UFC Data Set

Collection: The information was gathered and extracted from the UFCStats website, so it should be quite reliable.

Contents: Included below is a comprehensive compilation of all UFC fights throughout the organization's history. Each row presents details about the fighters, fight specifics, and the ultimate winner. The Kaggle author's source comprises raw data files, preprocessed data, and a conclusive merged dataset. Initially, the final merged dataset featured 144 columns. However, I narrowed down the contents significantly.

Q 8) Limitations: An external entity conducted the data scraping. While the origin of the scraping is the UFC site, a thorough review of columns and pertinent data cleaning is imperative to ensure complete accuracy and proper formatting before delving into further analysis. I am inclined to believe that bias is minimal since the calculations are purely statistical and derived from actual fights. However, there remains a potential concern for human-made errors during the translation and interpretation of results, including variations in the analysts' level of experience.

Source: This data originates from Kaggle, a community comprising data scientists and developers. Kaggle stands out as the world's largest data science community, boasting over one million registered users. Within this platform, members engage in discussions about model development, explore various datasets, and network with individuals from 194 different countries across the globe.

Reason For Picking: I don't have a strong interest in UFC but I think sports analysis is interesting and I wanted to do something out of my comfort zone since we already did a project on games and healthcare. I am also confident that this dataset aligns with the essential project requirements, as it is open source, incorporates a geospatial component, and satisfies the specified size and variable criteria. It was also in the CareerFoundry brief where it was discovered.

Q 7)Understanding Your Data: Check the strips folder where it says question 7 for a full stats

	B_wins	B_losses	B_win_by_Decision_Majority	B_win_by_Decision_Split	B_win_by_Decision_Unanimous	B_win_by_KO/TKO	B_win_by_Submissio
count	6012.000000	6012.000000	6012.000000	6012.000000	6012.000000	6012.000000	6012.00000
mean	2.584997	1.540419	0.016134	0.218397	0.838157	0.895709	0.56952
std	3.317241	1.940694	0.127316	0.525232	1.370454	1.580829	1.16125
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
50%	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.00000
75%	4.000000	2.000000	0.000000	0.000000	1.000000	1.000000	1.00000
max	23.000000	16.000000	2.000000	5.000000	10.000000	11.000000	14.00000

8 rows × 24 columns

```
R_fighter                      object
B_fighter                      object
fight_date                     object
location                       object
Winner                         object
weight_class                   object
B_wins                          int64
B_losses                        int64
B_win_by_Decision_Majority      int64
B_win_by_Decision_Split         int64
B_win_by_Decision_Unanimous     int64
B_win_by_KO/TKO                 int64
B_win_by_Submission             int64
B_win_by_TKO_Doctor_Stoppage    int64
B_Stance                       object
B_Height_cms                  float64
B_Reach_cms                   float64
B_Weight_lbs                  float64
R_wins                          int64
R_losses                        int64
R_win_by_Decision_Majority      int64
R_win_by_Decision_Split         int64
R_win_by_Decision_Unanimous     int64
R_win_by_KO/TKO                 int64
R_win_by_Submission             int64
R_win_by_TKO_Doctor_Stoppage    int64
R_Stance                       object
R_Height_cms                  float64
R_Reach_cms                   float64
R_Weight_lbs                  float64
B_age                         float64
R_age                         float64
dtype: object
```

Q 10)

1) Which fighting stance has the highest win rate?
2) What is the most common way of victory?
3) Is Red or Blue side more likely to win?
4) What was the most popular year for fighting?
5) What was the most popular weight class?
6) What's the difference in wins form the tallest and smallest
7) What's the difference in wins from a heavy to light person
8) What factors tend to lead to the highest success rate
9) Which fighters maintained the highest win rate over many games
10) Which fighters have the highest number of knockouts?
11) Does the highest knockout fighter also have the highest number of wins?