

Mark Grillo

Career Foundry Project 6

Sourcing Open Data

11.11.2023

UFC Fights 1993 – 2021 Data Set

Source: The source of this data is from [Kaggle](#). Which is formed of a community of data scientists and developers. It is the world's largest data science community with over one million registered users. Members can talk about development models, explore data sets, or network across 194 separate countries around the world.

Collection: The data was compiled and scraped from the ufcstats website so should be fairly dependable.

Contents: This is a list of every UFC fight in the history of the organization. Every row contains information about both fighters, fight details and the winner. The original source from the Kaggle author does contain the raw data files, preprocessed data, and a final merged data set. There were originally 144 columns in the final merged data set. For the purpose of my project many columns were not relevant and reduced this to 32 Columns.

Limitations & Ethics: The data was scraped by an independent party. Although the source of the scraping is from the UFC site itself full review will need to be done of columns and relevant data cleaning to confirm it is fully cleaned and correctly formatted before further analysis. I do not believe much bias is possible as these are purely statistical calculations from actual fights. Although there is always concern over human made error translating and interpreting the results and methods used including their level of experience as an analyst.

Relevance:

I believe this data set meets the necessary requirements for this project as it is open source, includes a geospatial component, and meets the size and variable requirements. The project is a little older than 3 years old but was found in the career foundry brief.

Data Profile

Variable	Description	Time Variant/ Invariant	Structured/ Unstructured	Quantitative/Qualitative	Nominal/Ordinal /Discrete/Continuous
r_fighter	Name of the fighter assigned to the red corner. Typically considered higher ranked fighter.	invariant	structured	Qualitative	ordinal
b_fighter	Name of the fighter assigned to the blue corner. Typically ranked lower opposing fighter.	invariant	structured	Qualitative	ordinal
date_of_fight	Date the fight took place	invariant	structured	Quantitative	discrete
location	Geographic location the fight was held	invariant	structured	Qualitative	nominal
winner	Winner of the fight	invariant	structured	Qualitative	ordinal
weight_class	Division assigned to match even competitors by weight	variant	structured	Quantitative	continuous
b_wins	Blue corner wins	variant	structured	Quantitative	discrete
r_wins	Red corner wins	variant	structured	Quantitative	discrete
b_losses	Blue corner losses	variant	structured	Quantitative	discrete
r_losses	Red corner losses	variant	structured	Quantitative	discrete
b_win_by_decision_majority	Win for the blue corner. When 2 of 3 judges pick the same competitor as the winner and the third judge scores a draw	variant	structured	Quantitative	discrete
r_win_by_decision_majority	Win for the red corner. When 2 of 3 judges pick the same competitor as the winner and the third judge scores a draw	variant	structured	Quantitative	discrete

b_win_by_decision_split	Win for the blue corner when 2 of the 3 judges score the win for one competitor and the third judge scores the win for the other red corner	variant	structured	Quantitative	discrete
r_win_by_decision_split	Win for the red corner when 2 of the 3 judges score the win for one competitor and the third judge scores the win for the other blue corner	variant	structured	Quantitative	discrete
b_win_by_decision_unanimous	Win for the blue corner when all 3 judges score blue the winner	variant	structured	Quantitative	discrete
r_win_by_decision_unanimous	Win for the red corner when all 3 judges score red the winner	variant	structured	Quantitative	discrete
b_win_by_ko/tko	Win for blue fighter when judge decides the other fighter can't safely continue	variant	structured	Quantitative	discrete
r_win_by_ko/tko	Win for red fighter when judge decides the other fighter can't safely continue	variant	structured	Quantitative	discrete
b_win_by_submission	When the blue fighter gets the red fighter to verbally give up and yield the fight or tap the hand on the mat in a sign of submission	variant	structured	Quantitative	discrete
r_win_by_submission	When the red fighter gets the blue fighter to verbally give up and yield the fight or tap the hand on the mat in a sign of submission	variant	structured	Quantitative	discrete
b_win_by_tko_doctor_stoppage	When the ring side doctor determines the red corner fighter cannot safely continue the fight and notifies the referee to stop by tko	variant	structured	Quantitative	discrete

r_win_by_tko_doctor_st oppage	When the ring side doctor determines the blue corner fighter cannot safely continue the fight and notifies the referee to stop by tko	variant	structured	Quantitative	discrete
b_stance	Stance of the blue fighter. Orthodox, switch or southpaw. This is what dominant hand you lead with a jab. Orthodox is left hand, southpaw is right hand, switch is equally using both hands.	variant	structured	Qualitative	nominal
r_stance	Stance of the red fighter. Orthodox, switch or southpaw. This is what dominant hand you lead with a jab. Orthodox is left hand, southpaw is right hand, switch is equally using both hands.	variant	structured	Qualitative	nominal
b_height_cms	Blue fighter height mentioned in centimeters	invariant	structured	Quantitative	continuous
r_height_cms	Red fighter height mentioned in centimeters	invariant	structured	Quantitative	continuous
b_reach_cms	Blue fighter total arm length in centimeters	invariant	structured	Quantitative	continuous
r_reach_cms	Blue fighter total arm length in centimeters	invariant	structured	Quantitative	continuous
b_weight_lbs	Blue fighter total weight in lbs	variant	structured	Quantitative	continuous
r_weight_lbs	Red fighter total weight in lbs	variant	structured	Quantitative	continuous
b_age	Blue fighter age	variant	structured	Quantitative	continuous
r_age	Red fighter age	variant	structured	Quantitative	continuous

Questions to Explore

- Is the red or blue corner fighter more dominant with wins?
- What is the most popular location for fights?
- What is the most popular stance for winning percentage (orthodox, southpaw, or switch)?
- What is the most popular weight class for fights?
- What is the highest rated way of winning a fight without referee voting between submission, tko or doctor stoppage?
- What is the highest ranked way of winning overall by submission type or referee votes?
- What were the most popular and least popular years for fighting?
- What is the highest and lowest range age group for fighting?
- Who are the top 3 fighters with the most fights from 1993-2021?