



# Data Science Final Project

ANALYSIS OF UFC FIGHTS FROM 1993 TO 2019

Arno Reinhofer | 21.05.2020

# Introduction

Adrian Spataru and Bohdan Andrusyak held an online course “Learn Data Science at Home” for beginners in the field of data science. The last part of this 5-week course is to create a project based on public data that uses many of the learned techniques.

I decided to use UFC fight statistics from Kaggle:

<https://www.kaggle.com/rajeevw/ufcdata>

## Objectives

The following questions seem interesting:

- Is data.csv calculated correctly?
  - Is the aggregate damage average correct?
  - Is the aggregated opponent average correct?
- Who won the most fights?
- What is the best age to win fights?
- Who won the most title fights?
- Which accuracy can be reached for the prediction of the winner?

## Methods

### GET CLEAN DATA

UFC-01-clean\_data.ipynb

1. Examine the data:  
Get an overview about the content. Try to understand all data in the dataset.
2. Handle missing values:  
Delete missing data if it is not needed for analysis. Try to best guess values for the rest by correlating to other metrics, using the median or the most frequent occurrences.
3. Delete unnecessary data:  
Drop columns not needed and rows with certain values.
4. Write an output file for further processing

### EXPLORATORY DATA ANALYSIS (EDA)

UFC-02-eda.ipynb

1. Correlations:  
Find correlations in the data. Use a correlation matrix and a table view to get insights.

2. Prediction:  
Define one hot encoding categories and use Random Forrest to predict winners.
3. Get winner statistics:  
Calculate bar charts and histograms to answer the questions.

## Results

### IS DATA.CSV CALCULATED CORRECTLY?

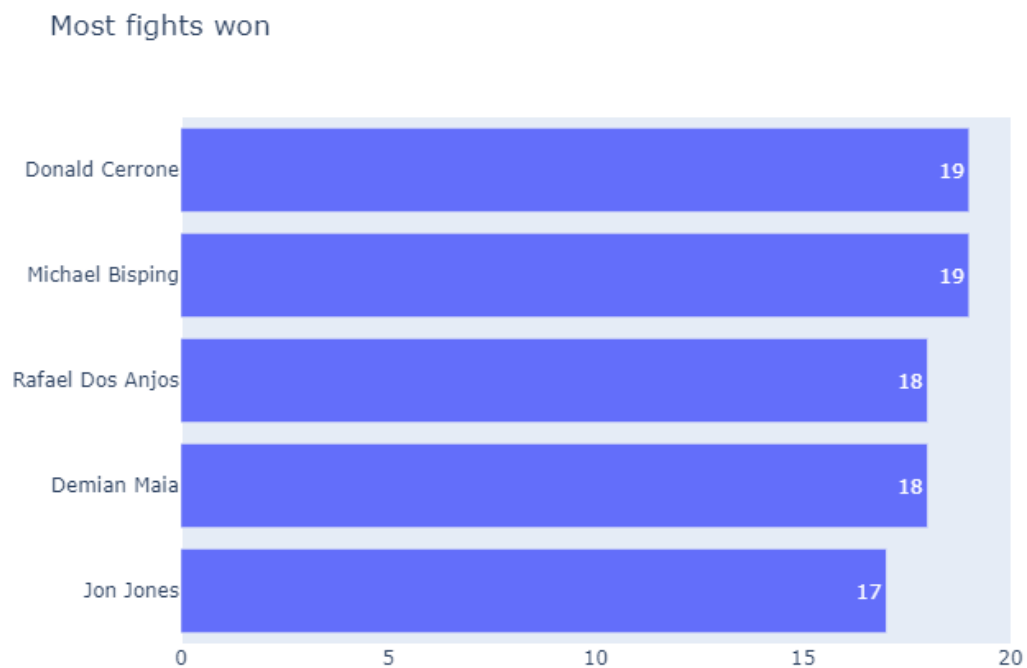
This question was answered by recalculating random samples by hand. E.g. for Henry Cejudo's first 3 fights. **The work of Rajeev Warriar is OK.**

Thus, all **further work is based on data.csv** which is partly processed. It already has aggregations of previous fights.

No proof in form of documents is contained in this project though.

## WHO WON THE MOST FIGHTS?

**Donald Cerrone** and **Michael Bisping** won 19 fights.

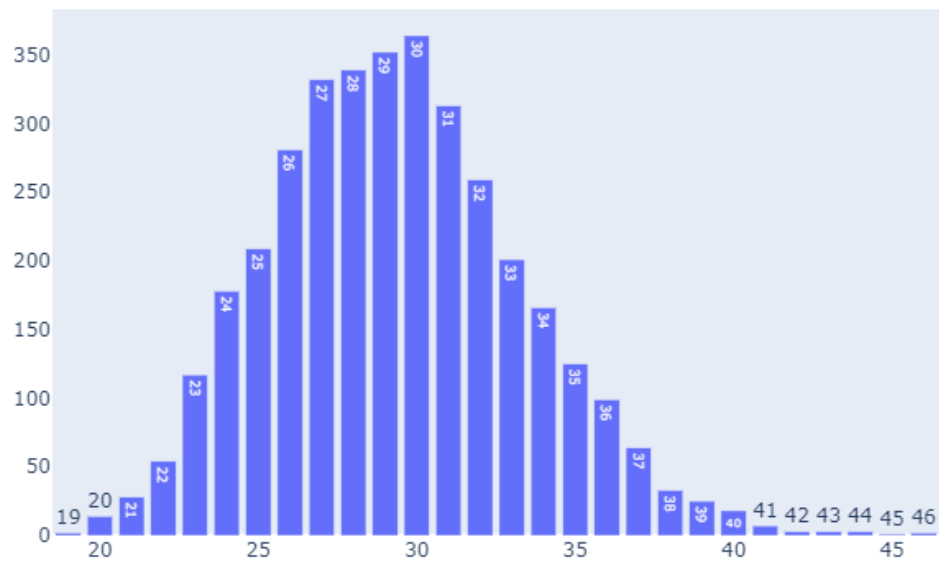


## WHAT IS THE BEST AGE TO WIN FIGHTS?

The best age to win fights is from **27 years** to **30 years**.

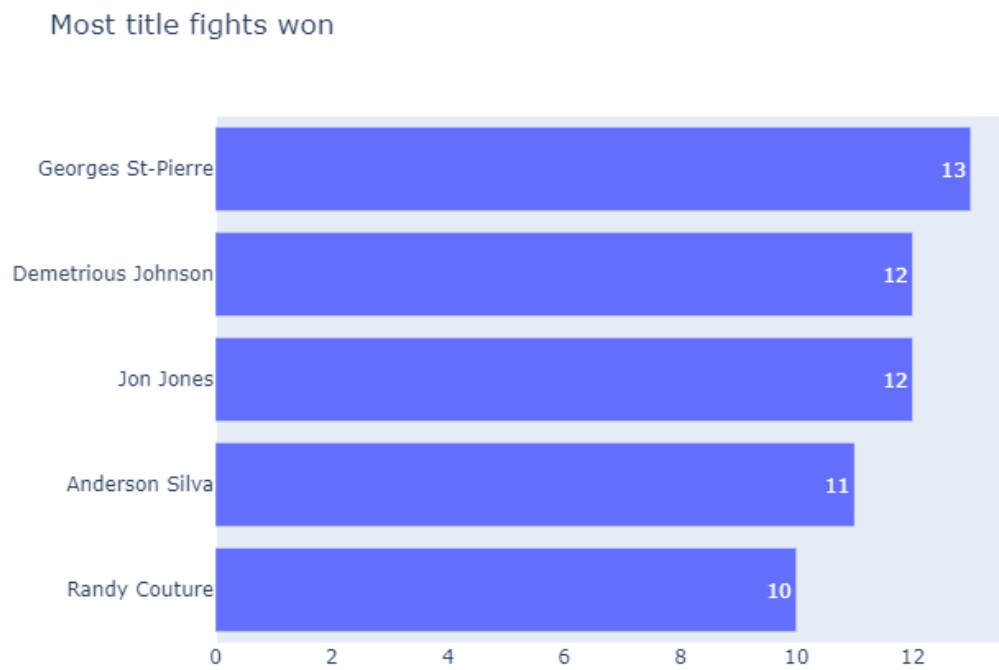
Fun fact: The oldest winners were already 46!

Age of Winners



## WHO WON THE MOST TITLE FIGHTS?

**Georges St-Pierre** won 13 title fights.



## WHICH ACCURACY CAN BE REACHED FOR THE PREDICTION OF THE WINNER?

The prediction accuracy is only around **60% to 70%**, which is not enough to get rich through professional betting...

No clear correlations/feature importances were found that have an impact on the prediction.