Project 7: UFC Fights

You've just finished watching a string of UFC pay-per-views, noticing how some fighters climb through brutal matches to championship glory — and others fall just short. You realize the numbers behind these fights tell their own story: how many strikes landed, knockdowns suffered, or fights won by submission.

Inspired by the way data can reveal unseen patterns in sports, you've decided to bring that same analytical mindset to mixed martial arts (MMA). Your mission: build a clean, structured dataset of UFC fighter records from UFCStats.com — the official archive of UFC fight data.

The catch? UFCStats has no open API. All those rich statistics are tucked inside webpages, waiting for your web scraping skills to bring them to life.

Your goal? Scrape, clean, and organize fight records for as many UFC fighters as you can find — capturing the details of each bout, the opponent, fight result, and performance stats. From there, you'll assemble these records into one unified, analysis-ready dataset.

You might uncover which fighters rack up the most knockdowns, which weight classes have the highest strike rates, or how title fight outcomes differ from undercard scraps.

For this project, you'll:

- Scrape fight history data from individual fighter profile pages.
- Collect links to as many fighter profiles as possible.
- Tidy and standardize the collected data.
- Aggregate everything into a single dataset.
- Explore trends, patterns, and notable outliers through visualizations.

Part 1: Data Collection

Step 1: Access UFCStats.com

Head over to UFCStats.com.
Familiarize yourself with:

- **Event listings** or
- **Fighter profile pages**

Identify where fighter profiles are linked — either from event cards or rankings pages.

Step 2: Collect Fighter Names and Profile Links

On event or division pages, you'll find links leading to individual fighter profiles.

Extract for each fighter:

- Fighter's name
- URL to their profile page

Step 3: Scrape Fight Histories from Each Profile

Loop through the list of fighter profile links you collected.

For each fighter's profile page:

- Locate the **fight history table**.
- Scrape details for each fight: Win/Loss outcome, Opponent's name, Knockdowns, Strikes landed, Takedowns, Submission attempts, Event name, Method of victory, Round fight ended, Time fight ended

For each record, attach the fighter's name so you can later distinguish between fighters in your dataset.

Step 4: Store the Collected Data

Organize your scraped records into a structured format, like a CSV file or DataFrame.
Ensure all fighters' data follow the same structure so they can be combined smoothly.

Part 2: Data Cleaning & Preparation

Step 1: Standardizing Missing Data

Review your dataset for any placeholder symbols that indicate missing values.
Replace these with appropriate placeholders (like NaN), depending on the column's context.

Step 2: Renaming Ambiguous Columns

If your scraped data has abbreviated or unclear column names, rename them to be more descriptive.
Examples:

- KD → Knockdowns
- SLpM → Strikes Landed per Minute
- TD → Takedowns

Step 3: Handling Missing or Inconsistent Values

Identify columns with missing or inconsistent values.

- Decide whether missing values should be filled, left as-is, or removed.
- Replace numerical placeholders (like empty strings) with appropriate numbers (often zero).

Step 4: Removing Duplicate Records

Check for duplicate fight records.

- If exact duplicates exist, remove them.
- Keep one clean copy of each fight result.

Step 5: Splitting Complex Columns

If any columns contain combined information (like "3 of 7" for strikes landed out of attempted), split them into separate fields:

- Strikes_Landed
- Strikes_Attempted

Remove the original combined column once separated.

Step 6: Correcting Data Types

Verify that each column's data type matches its contents.

- Convert columns like Knockdowns, Round, or Takedowns into integers.
- Ensure Time is formatted consistently (e.g., in seconds or mm:ss format).

Clean any symbols (like asterisks or stray characters) that might disrupt conversions.

Step 7: Change the units of time

- Convert Time column to seconds and multiply by number of rounds)

Part 3: Data Analysis

Now that your dataset is clean and unified, explore the data through visualizations to uncover trends and standout performances. Here are some ideas to start with:

- **Distribution of Knockdowns:** Plot a histogram to see how often knockdowns occur.

- **Top 10 Fighters by Total Strikes Landed:** Use a bar chart to showcase fighters with the highest cumulative strikes.

- **Average Fight Length by Weight Class:** Compare fight durations across divisions.
- **Win Method Trends:** Create a pie chart or bar chart showing the proportion of wins by KO, submission, or decision.
- **Fight Frequency Over Time:** Visualize how active fighters are across their careers.