeightClassMiddleweight                     1.42e-06 ***
WeightClassWelterweight                     4.50e-14 ***
WeightClassWomen's Bantamweight             4.27e-08 ***
WeightClassWomen's Featherweight            0.000381 ***
WeightClassWomen's Flyweight                 < 2e-16 ***
WeightClassWomen's Strawweight               < 2e-16 ***
```

```
log(Height)                                    0.087339 .
WinStreak:WeightClassCatch Weight              0.902230
WinStreak:WeightClassFeatherweight             0.901824
WinStreak:WeightClassFlyweight                 0.196284
WinStreak:WeightClassHeavyweight               0.894523
WinStreak:WeightClassLight Heavyweight         0.640726
WinStreak:WeightClassLightweight               0.879649
WinStreak:WeightClassMiddleweight              0.644412
WinStreak:WeightClassWelterweight              0.829044
WinStreak:WeightClassWomen's Bantamweight      0.095757 .
WinStreak:WeightClassWomen's Featherweight     0.785675
WinStreak:WeightClassWomen's Flyweight         0.110345
WinStreak:WeightClassWomen's Strawweight       0.000781 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.21 on 9241 degrees of freedom
Multiple R-squared:  0.07087,   Adjusted R-squared:  0.06816
F-statistic: 26.11 on 27 and 9241 DF,  p-value: < 2.2e-16
```

```r
par(mfrow = c(2, 2), mar = c(2,2,2,2))
plot(model_q1_clean)
```



```r
vif(model_q1_clean)
```

```
there are higher-order terms (interactions) in this model
consider setting type = 'predictor'; see ?vif
```

```
                             GVIF Df GVIF^(1/(2*Df))
log(ReachCms)            5.706517  1        2.388832
WinStreak               11.784558  1        3.432864
WeightClass           1301.454686 12        1.348242
log(Height)              6.579795  1        2.565111
WinStreak:WeightClass 3430.026689 12        1.403796
```

```r
r_squared <- summary(model_q1_clean)$r.squared
cat("R-squared:", r_squared, "\n")
```

```
R-squared: 0.07087118
```

A multiple linear regression (MLR) model was applied to examine the relationship between the average significant strikes landed and predictors including log-transformed Reach, log-transformed Height, Win Streak, and the interaction between Win Streak and Weight Class. Log transformations were performed on Reach and Height to address non-linearity and non-constant variance, as indicated by diagnostic plots in Appendix 1.

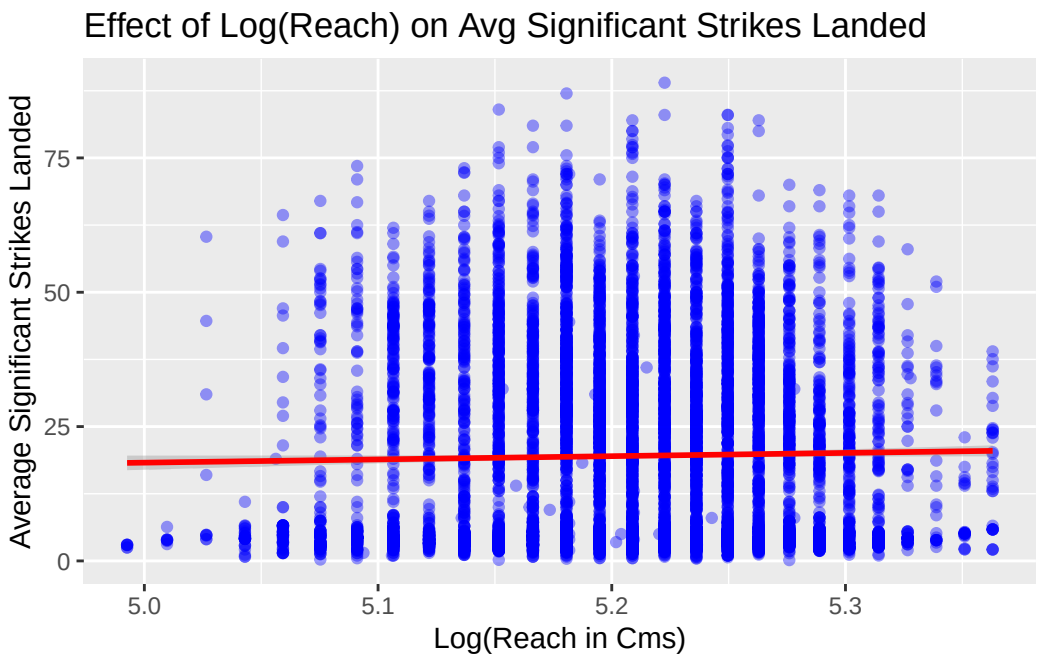
The adjusted R^2 value for the model was **0.068**, suggesting that the predictors explain approximately **6.8%** of the variability in average significant strikes landed.

The analysis revealed several important findings regarding the relationship between fighters' physical attributes and the number of significant strikes landed. The log-transformed Reach variable showed a negative relationship with average strikes landed ($\beta = -52.68, p < 0.001$), indicating that as reach increases, the average number of strikes landed decreases. The Weight Class variable also had significant effects for several divisions. Specifically, the fighters in the Featherweight class had a significantly higher number of strikes landed ($\beta = 3.15, p = 0.001$), while Flyweight fighters landed fewer strikes ($\beta = 5.67, p < 001$). On the other hand, Women's Flyweight fighters had a significant negative relationship with strikes landed ($\beta = -14.95, p < 0.001$). In terms of interaction effects, the model included interaction terms between Win Streak and Weight Class, but most of these were not significant. However, there was a marginally significant positive interaction observed for Women's Strawweight ($\beta = 3.88, p = 0.0008$), suggesting that an increasing win streak slightly positively impacts the number of strikes landed in this weight class. Finally, the logarithm of the height variable showed a weak, marginally significant negative relationship with strikes landed ($\beta = -15.59, p = 0.087$), suggesting that taller fighters might land fewer strikes on average,

although the effect is not strong. These findings underscore the complex interplay between a fighter's physical characteristics, weight class, and performance outcomes, with reach and weight class being the most influential factors in predicting the number of strikes landed.

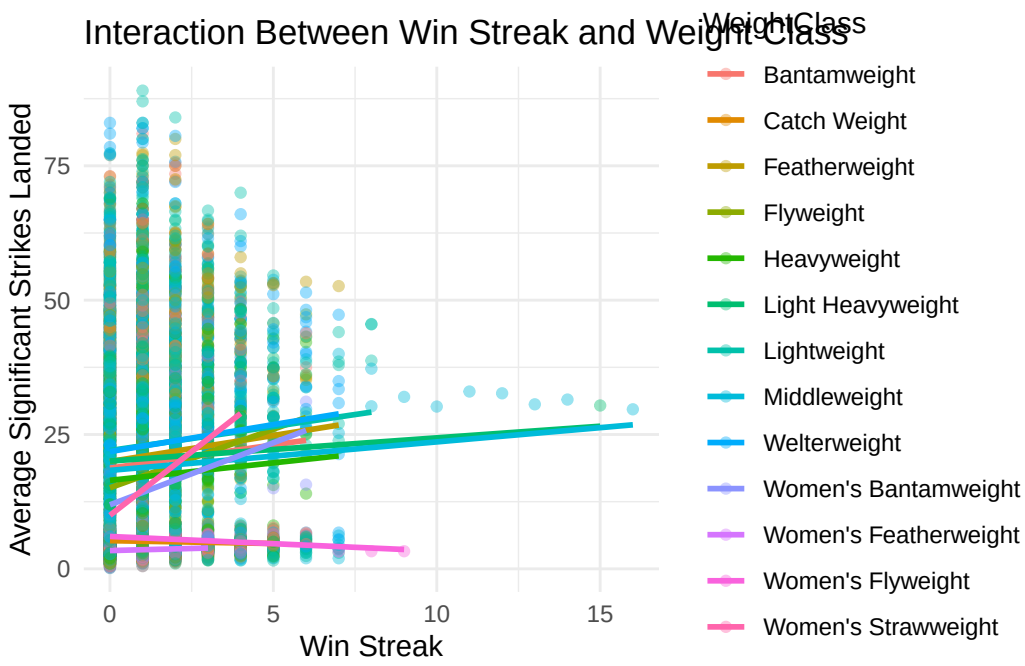
```
# Plot the relationship between log(Reach) and AvgSigStrLanded
ggplot(ufc_q1_clean, aes(x = log(ReachCms), y = AvgSigStrLanded)) +
  geom_point(alpha = 0.4, color = "blue") +
  geom_smooth(method = "lm", se = TRUE, color = "red") +
  labs(title = "Effect of Log(Reach) on Avg Significant Strikes Landed",
       x = "Log(Reach in Cms)", y = "Average Significant Strikes Landed")
```

```
`geom_smooth()` using formula = 'y ~ x'
```

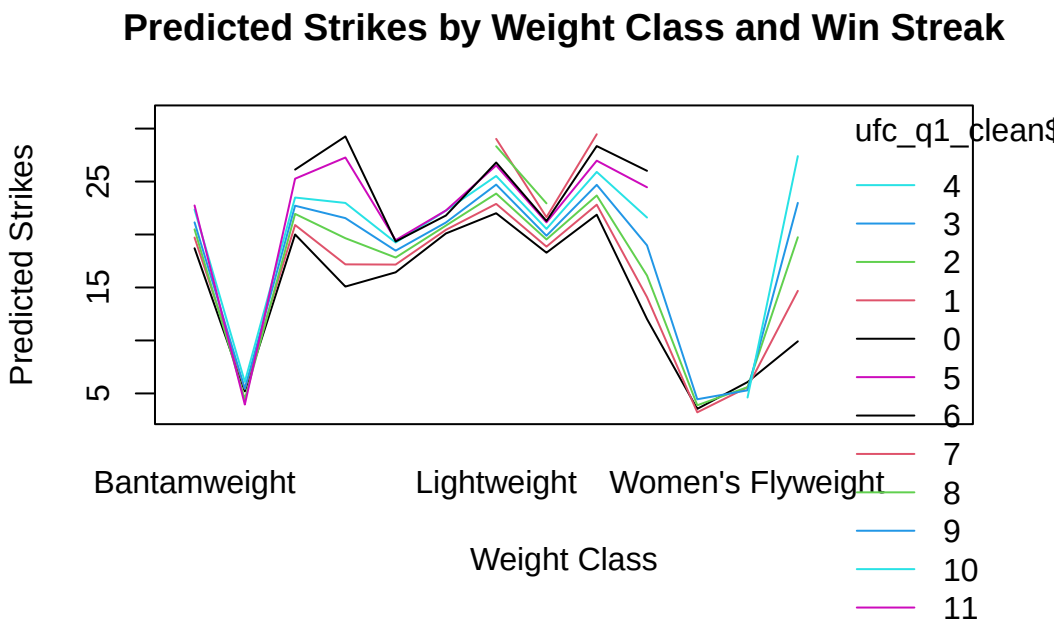



```
# Interaction plot
ggplot(ufc_q1_clean, aes(x = WinStreak, y = AvgSigStrLanded, color = WeightClass)) +
  geom_point(alpha = 0.4) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Interaction Between Win Streak and Weight Class",
       x = "Win Streak", y = "Average Significant Strikes Landed") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```




```
interaction.plot(ufc_q1_clean$WeightClass, ufc_q1_clean$WinStreak,
                 fitted(model_q1_clean), col = 1:6, lty = 1,
                 main = "Predicted Strikes by Weight Class and Win Streak",
                 xlab = "Weight Class", ylab = "Predicted Strikes")
```



Research Question 2: Is the fight outcome associated with the number of submission attempts made by a fighter?

```
ufc_q2 <- ufc_clean %>%
  mutate(
    Outcome = ifelse(Winner == "Red", 1, 0),  # Binary outcome: 1 for Red win, 0 for Blue wi
    WeightClass = as.factor(WeightClass),
    TotalRedSubAttempts = RedAvgSubAtt,        # Red's submission attempts
    TotalBlueSubAttempts = BlueAvgSubAtt
  ) %>%
  mutate(
    LogRedSubAttempts = log1p(TotalRedSubAttempts),
    LogBlueSubAttempts = log1p(TotalBlueSubAttempts),
    LogBlueReach = log1p(BlueReachCms),
    LogRedReach = log1p(RedReachCms),
    LogBlueSigStr = log1p(BlueAvgSigStrLanded),
    LogRedSigStr = log1p(RedAvgSigStrLanded),
    LogFightTime = log1p(TotalFightTimeSecs)
  )

# Check dimensions of the cleaned dataset
dim(ufc_q2)
```

```
[1] 4895  100
```

```r
sim_logistic_model <- glm(
  Outcome ~
    LogRedSubAttempts +
    LogBlueSubAttempts +
    LogBlueReach +
    LogRedReach +
    LogBlueSigStr +
    LogRedSigStr +
    LogFightTime +
    WeightClass,
  data = ufc_q2,
  family = binomial
)


step_model <- step(sim_logistic_model, direction = "both")
```

```
Start:  AIC=6606.31
Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
    LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime +
    WeightClass

                     Df Deviance    AIC
- WeightClass        12   6576.9 6592.9
- LogFightTime        1   6567.5 6605.5
<none>                    6566.3 6606.3
- LogBlueReach        1   6569.7 6607.7
- LogRedReach         1   6573.9 6611.9
- LogBlueSubAttempts  1   6578.2 6616.2
- LogRedSubAttempts   1   6587.2 6625.2
- LogRedSigStr        1   6624.5 6662.5
- LogBlueSigStr       1   6625.8 6663.8

Step:  AIC=6592.94
Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
    LogRedReach + LogBlueSigStr + LogRedSigStr + LogFightTime

                     Df Deviance    AIC
- LogFightTime        1   6577.8 6591.8
<none>                    6576.9 6592.9
```

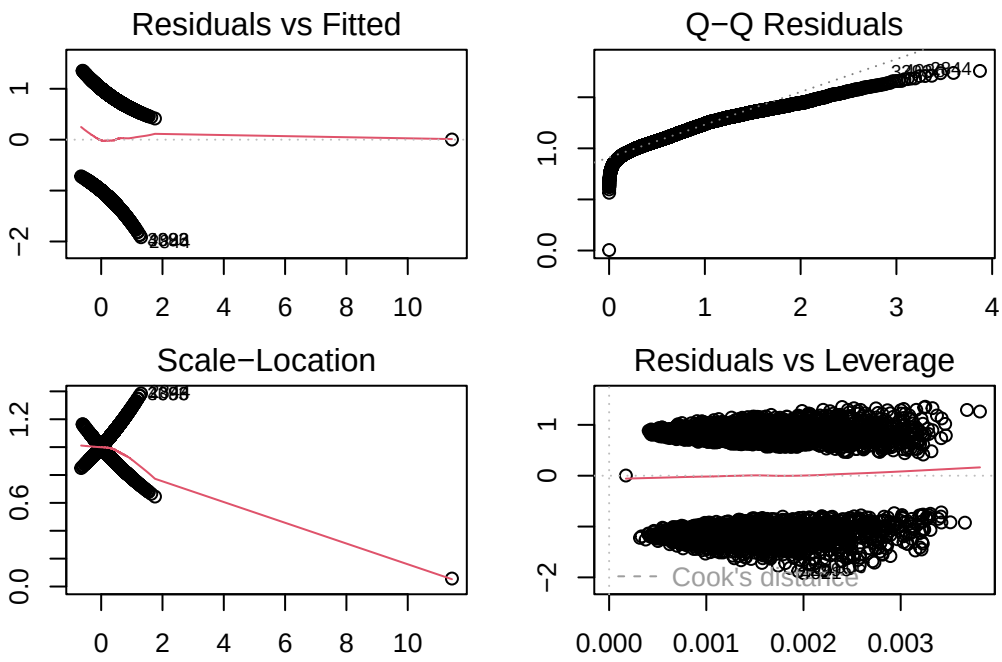
```
- LogRedReach          1   6584.6 6598.6
- LogBlueReach         1   6584.8 6598.8
- LogBlueSubAttempts   1   6589.3 6603.3
+ WeightClass         12   6566.3 6606.3
- LogRedSubAttempts    1   6597.5 6611.5
- LogRedSigStr         1   6634.1 6648.1
- LogBlueSigStr        1   6637.7 6651.7

Step:  AIC=6591.77
Outcome ~ LogRedSubAttempts + LogBlueSubAttempts + LogBlueReach +
    LogRedReach + LogBlueSigStr + LogRedSigStr

                     Df Deviance    AIC
<none>                    6577.8 6591.8
+ LogFightTime        1   6576.9 6592.9
- LogRedReach         1   6585.2 6597.2
- LogBlueReach        1   6586.0 6598.0
- LogBlueSubAttempts  1   6590.4 6602.4
+ WeightClass        12   6567.5 6605.5
- LogRedSubAttempts   1   6597.9 6609.9
- LogRedSigStr        1   6635.0 6647.0
- LogBlueSigStr       1   6638.4 6650.4
```

```r
# Calculate Cook's distance and leverage
cooks_distance <- cooks.distance(step_model)
hat_values <- hatvalues(step_model)
residuals <- residuals(step_model, type = "deviance")

# Thresholds
n <- nrow(ufc_q2)
p <- length(coef(step_model)) - 1
cooks_threshold <- 4 / n
leverage_threshold <- 2 * (p + 1) / n

# Identify influential points
influential_points <- which(cooks_distance > cooks_threshold |
                            hat_values > leverage_threshold |
                            abs(residuals) > 2)

# Remove influential points
ufc_q2_filtered <- ufc_q2[-influential_points, ]
```

```r
# Refit the model
final_model <- glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
    LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr,
    family = binomial, data = ufc_q2_filtered)

# Model summary
summary(final_model)
```

```
Call:
glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
    LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr,
    family = binomial, data = ufc_q2_filtered)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.68061    2.87556   0.237 0.812897
LogRedSubAttempts   0.43736    0.09732   4.494 6.98e-06 ***
LogBlueSubAttempts -0.34346    0.09384  -3.660 0.000252 ***
LogBlueReach       -2.10785    0.76588  -2.752 0.005920 **
LogRedReach         2.03037    0.74314   2.732 0.006292 **
LogBlueSigStr      -0.46782    0.05704  -8.201 2.37e-16 ***
LogRedSigStr        0.46156    0.05861   7.876 3.39e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6324.0  on 4643  degrees of freedom
Residual deviance: 6221.7  on 4637  degrees of freedom
AIC: 6235.7

Number of Fisher Scoring iterations: 4
```

```r
par(mfrow=c(2,2), mar = c(2,2,2,2))

plot(final_model)
```



Appendix

1. Research Question 1

```
summary(model_q1)
```

```
Call:
lm(formula = AvgSigStrLanded ~ ReachCms + WeightClass * WinStreak +
    Height, data = ufc_q1)

Residuals:
    Min      1Q  Median      3Q     Max
-40.690 -16.289  -5.282  12.193 130.324

Coefficients:
                                          Estimate Std. Error t value
(Intercept)                             70.464514   7.601471   9.270
ReachCms                                -0.291116   0.044133  -6.596
WeightClassCatch Weight                 -9.095782   3.143148  -2.894
WeightClassFeatherweight                 1.876641   1.086578   1.727
```

```
WeightClassFlyweight                          -4.529124   1.338491  -3.384
WeightClassHeavyweight                         3.178827   1.502231   2.116
WeightClassLight Heavyweight                   6.209295   1.391018   4.464
WeightClassLightweight                         4.596201   1.021647   4.499
WeightClassMiddleweight                        2.927605   1.219982   2.400
WeightClassWelterweight                        6.033458   1.113355   5.419
WeightClassWomen's Bantamweight               -0.802552   1.602920  -0.501
WeightClassWomen's Featherweight             -12.811029   3.967100  -3.229
WeightClassWomen's Flyweight                  -9.880458   1.540134  -6.415
WeightClassWomen's Strawweight                -4.992836   1.487524  -3.356
WinStreak                                      0.870168   0.412687   2.109
Height                                         0.003184   0.057040   0.056
WeightClassCatch Weight:WinStreak             -1.423668   1.632883  -0.872
WeightClassFeatherweight:WinStreak             0.554692   0.544312   1.019
WeightClassFlyweight:WinStreak                 1.981276   0.668963   2.962
WeightClassHeavyweight:WinStreak               0.396386   0.606659   0.653
WeightClassLight Heavyweight:WinStreak        -0.306468   0.565002  -0.542
WeightClassLightweight:WinStreak              -0.348661   0.499440  -0.698
WeightClassMiddleweight:WinStreak             -0.051047   0.525470  -0.097
WeightClassWelterweight:WinStreak             -0.279404   0.510292  -0.548
WeightClassWomen's Bantamweight:WinStreak     -0.246415   0.878204  -0.281
WeightClassWomen's Featherweight:WinStreak    -1.308624   1.961281  -0.667
WeightClassWomen's Flyweight:WinStreak        -1.583140   0.826751  -1.915
WeightClassWomen's Strawweight:WinStreak       3.041336   0.843915   3.604
                                             Pr(>|t|)
(Intercept)                                   < 2e-16 ***
ReachCms                                     4.43e-11 ***
WeightClassCatch Weight                      0.003814 **
WeightClassFeatherweight                     0.084179 .
WeightClassFlyweight                         0.000718 ***
WeightClassHeavyweight                       0.034364 *
WeightClassLight Heavyweight                 8.14e-06 ***
WeightClassLightweight                       6.91e-06 ***
WeightClassMiddleweight                      0.016427 *
WeightClassWelterweight                      6.13e-08 ***
WeightClassWomen's Bantamweight              0.616607
WeightClassWomen's Featherweight             0.001245 **
WeightClassWomen's Flyweight                 1.47e-10 ***
WeightClassWomen's Strawweight               0.000792 ***
WinStreak                                    0.035010 *
Height                                       0.955490
WeightClassCatch Weight:WinStreak            0.383299
WeightClassFeatherweight:WinStreak           0.308195
```

```
WeightClassFlyweight:WinStreak              0.003067 **
WeightClassHeavyweight:WinStreak            0.513519
WeightClassLight Heavyweight:WinStreak      0.587541
WeightClassLightweight:WinStreak            0.485130
WeightClassMiddleweight:WinStreak           0.922613
WeightClassWelterweight:WinStreak           0.584022
WeightClassWomen's Bantamweight:WinStreak   0.779031
WeightClassWomen's Featherweight:WinStreak 0.504642
WeightClassWomen's Flyweight:WinStreak      0.055535 .
WeightClassWomen's Strawweight:WinStreak    0.000315 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.83 on 9707 degrees of freedom
Multiple R-squared:  0.03763,   Adjusted R-squared:  0.03495
F-statistic: 14.06 on 27 and 9707 DF,  p-value: < 2.2e-16
```

```r
# 1. Check Variance Inflation Factor (VIF) for collinearity
vif_values <- vif(model_q1)
```

```
there are higher-order terms (interactions) in this model
consider setting type = 'predictor'; see ?vif
```

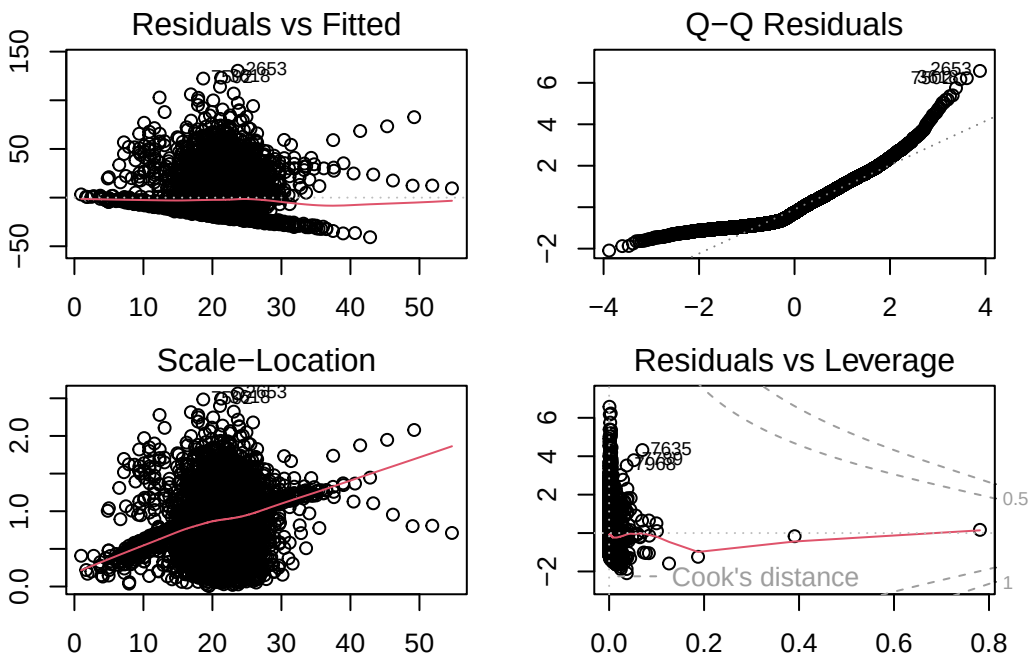
```r
print("Variance Inflation Factor (VIF):")
```

```
[1] "Variance Inflation Factor (VIF):"
```

```r
print(vif_values)
```

```
                             GVIF Df GVIF^(1/(2*Df))
ReachCms                 5.823552  1        2.413204
WeightClass            476.978854 12        1.293017
WinStreak               11.703017  1        3.420967
Height                   6.653360  1        2.579411
WeightClass:WinStreak 1247.323837 12        1.345858
```

```r
# 2. Residuals vs Fitted Plot for Linearity
par(mfrow = c(2, 2), mar = c(2, 2, 2, 2))  # Set plotting layout
plot(model_q1)
```




```
r_squared <- summary(model_q1)$r.squared
cat("R-squared:", r_squared, "\n")
```

```
R-squared: 0.03763049
```

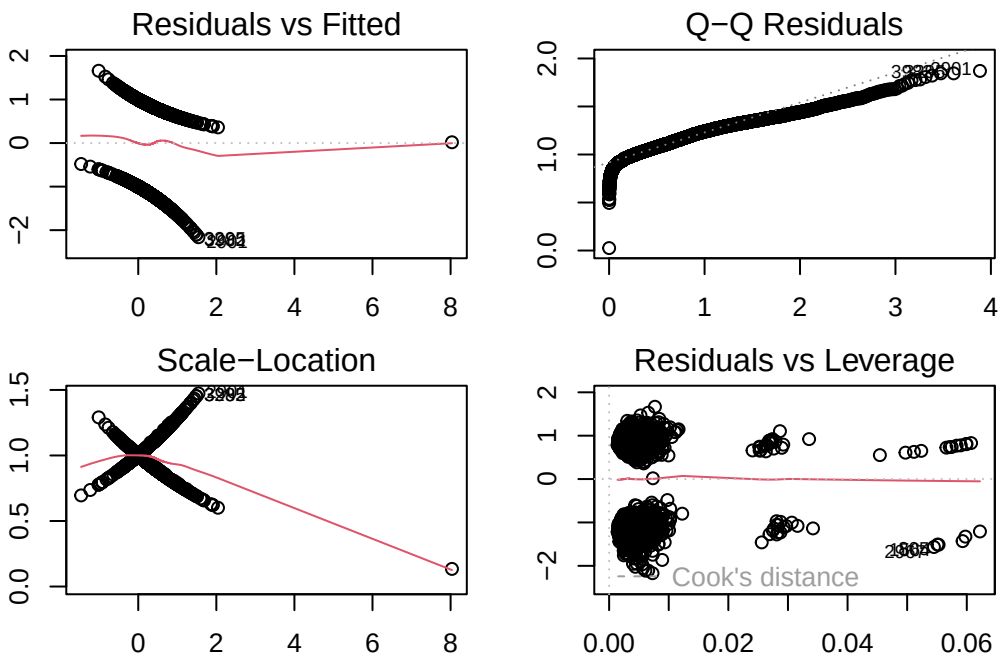
2. Research Question 2

```
# Model summary of the initial simple logistic model
summary(sim_logistic_model)
```

```
Call:
glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
    LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr +
    LogFightTime + WeightClass, family = binomial, data = ufc_q2)

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                -5.351057   6.459545  -0.828 0.407447
LogRedSubAttempts           0.405680   0.089319   4.542 5.57e-06 ***
LogBlueSubAttempts         -0.292197   0.084704  -3.450 0.000561 ***
LogBlueReach               -1.439347   0.907894  -1.585 0.112883
```

```
LogRedReach                          2.491817   0.903558   2.758 0.005819 **
LogBlueSigStr                       -0.365003   0.048374  -7.545 4.51e-14 ***
LogRedSigStr                         0.373853   0.050081   7.465 8.33e-14 ***
LogFightTime                         0.035227   0.032697   1.077 0.281313
WeightClassCatch Weight             -0.057436   0.343726  -0.167 0.867293
WeightClassFeatherweight            -0.027917   0.131144  -0.213 0.831425
WeightClassFlyweight                -0.003566   0.159466  -0.022 0.982161
WeightClassHeavyweight              -0.045397   0.208875  -0.217 0.827942
WeightClassLight Heavyweight        -0.176252   0.192860  -0.914 0.360776
WeightClassLightweight              -0.136914   0.126242  -1.085 0.278128
WeightClassMiddleweight             -0.276099   0.164307  -1.680 0.092883 .
WeightClassWelterweight             -0.281531   0.145431  -1.936 0.052888 .
WeightClassWomen's Bantamweight     -0.177841   0.191071  -0.931 0.351979
WeightClassWomen's Featherweight     0.230400   0.512452   0.450 0.652996
WeightClassWomen's Flyweight        -0.122585   0.184749  -0.664 0.506998
WeightClassWomen's Strawweight       0.011955   0.187268   0.064 0.949100
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6674.5  on 4894  degrees of freedom
Residual deviance: 6566.3  on 4875  degrees of freedom
AIC: 6606.3

Number of Fisher Scoring iterations: 4
```

```r
par(mfrow=c(2,2), mar = c(2,2,2,2))

plot(sim_logistic_model)
```




```
# Model summary of the step model
summary(step_model)
```

```
Call:
glm(formula = Outcome ~ LogRedSubAttempts + LogBlueSubAttempts +
    LogBlueReach + LogRedReach + LogBlueSigStr + LogRedSigStr,
    family = binomial, data = ufc_q2)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.36477    2.73096   0.134 0.893744
LogRedSubAttempts   0.39057    0.08768   4.454 8.41e-06 ***
LogBlueSubAttempts -0.29815    0.08385  -3.556 0.000377 ***
LogBlueReach       -1.90853    0.71404  -2.673 0.007521 **
LogRedReach         1.88597    0.69892   2.698 0.006967 **
LogBlueSigStr      -0.36729    0.04821  -7.618 2.57e-14 ***
LogRedSigStr        0.36873    0.04979   7.405 1.31e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

```
    Null deviance: 6674.5  on 4894  degrees of freedom
Residual deviance: 6577.8  on 4888  degrees of freedom
AIC: 6591.8

Number of Fisher Scoring iterations: 4
```

```r
par(mfrow=c(2,2), mar = c(2,2,2,2))

plot(step_model)
```