**UFC Fighters Statistics: Exploratory Data Analysis (EDA)**

**Introduction**

* **Overview**: This project conducts an exploratory data analysis (EDA) of UFC fighters' statistics to uncover trends, patterns, and insights that can inform strategies within the sport. The analysis includes data preprocessing, visualization, and statistical analysis to provide a comprehensive understanding of the dataset.
* **Python Libraries Used**:
  + **numpy**: Numerical computations.
  + **pandas**: Data manipulation and analysis.
  + **matplotlib.pyplot** & **seaborn**: Data visualization (static, animated, interactive).
  + **requests**: Handling HTTP requests, commonly used for web scraping and API interaction.
  + **BeautifulSoup**: Parsing HTML and XML documents for web scraping.

**Dataset Overview**

* **Primary Dataset**: Contains statistics of UFC fighters, including physical attributes (height, weight, reach), fight records, and demographics.
* **Supplementary Dataset**: Provides gender-specific information for targeted analysis (male vs. female fighters).

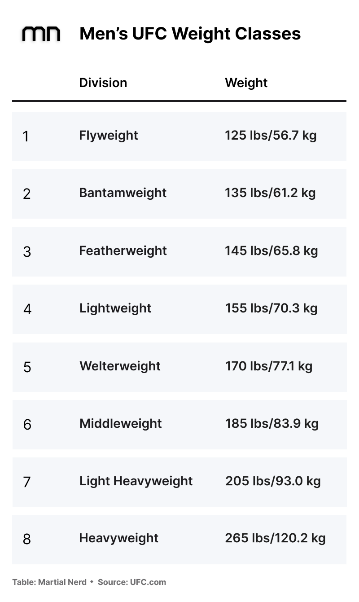
**Data Inspection and Preparation**

1. **Initial Inspection**:
   * **Commands Used**:
     + df.head(), df.info(), and df.describe() were employed to inspect the dataset structure, check for null values, data types, and obtain descriptive statistics.
2. **Handling Missing Values**:
   * Missing values were addressed either by removing irrelevant entries or imputing values where necessary to avoid skewing results.
3. **Data Cleaning**:
   * **Duplicates**: Removed duplicate records to ensure data integrity.
   * **Female Fighters**: Female fighters were excluded from certain analyses when focusing on male competitors to maintain relevance and consistency in the study.
4. **Date of Birth to Age Conversion**:
   * **Rationale**: Age is more practical for analysis than the date of birth. It helps in comparing fighter performance and making assessments across age groups.
   * **Transformation**: The date of birth was converted into age, which enhances the dataset’s utility in evaluating fighter performance.

**Feature Engineering**

To extract deeper insights, additional performance metrics were introduced:

1. **Performance Metrics Enhancement**:
   * **Total Matches Played**: Calculated by summing the wins, losses, and draws.
   * **Win and Loss Percentages**: To evaluate performance efficiency:
     + Win % = (wins / matches\_played) \* 100
     + Loss % = (losses / matches\_played) \* 100
2. **Weight Class Assignment**:
   * **Weight Rounding**: Weight values were rounded to one decimal place for consistency.
   * **Weight Class Mapping**: Fighters were categorized into specific weight classes based on their weight in kilograms.



1. **BMI Calculation**:
   * **Formula**: BMI=Weight (kg)(Height (m))2BMI = \frac{\text{Weight (kg)}}{(\text{Height (m)})^2}BMI=(Height (m))2Weight (kg)​
   * **Purpose**: The Body Mass Index (BMI) provides insights into the relationship between a fighter’s physical attributes (weight and height) and their overall fitness, which may influence their performance.

**Analyzing Age and Weight Demographics with Boxplots**

1. **Age Analysis**:
   * A boxplot of the age column was created to identify the distribution of fighter ages and highlight potential outliers.
   * **Observation**: The data shows that most fighters fall within a particular age range, with some notable outliers.
   * **Outliers**: Fighters older than 80 years are considered outliers. The analysis notes that Andrei Arlovski, at 45 years and 4 months, is the oldest active UFC fighter, with many fighters over 45 being inactive or retired.
2. **Weight Analysis**:
   * A boxplot of the weight\_in\_kg column was generated to show the distribution of fighters' weights.
   * **Observation**: The analysis highlights the distribution of weights among fighters.
   * **Outliers**: Fighters exceeding 120.2 kg are classified as outliers, consistent with the UFC’s heavyweight division limit (93-120 kg).

**Filtered Dataset**

* **Filtered Dataset**:
  + Fighters with weights under 120.2 kg were retained to focus the analysis on active fighters and avoid skewing results with heavyweight outliers.
* **Dataset Preservation**:
  + The filtered dataset was saved as a CSV file for future analyses. This ensures data preservation and easy access for subsequent examinations.

**Distribution Analysis**

**Physical Attribute Distribution Analysis**

**Objective:**

Analyze the distribution of UFC fighters' physical attributes: **age, height (cm), weight (kg),** and **reach (cm)**.

**Visualizations:**

* **Age Distribution**: Displays the frequency of different age groups among fighters.
* **Height Distribution**: Shows how fighter heights are distributed.
* **Weight Distribution**: Highlights the range of weights among fighters.
* **Reach Distribution**: Visualizes the variation in fighters' reach.

**2. Outcome and Match Participation Distribution**

**Objective:**

Examine the distribution of UFC fighters' fight outcomes, including **wins, losses, draws,** and **matches played**.

**Visualizations:**

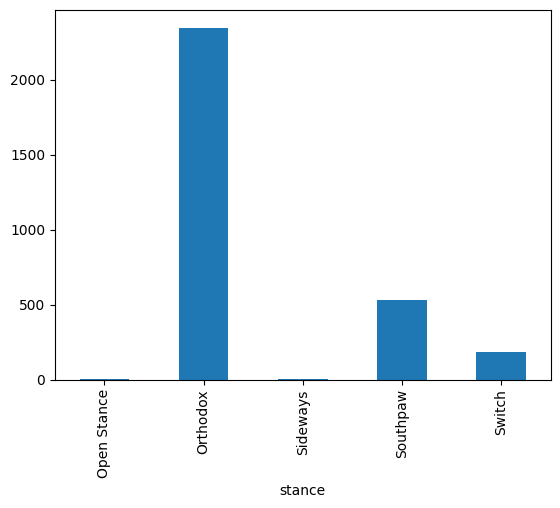
* **Win Distribution**: Displays how often fighters win across the dataset.
* **Loss Distribution**: Shows the distribution of losses.
* **Draw Distribution**: Visualizes the frequency of draws.
* **Total Matches Fought Distribution**: Highlights the total number of matches played by fighters.

**3. Fighter Stance Distribution**

**Objective:**

Analyze the distribution of fighters based on their fighting stance.

**Visualization:**



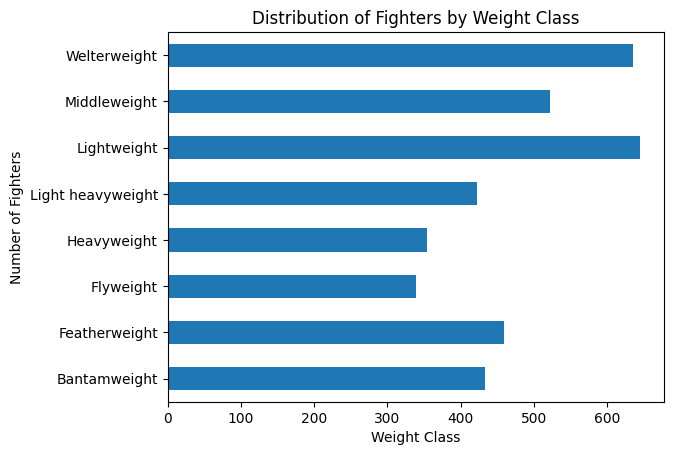
* **Orthodox** stance is by far the most popular among UFC fighters, with over 2,000 fighters adopting it.
* **Southpaw** is the second most common stance, with a significantly smaller number (approximately 500 fighters).

**4. Fighter Distribution by Weight Class**

**Objective:**

Examine how fighters are distributed across different UFC weight classes.

**Visualization:**



* **Welterweight** has the highest number of fighters, with over 600 fighters.
* **Heavyweight** and **Flyweight** divisions have fewer fighters, with Heavyweight around 300 and Flyweight just under 200 fighters.

**Correlation Analysis**

**Objective:**

To explore correlations between UFC fighters' physical attributes, performance metrics, and fight outcomes. The analysis includes calculating and visualizing the correlation matrices to uncover patterns and relationships.

**Physical Attributes:**

* Age
* Height (cm)
* Weight (kg)

**Performance Metrics:**

* Significant Strikes Landed per Minute
* Average Takedowns Landed per 15 Minutes
* Average Submissions Attempted per 15 Minutes

**Fight Outcomes:**

* Wins
* Losses
* Draws

**Physical Attributes Correlation Plot**

The correlation matrix was generated for physical attributes (age, height\_cm, weight\_in\_kg) and fight outcomes (wins, losses, draws) to investigate potential relationships.

**Insights:**

* The heatmap visualizes the correlations between fighters' physical characteristics and their fight results.
* The darker areas represent stronger correlations, while lighter areas show weaker or no correlation.
* Observations like a positive correlation between weight and wins or age and losses can guide further analysis.

**Performance Metrics Correlation Plot**

Similarly, the performance metrics (significant strikes landed per minute, average takedowns landed per 15 minutes, and average submissions attempted per 15 minutes) were correlated with fight outcomes (wins, losses, draws) to study how different aspects of performance influence results.

**Insights:**

* The heatmap reveals how specific fight metrics (striking, takedowns, submissions) relate to win/loss rates.
* Higher significant strikes per minute may show a strong correlation with wins, indicating the effectiveness of striking in UFC victories.

**Visualization: Comparison of Fighters' Average Wins by Stance**

Fighters were grouped by stance, and their average wins were compared across different stances (Orthodox, Southpaw, etc.) using a bar plot.

**Insights:**

* Orthodox stance is the most common, with a large proportion of fighters adopting it.
* Analyzing the average wins by stance offers insights into whether a particular fighting style is more advantageous.

**Scatter Plot: Relationship Between Weight and Wins Across Weight Classes**

To further explore how fighter weight affects performance, a scatter plot was created for each weight class, displaying the relationship between weight and wins.

**Insights:**

* The red vertical line indicates the average weight within each weight class.
* Correlation values between weight and wins help determine whether heavier or lighter fighters within a weight class perform better.

**Scatter Plot: Relationship Between Height and Wins Across Weight Classes**

Similarly, the relationship between a fighter's height and their wins was visualized with scatter plots for each weight class.

**Insights:**

Similar to weight, the correlation between height and wins can reveal patterns, such as whether taller fighters in a division have an advantage

**Striking Analysis Documentation**

**Objective:**

The goal of this analysis is to identify the top strikers in each weight class based on key striking metrics. These metrics include significant strikes landed per minute, striking accuracy, strikes absorbed per minute, and striking defense. Additionally, scatter plots are used to explore the relationship between these metrics and fight outcomes such as wins.

**Key Metrics:**

* **Significant Strikes Landed per Minute:** Number of significant strikes a fighter lands in one minute.
* **Significant Striking Accuracy (%):** The percentage of significant strikes landed successfully.
* **Significant Strikes Absorbed per Minute:** Number of significant strikes a fighter absorbs per minute.
* **Significant Strike Defense (%):** The percentage of significant strikes a fighter successfully defends.

**Scatter Plots for Striking Metrics vs Wins**

The scatter plots visualize the relationship between significant striking metrics and the number of wins. Each metric is plotted individually against the number of wins.

**Insights:**

* **Significant Strikes Landed per Minute:** Fighters who land more significant strikes per minute generally have higher win counts, suggesting aggressive striking leads to better outcomes.
* **Striking Accuracy:** Accurate strikers tend to win more matches, highlighting the importance of precision over volume.
* **Strikes Absorbed per Minute:** Higher absorption of strikes tends to correlate with fewer wins, indicating the defensive vulnerabilities of fighters.
* **Strike Defense:** A higher percentage of strikes defended correlates positively with the number of wins, showing the importance of strong defense in MMA success.

**Wrestling Analysis**

**Key Metrics:**

* **Average Takedowns Landed per 15 Minutes:** The number of successful takedowns a fighter lands in a 15-minute fight.
* **Takedown Accuracy (%):** The percentage of takedown attempts that are successful.
* **Takedown Defense (%):** The percentage of opponents' takedown attempts successfully defended.

**Scatter Plots: Wrestling Metrics vs Wins**

This section visualizes the relationship between key wrestling metrics (takedown accuracy, takedown defense) and wins.

**Grappling Analysis**

**Key Metrics:**

* **Average Submissions Attempted per 15 Minutes:** The number of submission attempts made by a fighter in a 15-minute fight.

**Web Scraping: Extracting UFC Fighters' Control Time Data**

**Objective**

The control time, a critical metric in UFC, was not available in the provided dataset. Therefore, a web scraping approach was used to extract this data from an external source. This section explains how control time data was retrieved using Python’s requests and BeautifulSoup libraries.

**Rationale for Scraping Control Time Data**

* **Control Time** is a critical performance metric in UFC, representing the amount of time a fighter spends in dominant positions during a match. Fighters with high control time often demonstrate superior grappling and ground control skills, which can significantly influence match outcomes.
* Since this metric was not available in the provided dataset, it was crucial to scrape this data to include it in the overall analysis and ensure a comprehensive evaluation of fighter performance.

**Web Scraping: UFC Pound-for-Pound Rankings**

**Objective**

To supplement the dataset with the current UFC pound-for-pound (P4P) rankings, web scraping was used to extract the top 15 fighters from the official UFC rankings page. These rankings help assess the overall performance and prestige of the fighters across weight classes.

**Rationale for Scraping Pound-for-Pound Rankings**

* **Pound-for-Pound (P4P)** rankings represent the top fighters irrespective of weight class, reflecting overall skill, dominance, and recent performance.
* Including this information in the analysis offers valuable context when comparing fighters' performance metrics and assessing where they stand relative to the best fighters across all divisions.

**GOATS (Greatest of All Time) in UFC**

**Objective**

This section focuses on the fighters widely recognized as the **Greatest of All Time (GOATs)** in UFC history. These fighters have made significant contributions to the sport and have consistently been at the top of discussions among MMA experts and communities.

**Selection Criteria**

Fighters who are widely recognized within MMA communities, discussions, and expert analyses.

**Importance of the GOATs in UFC History**

* **Anderson Silva**: Dominated the middleweight division with a record-breaking title reign.
* **Georges St-Pierre (GSP)**: Known for his unparalleled versatility and fight IQ, holding titles in two weight classes.
* **Jon Jones**: Considered one of the most skilled fighters ever, dominating the light heavyweight division for years.
* **Khabib Nurmagomedov**: Retired undefeated with a flawless record and is revered for his grappling dominance.
* **Demetrious Johnson**: Known for his agility, speed, and technical prowess, especially in the flyweight division.
* **José Aldo**: Dominated the featherweight division and is respected for his striking and longevity.
* **Chuck Liddell**: A pioneer of the light heavyweight division and instrumental in bringing mainstream attention to UFC.
* **Matt Hughes**: A two-time welterweight champion known for his wrestling and ground game.
* **Daniel Cormier**: Held titles in both the light heavyweight and heavyweight divisions.
* **Randy Couture**: A five-time UFC champion, known for his adaptability across weight classes.

**Rationale for GOAT Inclusion**

Including GOAT fighters in the analysis offers:

* A focused study on elite fighters' attributes and performance.
* Comparative insights into the evolution of MMA skills and strategies.

**GOATs with Specialized Skills**

**Objective**

This section analyzes the GOAT fighters who excel in specific fighting disciplines such as striking, wrestling, grappling, and control time. By cross-referencing the GOATs with these specialized categories, we aim to explore how certain elite fighters dominate not only overall but also in particular areas of MMA.

**Filtering for Specialized Skills**

The goat\_df is filtered to include fighters from the following specialized categories:

* **Striking**: Fighters who are renowned for their striking capabilities.
* **Wrestling**: Fighters who excel in takedowns and control through wrestling.
* **Grappling**: Fighters proficient in submissions and ground control.
* **Control Time**: Fighters who dominate opponents by controlling them on the ground or against the cage.

**Insights**

* **Matt Hughes**: A wrestling powerhouse, Matt Hughes dominated opponents through relentless takedowns and ground control. His win percentage of 83.0% showcases his effectiveness in the welterweight division.
* **Georges St-Pierre (GSP)**: Known for his well-rounded skills, GSP not only excelled in striking but was also a master of takedowns, with a high takedown accuracy of 74.0% and a remarkable defense rate of 83.0%. His ability to adapt and control fights helped him achieve a win percentage of 93.0%.
* **Randy Couture**: One of the oldest competitors to hold a UFC title, Randy Couture was a versatile fighter who utilized his wrestling background to great effect. His balanced striking and wrestling skills earned him titles in multiple weight classes.

**Importance of Specialization**

By analyzing these GOATs' specialized skills, we gain insights into how fighters can use their strengths to control fights and dominate their divisions. Their adaptability across striking, wrestling, and grappling disciplines has contributed significantly to their legacy in MMA.

**Undisputed Undefeated UFC Fighters**

**Objective**

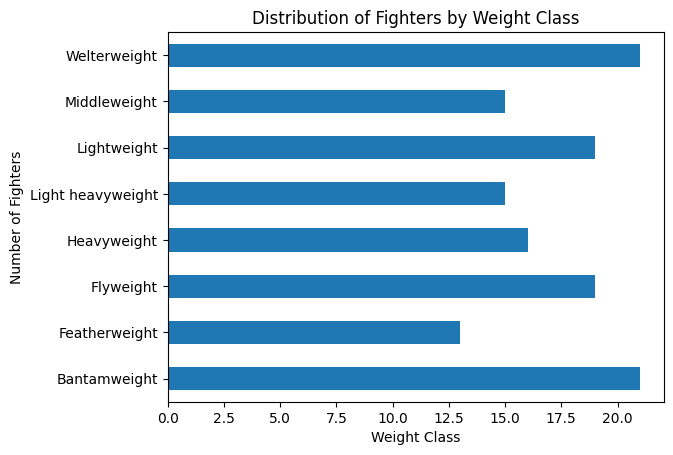
In this section, we analyze UFC fighters who have remained undefeated throughout their careers. This includes fighters with no losses or draws in their professional fight records. The data is further filtered to include only those fighters who have competed in more than 10 matches, thus ensuring a more robust representation of dominance.

**Filtering Undefeated Fighters**

We start by filtering the dataset to include fighters with 0 losses and 0 draws. Additionally, we append Jon Jones to this list, who is considered undefeated due to his controversial disqualification loss, often debated among MMA fans.

**Distribution of Undefeated Fighters by Weight Class**

We group the undefeated fighters by their respective weight classes to visualize the distribution across divisions. This provides insights into which weight classes tend to have more undefeated fighters.



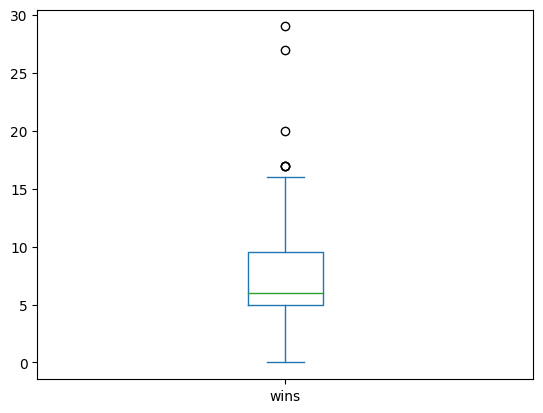
**Specialized Skill Analysis of Undefeated Fighters**

After cross-referencing the undefeated fighters with the datasets of specialized skills (striking, wrestling, grappling, and control time), we found **0 fighters** who matched any of the categories. This lack of overlap suggests a few possibilities:

1. **Well-Rounded Fighters**: Undefeated fighters tend to have a more balanced and adaptable skillset rather than being specialists in a particular domain.
2. **Data Coverage**: The dataset for specialized skills might not include all the attributes of the undefeated fighters, or these fighters excel in areas not covered by the current metrics (e.g., game planning, mental toughness, adaptability).

This finding indicates that while specialized skills can be important, overall versatility and adaptability might play a bigger role in maintaining an undefeated record in the UFC.

**Visual Representation**



The boxplot above represents the number of wins of undefeated UFC fighters. From the boxplot, we can see several outliers in the win distribution, with the most notable ones being **Khabib Nurmagomedov** and **Jon Jones**, who both have over 20 wins.

**Key Insights:**

1. **Khabib Nurmagomedov** retired undefeated with an extraordinary record of 29 wins.
2. **Jon Jones**, often considered one of the greatest, has also been undefeated with only 1 controversial loss, having won 27 matches.

Both of these fighters are considered **GOATs (Greatest of All Time)** due to their undefeated records, dominance in their respective weight classes, and overall impact on the sport.

**Almost Undefeated UFC Fighters - Exploratory Data Analysis**

The following section details an analysis of UFC fighters with nearly undefeated records. The focus is on identifying potential "future GOATs" (Greatest of All Time) by evaluating fighters with very few losses (≤ 1) and a minimum of 10 matches played.

**Data Selection Process**

We filtered the dataset to include only those fighters who meet the following criteria:

* Maximum of 1 loss
* At least 10 matches played

The filtered dataset contains fighters who have performed exceptionally well, with minimal losses despite a significant number of matches played. This helps identify consistent performers who could be considered the best in the future.

**Box Plot of Wins**

The box plot below shows the distribution of wins among these fighters, highlighting potential outliers at the higher end of the win spectrum.

**Key Observations:**

* The median number of wins is approximately 15.
* A few fighters have wins in the range of 25 to 30, making them notable outliers.

**Top Fighters with More than 22 Wins**

We now focus on fighters with more than 22 wins, as they demonstrate significant experience and success in the UFC.

**Top Fighters:**

* **Julio Cesar Neves Jr.**: 33 wins and only 1 loss
* **Khabib Nurmagomedov**: 29 wins with no losses
* **Jon Jones**: 27 wins and 1 loss
* **Islam Makhachev**: 25 wins and 1 loss
* **Khusein Askhabov**: 23 wins and 1 loss

These fighters represent the pinnacle of the UFC, with high win counts and minimal losses.

**Analyzing Specific Skill Sets:**

To refine the analysis, we cross-referenced the list of almost undefeated fighters with specific fighter skill sets (strikers, wrestlers, grapplers, etc.). This helps us identify whether these top performers are also excelling in specialized domains.

**Fighters excelling in specialized domains:**

1. **Zabit Magomedsharipov** (Featherweight): 18 wins, 1 loss
   * Significant strikes landed per minute: 4.89
   * Takedown accuracy: 56.0%
   * Takedown defense: 77.0%
2. **Sean O'Malley** (Bantamweight): 17 wins, 1 loss
   * Significant strikes landed per minute: 7.25
   * Striking accuracy: 61.0%
   * Takedown defense: 62.0%
3. **Ikram Aliskerov** (Middleweight): 15 wins, 1 loss
   * Significant strikes landed per minute: 8.24
   * Takedown accuracy: 33.0%

These fighters stand out for their striking, grappling, and takedown skills, making them well-rounded UFC athletes.

**Future GOATs Prediction**

Based on the analysis of fighters with fewer than 2 losses, we can identify potential future GOATs who exhibit exceptional win records and high proficiency in diverse fighting styles:

**Potential GOAT Candidates:**

* + skillset is unmatched.

1. **Islam Makhachev (Lightweight)**
   * **Wins**: 25
   * **Losses**: 1
   * **Style**: A successor to Khabib, Makhachev excels in grappling with a takedown accuracy of 60% and an impressive defense of 90%.
2. **Sean O'Malley (Bantamweight)**
   * **Wins**: 17
   * **Losses**: 1
   * **Style**: O'Malley's striking game is superior, with an incredible 7.25 strikes landed per minute and 61% accuracy. His rising profile in the UFC makes him a fan favorite for future stardom.
3. **Zabit Magomedsharipov (Featherweight)**
   * **Wins**: 18
   * **Losses**: 1
   * **Style**: Zabit's versatility in striking and grappling, coupled with a solid takedown accuracy of 56%, makes him another name to watch in the GOAT conversation.