**Find the data set**

<https://www.kaggle.com/rajeevw/ufcdata> - historical fight data

<https://www.kaggle.com/calmdownkarm/ufcdataset-> UFC Fighter data

Above are link to the two data sets I have used for this CA

I have a huge interest in the sport of mixed martial arts so I wanted to do my assignment based on data from the biggest organisation for it called the UFC. This is sport where two athletes competed against one another to knockout ,submit or out point one another. The UFC is the most well-known and biggest organisation for the sport of mixed martial arts world-wide.

**1st data set**

**historical fight data**

This data was large it contained 145 columns and over 1000 rows .It details got to do with every single fight the organisation had held in from the year 1993 until 2019.This meaning the data was pretty update missing only a couple of months .The data was taken from the UFC official website. Each row of the data set is a mixer of stats between the two fighters R\_fighter (red gloved fighter) and B\_fighter(blue gloved fighter).

Attributes that should be include( Example columns in data set )

Continuous Numeric data == Reach : The arm length of every fighter

Categorical data == Weight Class : Every fight that happened was under a specific weight class

Temporal data == date : This would store the date each fight took place

Location data == location :This was where each fight took place

The data is extremely detailed listing various things such as damage taken, leg kicks landed and takedowns completed. The location of

**2nd data set**

**UFC Fighter data**

This was also a ufc data set it had over 894 columns and over 500 rows .Unlike the first data set this gave me valuable information about each fighter in the organisation like stats and achievements .The focus was more about fighter information rather than the fight itself. Some example of columns are if the fighter is a champion, fighting stance and where they are from.

Attributes that should be include( Examples in data set )

Continuous Numeric data == height : The height of each fighter

Categorical data == stance : The way the fighter stands whilst fighting

Temporal data == date : year they work born

Location data == Hometown :Where the fighter actually lives

Merging

Merging these two data sets was crucial to my data investigation .It allowed me to link up fighter attributes, performance stats and personal information. Using this I was able to analyse their perform in their fights. This analyse allowed me to compare how they fight, train and how their physical attributes effected them come fight time.

**Explore the data**

Wrangling

I did tonnes of data wrangling which can be seen in my Jupyter notebook. This allowed me to break down and understand the data for what it was .

Restructuring the data into smaller data frames gave me a clear view of the data and how I could go about using it in my assignment .I created many smaller data frames had about 8 columns. This helped me to understand the information and figure out what it was telling me.

This picture below is an example of some data wrangling I done. Here we can see that I am merging to different data frames and then taking out the columns which I require.

I then go onto sort the data according to the most tko/ko victories.

A screenshot of a social media post

Description automatically generated

generating graphs

As you can see from my notebook I Generated many different graphs to get better picture of what the information was telling me. I use the graphs to help me find interesting statistics that I could later form my story about.

Below are just two of the graphs that I made whilst wrangling the data .The pie chart is a representation of the ages of ever fighter registered with the UFC. We can quickly see by the chart that majority of fighters are between the age of 30-35 and only 4.1% is less than the age of 20.

The column chart details information about where all of the fights take place around the world. We can clearly see that the most amount of fights take place in Las Vegas. The second most popular city for the organisation to host fights is in London.

A picture containing screenshot

Description automatically generated A screenshot of a cell phone

Description automatically generated

correlations between attributes.

It was interesting to see the correlation between the attributes throughout the data. I seen this on many occasions for example the weight of the fighter had a real correlation as to how each fight would finish. If the two athletes fighting were in the heavier weight divisions the chances of that the fight would end early would go up dramatically. Alternatively if the Two athletes fighting were in the smallest weight category the chances of the fight ending early are very slim.

Fix errors and decide on how errors and omissions will be handled.

A couple of errors were found throughout the wrangling process these included nonsense, empty fields and N/a fields .

When these errors according I would clean up the dataset by creating newer cleaner set by filtering out o.

This screenshot below is a great example of me fixing and handling data errors. Here I merge data, take out useless/meaningless columns and also remove any rows containing null values within the data set.

A screenshot of a social media post

Description automatically generated

Justify the chart types

The charts I used during the wrangling process were scatterplots ,bagplots and pie charts .I used these as I find them easy to construct and straight forward understand .A lot of more complex charts may sometimes confuse you more. When data wrangling its best to keep it as simple as is possible.

These types 3 charts are very good for visualising different types of data. The pie chart was used where a percentage was needed .The scatter plot was used to see how one variable effects another .The bar charts were used when displaying different data frequencies.

Develop a meta-data description of your dataset

I used many columns in my data visualisations such as :

Height,age,stance,name,weight,champion\_status,submission\_victories,tko/ko\_victories,ref\_data,date,location,city,country

**Form your Big Idea**

what the data is telling you

The data told me information about certain fights attributes and how they correlate to the way that particular fighter wins.

establish a message

my message can be broken up into four stages .

1)if you want to be a UFC champion one day you need ideally you need to be competing at the highest level in your early 20s

2)The most successful fighters in the UFC are fighting using the orthodox stance. This is a good indication that this stance is the most effective when it comes of becoming a champion.

3)If you one of the smaller guys you should focus more on your submission rather than knocking people out as you are more likely to win fights that way

4)If you are one of the bigger guys you should focus more on striking rather than submissions as you have much more chances of winning fights be knocking guys out.

an audience

My audience would be to current fighters, coaches and young martial artists looking to make a name for themselves in the sport of MMA.

what do you want from the story

I want the people mentioned above to benefit from the story for it to help guide them to becoming better at the sport .

I want to give a realistic blue print ,guide and reference in terms of a fighters attributes. How they can better achieve their goals of becoming a top fighter and where the window of opportunity lies especially when it comes ones age.

what is at stake for them

Their job/money is at stake nearly all fighters in the UFC are full time fighters. If they do not have the right game plan, they could begin to lose and not get paid what they otherwise could for winning.

This could be a coaches fighter that he is concerned about. If the coach has the for his fighter he can better caterer to their fighters strengths and direct them on the path to victory.

A sentence expressing your big idea

How martials artists can improve their performances and get the most out of their attributes.

**Develop your Storyboard**

At least 4 charts in the story

A lot of data wrangling and manipulation was done to achieve these graphs, the work can be seen in the jupyter notebook file attached to this folder.

See juypter notebook for more details and clearer graphs

Chart1-Age percentages of current champions

A close up of a logo

Description automatically generated

This graph represents the all of the current champions and there age ranges .Here we can see that most of the champions are between the ages of 20 and 25.The age group with lowest amount of champions is the age bracket of 35 -40 years old. This data tells us that if you want more of a chance of becoming a UFC champion you need to be competing at a high level at quite a young age.

Chart2- Fighting stances of all current champions

A screenshot of a cell phone

Description automatically generated

The graph above is also represent data from current UFC champions. This data is the stance that each of the champions choose to use .The charts show the majority of the champions fight using an orthodox stance .We can conclude that if you are trying to become a UFC champion your probably best off fighting using the same stance as many of the champions in the UFC .

Chart3- TKO/KO statistic for the most top 10 fighters with the most TKO/KO according to their height, weight and amount of TKO/KOs

A picture containing computer

Description automatically generated

This chart above details if from the top ten fighters for TKO/ko wins in the UFC. This graph takes in each ones attributes such as height ,age, weight and number of TKO/KO victories. We can see from the graph that if you are an athlete that weighs in the ball park of 70 to 90 kg you should be looking for the TKO/ko in fights to increase your chances of winning

Chart4- Submission statistic for the most top 10 fighters with the most submissions according to their height and amount of submissions

A screenshot of a cell phone

Description automatically generated

This chart above details if from the top ten fighters for submission wins in the UFC. This graph takes in each ones attributes such as weight and number of submission victories. We can see from the graph that if you are an athlete that weighs 75 kg or below you should be looking for the submissions in fights to increase your chances of winning.

Story - bring it all together

Using these visualisations we can begin to create a plan to show athletes how to use their game in order to give them a competitive advantage .We can see from analysing data from the very best of the best sport exactly what tends to work and what does not .

If you want to become a UFC champion you need to be fight competitively at a young age to increases your chances . Using the orthodox fighting stance will increase your chances of winning .If you are smaller guy in weight focus more on submission for better chances .If you are a heavier guy focus less on submissions and more on knocking your opponent and you will have more victories