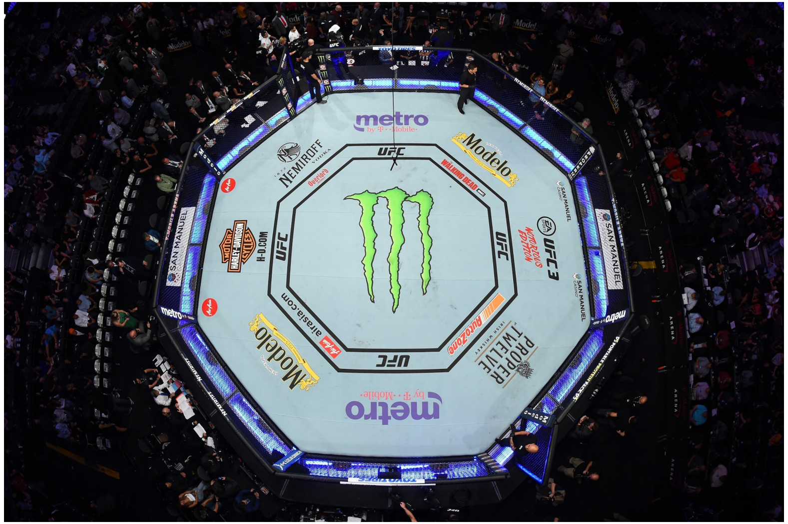
Fighter Analysis

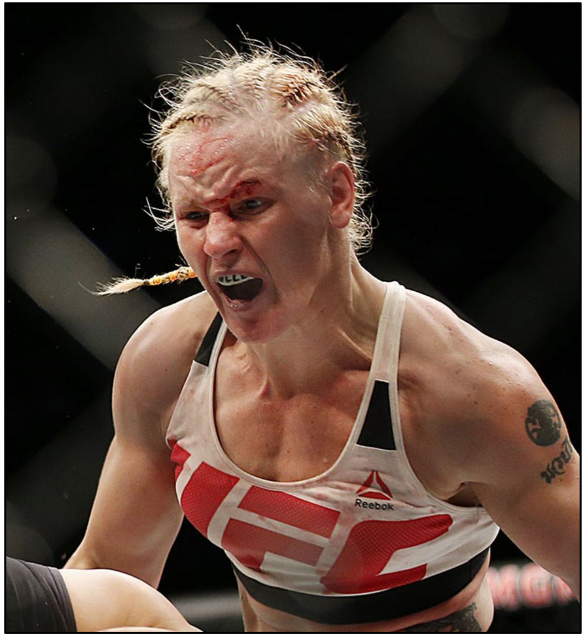
Predicting Match Winners

Overview of Project

Our group members are UFC and MMA fans and would like to use the skills we learned in the course to examine fighting techniques to determine which have the most effect the win in a match.



Mixed martial arts (MMA) is a full-contact combat sport based on striking, grappling and ground fighting, incorporating techniques from various combat sports and martial arts from around the world.



Ultimate Fighting Championship (UFC) is a Las Vegas based promotion company that has revolutionized the fighting business since 1993. UFC features some of the highest-level fighters in the sport on its roster and produces events worldwide that showcase twelve weight divisions (eight men's divisions and four women's divisions). As of 2020, the UFC has held over 500 events and grown into a globally popular multi-billion-dollar enterprise.

Purpose of Analysis

Using a Kaggle dataset containing various attributes of UFC fighter stats, fighting techniques and body metrics, we will predict winning fighters with machine learning.

Our CSV file is small (23 columns and 8.990 rows) but complete as it contains roughly every match under the UFC umbrella.

- A Git Hub repository was created for the analysis so everyone in the group can contribute and review information.
- The group will meet twice a week during our scheduled class sessions on Zoom to work on the project and use our team Slack channel to communicate during the week.

Role Distribution

After establishing the communication structure, we created the foundation for our UFC fighter analysis project by defining roles that play to our individual strengths.

	Segment 1	**Segment 2**	**Segment 3**	**Segment 4**
Square	Mohammed	Alexandra	Mohammed	Felicia
Triangle	Alexandra	Mohammed	Felicia	Oybek
Circle	Oybek	Oybek	Oybek	Alexandra
X	Felicia	Felicia	Alexandra	Mohammed

Machine Learning Model

The Dataset:

- **Source:** Kaggle dataset containing UFC fighter winners/losers and their stats (weight, height, stance, technique, odds, gender, finish, etc.)
- **CSV file:** 23 columns and 8,990 rows
- **Model Goal:** use fighter metrics to predict win or loss (winner_total column)

	match_id	fighter_total	odds_total	ev_total	date_total	location_total	country_total	winner_total	weight_class_total	gender_total
0	1	Brandon Vera	215	215.000000	2010-03-21	Broomfield, Colorado, USA	USA	0	Light Heavyweight	MALE
1	2	Junior Dos Santos	-250	40.000000	2010-03-21	Broomfield, Colorado, USA	USA	1	Heavyweight	MALE
2	3	Cheick Kongo	-345	28.985507	2010-03-21	Broomfield, Colorado, USA	USA	1	Heavyweight	MALE
3	4	Alessio Sakara	-120	83.333333	2010-03-21	Broomfield, Colorado, USA	USA	1	Middleweight	MALE
4	5	Clay Guida	-420	23.809524	2010-03-21	Broomfield, Colorado, USA	USA	1	Lightweight	MALE

Machine Learning Model

Step 1: Pre-Processing the Model

- **Data Cleaning** - dropped unnecessary columns (fighter name, date of match, and location of match), dropped rows with NaN (754 total rows)
- **Data Encoding** - used OneHotEncoder to encode and read the data into the model

Step 2: Feature Engineering and Feature Selection

- **Define and split the data into training and testing sets -**
  - **y,** or the target variable indicates whether or not the fighter won the match and
  - **X,** or the independent variables are the metrics that the model uses to predict whether the fighter would win (i.e. gender, class, weight, reach height or odds)

- **Random Forest Model -** We chose the Random Forest Classification Model. Random Forest is our preferred modeling tool because it:
  - Runs efficiently on large data sets
  - Works against overfitting
  - Can be used to rank input variables

Machine Learning Model

Step 3: Model Results

- We performed an exploratory analysis and established a baseline accuracy score of 64%

- We decided to use feature selection to find the best attributes to explain the relationship between a fighter's characteristics and winning matches

- A linear regression model helped us identify which variables were most significant.
    - This removed the noise in our model but the accuracy didn't improve - maintained 63% accuracy level
    - We were able to improve the false negatives and positives - making our model more precise
    - However, there wasn't enough variation in our dataset to determine which UFC fighter would win the match

```
Accuracy Score : 0.6348958333333333
Classification Report
              precision    recall  f1-score   support

           0       0.62      0.64      0.63       926
           1       0.65      0.63      0.64       994

    accuracy                           0.63      1920
   macro avg       0.63      0.64      0.63      1920
weighted avg       0.64      0.63      0.64      1920
```

Entity Relationship Diagram (ERD)

Entity Relationship Diagram. First we created a Entity Relationship Diagram and includes two tables, UFC Dataset and Mastertable_Text, and is joined using Match ID.

```
//// -- LEVEL 1
//// -- Tables and References

Table UFC_DATASET {
  Match_ID int [pk]
  no_of_rounds_total INT
  total_draw INT
  total_losses INT
  total_win_by_KO_TKO INT
  total_win_by_submission INT
  total_win_by_TKO_Doctor_stoppage INT
  total_wins INT
  total_stance VARCHAR
  total_height FLOAT
  total_reach FLOAT
  total_weight INT
  total_age INT
  finish VARCHAR
  }

Table Mastertable_Text {
  Match_ID INT [pk]
  Fighter_total VARCHAR
  odds_total INT
  ev_total float
  date_total date
  location_total VARCHAR
  country_total VARCHAR
  winner_total INT
  weight_class_total VARCHAR
  gender_total VARCHAR
  }



// Creating references
// You can also define relaionship separately
// > many-to-one; < one-to-many; - one-to-one
Ref: UFC_DATASET.Match_ID > Mastertable_Text.Match_
```

Postgres (SQL)

I Created a UFC Database in SQL and connected it with our Notebook in order to use a modified database. I joined two tables (UFC Dataset and Mastertable_Text) and created a final joined table. With that table i made two data outputs. One to show our audience the fighter and their respective countries weight class height and more. The other data i made is to show the wins and losses.

Fighter totals

Data Output

	fighter_total character varying	gender_total character varying	weight_class_total character varying	country_total character varying	total_height double precision	total_weight integer	total_age integer
1	Brandon Vera	MALE	Light Heavyweight	USA	190.5	230	32
2	Junior Dos Santos	MALE	Heavyweight	USA	193.04	238	26
3	Cheick Kongo	MALE	Heavyweight	USA	193.04	240	34
4	Alessio Sakara	MALE	Middleweight	USA	182.88	185	28
5	Clay Guida	MALE	Lightweight	USA	170.18	155	28
6	Eliot Marshall	MALE	Light Heavyweight	USA	187.96	205	29
7	Duane Ludwig	MALE	Lightweight	USA	177.8	170	31
8	John Howard	MALE	Welterweight	USA	170.18	170	27
9	Brendan Schaub	MALE	Heavyweight	USA	193.04	245	27
10	Mike Pierce	MALE	Welterweight	USA	172.72	170	29
11	Eric Schafer	MALE	Light Heavyweight	USA	190.5	185	32
12	Georges St-Pierre	MALE	Welterweight	USA	180.34	185	28
13	Frank Mir	MALE	Heavyweight	USA	190.5	264	30
14	Kurt Pellegrino	MALE	Lightweight	USA	172.72	155	30
15	Ben Saunders	MALE	Welterweight	USA	187.96	170	26
16	Jim Miller	MALE	Lightweight	USA	172.72	155	26
17	Nate Diaz	MALE	Welterweight	USA	182.88	170	24
18	Ricardo Almeida	MALE	Welterweight	USA	182.88	170	33
19	Rousimar Palhares	MALE	Middleweight	USA	172.72	170	30
20	Rodney Wallace	MALE	Light Heavyweight	USA	175.26	205	28

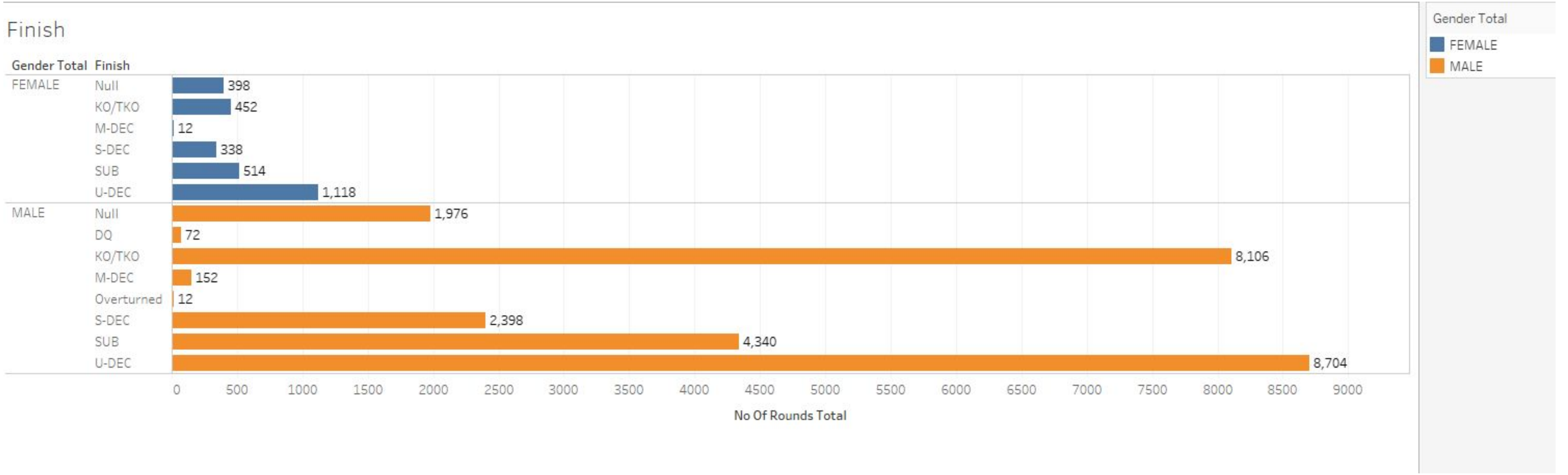
Fighter Wins and Losses

Explain Query History **Data Output** Messages

	fighter_total character varying	total_wins integer	total_losses integer	total_draw integer
1	Brandon Vera	7	4	0
2	Junior Dos Santos	4	0	0
3	Cheick Kongo	7	4	0
4	Alessio Sakara	5	5	0
5	Clay Guida	5	5	0
6	Eliot Marshall	3	0	0
7	Duane Ludwig	2	1	0
8	John Howard	3	0	0
9	Brendan Schaub	0	1	0
10	Mike Pierce	1	1	0
11	Eric Schafer	3	3	0
12	Georges St-Pierre	13	2	0
13	Frank Mir	11	4	0
14	Kurt Pellegrino	6	3	0
15	Ben Saunders	4	1	0
16	Jim Miller	5	1	0
17	Nate Diaz	6	3	0
18	Ricardo Almeida	4	3	0
19	Rousimar Palhares	3	1	0
20	Rodney Wallace	0	1	0

Dashboard / Visualization

KO/TKO - Knockout / Technical Knockout
DQ - Disqualified
M-DEC - Majority decision
Overturned - Result of a fight can be overturned by the athletic commission
Null - No cotest
S-DEC - Split Decision
SUB - A submission is a combat sports term for yielding to the opponent
U-DEC - Unanimous Decision



Dashboard / Visualization

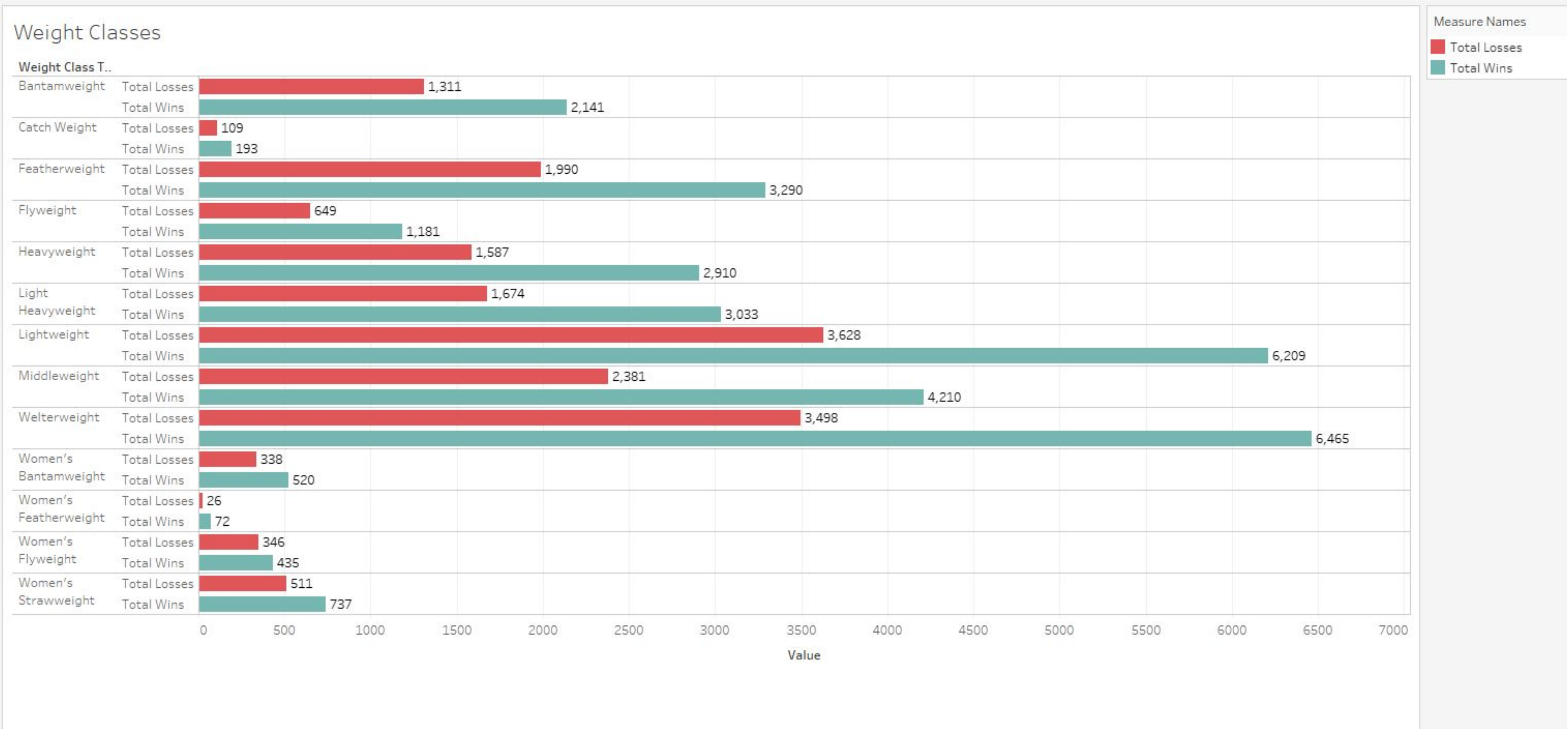
MMA Weight Classes for Men

Class	Weight Range
Flyweight	Up to 105 lbs
Super flyweight	105.1–115 lbs
Bantamweight	115.1–125 lbs
Super bantamweight	125.1–135 lbs
Featherweight	135.1–145 lbs
Lightweight	145.1–155 lbs
Super lightweight	155.1–165 lbs
Welterweight	165.1–175 lbs
Super welterweight	175.1–185 lbs
Middleweight	185.1–195 lbs
Super middleweight	195.1–205 lbs
Light heavyweight	205.1–225 lbs
Heavyweight	225.1–265 lbs
Super heavyweight	Over 265 lbs

MMA Weight Classes for Women

Class	Weight Range
Flyweight	Up to 95 lbs
Bantamweight	95.1–105 lbs
Featherweight	105.1–115 lbs
Lightweight	115.1–125 lbs
Welterweight	125.1–135 lbs
Middleweight	135.1–145 lbs
Light heavyweight	145.1–155 lbs
Cruiserweight	155.1–165 lbs
Heavyweight	165.1–185 lbs
Super heavyweight	Over 185 lbs

Dashboard / Visualization



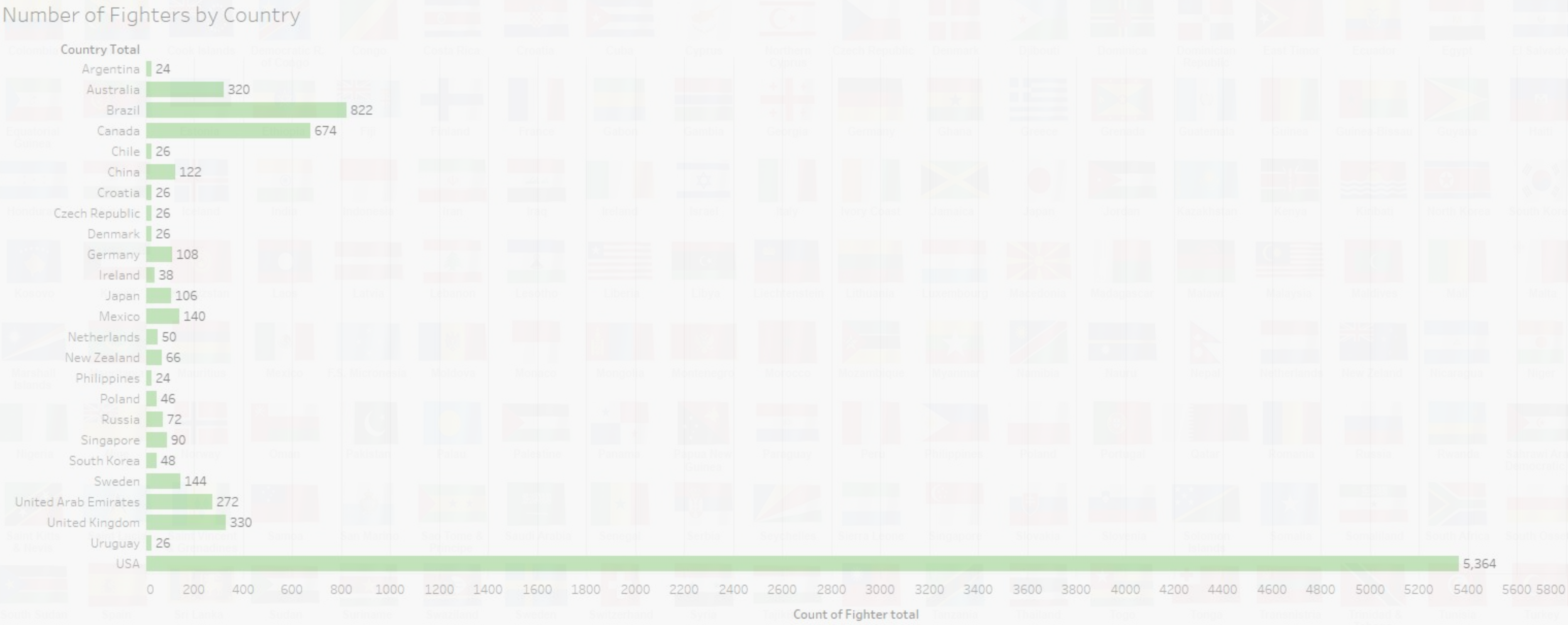
Dashboard / Visualization

MMA no-no's in fighting

Although every MMA fighting organization has its own specific rules, some universal no-no's do exist. They're listed in the *Unified Rules of MMA*, but here's a quick look at what's not allowed:

- No groin attacks.
- No knees to the head on a grounded opponent.
- No strikes to the back of the head or the spine.
- No head butts. (Sorry, soccer fans.)
- No eye gouging.
- No fish hooking.
- No fingers in an opponent's orifices. (Eww!)
- No biting.
- No hair pulling. (Besides, that's so second grade.)
- No strikes or grabbing of the throat.
- No manipulation of the fingers or toes.
- No intentional grabbing of the ring or cage.
- No intentional throwing of your opponent outside of the ring or cage. (That stuff belongs in professional wrestling.)

Dashboard / Visualization



Summary

- We created the foundation for our UFC fighter analysis project by defining roles that play to our individual strengths and establishing the communication structure. We preprocessed our data for easier encoding and modeling.
- We performed an exploratory analysis to establish a baseline accuracy score, created a database and used a linear regression to fine tune our model's accuracy.
- While we were able to remove the noise and improve the model's precision, we didn't have enough data to explain the variance so we used our dashboard to visualize interesting observations from the dataset such as countries with the most winning fighters, winning stances and winning finishes.