Luis Gil

Data Source

- o **Source:** [Kaggle UFC Data](#)

- o **Summary:** The dataset includes detailed information on UFC fights, including fighter statistics, fight outcomes, and fight locations from 1993 to 2021.

- o **Reason for Choice:** This dataset offers a comprehensive look at various aspects of UFC fights, providing ample opportunities for exploratory and advanced data analysis. The dataset meets the project criteria, including having geographical data and a mix of continuous and categorical variables.

Data Cleaning

- o **Handled Missing Values:** Missing values were identified and handled appropriately.

- o **Removed Duplicates:** Duplicate rows were removed to ensure data integrity.

- o **Data Type Conversions:** Ensured all columns have the correct data types.

Data Profiling:

- o **Variables Review:** Detailed review of all variables in the dataset.

- o **Descriptive Statistics:** Performed basic descriptive statistical analysis to understand data distributions and central tendencies.

- o **Key Findings:**

  - ▪ Average and range of fight statistics (e.g., knockdowns, submissions).

  - ▪ Distribution of fight outcomes across weight classes.

Limitations:

- o **Data Completeness:** Some fights have missing or incomplete data.

- o **Temporal Range:** Data spans from 1993 to 2021, so some trends might be outdated.

Luis Gil

Ethical Considerations:

- o **Data Source:** Ensure the data was collected and shared ethically.

- o **Use of Data:** The analysis should not promote violence or unethical behavior.

Analysis Questions:

- o What factors most significantly influence the outcome of a fight (win/loss)?

- o How do fight statistics vary across different weight classes?

- o Are there geographical trends in fight outcomes?

- o Can we predict the winner of a fight based on historical data?

- o How have fight outcomes evolved over time?