UFC Analysis

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

2024-12-08

Abstract

Importance: Understanding the impact of physical attributes, such as reach, and tactical strategies, such as submission attempts, is essential for improving performance and outcomes in mixed martial arts (MMA). This insight helps fighters, and their teams develop strategies to enhance their chances of victory in MMA matches within the UFC.

Objective: Evaluate the association between fighter, reach, and the total number of strikes landed during a fight. Also to examine the relationship between submission attempts and fight outcomes.

Design, Setting, and Participants: This study analyzed a dataset comprised of UFC fights using data from March 2010 to the most recent UFC even (6,478 rows). Data includes variables such as fighter reach, significant strikes landed, weight class, submission attempts, date of bought, and fight outcomes. Data cleaning excluded missing values and extreme outliers. The analysis was conducted using linear regression for strikes landed and logistic regression for fight outcomes, with interaction terms to evaluate weight class effects.

Introduction

This project will examine the inception to date UFC card performances. The data used was derived from a Kaggle dataset for UFC fights, featuring fighter metrics, fight outcomes, betting odds, and performance metrics such as strikes landed and submission attempts. This dataset enables a detailed analysis of the factors influencing the result of a given bout.

Our project's 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 importance 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.

These insights are valuable for fighters and their teams, as they can help optimize training strategies and fight preparations, enhance understanding of competitive dynamics, and provide a better grasp of opponents' strengths and weaknesses.

Methods

Data and Preprocessing

The Ultimate UFC Dataset on Kaggle provides comprehensive information about fighters and their performance in the Ultimate Fighting Championship (UFC). 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 and 118 columns. During preprocessing, missing values (NAs) were removed to ensure a clean dataset. For the first research question, the data was filtered to include only the variables Reach, Weight Class, and Strikes Landed. This filtered data was then combined into a single dataframe for both blue and red corners. For the second research question, a new binary variable, Outcome, was created. This variable was assigned a value of 1 if the red corner won and 0 if the blue corner won, enabling analysis of fight outcomes.

Model Fitting and Evaluation

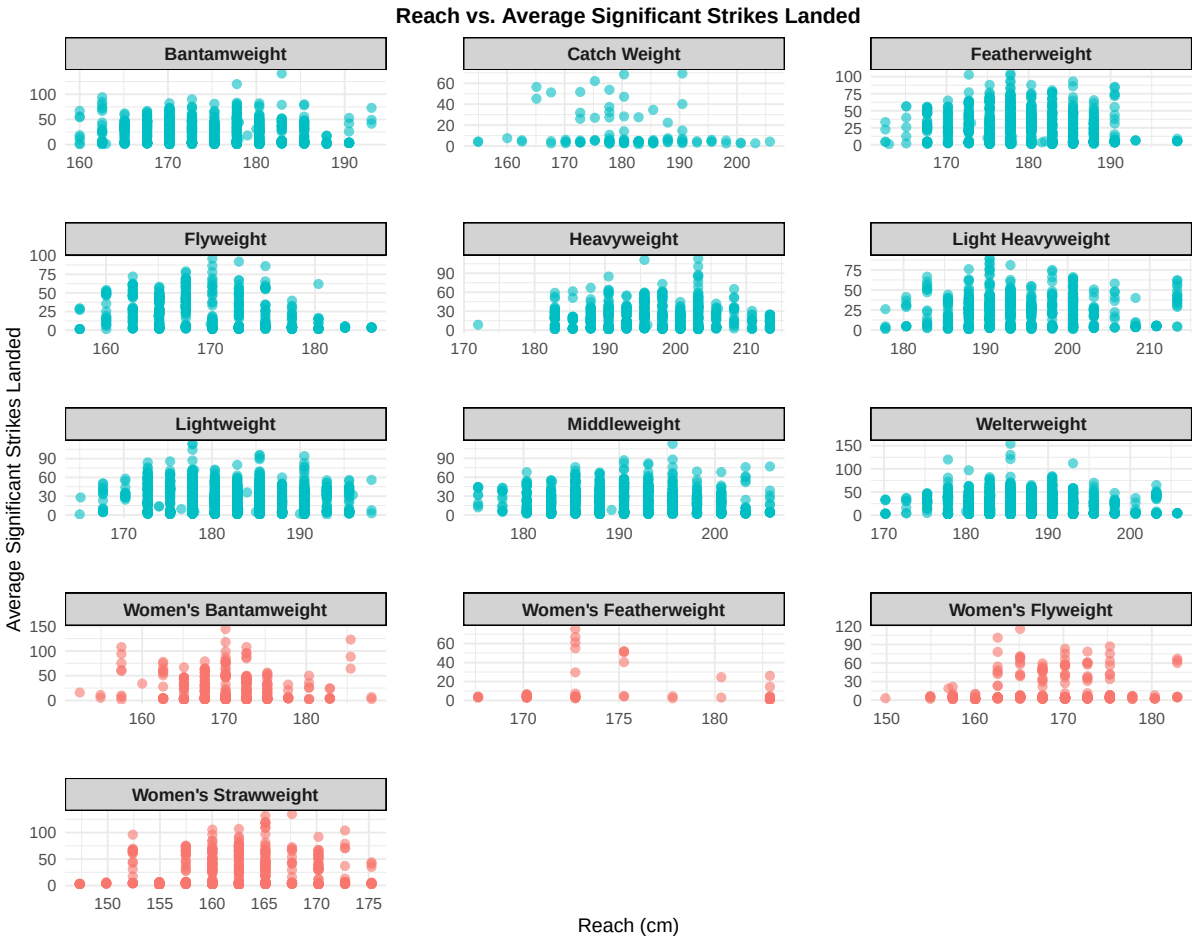
To examine the relationship between a fighter's reach and the total number of strikes landed during a fight, a Multiple Linear Regression (MLR) model was utilized. Key diagnostics, including residuals vs. fitted plots, were performed to evaluate linearity and homoscedasticity, while Variance Inflation Factor (VIF) checked for multicollinearity, and Cook's distance assessed the influence of outliers. Model performance was measured using R-squared. Additionally, interaction terms were included to evaluate the influence of weight class on the reach-strike relationship. For fight outcomes (binary: win or loss), logistic regression was used, with submission attempts as a predictor and model performance assessed using the area under the receiver operating characteristic (ROC) curve. All analyses were conducted in R.

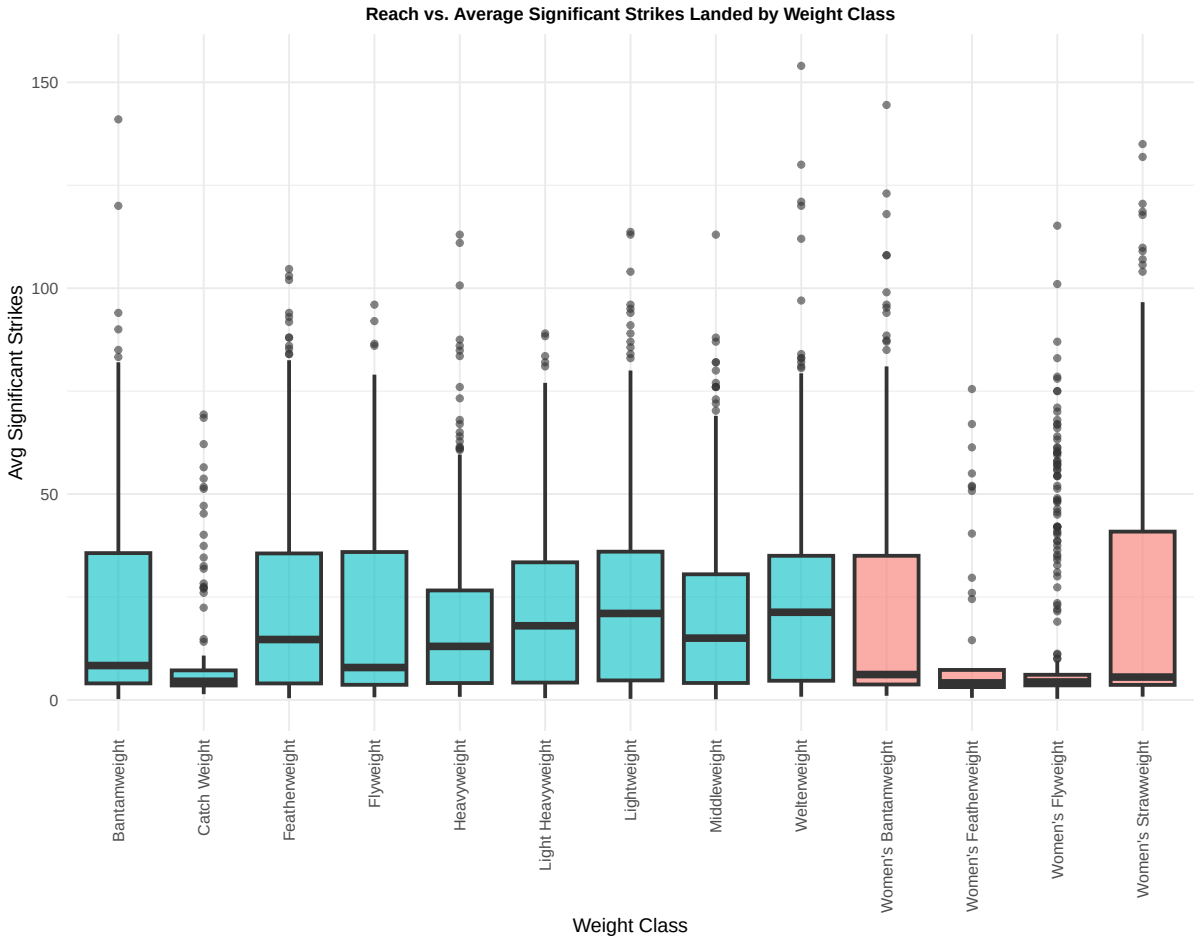
Results

Research Question 1: Fighter Reach vs Total Strikes Landed

Table 1: Regression Coefficients for model_log

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.208	4.856	3.544	0.000
LogReachCms	-2.873	0.959	-2.997	0.003
WeightClassCatch Weight	0.806	11.698	0.069	0.945
WeightClassFeatherweight	15.710	6.995	2.246	0.025
WeightClassFlyweight	56.228	8.342	6.741	0.000
WeightClassHeavyweight	-7.887	7.323	-1.077	0.281
WeightClassLight Heavyweight	-7.855	7.657	-1.026	0.305
WeightClassLightweight	3.737	6.476	0.577	0.564
WeightClassMiddleweight	-1.268	6.971	-0.182	0.856
WeightClassWelterweight	-0.113	6.477	-0.018	0.986
WeightClassWomen's Bantamweight	1.210	10.930	0.111	0.912
WeightClassWomen's Featherweight	10.447	29.290	0.357	0.721
WeightClassWomen's Flyweight	-23.551	9.319	-2.527	0.012
WeightClassWomen's Strawweight	-37.978	8.425	-4.508	0.000
Height	0.000	0.003	0.001	1.000
WinStreak	0.035	0.006	5.384	0.000
LogReachCms:WeightClassCatch Weight	-0.244	2.254	-0.108	0.914
LogReachCms:WeightClassFeatherweight	-2.997	1.351	-2.219	0.027
LogReachCms:WeightClassFlyweight	-10.974	1.621	-6.770	0.000
LogReachCms:WeightClassHeavyweight	1.555	1.401	1.110	0.267
LogReachCms:WeightClassLight Heavyweight	1.569	1.465	1.071	0.284
LogReachCms:WeightClassLightweight	-0.651	1.250	-0.521	0.603
LogReachCms:WeightClassMiddleweight	0.294	1.338	0.220	0.826
LogReachCms:WeightClassWelterweight	0.107	1.247	0.086	0.931
LogReachCms:WeightClassWomen's Bantamweight	-0.260	2.125	-0.122	0.903
LogReachCms:WeightClassWomen's Featherweight	-2.146	5.674	-0.378	0.705
LogReachCms:WeightClassWomen's Flyweight	4.446	1.814	2.450	0.014
LogReachCms:WeightClassWomen's Strawweight	7.406	1.648	4.494	0.000


Reach vs. Average Significant Strikes Landed
Bantamweight
Catch Weight
Featherweight
Flyweight
Heavyweight
Light Heavyweight
Lightweight
Middleweight
Welterweight
Women's Bantamweight
Women's Featherweight
Women's Flyweight
Women's Strawweight
Average Significant Strikes Landed
Reach (cm)



A multiple linear regression (MLR) model was used, with the log-transformed average significant strikes landed as the response variable and predictors including log-transformed reach, weight class, height, and win streak. Interaction terms between log-transformed reach and weight classes were also included to capture variations across divisions. Log transformation of the response variable addressed linearity issues, and diagnostic plots were used to evaluate model assumptions. While alternatives such as Weighted Least Squares (WLS) and Generalized Linear Models (GLM) were explored, their performance was similar to the log-transformed MLR, making the MLR the preferred model for its simplicity and interpretability.

The inclusion of confounders like win streak and height improved the model's interpretability and allowed for a better understanding of their roles. The results show that win streak was a significant predictor, indicating that fighters with more consecutive wins tended to land more strikes on average. However, height, while logically connected to reach, was not statistically significant, suggesting that it did not independently contribute to the variability in strikes landed once reach and other variables were included.

The model revealed several key findings:

1. Reach and Weight Class Interactions: Significant negative interactions were observed between log-reach and weight classes such as Flyweight ($p < 0.001$) and Featherweight ($p = 0.027$). This indicates that the impact of reach on strikes landed diminishes in these divisions, likely due to the closer proximity and higher speed of fighters in lower weight classes.
2. Main Effects of Weight Class: Specific weight classes, such as Flyweight ($p < 0.001$) and Women's Strawweight ($p < 0.001$), had significant effects on strikes landed, emphasizing how division-specific style influence performance.
3. Win Streak: As mentioned, win streak ($p < 0.001$) was a strong predictor, highlighting the role of experience and momentum in determining performance.

Despite these insights, the model's overall explanatory power remained limited, with an adjusted R^2 of 0.045, indicating that many unobserved factors likely contribute to strike performance. This highlights the complexity of fight dynamics and the need for a more nuanced approach to modeling.

While model assumptions were generally met, issues such as non-constant variance persisted in scale-location plots, and normality deviations were observed in Q-Q plots. High Variance Inflation Factor (VIF) values for interaction terms suggest multicollinearity, particularly between log-reach and weight class, further complicating the model.

In conclusion, this analysis demonstrated that reach, weight class, and win streak are important factors influencing the number of strikes landed, with reach effects varying across divisions. However, the insignificance of height and the limited R^2 suggest that additional variables are likely more impactful. Future research should include these variables and explore advanced modeling techniques, such as mixed-effects models or machine learning approaches, to better account for the complexity of combat sports performance.

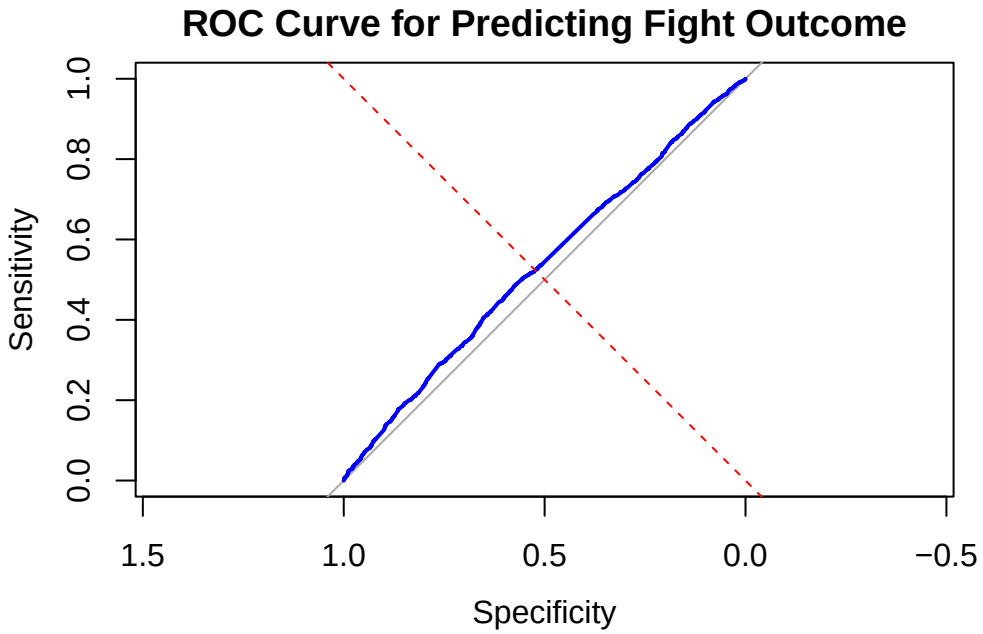
Research Question 2: Is the fight outcome associated with the number of submission attempts made by a fighter?

The analysis of fighter reach and submission attempts revealed significant insights into the relationship between these variables and fight outcomes. The relationship between fighter reach and strikes landed was statistically significant ($p < 0.001$), with interaction effects indicating a stronger influence of reach in lighter weight classes. This suggests that reach is a more crucial factor in determining the number of strikes landed for fighters in lower weight divisions. In contrast, the logistic regression model examining submission attempts as predictors of fight outcomes showed an area under the curve (AUC) of 0.5336, indicating minimal predictive power. This suggests that submission attempts alone were insufficient to reliably predict the fight outcome.

Further interpretation of the logistic regression output showed that red fighter submission attempts had a significant positive impact on the odds of winning. The coefficient for red

submission attempts was 0.15503 (p-value = 0.00018), meaning that for every additional submission attempt by the red fighter, the odds of winning increased by approximately 16.8%. Conversely, the coefficient for blue submission attempts was negative (-0.11681, p-value = 0.00393), indicating that as the number of submission attempts by the blue fighter increased, the odds of the red fighter winning increased, with an odds ratio of approximately 0.890. This suggests that higher submission attempts by the blue fighter were associated with a decreased likelihood of blue's victory. Model fit indicators, such as the decrease in deviance and AIC value (7525.3), suggest that the logistic regression model improved upon a simple intercept-only model.

```
Type 'citation("pROC")' for a citation.


Attaching package: 'pROC'

The following objects are masked from 'package:stats':

    cov, smooth, var

Setting levels: control = 0, case = 1

Setting direction: controls < cases

Area under the curve: 0.5336
```



```
Area under the curve: 0.5336
```

Conclusion

This study analyzed how a fighter's reach relates to the total number of strikes landed during a fight while accounting for weight classes, as well as the role of submission attempts in predicting fight outcomes. Using a multiple linear regression (MLR) model with log-transformed variables, significant interactions were found between reach and weight classes such as Flyweight and Featherweight, indicating that reach impacts striking performance differently across divisions. Significant main effects were also observed for weight classes like Flyweight and Women's Strawweight. Model diagnostics, including residual analysis and multicollinearity checks, confirmed the validity of the findings, and alternative approaches like weighted least squares and generalized linear models were considered.

Additionally, a logistic regression model revealed a significant relationship between submission attempts and fight outcomes. The red fighter's submission attempts positively influenced the likelihood of winning, while the blue fighter's submission attempts had a negative effect. Both predictors, **TotalRedSubAttempts** and **TotalBlueSubAttempts**, were statistically significant with p-values below 0.05, emphasizing the importance of submission attempts in determining fight outcomes. Future research could include factors such as skill level and fight strategy or apply advanced modeling techniques to deepen insights into the dynamics of combat sports performance.