|  |
| --- |
| IT Tallaght |
| Interactive Media Design and Data Visualisation |
| Continuous Assessment 1 |
|  |
| **Conor Griffin X00111602** |
| **10/26/2017** |

|  |
| --- |
| This document contains data analysis and snapshots of visualizations I have created using Tableau. The document is about UFC data which I gathered online and used to then create visualizations. My goal in creating this document is for the user to gain an understanding of the UFC and show how data analysis can be used to make future decisions. |

# 

Contents

[Summary 1](#_Toc497438168)

[Background 0](#_Toc497438169)

[Data Sets 0](#_Toc497438170)

[Seven Stages 0](#_Toc497438171)

[Parsing, Filtering and Mining the data 0](#_Toc497438172)

[Represent 1](#_Toc497438173)

[Fights and Fighter analysis 1](#_Toc497438174)

[Seven Stages 0](#_Toc497438175)

[Parsing, Mining and Filtering the data 0](#_Toc497438176)

[Represent 3](#_Toc497438177)

[Fighters and PPV draw in Countries 3](#_Toc497438178)

[Problems & Solutions 0](#_Toc497438179)

[Conclusion 0](#_Toc497438180)

# Summary

For the purpose of this Continuous Assessment (CA) I will be doing analysis on the Ultimate Fighting Championship (UFC). I have gathered information from multiple data sets online. Each data set has data on UFC events, UFC fighters, UFC pay per view buys and other interesting information. I have gathered the information from multiple online sources which I have referenced in the Data sets section of the document.

My goal in this project is to display to the reader information on the UFC. The task is to gather as much information on the UFC as possible, then sort the data by organising it into separate excel spread sheets and finally data mine the data so to extract what I need from each spread sheet then use it to create visualisations using tableau and display it to the user.

Another objective is to gather the data and make visualisations that can then be used to make future decisions, the data will show to the user patterns in fighters, their stats, fighting style, wins, weight class, height, etc. I will also show data on every UFC fight/event up to UFC 195, I am hoping to identify patterns in fights, methods of victory (KO, TKO, Decision, Submission, etc), what round most fights were won in, what a fighters most frequent type of victory is. Using the data on UFC fighters I will show what country most fighters are from and show where each of the UFC belt holders are from and provide stats on them. I will try to show a link in how the number of fighters from a country can be linked to the number of gyms in a country. Then using this data display visualizations to the user in order for them to see how the data can be used to make future decisions on UFC events and fighters.

I will gather information on each of the UFC events up to UFC 217 and show the number of Pay per view (PPV) buys on that event, I will try to use this data to then show to the user how certain fighters are a bigger draw then others and that maybe there is a correlation in the number of buys increasing as the UFC has gotten older and more popular.

I am hoping to discover many things; I want to be able to make wise and objective decisions on the UFC when watching its upcoming events. I want to know what fighters are more popular than others and be able to make a good estimate on PPV buys on an event based on fighter popularity, this data I am sure is used by the UFC to also know what fighters and fights are more popular and what entices more PPV buys. This data can be used for promotion handling, fighter pay and also for greater match ups.

I also hope to discover through data mining and creating visualisations patterns in the data and how they can be used to make decisions whether on betting on fights or even to have an advantage in making decisions on fights. I also want to be able to use Tableau confidentially so that in the future I will be able to use this tool in the workplace. Data analysis and Information Management has always been something that I’ve been interested in and it is definitely something that is used more and more in IT and the workplace, I feel through data mining and analysis and creating visualisations this will increase my knowledge and skills in this area.

# Background

The Ultimate Fighting Championship is an American mixed martial arts organization based in Las Vegas, Nevada, that is owned and operated by parent company WME–IMG. It was founded in 1993.

I came up with the idea because I have always been interested in the UFC and been a fan for years. I have always been a fan of the fight game and the UFC represents the best fighting athletes in the world, some of my favourites include Jon Jones, Conor McGregor, Anderson Silva, Robbie Lawler, Ronda Rousey and many others. I have always wanted to find out more information about these fighters so choosing the UFC for this data analysis assignment was easy. There have been hundreds of fights since the UFC was set up so gathering data on these events has allowed me to gain an insight into how fights usually end, the result, method of victory, fighter popularity, PPV popularity, etc. This allows me to then make future decisions on events and fights based on data I’ve gathered and visualisations I’ve made.

I also picked this area because from my research I found that there doesn’t seem to be much analysis or data visualisations made on the UFC, through looking on the web I could only find a couple websites where the author has done data analysis and made visualisations on UFC data. By creating visualisations on UFC data I’m doing something that hasn’t been done yet to a certain extent, I’m not copying anyone else’s work since there is no work to copy. These factors make it a challenging and interesting area to do my data analysis and visualisations on.

I watch the UFC events with my friends when it’s on the television so doing this analysis on the events since the beginning of the UFC will help me to make decisions on future events. Creating visualisations on the UFC events data will allow me to know how a fight might play out, the method of victory, fighter’s preferred style, and other key factors in an event. If I was a betting man, the data I will create visualisations off will allow a person who likes to bet on sport to use these visualisations to make a confident bet on how a fight will likely play out, or maybe the amount of PPV buys on an event based on fighter popularity.

I want the reader to also gain an understanding of the UFC and for them to see how analysis in this area can be important for all UFC fans in terms of knowing what fighters attract more fans because of popularity, there fighting style, why there are more fighters from one country than others. I hope that this data can be used by all UFC fans so that they can look at the data and visualisations and gain greater knowledge of the sport and also make decisions on future events. I hope by doing this assessment I will show myself and the reader some important facts and patterns in the UFC and how there is correlation in how fights and events usually play out.

# Data Sets

* UFC Fighters and UFC Events:

Redditcom. 2017. *Reddit.* [Online]. [31 October 2017]. Available from: https://www.reddit.com/r/datasets/comments/47a7wh/ufc\_fights\_and\_fighter\_data/

In-text citation: (Redditcom, 2017)

* UFC PPVBuys

Thesportsdailycom. 2017. *Thesportsdailycom.* [Online]. [31 October 2017]. Available from: http://thesportsdaily.com/mma-manifesto/all-time-ufc-ppv-sales-data/

In-text citation: (Thesportsdailycom, 2017)

* Amount of Gyms in countries

I created this spread sheet myself based off data I found through multiple google searches.

Statistacom. 2017. *Statista.* [Online]. [31 October 2017]. Available from: https://www.statista.com/statistics/308831/health-club-amount-in-european-countries/

In-text citation: (Statistacom, 2017)

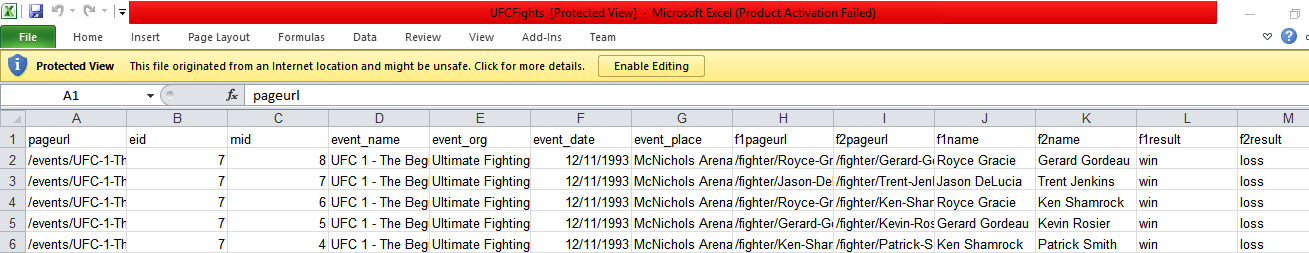
# Seven Stages

First I will discuss the seven stages for completing the visualisations on the UFC Events dataset. I acquired the dataset online by searching “UFC datasets,” and found a website with two excel spread sheets, one on UFC Events and another on UFC Fighters. The data is from the first UFC event (UFC 1, November 1993) to UFC 195 (January 2016) and contains information on every UFC event and fight up to UFC 195.

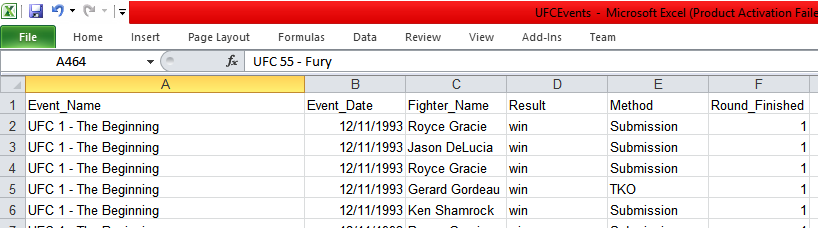
## Parsing, Filtering and Mining the data

I began by analysing the UFC Events data, highlighting its features in order of their importance. This revealed patterns in the data such as identifying that although this data was on every UFC event/fight in the UFC from the beginning I could now use the data to identify other features such as the amount of wins a fighter has, most dominant method of victory (KO, TKO, Submission, etc.) and the round a fight was won in the most. By identifying that these features existed within the data my goal was to then make use of the data to then build nice visualisations off.

The original spreadsheet looked like the following



I parsed the data by cleaning out unnecessary columns and rows. I renamed the Spread sheet to UFC Events.



There were some rows that I didn’t need such as certain types of Methods like “No Contest – Failed Drug Test”, or “No Contest – Overturned by Commission,” I found that I didn’t need these rows as there were only one of each and they didn’t help reaching my intended goals.

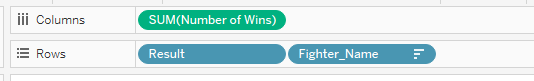


By filtering out the data that I didn’t need this gave me a clearer picture of my goal in identifying some of the key factors in a UFC events and fights.

## Represent

### Fights and Fighter analysis

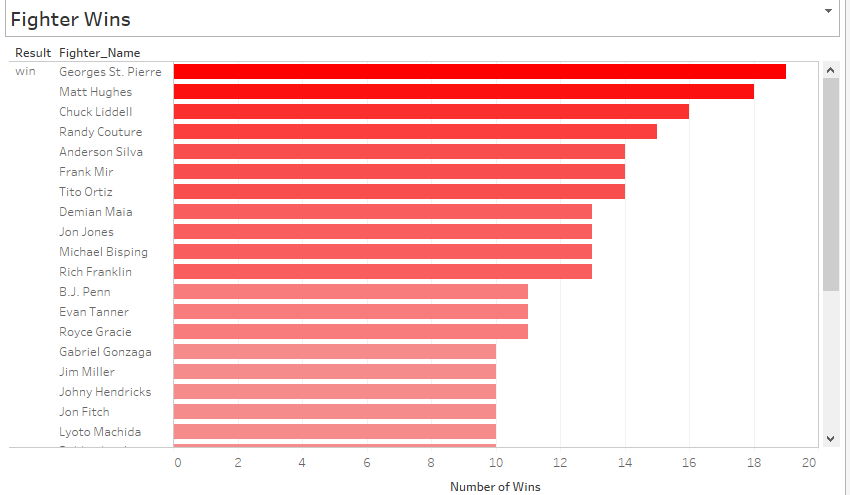
I used Tableau to represent the data into visualisations. I broke the visualisations up into categories based on the data. The spread sheet UFC Events was imported into Tableau and I broke the visualisations for this data set into three categories. First Fighter data, by using the data I was able to gain information on what fighter has the most wins, by using the data I could make Tableau sort the fighters and sum up the number of wins a fighter has.



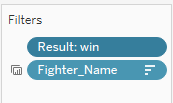
I chose a horizontal bar chart as my graph type. I think this was the best graph type as it allowed me to show the Fighter name on the rows clearly and then to show the number of wins on the columns clearly.

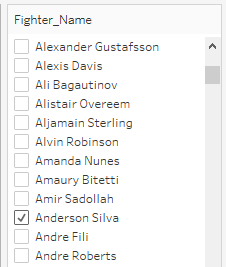


I gave the graph the heading Fighter Wins as this graph will be used to represent the amount of wins a fighter has based on every UFC event and fight gathered from the data set. I added a a colour to the marks card for the number of wins, the more wins the darker the colour.

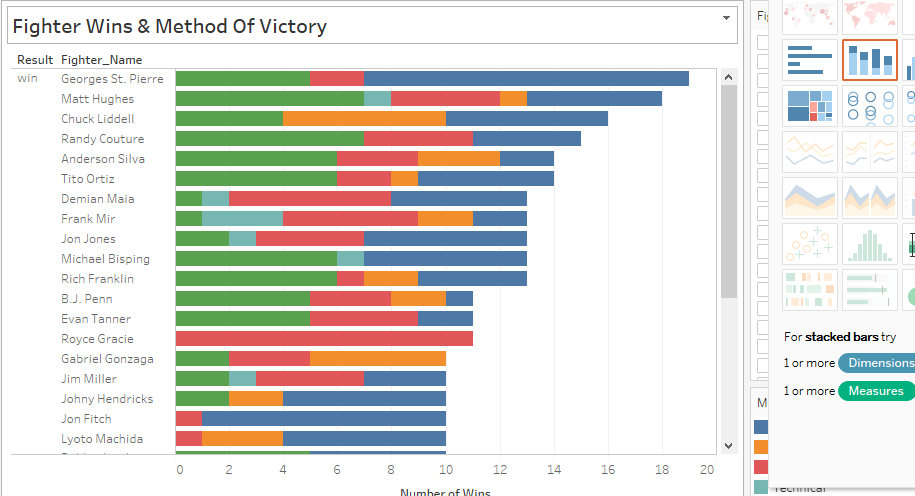


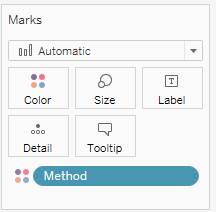
I also added a filter to the visualisation because there is so many fighters in the UFC so by adding a filter for the fighter name the user can then interact with the visualisation and decide what fighter(s) they want to see data on. I also filtered out draws as I only want to see the number of wins a fighter has.



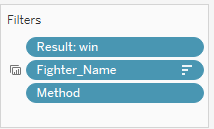


The next visualisation was based on fighter wins and method of victory categories. Similar to the first visualisation but different in the way the user can now see how many fights a fighter has won and how he won that fight. When looking at the dashboard the user will be able to interact with both visualisations. First I selected result, fighter name and number of wins and picked a stacked bar chart as my graph type because I want to be able to see the different method types (KO, TKO, submission, etc.). I gave the graph a title Fighter wins and method of victory. I also used a colour marks function to colour differentiate each of the different methods.

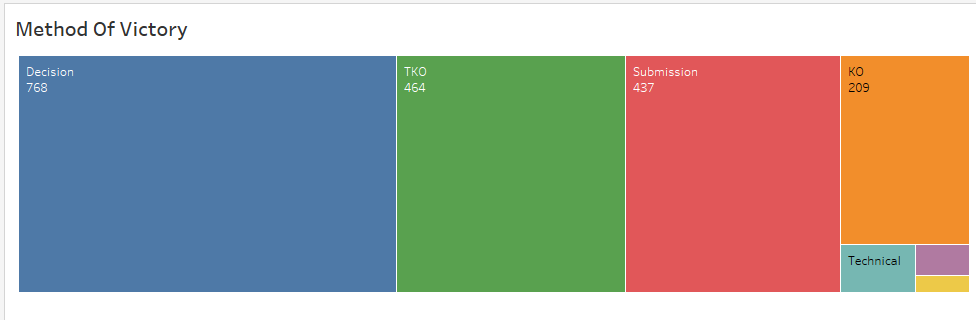




I added filters to the graph, like the first visualisation I only want to see amount of wins, also a filter for fighter name so that the user can interact with the visualisation and select fighter(s) name and I also added a filter for the method because I don’t want to see if the fighter won by disqualification or a no contest bout I only want to see KO’s, decision wins or submissions, etc.

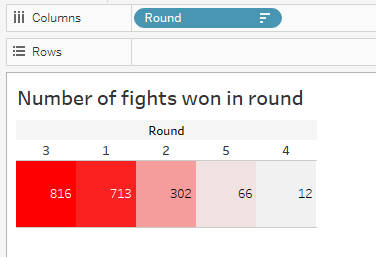


The next visualisation was created to represent the method of victory for all UFC events/fights from the beginning. I wanted to be able to see what the most dominant victory method is so that this data could maybe be used to make future decisions when watching the UFC. I used a treemap for this visualisation. I selected the number of records (fights) from the Measures section and Method from Dimensions and then selected treemap. Tableau then added these fields to the marks card rather than the columns and rows because this graph type doesn’t have them. I gave the graph a title Method of victory.



From the tree map I could easily identify what method of victory is most common and I also I added the number of records from the measures to the marks card to see in detail the number of fights that type of victory was won for example fights have been won by decision 768 times.

The final visualisation I created was based on the number of fights won in a certain round. By organising this data I could then use the results to make confident decisions on future fights and know how what round a fight is most likely to be won in. I gave this visualisation the title Number of fights won in round and I used a highlight table to represent the data. This table has only a column representing the round number (1-5, since there are only 5 rounds in a UFC fight) and has no rows.



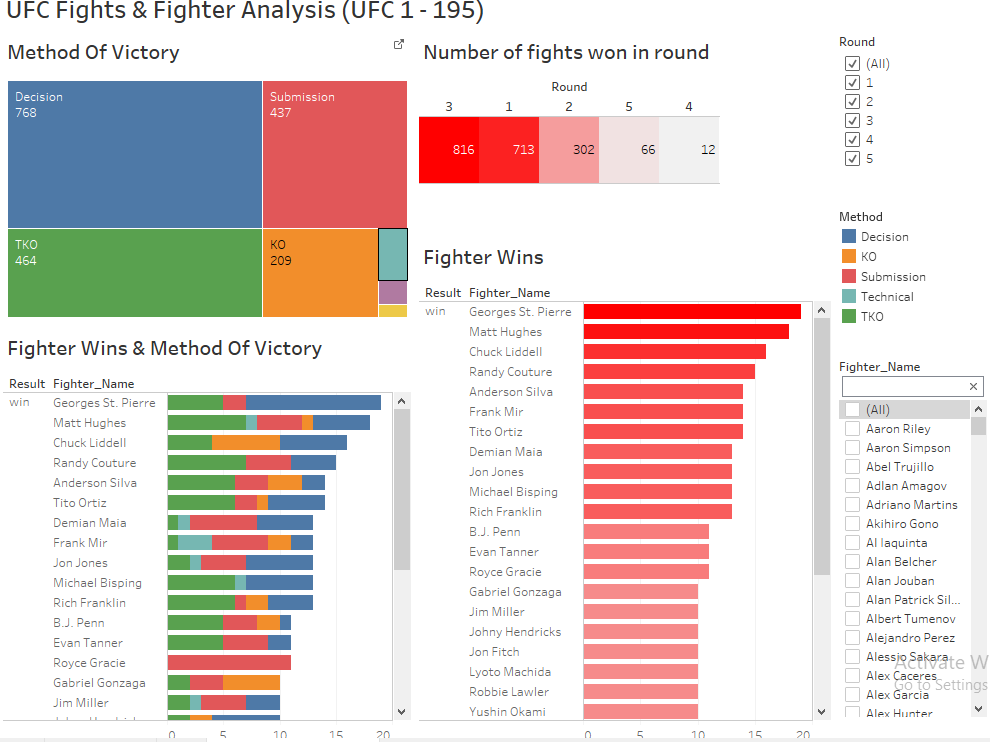
I added the number of records from the measures section to the marks colour card so that the user can see the more fights won in a certain round, the darker/stronger the colour. I also added the number of records to the marks detail card so that the user can see the exact number of times a fight was won in that round. I added a filter to the round so that the user can select what round they wish to see the amount of wins in.



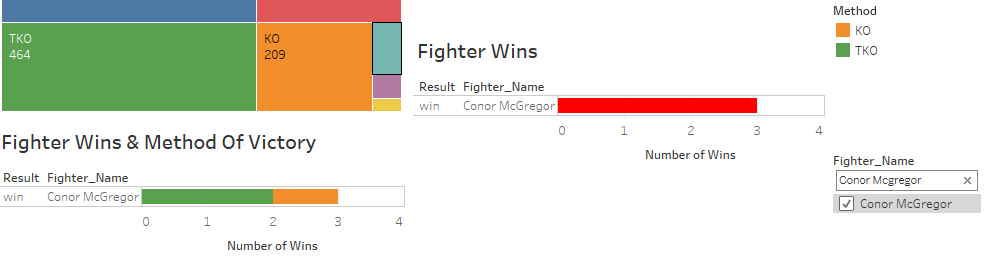


#### Dashboard – UFC Fights & Fighter Analysis

Using the 4 visualisations I created a Dashboard where the user is able to see each visualisation clearly and is also able to interact with each sheet using the filters. I wanted to create a dashboard that could be used by any UFC fan around the world, I wanted it to be fun but professional so that anyone could look at it and gain a better and easy understanding of the UFC, its events and fighters. The user is able to interact with the dashboard by looking at the number of fights won in a certain or all rounds depending on the users choice, the colour of the method of victory can be looked at in more detail by selecting a method type and also the user can select all, or certain fighters or even search for a particular fighter and look at their stats such as number of wins, and there most dominant victory method.



##### Using Filters

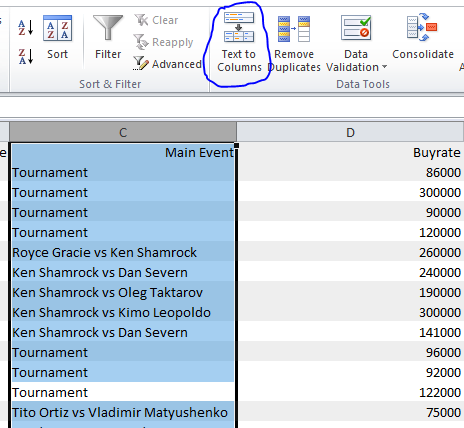


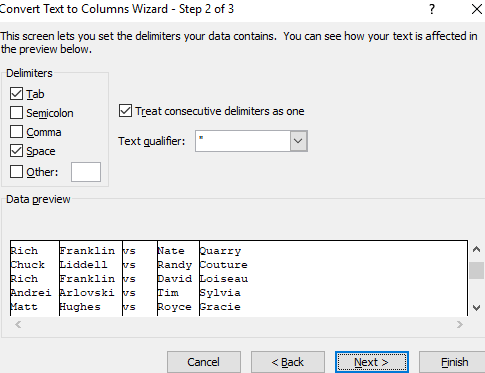
# Seven Stages

When creating visualisations for my next dashboard my goal was to find data sets on UFC Pay Per View (PPV) buys and fighters that are a draw. I also wanted to know why certain countries had a bigger buy rate of PPV buys then other countries and what country has the most fighters with a PPV draw. To do this I acquired 3 data sets online, the first was data on PPV buys from the first UFC event (UFC 1, 1993) to UFC 216 (October 2017). I also acquired another data set on every UFC fighter and used this data to identify the number of fighters from each country. And finally I made another data set on the amount of gyms in each country, the idea here was to see is there a correlation between the number of fighters from a particular country and the number of gyms in a country, one could say the more gyms a country has the more athletes/fighters it will have.

## Parsing, Mining and Filtering the data

For the PPV Buys data set I didn’t filter out any of the data as I needed all rows and columns. When looking at the data I noticed that the Main Event column represented the two fighters that headlined the main event and also represented the amount of PPV buys for that fight, I needed each fighter to be in their own columns so I could link it to my other data set UFC fighters. To do this I selected all the data in excel for the Main Event column and used the split text to columns functionality in Excel.





This spit the test into 4 rows which is not exactly what I wanted. I wanted to split the Main Event column into Fighter Name1 and Fighter Name2.



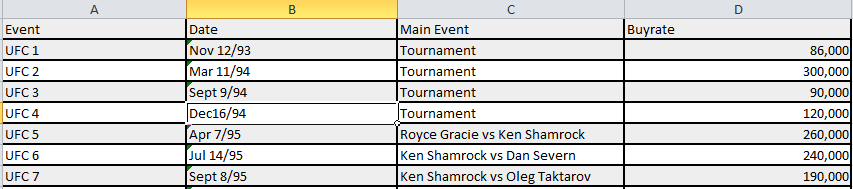
To combine these columns into fighter 1 and fighter 2 I had to concatenate both columns into one. I wrote a function in excel like so,



This turned both columns into,



This is what I wanted and I could now copy this function into all of the other rows and delete the two single columns. The spread sheet previously looked like the following,

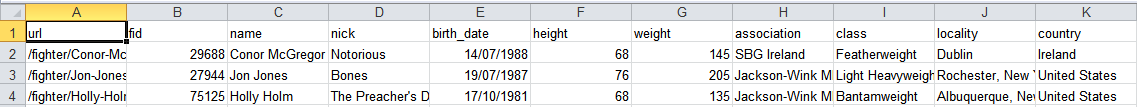


Now looks like,



Now that the fighters were separated I am able to link it to other excel spread sheets to do with UFC fighters.

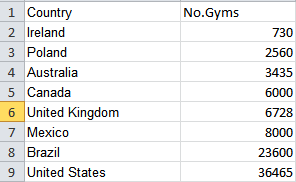
There was a lot of to sort for the UFC fighter’s data set. I filtered out a lot of unnecessary columns and extracted the data I needed for building visualisations off based off my first data set on PPV buys.



I then changed this into, Fighter\_Name and Country since I only need these columns for building my visualisations off.



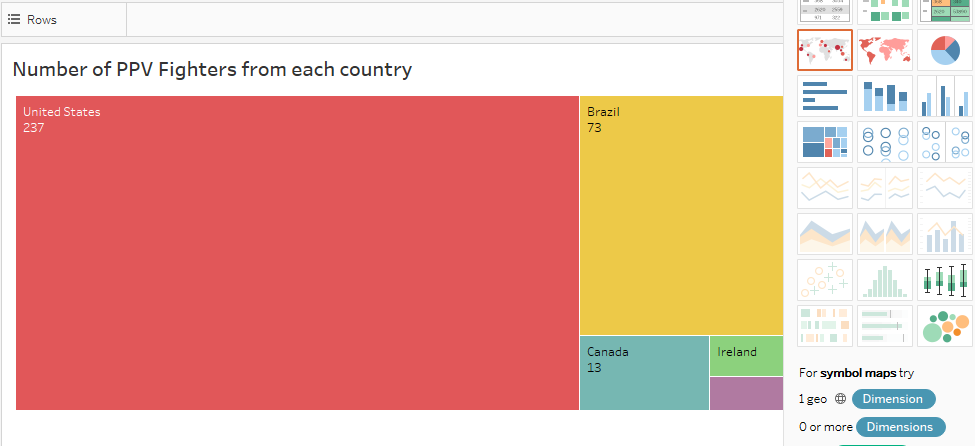
For my final data set on UFC gyms I made this data set myself because I wanted to show how there is greater number of PPV fighters from certain countries due to the fact that there is also a much higher number of gyms in that country, I think that there is a direct correlation between the two. Since there are so many fighters from different countries I decided to find out the number of gyms in each of the countries with fighters that have the biggest PPV draw (USA, UK, Brazil, Ireland, etc.). The data set when structured looked like the following.



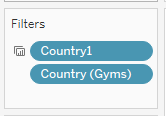
## Represent

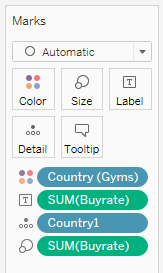
### Fighters and PPV draw in Countries

For the first visualisation I based it on the number of PPV fighters from each country (USA, UK, Brazil, Poland, Australia, and Ireland). The idea behind the visualisation was for the user to see the country with the largest number of PPV fighters. I chose a treemap for this visualisation because I think it shows clearly each countries size in terms of the number of PPV fighters it has.

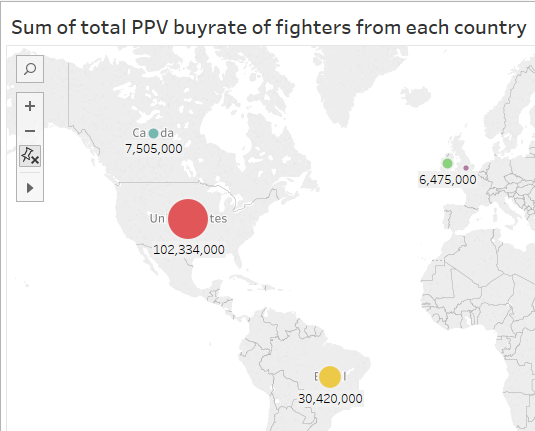


For the second visualisations I wanted to use a map to show the reader the sum of total PPV buy rate of fighters from each country. The map shows the total of all fighters from each country and there PPV rate is added together to give the total. I added a filter for the country as I only wanted to show the countries with the largest PPV totals. I also added colour for each country and a label for the total buy rate to the marks card so that the reader could see the information clearly.

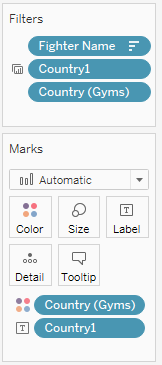




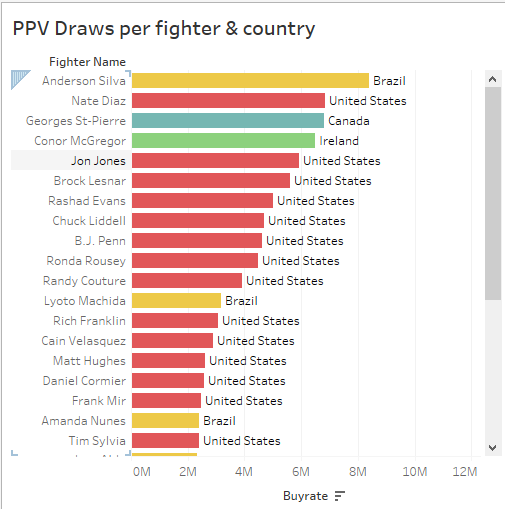
The map when completed looked like the following,



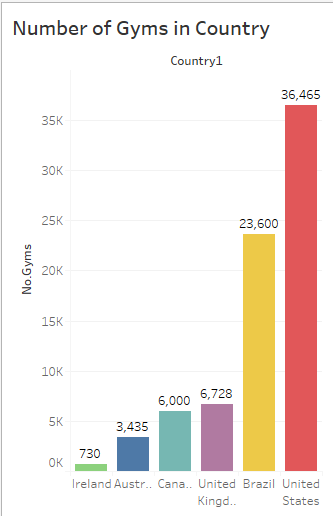
The third visualisation was based on PPV Draws per fighter and displays the country they come from so as to identify the country with the most fighters and PPV draw. For this visualisation I used a horizontal bar chart because I feel it displays to the user clearly the fighter name and their PPV draw number i.e. buy rate. Like the second visualisation I used a filter to only show the top countries and their fighters. I also used colour and a label on the marks card to show clearly what country that fighter is from.



The visualisation when completed looked like the following,

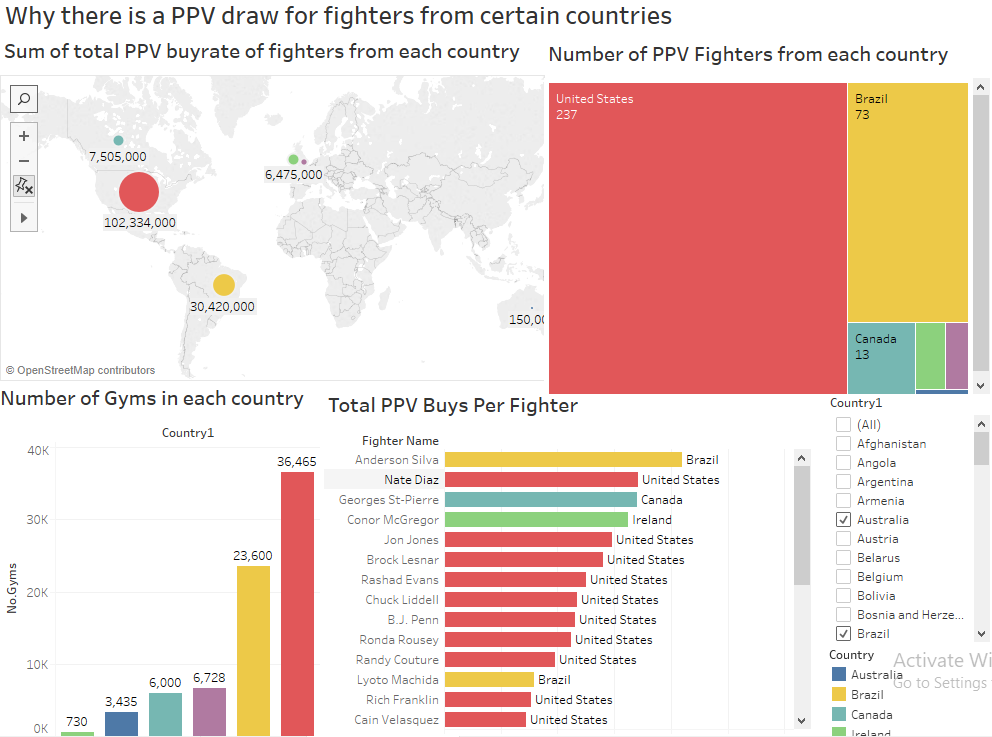


For the final visualisation I wanted to show the number of gyms in each of the top countries so as to show to the reader that there is a correlation between the number of fighters from a country because there is such a high number of gyms in that country, i.e. producing more athletes/fighters. I used a filter and colour to show each country and also a label so that the reader could see exact figures. The chart when finished looked like the following,

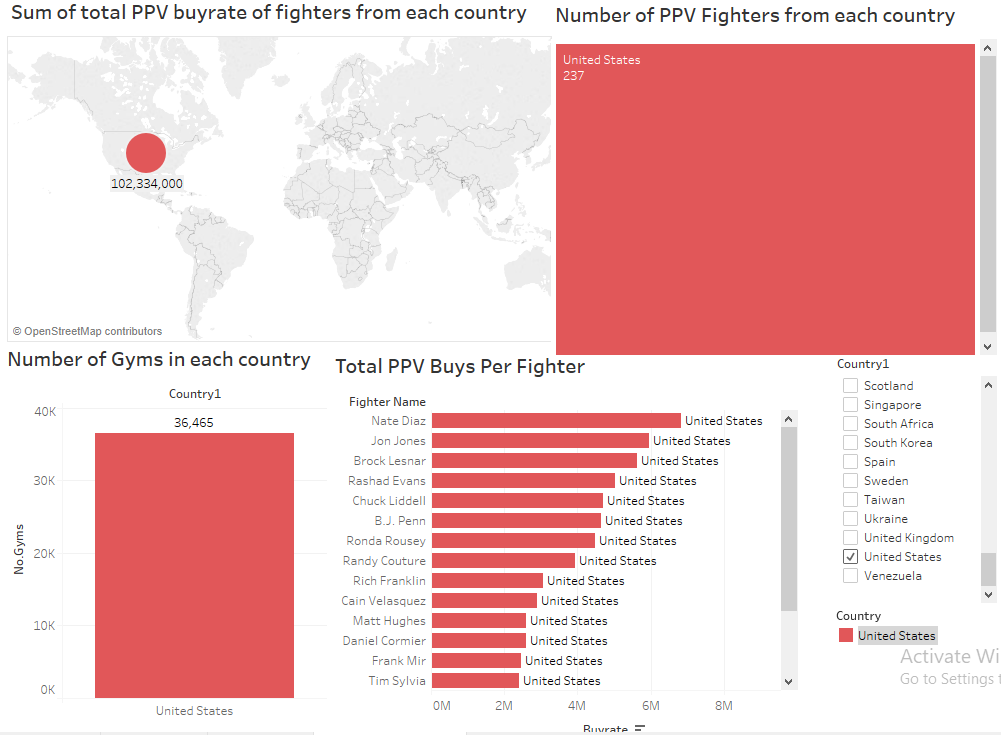


#### Dashboard – Why there is a PPV draw for fighters from certain countries

For the dashboard I wanted to use each visualisation to show to the user why there is PPV draw for fighters from certain countries. From the dashboard the reader can clearly see most fighters come from the United States because there are the most gyms in that country, resulting in a higher number of PPV buys for fighters from the US. The dashboard can be filtered and interacted with by the user to show different stats for each country.



##### Using Filter and Colour



# Problems & Solutions

While creating both visualisations I ran into a number of problems. It was tough to identify a solution to these problems because a lot of them came down to the data sets which I did not create, I acquired them online.

The first problem I ran into was with the first dashboard, my fourth visualisation on number of fights won in a round from rounds 1 to 5. The UFC was set up in 1993, when it first began there were no rounds meaning every fight finished in round 1. Up until UFC 21 every fight finished in round 1 even if it went on longer than 5 minutes (the time limit per round). When rounds were eventually implemented at UFC 21 then fights finished in different rounds. Also the UFC only implemented 5 round fights for all main events even if was not for a championship belt only after UFC 133. What this means in both situation is that not all of the data is in fact true because if every main event from UFC 1 was 5 rounds long then the data would look different, we would probably see less fights won in round 3 and more fights won in round 5. Since this cannot be proven to be fact I would be confident in the presentation of this visualisation.

Another problem with my first dashboard was with both bar charts. Each bar chart shows every single UFC fighter on the roster, meaning hundreds of rows. It is impossible for the reader to see every fighter on the bar chart so my solution to this problem was to only show the top around 40 fighters with the most wins but to also allow the user to use the filter options to pick specific fighters themselves.

The first problem I had with my second dashboard came in trying to use three different data sets together to try and build visualisations off. When using three different data sets together in tableau a lot of the time lots of the data is filtered out because it’s not possible to have all rows and columns of every data set in the same worksheet. I wanted to show data on every UFC fighter but my PPV buys data set only had the fighter name that headlined the event and not every fighter that fought on the card. A solution to this problem was to only show the fighters from each country that were the PPV headline.

# Conclusion

Throughout the CA I have made many discoveries and have gone on a journey in UFC data analysis that I have really enjoyed. I find that I am now interested in data mining and analysis and building dashboards using Tableau. I could definitely see myself in the future working in data analysis or information management. I have now learned to use tableau confidently which I feel will definitely help me in the future and when I graduate and am working next year my data analysis skills will have increased due to this assignment.

My dashboards tell interesting stories on the UFC and my goal was to tell a story on the UFC and show that its data can not only be used by UFC fans but also by the company to look at and make future decisions based off the visualisations. In terms of each of my dashboards I made a number of discoveries.

When building my first dashboard my goal was to gather data on UFC events and fights and take that data and build visualisations about fighters, methods of victory, rounds fights were won in, etc. My first discovery was finding out that most fights were won by decision and also that most fights have been won in round 3. Using this data I could bet against my friends or other UFC fans how a fight will be won and what round using my visualisations as support for my decisions.

I also found out a lot of interesting discoveries about some of my favourite fighters such as one of my favourite fighters George St-Pierre has the most wins in the UFC with 19 wins, and Royce Gracie another one of my favourite fighters has won all of his 11 fights by submission. My goal in creating these visualisations was to find out some interesting facts about fighters.

For the second dashboard I made a lot of discoveries to do with PPV buys and UFC fighters from each of the top countries (USA, UK, Brazil, Canada, Ireland, and Australia). I found out that USA has the biggest PPV buys due to the fact that most UFC fighters come from the USA. One big discovery was finding out that there is a relation between the amount of fighters/athletes from a country and the amount of Gyms in that country, the fewer gyms a country has the fewer fighters from that country and the more gyms, the more fighters.

Another interesting discovery was finding out that Ireland has the 4th biggest PPV draw rate (with over 6 million PPV buys) despite its low number of fighters. This is due to the fact that Conor McGregor is such a global star that he on his own draws in more PPV buys than a lot of countries combined together such as the United Kingdom.

These visualisations will allow me to make decisions on the UFC in the future and has given me a greater insight into the sport. There is not a lot of data analysis done in this area and through searching online I found that by creating visualisations on the UFC and its events and fighters I have created research in areas that hasn’t been done yet. I couldn’t find any analysis on countries and there combined PPV buys so by creating visualisations on this area I have created dashboards that could be used by the UFC in making future decisions on planning where to hold UFC events and also what fighters to promote. They could also look at this data to see where they need to improve in terms of finding more fighters, and increasing their promotion in different countries.

My recommendation for future analysis in this area would be to do a lot more analysis and data mining and identifying patterns before creating visualisations. I think it’s more important to know what your goals are first rather than jumping straight into the visualisations. I am now confident in using Tableau and data analysis and have enjoyed completing this assessment.