**Section 4: Summary Statistics**

In this section, we perform a thorough examination of our dataset by calculating key summary statistics such as the mean, median, mode, and variance for all relevant numeric variables. This step allows us to understand the central tendencies, variability, and distribution of the data.

**Frequency Counts and Unique Counts**

We also counted the number of unique fighters, countries, and other categorical variables to get a sense of the diversity in our dataset. For example, there are 128 unique fighters from 36 different countries, with the USA and Brazil being the most represented. Such insights are crucial for understanding the scope and representativeness of the dataset.

**Interpreting the Correlation Matrix**

One of the most powerful tools in this section is the correlation matrix, which shows the relationships between different variables. For instance, we observed a strong positive correlation between Height and Reach, which is expected since taller fighters generally have a longer reach.

**Summary** **Statistics** - **Correlation matrix for numeric variables – Snippet**

                 Height       Weight        Reach          Rnd

Height                    1.000000000  0.809645711  0.927082356 -0.080466651 -0.0772000335

Weight                    0.809645711  1.000000000  0.773215350 -0.117097237 -0.1438516062

Reach                     0.927082356  0.773215350  1.000000000 -0.073869176 -0.0828932647

Rnd                      -0.080466651 -0.117097237 -0.073869176  1.000000000  0.4722461377

**10 Key Takeaways from the Correlation Matrix:**

1. **Strong Positive Correlation Between Height and Reach (0.93):**
   * **Why It Matters:** This strong correlation is expected, as taller fighters generally have a longer reach. Understanding this relationship is crucial for interpreting fighter statistics, particularly in striking and defensive metrics where reach plays a significant role.
2. **Moderate Negative Correlation Between Weight and Rounds Fought (-0.12):**
   * **Why It Matters:** This suggests that heavier fighters tend to have fewer rounds in their fights, potentially due to higher knockout power, leading to quicker finishes. This could influence strategies for fighters in heavier weight classes.
3. **High Correlation Among Striking Defense Metrics (e.g., SDBL and SDB\_A - 0.95):**
   * **Why It Matters:** These high correlations indicate redundancy among these metrics, which could lead to multicollinearity in predictive models. It’s important to consider combining or selecting only one of these variables to avoid instability in model training.
4. **Moderate Positive Correlation Between Total Strikes Landed and Rounds Fought (0.63):**
   * **Why It Matters:** This suggests that fighters who participate in longer fights tend to land more total strikes, which makes sense as more time allows for more opportunities to strike. This insight is crucial when analyzing fighter endurance and activity.
5. **Negative Correlation Between Total Strikes Attempted and Knockdowns (TS\_A and KD: -0.11):**
   * **Why It Matters:** This counterintuitive result suggests that fighters who attempt more strikes are less likely to achieve knockdowns. It could imply that more accurate or powerful striking, rather than sheer volume, is more important for knockdowns.
6. **Positive Correlation Between Significant Strikes Landed and Total Strikes Landed (SSL and TSL: 0.83):**
   * **Why It Matters:** This strong correlation shows that fighters who land more total strikes tend to land more significant strikes. This could influence how we evaluate a fighter’s effectiveness in striking, as it indicates consistency in landing impactful strikes.
7. **Negative Correlation Between Height and Knockouts (KO/TKO: -0.50):**
   * **Why It Matters:** Taller fighters are less likely to win by knockout, which could be due to a reliance on reach and technique over power. This insight might influence how we assess fighters' finishing ability in different weight classes.
8. **Positive Correlation Between Total Strikes Landed and Fight Outcome (TSL and Results: 0.20):**
   * **Why It Matters:** This suggests that landing more strikes is generally associated with winning a fight. While not a very strong correlation, it reinforces the importance of striking volume in determining fight outcomes.
9. **Weak or No Correlation Between Ground Strikes and Fight Outcomes:**
   * **Why It Matters:** The lack of a strong correlation between metrics like Ground Strikes and the result of the fight suggests that ground striking might not be as decisive in determining outcomes as other factors, like striking defense or overall striking.
10. **High Correlation Between Height and Weight (0.81):**
    * **Why It Matters:** This is an expected correlation since heavier fighters are often taller, particularly within the same weight class. This relationship is important when analyzing weight distribution across different fighter profiles and could be a factor when adjusting for weight in models.