

# ‘Ultimate Fighting Charts’

## *An analysis of UFC fighters*

COMP 30780 Data Science in Practice

Lorcan Rooney (16413092), Michael Roche (16408964)

School of Computer Science

University College Dublin

May 11, 2018

**Abstract.** The main purpose of this project is to analyse UFC fighters, weight class to weight class and to compare the champions with the rest of the pack. Hopefully some pattern will emerge between fighters in different weight classes and between the various champions. We plan to accomplish this by scraping relevant data from the web and analyse it using charts and graphs.

**Declaration.** We (Lorcan Rooney, 16413092 and Michael Roche, 16408964) declare that this assignment is our own work and that we have correctly acknowledged the work of others. This assignment is in accordance with University and School guidance[1] on good academic conduct in this regard.

## 1. Introduction

UFC is a mixed martial arts (MMA) organisation. It is to MMA what NBA is to basketball or WWE to professional wrestling. There are other organisations but the UFC is home to the majority of the top fighters in the world. Fighters are split into weight classes, much like boxing. Fighters physical stats (i.e height and age) and more technical stats on how they fight will be analysed in this project. We will compare these stats between the champions (“champs”) and the rest of the weight class, and between different weight classes.

MMA is a fast growing sport and the UFC is a multi-million dollar company. Michael didn’t have any previous knowledge of MMA but was curious after I showed him the amount of data related to the sport on the web. Michael plays hurling at a high level and has seen the growth in data in sports first hand. When the rules of the sport were explained to him and when we found the statistics online, it was clear how much skill and tactics are needed to succeed in UFC.

Data and data analysis have become huge in sports in recent years and MMA is no different. There is an abundance of data available on numerous websites, not least the official UFC one. Wikipedia, Tapout and Sherdog are great resources also. MMA and specifically the UFC has an obsession with stats. American sports/organisations in general seem to be ahead of the rest of field in sports stats, with the NFL and NBA pioneers in the field.

In this report, we will investigate three main research questions. The first one will analyse physical differences between fighters. We found out that height and reach aren't indicative of a successful fighter and that the average age of a fighter in the UFC is thirty.

The second question revolves around which is more important in MMA, striking or grappling. We discovered that striking is key to have any success, whereas grappling can be effective if used correctly.

Lastly, we looked at how the champions of each division differ from the average fighter. These results were not as straightforward as the other two question, as you will see below.

In this report we will discuss our motivations and objectives for this project. Then our three research questions will be expanded on fully. Next we will explain how we obtained the data used in this project and any cleaning that was necessary. We will then delve into the brunt of the project, explaining our results. Finally, we will finish with limitations, reproducibility, ethical considerations and our conclusion.

## **2. Motivations & Objectives**

### **2.1. Background & Motivations**

A project was discovered on Kaggle that did detailed analysis on UFC fights. "<https://www.kaggle.com/rishpande/ufc-data-analysis-visualization-beginner>". They went into specific detail on each fighters statistics in the fights, round by round. We decided that looking at the fighters overall stats would give us a better understanding of the fighter's ability. There was a lot of pointless analysis as well such as "leg kicks thrown in the third round" . It turns out there wasn't many thrown because the fighters were tired.

Also they analysed win percentage of red corner vs blue corner. It's random which corner a fighter is put in so this is pretty irrelevant. Overall we thought we could analyse different aspects of the sport and go into more detail to questions asked in that project.

As stated in the introduction, the sport is very rich in statistics so it looked like a good choice for this project. Originally , the plan was to analyse soccer players and try and predict future stars. The data for this proved hard to come by. The wealth of data available on this sport made it too good of an opportunity to pass up on.

## 2.2. Research Questions

### 2.2.1. RQ1: How do fighters differ physically, weight class to weight class?

Are physical attribute of UFC fighters a telling factor in their success? How much does height, reach and age decide the outcome of a UFC fight? As well as that, how does the performance of fighters differ as their physique differs? In other words, how do fighters differ between weight class?

### 2.2.2. RQ2: Striking vs Grappling - Which is more popular and which is more successful?

The two main aspects of MMA are striking and grappling. Striking refers to throwing punches, kicks etc. The martial arts that fighters mainly utilise while striking are boxing or muay thai kickboxing . Grappling involves fighters using wrestling or brazilian jiu-jitsu. We wanted to know whether striking or grappling was more generally used and which was more widely used by those at the top of their respective division.

### 2.2.3. RQ3: How do champions differ from the rest?

In the UFC there are twelve weight classes, each with their own champion (currently there are ten as Amanda Nunes is a champion in two divisions and bantamweight is vacant). For this question we wish to find out how the best of the best differ from the rest of the pack? This question will be touched on in question two but fully elaborated on here.

## 3. Data Wrangling

### 3.1. Data Acquisition

There were two main datasets that we used for our analysis – We scraped a Wikipedia page that contained a list of current UFC fighters & we scraped the UFC official website. The wikipedia page gave us physical/personal details of the fighters that helped us with our analysis for our first research question. With the wiki page, we also had access to a table where the top 15 fighters in each weight class were listed. The UFC website provided us with more technical statistics on the fighters, such as the strikes attempted and landed by the fighter as well as what part of the body was hit by the strike.

The Wikipedia page ‘ List of current UFC fighters’ -

[https://en.wikipedia.org/wiki/List\\_of\\_current\\_UFC\\_fighters](https://en.wikipedia.org/wiki/List_of_current_UFC_fighters) was scraped first. A list of names of all UFC fighters was needed to scrape the UFC website. This will be explained why shortly.

The Wikipedia page contained tables with every fighter in the UFC, divided by weight class. It had their names, heights and records among other things. This was scraped and put in a dataframe.

Next the UFC website was scraped. This website has a page for every fighter containing technical stats about their fights including strikes attempted and landed, takedowns attempted and landed and average fight type. However there was no contents of all fighters to make scraping easy. Instead we cycled through urls. Every page has the same base url, followed by the fighters name. For example '<https://www.ufc.com/athlete/alistair-overeem>' is UFC heavyweights Alistair Overeem's page.

The list of names acquired from Wikipedia was used to scrape the UFC website. Some things on the website proved hard to scrape such as records and physical attributes such as height and reach. Luckily records and heights were in the first dataset and a workaround was found for reach.

Reach was awkward as some fighters had an 'octagon debut' date while others didn't. This meant that reach was in one of two places in an array. Thus reach was scraped into its own dictionary separate from the rest. Half of the fighter's reaches were in the first column and the other half were in the second column, with their octagon debut in the first. The first row was iterated through and any value with a comma (there was a comma in the date format) was replaced with the value in column two on the same row. The second column was then dropped and we had our reaches.

Another Wikipedia page, similar to the first one was also scraped. This page contained the name of every UFC champion ever, this was cross referenced with the data already obtained. This meant we had every former champion active on the roster. (No retired champions, explained in 'Limitations' section).

Some names on the Wikipedia page were different to those on the website for various reasons. These were scraped manually with the proper urls being put in a dictionary and merged with the rest later.

## **3.2. Data Cleaning & Preparation**

Both main datasets needed to be cleaned. Any rows that didn't have at least ten non empty values from the UFC website dataset were dropped. Some of the data had percentages beside them (standing, clinch and ground strikes had a percent beside them, adding up to 100). These were stripped as while useful, they would get in the way and could be recalculated later.

The Wikipedia dataset had the name of each column in the first row so that was restored. Many of the fighters names had asterisks beside them so they were removed. The champs had a (C) beside their names so these were also removed. Irrelevant columns were dropped.

The hardest part about cleaning this dataset however was the records. A fighters record is their wins and losses. So 9-4 is nine wins and four losses. However while exceedingly rare, draws do happen (9-4-1 is the same record as above with one draw). Then to make matters even more complicated, some fighters have “no contests” for one reason or another. (9-4-1 1NC) . Separating each of these into their own column was challenging. We decided to drop draws and no contests as neither had much relevance to this project. The records were cleaned by dropping the dashes and separating the strings.

The ‘Height’ column contained both imperial and metric measurements. Since the reach column was in inches and ‘Height vs Reach’ was one of the planned graphs, it was decided to put Height in inches too. The metric measurement was stripped. Height was in feet as opposed to inches so the number was multiplied by 39.37.

All datasets were then merged. Firstly the UFC website data with the ‘forgotten’ dataset (the fighters whose names were different on Wikipedia and the UFC website). Next the two main datasets were merged (wikipedia and UFC website). These were merged on name.

We then merged this dataframe with a rankings dataframe that we created manually. This gives us the rank of every fighter. Every division has a ‘top 15’ with fighters ranked from one to fifteen. The champions were given a zero.

Finally the data frame was merged with the dataset of ex-champions , also scraped from Wikipedia. This added a new column called ‘ex\_champ’ with a ‘True’ if the fighter has held the belt, or ‘False’ otherwise.

## **4. Data Analysis & Results**

### **4.1. RQ1**

**How do fighters differ physically, weight class to weight class?**

#### **4.1.1. Datasets**

For this research question, we used data on the physical attributes of the fighters. In MMA, fighter’s physique and conditioning differs from weight class to weight class. While fighters weight obviously stays constant throughout each weight class, other physical attributes such as height, reach and age differ amongst the fighters.

In order to provide an in depth analysis on this research question, it would be essential to inspect how these physical attributes might affect the outcome of UFC fights. Assessing the fighter’s win rate during this analysis would be important to look at here.

The dataset contains strikes landed, strikes attempted and the positions that the strikes were landed in. This data was used to look at what type of punches the fighters throw during fights, and how that differs between weight class.

#### **4.1.2. Approach**

How the age of the fighters differs per weight class, was the first attribute inspected. An assumption could be made, based on most professional sports, that athletes enter the peak of their career midway through their twenties (24-28).

The age column was in a string format, so a new column called 'age\_int' was created and the average was acquired for each weight class. The 'Rank' column has numbers 0 – 15, which represent the top 15 fighters in each weight division (the zero represents the champion). From this, a 'top 5' data frame was created to get the average age of the top 5 fighters in each weight class. A bar chart was used to represent this data.

The next attribute analysed, Height, was graphed similarly to age. The average height of fighters in each weight class was compared to the average height of the top 5 fighters in those weight classes.

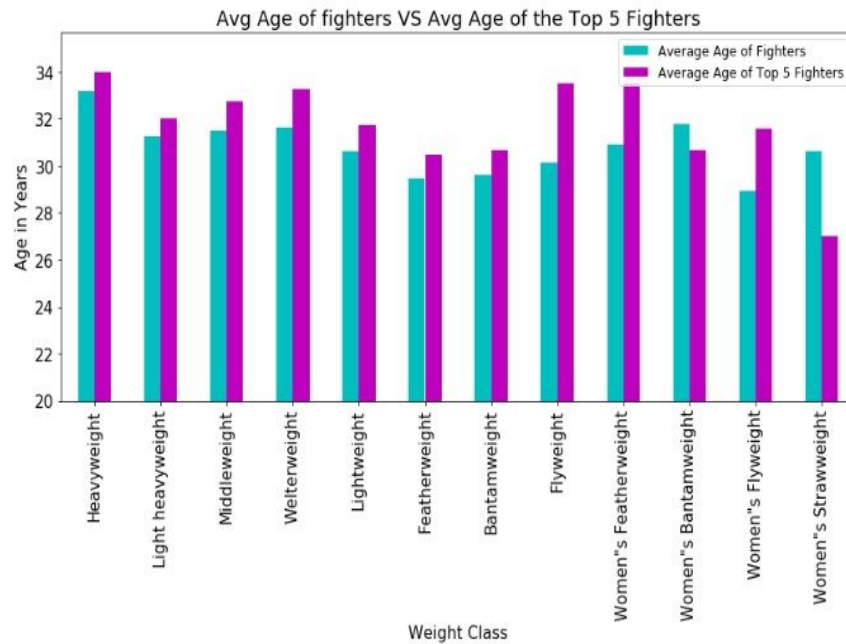
Thirdly, the reach data was used to calculate the ape index of each fighter in the dataset. Ape index is reach divided by height. We wanted to see if there was any correlation between ape index and win percentage. The champs of each weight class were used as part of this analysis as a way of confirming the effect of ape index on win rate as the champs would guarantee us a good win percentage.

Next we wanted to examine what position the fighters in each weight class would commonly throw strikes. These positions are split up into three types - standing strike (standing position), clinch strike (strike while grappling at close quarters) & ground strike (both fighters on the ground). The percentage of each type was calculated by dividing each strike position into all three strike positions and multiplying by 100 to get percentage. This was then grouped by weight class.

Heavier weight classes are generally considered to have worse fitness than lighter weight classes. Many fighters use the clinch as a means to rest, similar to boxing. Analysing the position the fighters throw strikes from will tell us useful information about their fitness potentially.

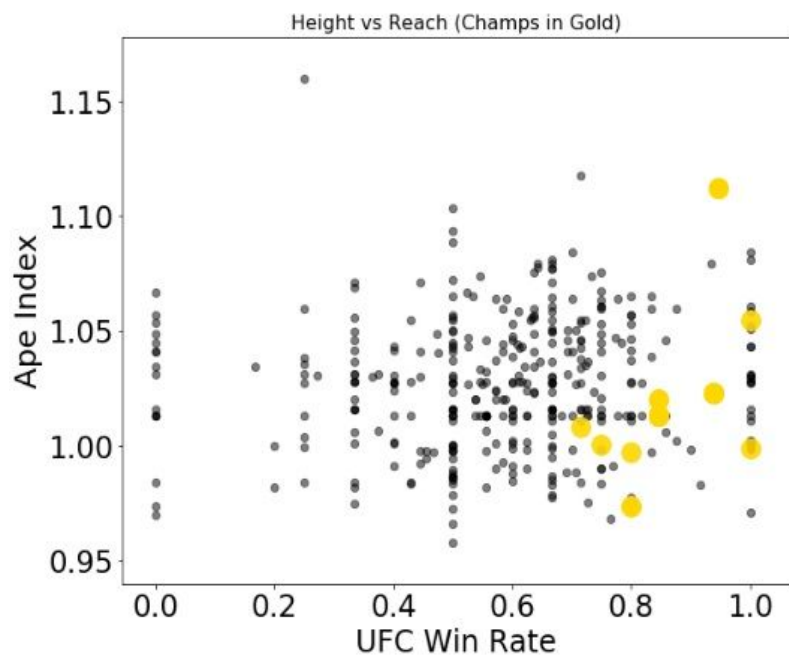
### 4.1.3. Results

Fig 1.1



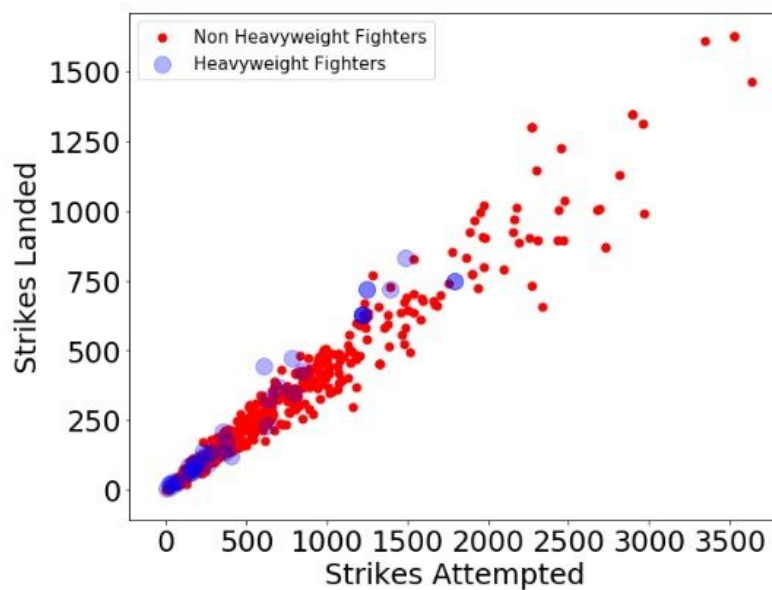
The x-axis has the twelve weight classes, from mens heaviest to lightest, and then female heaviest to lightest. The y axis represents the age in years. The blue bars are the average age of all fighters. The magenta bars are the average age of the 'Top 5' fighters.

Fig. 1.2



The x-axis contains the UFC win rate in decimal. The y-axis represents the ape index ranging from 0.95 to 1.15. The yellow circles represent the champs from each weight class while the black dots represent the rest of the fighters in the UFC.

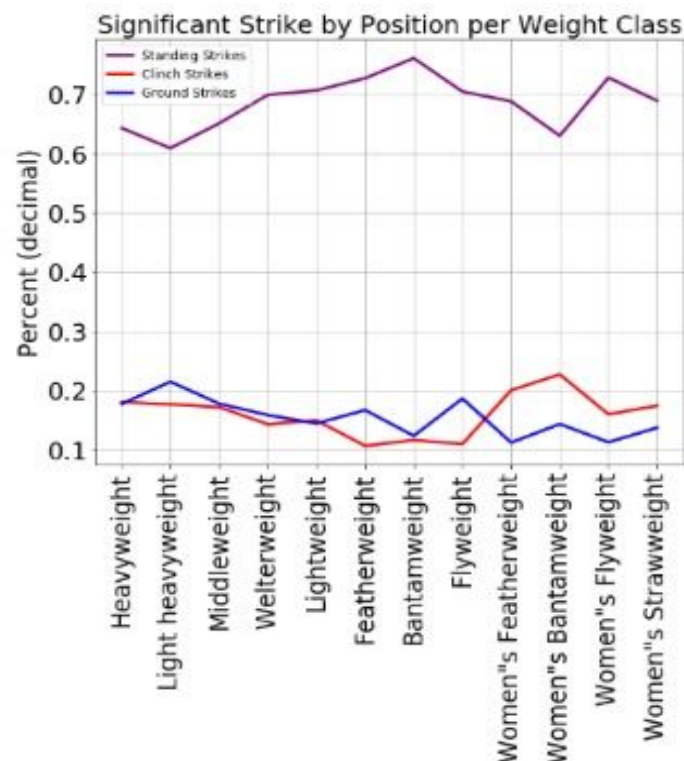
Fig 1.3



The x-axis contains the total number of strikes attempted. The y-axis represents the strikes landed. The blue circles represent the heavyweights from while the red dots represent the rest of the fighters in the UFC.



Fig. 1.4



The x-axis has the twelve weight classes, from mens heaviest to lightest, and then female heaviest to lightest. The y axis represents the strike type in decimal.

#### 4.1.4. Discussion

Regarding age in UFC, Figure 1.1 disproves the assumption that the UFC top athletes reach their career peak in their mid twenties. The average age of the top 5 athletes came to almost 32 years old. In fact the average age of fighters is over 30 for the most of the weight classes.

There was no evidence to suggest that height advantage has any effect on the success of UFC fighters. The top 5 fighters average height in each weight class was very similar to the average height of all fighters in those weight classes. This concludes that height advantage isn't important in order to be successful in UFC while success is more promising for fighters over the age of thirty. This suggests that experience is an important factor in succeeding in UFC.

Moving on to the effect of reach of UFC fighters, it's clear from Figure 1.2, that there's no correlation between win rate in UFC and ape index.

Next we analysed the the heavyweight fighters. We analysed the strikes attempted and strikes landed for heavyweight fighters against all other fighters. We thought that the heavier fighters would throw less strikes on account of them being less fit than other weight class fighters. However, as shown in Figure 1.3, the majority of heavyweights throw and land the same amount of strikes as the majority of the rest of the fighters in UFC. As displayed in

Figure 1.3, the cluster of dots in the bottom left corner represents the majority of fighters. As a result, there's no conclusive evidence to our theory that heavyweight fighters throw less strikes than other weight classes. To further analyse this theory, it might be relevant to look at the strike rate per fight. As we discovered in our analysis, the average fight time decreases as weight increases. Therefore, further analysis into strike per fight or per minute ratio might be a better way of studying this theory.

Penultimately, the significant strikes were analysed by the position the fighter threw that punch in. The fighters strike type percentage was compared weight class to weight class. Following on from our above theory about heavier fighters being less fit than other weight classes, we expected the heavyweights to have a strong clinch strike percentage as this is a position where fighters can rest and gather their breath. The graph that represented this analysis, Figure 1.4, showcases the gradual decline of clinch strikes for the men's weight classes as they range from heavy to light.

Overall, not as concrete evidence as hoped for, but certainly indicators that the heavy weight classes are less fit than the lighter classes.

## **4.2. RQ2**

### **Striking vs Grappling - Which is more popular and which is more successful?**

#### **4.2.1. Datasets**

The first thing to analyse was win type per weight class. This will inform us whether the effectiveness of striking/grappling changes weight class to weight class. For this, data on win type was used.

In MMA there are three ways to win. Knockout(KO/TKO) , submission and decision. Knockout artists are striking focused while submission specialists are grappling focused.

Next, champions striking accuracy vs takedown accuracy was looked at. Although this question doesn't specifically involve champions, trends in how they fight are very relevant as they are at the top of the sport.

The dataset contains strikes landed, strikes attempted, takedowns landed and takedowns attempted.

Lastly, win type per champion was examined. This will highlight how those at the top of the sport win their fights, through striking or grappling. The break down of their win types will also be compared to the average fighter.

#### **4.2.2. Approach**

How fighters are winning fights, weight class to weight class, was the first thing analysed. It is commonly argued that the heavier weight classes are less 'technical' - not as many submission victories, instead relying on brute strength. This could also tell us that striking is more important the heavier you get, and grappling less important. If submission victories do

decrease as weight increases then heavier fighters should probably focus more on their striking.

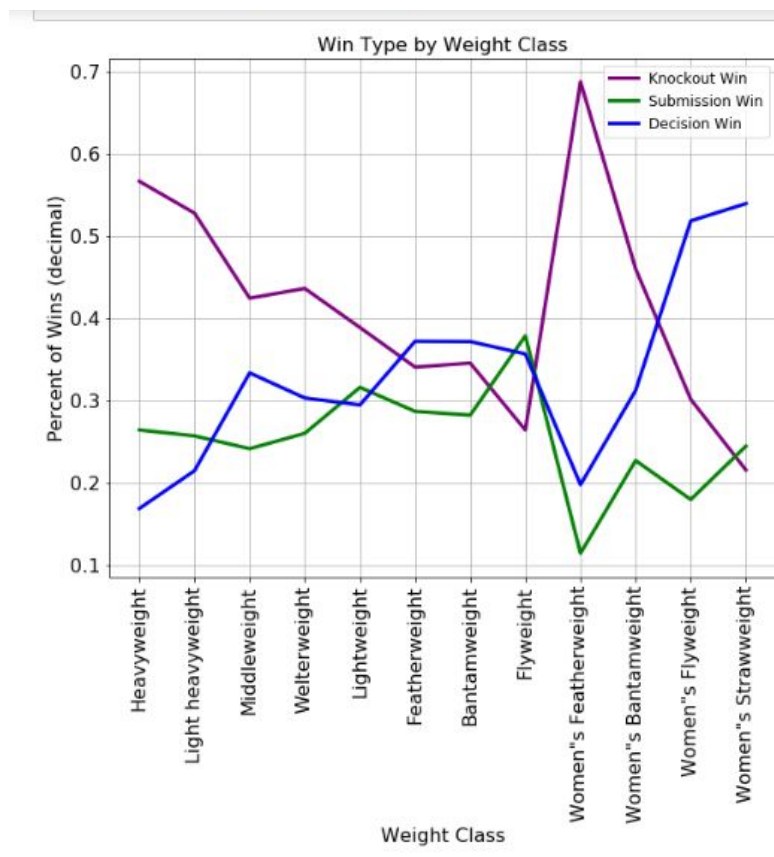
For this graph, the sum of each win type was found, per weight class. Each win type was then divided against total wins to get it as a percentage. The data frame was then ordered by weight class from mens heaviest to lightest, followed by womens heaviest to lightest.

Secondly we decided to compare the champions striking accuracy and takedown accuracy. This will tell us how important accuracy is among the best and help us answer the research question. To calculate striking accuracy we simply divided strikes landed by strikes attempted and we did a similar thing for takedown accuracy. All non-champs were then dropped.

Next we wished to design a graph breaking down how champions win their fights. Every champions wins of each type were divided by their total wins. Horizontal lines representing the breakdown of each win type for the average fighter were also drawn.

### 4.2.3. Results

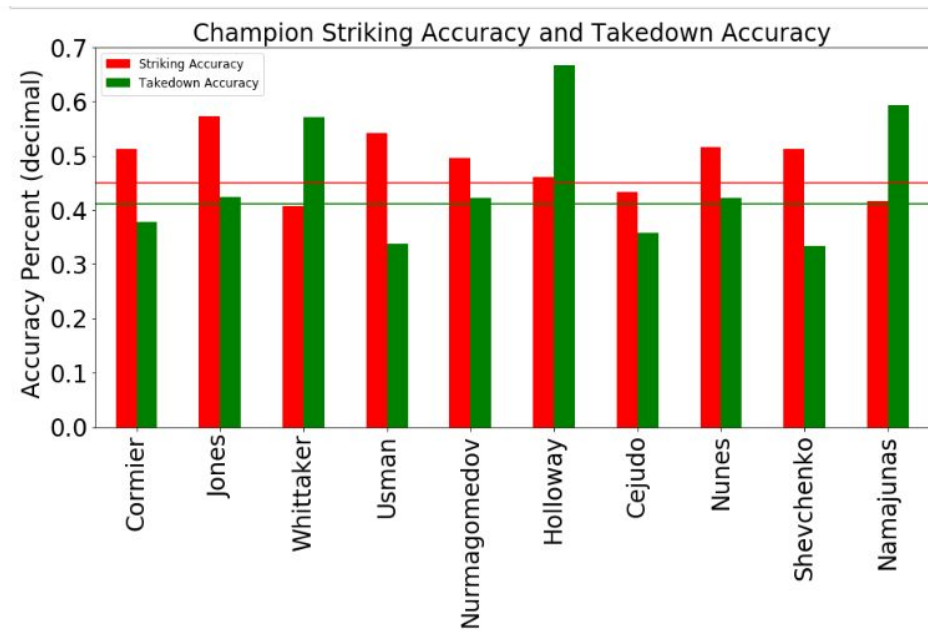
Fig 2.1



The x axis has the twelve weight classes, from mens heaviest to lightest, and then female heaviest to lightest. The y axis represents the percentage of each win type (as a decimal).

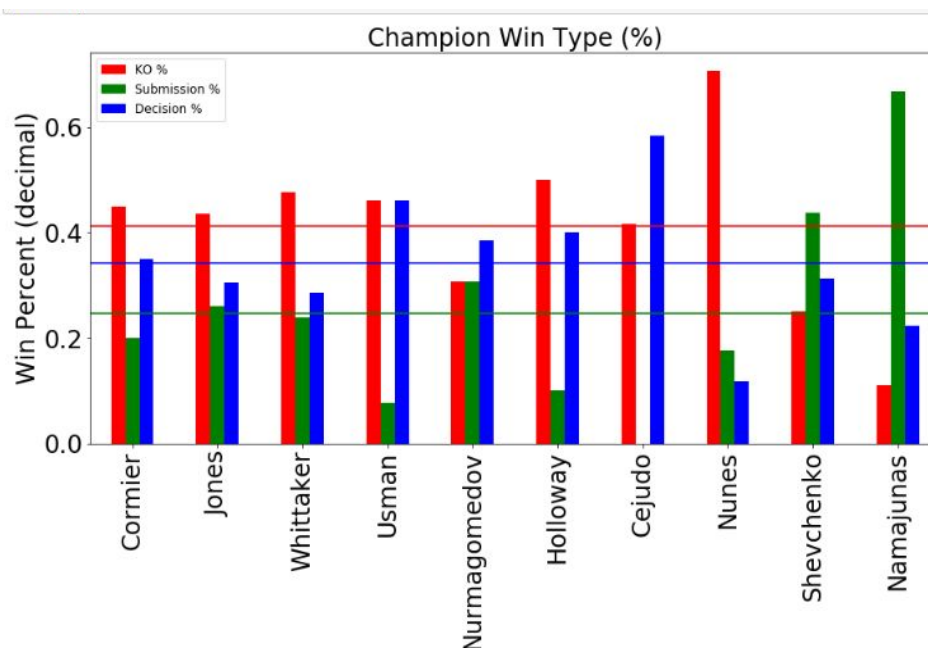
The purple line is knockout wins, the green line is submission wins and the blue line is decision wins.

Fig 2.2



The x-axis has the second names of the ten standing champions. (Nunes is champion of two divisions and bantamweight is vacant.) The y axis represents their accuracy as a percentage (decimal). The red bar represents striking accuracy, with the green bar representing takedown accuracy.

Fig 2.3



The x-axis has the names of the ten standing champions. The bars represent knockout, submission and decision victories as a percentage of total victories. The horizontal lines represent the mean win type percentage for the entire ufc i.e. the red line means that on average , knockout victories are 0.4 or forty percent of a fighters wins etc.

#### **4.2.4. Discussion**

##### **Win Type by Weight Class**

Figure 2.1 confirms the hypothesis that the importance of striking and grappling changes weight class to weight class. The proportion of knockouts increase consistently the heavier the weight class (mens). Submissions also decrease consistently as the weight class gets heavier. This tells us that striking is much more important at heavier weight classes than grappling, with there not being much of a threat of being submitted at heavyweight compared to lightweight.

Decisions also decrease consistently as the fighters get heavier, with relatively few decisions at heavyweight. This is one of the reasons that heavyweights will always draw a crowd as people like finishes. People will pay more to watch slower, more sluggish and less technical heavyweights than more athletic men at a lighter weight class, who are technically better fighters. Grappling fans should probably also stick to lighter weight classes as there is more of a chance of a submission.

Women's Featherweight and has a huge spike in knockout victories, this is an outlier as this division is relatively new, with few fights. Why Women's Bantamweight knockouts are also quite high, is unclear, however.

The main female divisions (Strawweight, Flyweight and Bantamweight) paint a different picture to the mens. The number of submission victories is quite low compared to their male equivalents. This indicates grappling is not as much of a factor as in mens MMA. This is likely because Women's MMA is in its infancy and the number of submission victories will increase as time goes on, with grappling improving in the women's divisions.

##### **Champion Striking and Takedown Accuracy**

Next we analysed champion striking accuracy vs takedown accuracy. This stat could be misleading as if you throw twice as many strikes as anyone else in the UFC, but only land half, you will have a poor striking accuracy but are probably still a dangerous striker.

Looking at Fig 2.2, interestingly seven of the ten champions have above average striking accuracy. (The red line indicates average striking accuracy for a UFC fighter, the green takedown accuracy etc). This is surprising because of the caveat I mentioned above. Being clinical with strikes seems to be important.

Takedowns paint a similar picture but don't have as much of a correlation, with six champions being above the mean.

## **Champion Win Type (%)**

Fig 2.3 is the most telling in the Striking vs Grappling question. Majority of fighters (seven out of ten) have above average knockout finishes. Only four of the ten have above average submission finishes. This indicates striking is more important than grappling in the UFC .

Striking is what most champion use to win fights in the UFC especially in the heavier weight classes. In Fig 2.1 only one weight class has more submission than knockout victories (flyweight). However when utilised correctly, and more so at lighter weights, grappling can be very deadly also, as highlighted by the champions with above average submission victories.

## **4.3. RQ3**

**How do champions differ from the rest?**

### **4.3.1. Datasets**

Firstly ,takedown and striking accuracy were analyzed in more detail than in research question two. As mentioned before, those stats could be misleading because they don't show the number of strikes/takedowns thrown. The actual number of strikes attempted and landed will be looked at now.

Secondly significant strikes by target will be examined. This is where a fighter lands their strikes, either head, body or legs. The mean for the entire UFC roster will also be calculated and compared to the champions

Finally significant strike by position will be graphed. This means where the fighter threw the strike from, be it standing, in the clinch or on the ground. This will be compared to the UFC average for each position also.

### **4.3.2. Approach**

For striking accuracy and takedown accuracy it was decided to plot both on two separate scatter graphs. Strikes/ Takedowns landed were put on one axis, with strikes/takedowns attempted on the other. Champions were highlighted as were ex-champions. These graphs also tie into research question 2 - Striking versus Grappling.

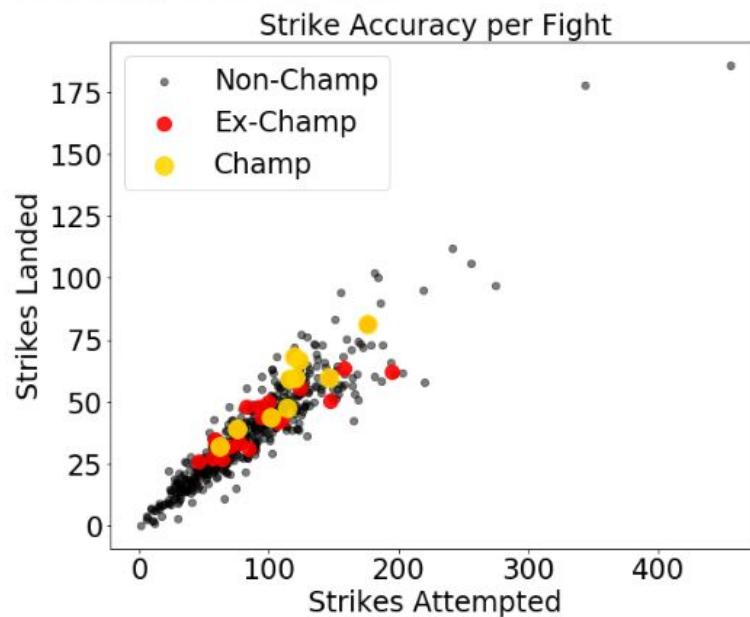
For significant strike by target, the data frame contained head strikes, body strikes and leg strikes for each fighter. Each of these three figures were divided by the total to get each as a percentage. This was then plotted on a barchart for the ten champions. The average was calculated for the entire UFC also and would be added to the bar chart.

Lastly, significant strike by position was displayed in a similar way to significant strike by target. The data frame had columns for each fighter containing standing, clinch and ground strikes thrown. These were divided by the fighters total strikes thrown to get each as a

percentage of the total. Once again these were plotted for each of the champions in a bar chart. The average for each was added to the chart once again.

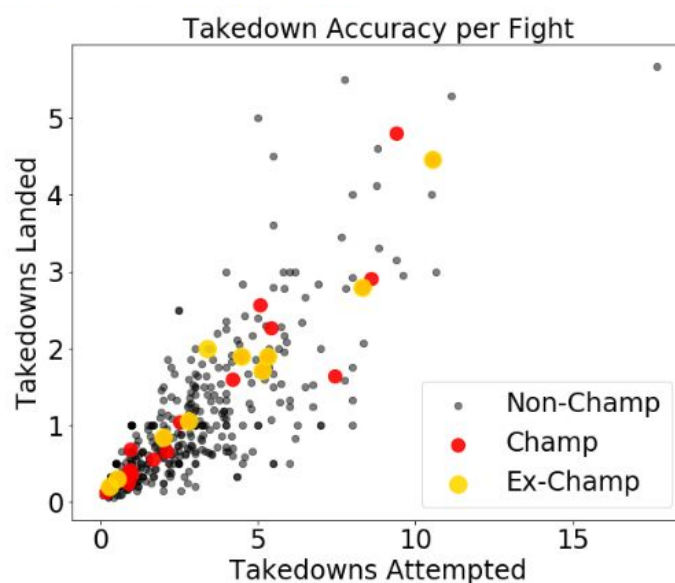
### 4.3.3. Results

Fig 3.1.1



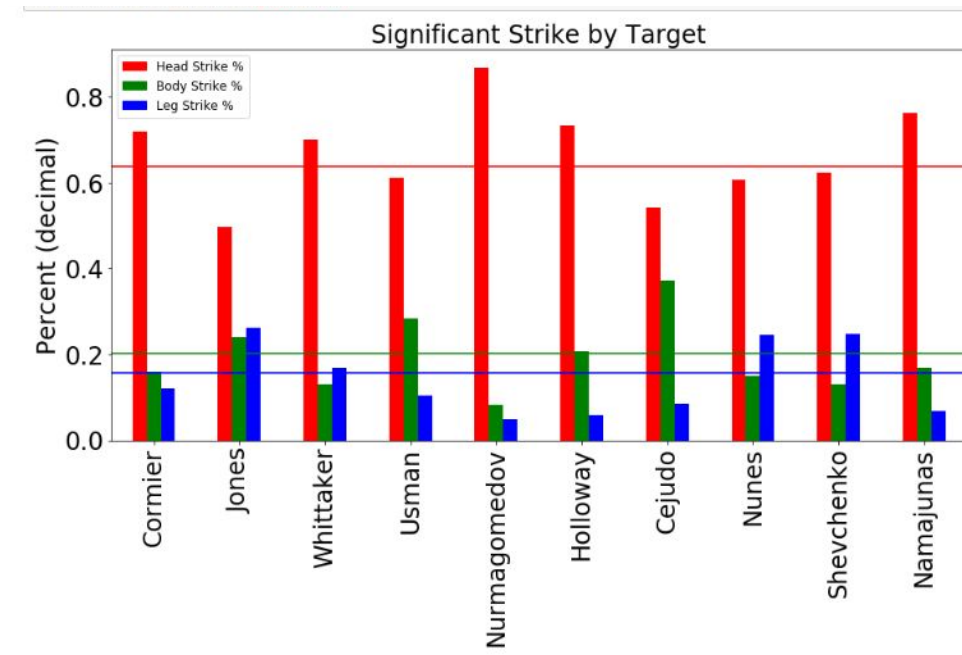
The x axis represents strikes attempted and the y , strikes landed. The large gold dots are champions, the medium red dots are former champions and the small black dots are everyone else.

Fig 3.1.2



This chart is very similar to the previous one. Takedowns Attempted are on the x axis, with takedowns landed on the y. Large gold dots are current champions, medium red dots are former champions and the small black dots represent every other fighter.

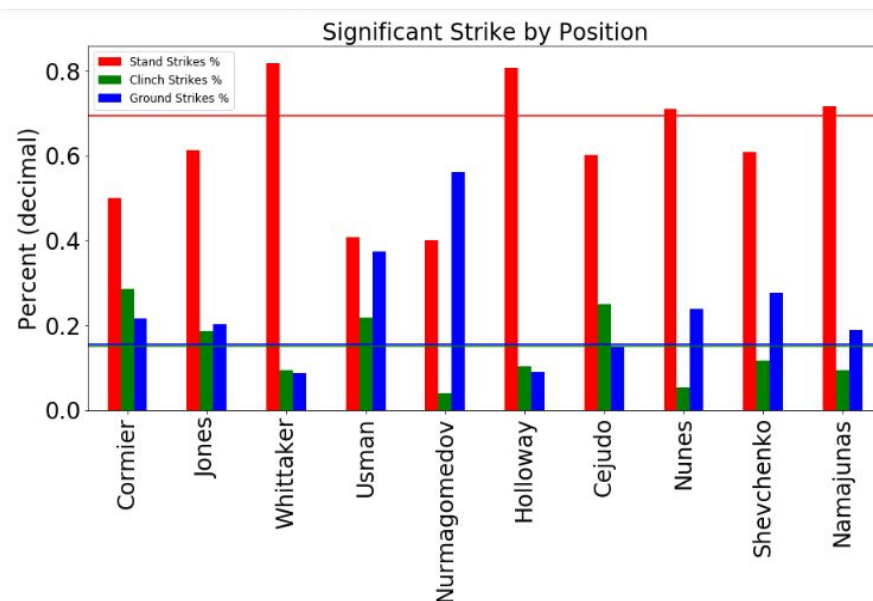
Fig 3.2



The ten current champions are on the x-axis. The red bars represent the percentage of their strikes that landed on the opponents head, green on the opponent's body and blue legs. The percentages are represented by decimals on the y axis.

The horizontal lines correspond to the bars, but for the entire UFC. The red line representing the average strikes landed to the opponent's head, for the entire roster.

Fig 3.3





This chart is similar to the one above. The ten champions are on the x axis. Each bar this time represents the position the fighter was in when they threw the strike. The red bar represents standing strikes, the green clinch strikes and the blue ground strikes. The horizontal lines once again represent the UFC average. The red line represents average percentage of strikes that were from standing etc. The blue and green lines are very close and hard to distinguish from one another.

#### **4.3.4. Discussion**

##### **Striking and Takedown Accuracy**

This graph doesn't tell us much in terms of striking accuracy, all the red and gold dots are in line with the rest. It does tell us two things however. Striking is essential in the UFC. All the champs, past and present have a certain level of striking with none of them appearing at the lower end of Fig 3.1.1. The current champions are higher up the chart than the bottom few former champions also. Not one champion, past or present in the dataset has a particularly poor striking output.

Secondly, this graph tells us that a good level of fitness or 'cardio' is essential to make it in the UFC. To throw more strikes than the average fighter requires very good fitness levels and it is clear from the graph that all champions have at very worst a decent striking output, with most near the upper right corner of this graph.

Fig 3.1.2 paints a different picture. Many of the champions and former champions hover in and around zero or one takedowns per fight. However four or five are in the upper tiers, with another four being in and amongst the best wrestlers statistically in the UFC.

Fig 3.1.1 and 3.1.2 reinforces the conclusion from research question 2. Striking is essential in the UFC and no fighter will get anywhere without it. Grappling isn't essential, with many champions and former champions not utilising it much or at all. However if used correctly it can be devastating as highlighted by those champions near the upper right of Fig 3.1.2.

##### **Significant Strike by Target**

Fig 3.2 shows that many of the champions have a much lower percentage of leg strikes than the average UFC fighter with six of the ten champions below the average on this count. Maybe leg strikes are a gimmick that should be abandoned to get to the upper echelons of the sport. On the other hand, the sport is changing all the time and maybe in a few years the champions will be using them more.

Cormier, Jones, Usman and Khabib (Nurmagomedov) are the only fighter who truly mixes up their strikes, with the lowest head strike ratio and high body and head strikes. Jones and Cormier are two of the more successful champions on this list, with Usman and Khabib being two of the more recent men to get the belt. (Although all champions on this list have been crowned relatively recently.) Maybe mixing up strikes is the future of the sport.

Whitaker and Holloway are the opposite with nearly no leg and body strikes and all head strikes.

For this reason this graph is inconclusive on what separates the champions from the rest, although I would argue based on the data, that a fighter should abandon body and leg strikes and focus on the head like Holloway and Whitaker , or try and mix up their strikes as much as they can like Cormier, Jones, Khabib and Usman.

As we saw in research question one, the heavier weight classes tire faster than the lighter classes. Body strikes and leg strikes are better at slowing an opponent than winning a fight. We also saw in research question two that knockouts are much more prevalent in heavier weight classes.

Both Jones and Cormier, the champions at the two heaviest weight classes, attack the legs and body frequently. Perhaps the plan is to tire these heavy powerful fighters at the start of the fight, avoid getting knocked out and finish them easily once they slow down near the end of the fight. Title fights are also twenty five minutes whereas non-title fights are fifteen so many of these heavy fighters could find themselves in big trouble if they can't finish their opponent early.

### **Significant Strike by Position**

Looking at Fig 3.3, it echoes Fig 3.2 in many ways. Cormier, Jones, Usman and Khabib (Nurmagomedov) are the best at mixing up their position of attack. These four men have a lot lower standing strike percentage compared to the average, but a lot more ground and clinch strikes (except for Khabib who has barely any clinch strikes).

However Holloway and Whitaker have much higher standing strike percentages than the average, and both have barely any clinch or ground strikes.

Once again these results are inconclusive, perhaps once again, it is better to either focus completely on standing strikes as much as possible or try to vary it up as much as you can. The same fighters who vary where they throw strikes form also varied up their targets, although I don't know why this is.

The two men at the heaviest weight classes, Jones and Cormier varied up their positions a lot. As we seen in research question two, there are not many submission victories at the heavier weight classes. Perhaps taking heavier men to the ground where they are uncomfortable is a winning combination. Jones is also notoriously dangerous in the clinch, and as we shown in research question one, heavier fighters spend the most time in the clinch. A good strategy at heavier weights therefore, might be to tire them out by switching positions and hit them hard when they go to clinch.

## 5. Discussion

### 5.1. Ethical Considerations

All the UFC athletes are in the public eye and most have huge online personas. All the data we used for the project is publicly available also.

### 5.2. Reproducibility

Reproducibility is important in ensuring that anyone that follows the same steps as outlined in this report and code, will achieve the same results.

We used markdown cells to explain exactly what was done in our notebooks, step by step. All the data we scraped is publicly available online.

The Wikipedia page of current UFC fighters is being updated all the time. Our scraper broke a few times during the course of the project because of this, with entries being added to the page that weren't formatted correctly. All the datasets we used are stored as csv files and these could be used if something goes wrong with the scraping process.

### 5.3. Limitations

Another question that we wished to answer was 'How has MMA has changed over time?'. Past champions were most likely less well rounded than the current champions. There were probably more 'specialists' i.e a champion with no submission victories, or a champion with only submission victories. It would have been interesting to analyse changes in the sport over time.

After finding the Wikipedia page containing the names of all past UFC champions, and noticing that the UFC website had a section named 'inactive fighters', we thought we could do this question justice. The website did contain pages for most ex-champions but sadly, most of these pages were devoid of stats, with zero in every column.

Research question 2 - 'Striking vs Grappling' was harder to answer than expected. It is hard to put striking and grappling into numbers. For example, we analysed knockout wins and correlated this to striking, yet some fighters could outmaneuver their opponent in a grappling exchange before finishing with strikes. On the other hand, a fighter could stun their opponent with strikes before locking in an easy submission, without much grappling involved.

Finally, for research question 3 - 'How do champions differ from the rest?', we wished to add back to back graphs to some of our graphs, based on champions title reigns and defences. For example for the champions win type percentage, we could have a bar chart of their title

defences back-to-back. This would further split the champions up and tell us who has been on top of the sport the longest.

Title defence information was present on the past champions Wikipedia page and was scraped and cleaned. However, unfortunately most of the current champions are new with one or zero title defences, so this didn't come to fruition.

## 6. Conclusions & Future Work

In summary this project was an analysis of UFC fighters, mainly looking at three things, the physical differences between fighters weight class to weight class, the effectiveness of striking vs the effectiveness of grappling and how UFC champions differ from the average fighter.

Physically we found that reach and height grant no advantage in the UFC. The average age for a fighter is in and around thirty which is much older than sports such as soccer when most players would be at their peak or declining at this stage. The average age of the top five fighters in each weight class was a year older also at thirty-one. Experience is therefore key in this sport.

In the striking vs grappling question, we found that striking is more and more important, the heavier the weight class, and grappling becoming more and more important the lighter the weight. We found that striking is necessary for achieving anything in this sport. Grappling is not as essential, with many champions not utilising at all, but if used effectively can be devastating.

Lastly we did not find many concrete reasons that separates the champions from the average fighter. Once again it was confirmed that striking is key with every champion having, at very worst, a decent level of striking.

Many champions mix up their attacks and the positions of their attacks and we can theorise why this would be successful, especially at the heavier weight classes. However, a few champions don't vary their attacks at all with above average head strikes and standing strikes.

There are many ways to be successful in MMA and this is why this sport has always fascinated me as opposed to boxing.

Both research questions two and three could benefit from more in depth analysis. As explained in the 'Limitations' section, knockout wins don't necessarily equal striking and submission victories don't necessarily equal grappling. It would be interesting to breakdown this area further.

The breakdown of the champions could benefit from more in depth examination also. Naming individual champions in the scatter plots and analysing the different levels of striking and grappling would be interesting. Are more successful champions good at both grappling and striking or is being more specialised more advantageous?

Also as mentioned in the 'Limitations' section, we only had access to active fighters. A bigger dataset of older fighters would be interesting to investigate. There was probably a lot more specialists in the past, with the average fighter being much less well rounded.

Some of the trends seen in the women's side of the sport , such as few submission victories was probably prevalent in the mens sport in the past. In the past fighters came to MMA with a boxing or wrestling background. More and more fighters are training MMA from a young age and it would be compelling to see how much the sport has changed in the last twenty years.

## 7. Bibliography

Kaggle Project - <https://www.kaggle.com/rishpande/ufc-data-analysis-visualization-beginner>

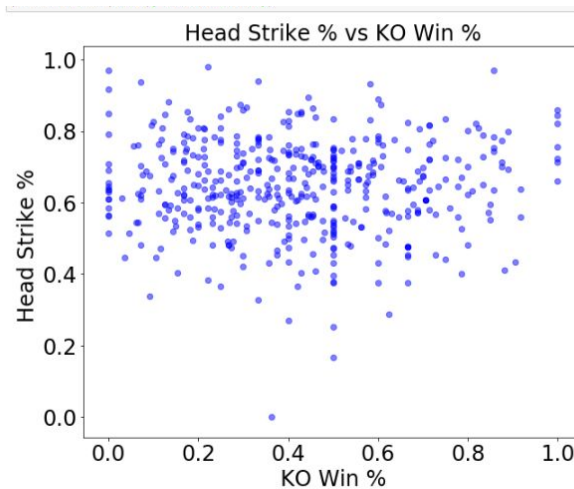
## 8. Appendix

A potential research question that didn't make it in the end was the effect of head strikes, strike accuracy and standing strikes on knockout victories

### **Head strike % vs K.O. Win %**

Head strike % is the percentage of a fighters strikes than land on the opponent's head. This was calculated by dividing head strikes by total number of strikes. K.O Win % is the percentage of a fighters wins that come by way of knockout. This was found similarly, by dividing knockout wins by total number of wins.

Fig 4.1

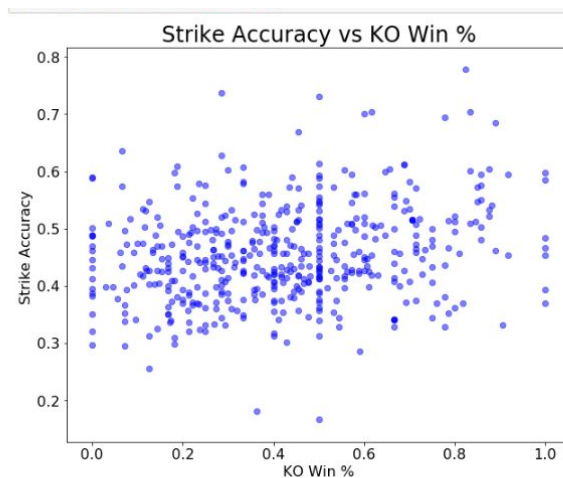


As seen in Fig 4.1 , there was no correlation.

### Strike Accuracy vs K.O Win %

Strike accuracy was calculated by dividing strikes landed by strikes attempted and K.O. win percentage was calculated the same way as it was above

Fig 4.2 (below)

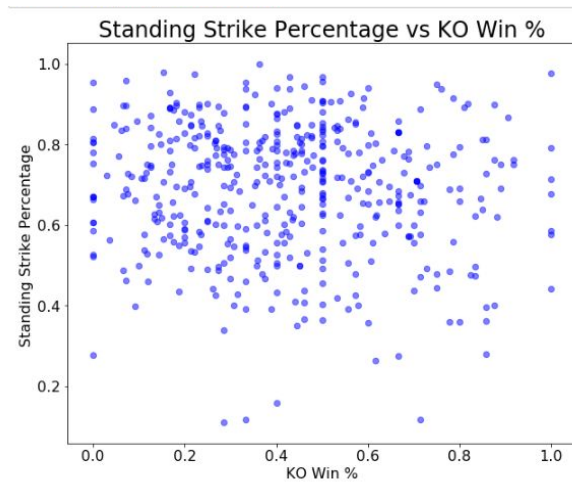


Once again, as seen in Fig 4.2, there was no correlation.

### Standing Strike % vs K.O. Win %

Finally we wished to compare standing strike percentage and knockout win percentage. Standing strikes are where the fighter throws the strike from a standing position, as opposed to in a clinch or on the ground. This was calculated by dividing standing strikes by total number of strikes. Knockout win percentage was calculated the same way as before.

Fig 4.3



Surprisingly once again, there was no correlation.