

DeQuWaun Barriento

CF Project 6

Sourcing open data

1/14/2024

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.

Collection: The data was compiled and scraped from the UFCStats website which should be dependable.

Contents: The data list every UFC fight in the history of the business. Each rows 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 set but for the purpose of this project, I have narrowed it down to 32.

Limitations and Ethics: The data set was put together by an independent party although the source of the data is from the UFC site itself. A review of the columns and relevant data that they are fully cleaned is needed before any further analysis is done. I don't believe any bias is possible as these are statistics from professionalized fights although human made error is always a possibility when interpreting the results due to the level of experience the analyst.

Relevance: The data set meets the necessary requirements as it was provided by CF in the project brief, includes a geospatial component and meets the size and variable requirements.

Data Profile:

Provided in Excel sheet.

Questions to Explore

- Who has more wins, red or blue corner?
- What location is the most popular for fights?
- What is the most popular stance for winning percentage (orthodox, southpaw, Or switch)?
- What is the most popular weight class?
- What is the highest and lower range age group for fighting?
- Who are the top 3 fighters with the most fights from 1993-2021?