**Final Project**

**(UFC Data Analysis)**

A person in boxing gloves

Description automatically generated Ref: https://medium.com/@raphael.schols/the-ultimate-fighting-championship-exploratory-data-analysis-python-1e7c962fe2e8

**Sriram Kannepalli**

**Northeastern University: College of Professional Studies**

**ALY 6010.**[**71820**](https://northeastern.instructure.com/courses/196161)**: Probability Theory and Introductory Statistics**

**Professor Antonius Breur**

**October 31st, 2024**

**1. Introduction**

The Ultimate Fighting Championship (UFC) dataset offers a unique view into how fighter attributes and performance metrics influence outcomes and betting odds. This report builds on prior exploratory analysis and hypothesis testing to explore deeper relationships between key variables, offering insights that go beyond standard assumptions. Our aim is to identify attributes and trends that contribute to winning strategies, beneficial for fighters, coaches, and analysts alike.

**2. Summary of Initial EDA**

The initial exploratory data analysis (Milestone 1 & 2) provided foundational insights:

* **Betting Odds Difference**: A significant difference in betting odds was found between fighters in the red and blue corners, with red corner fighters generally receiving more favorable (lower) odds.  
  A graph with a red and blue rectangle

  Description automatically generated
* **Corner Influence**: The chi-square test indicated that the red corner had a statistically significant advantage in match outcomes.  
  A graph showing a red and blue bar

  Description automatically generated
* **Height and Reach Correlation**: A positive correlation between height and reach was observed, suggesting that taller fighters generally have a reach advantage.  
  A comparison of different height bars

  Description automatically generated with medium confidence

These observations laid the groundwork for the advanced analysis conducted in the final milestone, aimed at understanding specific attribute relationships that affect winning probability and betting odds.

**3. Analytical Questions and Hypotheses**

Using insights from the initial EDA, we formulated specific questions to examine deeper relationships using regression analysis:

Note: These questions were formulated to understand not only the statistical relationships but also their strategic implications, making the insights relevant for UFC trainers and fighters.

1. **Question 1**: Does reach advantage and weight class affect win probability?
   * **Dependent Variable**: Win Probability (Winner)
   * **Independent Variables**: Reach Advantage (ReachAdvantage), Weight Class (WeightClass)
   * **Hypotheses**:
     + **H0**: Reach advantage and weight class do not affect win probability.
     + **H1**: Reach advantage and weight class positively affect win probability.
2. **Question 2**: How does fighter age relate to win probability?
   * **Dependent Variable**: Win Probability (Winner)
   * **Independent Variables**: Age (RedAge), Experience (RedTotalRoundsFought)
   * **Hypotheses**:
     + **H0**: Fighter age does not affect win probability.
     + **H1**: Younger age positively affects win probability.
3. **Question 3**: Does win streak impact betting odds?
   * **Dependent Variable**: Betting Odds (RedOdds)
   * **Independent Variable**: Win Streak (RedCurrentWinStreak)
   * **Hypotheses**:
     + **H0**: Win streak does not affect betting odds.
     + **H1**: A longer win streak improves betting odds.

**4. Methodology and Hypothesis Testing**

**Question 1: Reach Advantage and Weight Class on Win Probability**

A logistic regression model was used to test the impact of reach advantage and weight class on the probability of winning, with win probability as the dependent variable. This model allows us to estimate the change in odds of winning for each unit increase in reach advantage, controlling for weight class.

***(Refer to Appendix: Figure A1: Logistic Regression Output for Reach Advantage and Weight Class)***

**Plot: *(Refer to Appendix: Figure A2: Plot of Reach Advantage vs. Win Probability)***

**Model Summary:**

* **ReachAdvantage has a significant positive effect on win probability (p < 0.001), with each centimeter of reach advantage increasing the log odds of winning.**
* **WeightClass does not have a significant impact on win probability (p = 0.471).**

**Interpretation:**

* **Insight: Reach advantage significantly boosts win probability, likely due to better control of engagement range. The non-significance of weight class suggests that leverage is more impactful than weight in determining fight outcomes.**
* **Strategic Implication: Emphasize reach-based tactics in training, particularly for fighters lacking a natural weight advantage.**

**Question 2: Fighter Age on Win Probability**

Logistic regression was applied, with age and total rounds fought as predictors and win probability as the outcome. This model allowed us to observe the direction and significance of age on winning, considering experience.

***(Refer to Appendix: Figure A3: Logistic Regression Output for Age and Experience on Win Probability)***

***Plot: (Refer to Appendix: Figure A4: Plot of Age vs. Win Probability)***

**Model Summary:**

* **RedAge has a significant negative impact on win probability (p < 0.001), showing that younger fighters have higher chances of winning. RedTotalRoundsFought (experience) does not significantly affect win probability (p = 0.114).**

**Interpretation:**

* **Insight: Younger fighters exhibit better win probabilities, which may be attributed to their higher physical agility and endurance. Experience does not appear to counterbalance the physical declines associated with aging.**
* **Strategic Implication: Training for older fighters should focus on maximizing energy efficiency and injury prevention, while younger fighters should capitalize on their physical prime.**

**Question 3: Win Streak Effect on Betting Odds**

A linear regression was used, with betting odds as the dependent variable and win streak as the predictor. This test helped determine if there’s a statistically significant relationship between win streaks and betting odds.

***(Refer to Appendix: Figure A5: Linear Regression Output for Win Streak on Betting Odds)***

***Plot: (Refer to Appendix: Figure A6: Plot of Win Streak vs. Betting Odds)***

**Model Summary:**

* **The coefficient for RedCurrentWinStreak (-33.333, p < 0.001) reveals that longer win streaks are associated with more favorable (lower) betting odds.**
* **A low R-squared value (0.04618) indicates that win streak is only one of many factors influencing betting odds.**

**Interpretation:**

* **Insight: Win streaks are positively received in betting markets, reflecting greater public confidence. However, the low R-squared suggests that betting odds are affected by multiple factors, not solely by win streak.**
* **Strategic Implication: Fighters on long streaks may face heightened psychological pressure, emphasizing the need for mental resilience training.**

**5. Results and Interpretation**

**This analysis highlights the impact of three key factors—reach advantage, age, and win streak—on UFC fight outcomes and betting odds:**

* **Reach Advantage: Fighters with a greater reach relative to their opponents have a significantly higher probability of winning. This advantage likely allows fighters to control the distance, strike effectively, and defend more easily. Takeaway: Reach is a strategic asset that fighters and coaches should prioritize, as it can enhance both offensive and defensive tactics.**
* **Age: Younger fighters tend to have a higher win probability compared to older fighters, suggesting that physical fitness, agility, and endurance are critical advantages in UFC bouts. Experience alone does not offset the physical edge provided by youth. Takeaway: For older fighters, training should focus on maximizing efficiency and minimizing energy loss, while younger fighters should capitalize on their physical prime through intense conditioning.**
* **Win Streak and Betting Odds: A longer win streak leads to more favorable (lower) betting odds, reflecting increased confidence from both fans and bookmakers. However, this effect plateaus, indicating that win streaks are only one of many factors shaping public perception and betting odds. Takeaway: Consistent winning boosts a fighter’s market value and public confidence, but sustained performance pressure means that mental resilience is also crucial.**

**6. Conclusion and Insights**

**The analysis provides valuable insights into the factors that drive success in UFC fights:**

* **Physical Attributes Matter: Reach advantage and youth are pivotal in determining a fighter’s chances of winning. These findings suggest that while experience is valuable, physical advantages such as agility, stamina, and reach offer a tangible edge in the octagon. Trainers might focus on reach-based tactics and tailored conditioning programs to maximize these physical assets.**
* **Winning Momentum Influences Public Perception: A fighter’s win streak affects their betting odds, underscoring the importance of performance consistency in shaping public and bookmaker confidence. However, as win streak length alone doesn’t fully explain betting odds, factors like fighting style, opponent strength, and recent performance likely play a role. This insight suggests that fighters on winning streaks should balance media attention with sustained training focus to handle heightened expectations.**

**Overall, these findings guide coaches, fighters, and analysts on how to tailor training regimens, manage fighter expectations, and understand the public perception surrounding a fighter's performance. By optimizing strategies based on reach, youth, and winning consistency, UFC teams can maximize a fighter’s potential for success in both the octagon and the public arena.**

**7. Appendix:**

**Figure A1: Logistic Regression Output for Reach Advantage and Weight Class**A screenshot of a computer

Description automatically generated**Figure A2: Plot of Reach Advantage vs. Win Probability**A graph of a curve

Description automatically generated

**Figure A3: Logistic Regression Output for Age and Experience on Win Probability**A screenshot of a computer

Description automatically generated

**Figure A4: Plot of Age vs. Win Probability**

A graph with a line

Description automatically generated

**Figure A5: Linear Regression Output for Win Streak on Betting Odds**

A screenshot of a computer

Description automatically generated

**Figure A6: Plot of Win Streak vs. Betting Odds**

A graph of a graph with a line

Description automatically generated

**8. References**

1. Ultimate UFC Dataset. (n.d.). *GitHub repository*. Available at <https://github.com/UltimateUFC/Ultimate-UFC-Dataset> (Accessed: October 25, 2024).
2. Kaba(2022). *R in action: Data analysis and graphics with R and tidyverse* (3rd ed.). Manning Publications. Available at <https://www.manning.com/books/r-in-action-third-edition> (Accessed: October 26, 2024).
3. Apelgren, S., &(2024). *Predicting UFC matches using regression models*. Degree Project in Technology, First cycle, 15 credits. KTH Royal Institute of Technology. Available at <https://kth.diva-portal.org/smash/get/diva2:1878726/FULLTEXT01.pdf> (Accessed: October 27, 2024).
4. Tidyverse. (n.d.). *Tidy des*. Available at <https://tidyr.tidyverse.org/articles/tidy-data.html> (Accessed: October 26, 2024).