simplon MAGHREB

Jobintech

STATEMENT OF WORK

MMA SPORT DATA ANALYSIS PROJECT

START OF DATE

01/01/2026

END OF DATE

03/30/2026

1. Business Context

Mixed Martial Arts (MMA) has become one of the fastest-growing sports globally. Understanding the data behind fights and fighters can provide valuable insights for various stakeholders: promoters, trainers, and fans. This project aims to analyze data related to MMA fighters and fights to provide actionable insights, such as fighter performance analysis, fight trends, and prediction models for future match outcomes.

2. Problem Framing

The core objective of the project is to provide a data-driven solution for analyzing MMA fighters' performance and fight outcomes. Specifically, the data solution will focus on:
• Analyzing fighters' statistics and performance metrics.
• Identifying trends and patterns in MMA fights.
• Predicting the outcome of future fights based on historical data.
• Analyze the performance stats of fighters from martial arts like kickboxing, karate, sambo, and BJJ, etc. to identify key attributes that help them succeed in transitioning to professional MMA.
The central question is: What factors contribute to a fighter's success, and how can these insights be used to predict future fight outcomes?

3. Business Objectives & KPIs

• Objective 1: Evaluate fighter performance.
o KPI: Win/loss ratio, average fight duration, number of submissions, knockouts, etc.
• Objective 2: Detect trends in fight outcomes.
o KPI: Trends related to the effectiveness of different fighting styles, age, weight, location, etc.
• Objective 3: Build a predictive model for fight outcomes.
o KPI: Prediction accuracy, precision, recall, and F1 score.
• Objective 4: Automate the extraction, cleaning, and analysis of data.
o KPI: Time taken for data processing and automation effectiveness.

4. Data Sources

• CSV Files:
o Pro_MMA_Fighters.csv: Contains fighter profiles (e.g., age, weight, height, fighting styles).
o Pro_MMA_Fights.csv: Contains details of past MMA fights (e.g., date, event, opponent, outcome, statistics like strikes, submissions).

Both sources will be used for historical analysis of fighters and their performance in various fights.

5. Data Architecture and Process

• Data Pipeline:
o Extraction: Data will be extracted from the CSV files.
o Cleaning: Handling missing values, duplicate records, and errors in the dataset.
o Transformation: Standardizing formats and creating necessary relational structures for analysis.
o Storage: Cleaned data will be stored in a database or cloud platform for further analysis.
o Automation: Data extraction and cleaning will be automated through Python scripts or scheduled pipelines.
• Tools:
o Python for data cleaning, analysis, and automation.
o SQL for querying and relational data management.
o Tableau/Power BI for building interactive dashboards and visualizations.
• Data Model: A relational model or star schema to connect fighters' profiles, fight statistics, and outcomes.

6. Data Analysis & Statistical Methods

• Exploratory Data Analysis (EDA):
o Descriptive statistics to summarize fighter performance.
o Visualizations to identify key patterns in fight outcomes (e.g., how weight or age impacts fight success).
o Handle missing data, duplicates, and outliers.
• Predictive Analysis:
o Develop models (e.g., logistic regression, decision trees) to predict fight outcomes based on historical data.
o Perform statistical tests to assess relationships (e.g., correlation between fighter's attributes and match outcomes).

7. Visualizations

• Dashboard:
o KPIs: Fighter win rate, knockout statistics, prediction accuracy, etc.
o Filters: Fighters' attributes (weight, height, nationality), fight locations, event types.
o Storytelling: Presenting insights with charts, graphs, and tables to tell a cohesive story about MMA trends and fight outcomes.

• Tools: Tableau/Power BI for interactive dashboards and Jupyter Notebooks for data analysis.

8. Business Recommendations

• Fighter Development: Provide insights into which attributes (e.g., fighting style, weight, experience) are most strongly correlated with success, helping fighters and trainers focus on the right areas for improvement.
• Event Planning: Recommend matchups based on historical performance data, helping promoters schedule engaging and potentially profitable events.
• Fight Outcome Predictions: Build a prediction system to guide betting, audience engagement, and marketing strategies for upcoming fights.
• Become a Pro MMA Fighter: Analyze the performance statistics of fighters from other martial arts disciplines such as kickboxing, karate, sambo, and Brazilian jiu-jitsu (BJJ) to identify key attributes and techniques that contribute to their success in transitioning to professional MMA, with the aim of developing a data-driven approach to help aspiring fighters adapt and excel in MMA.

9. Deliverables

• Specifications Document: This document will include business context, problem framing, KPIs, data sources, and architecture.
• GitHub Repository: A structured repository containing:
o Data extraction and cleaning scripts.
o SQL scripts for data querying and relational management.
o Jupyter notebooks with data analysis and predictive models.
o A link to the project monitoring tool and documentation with clear README.
• Project Report: A comprehensive report structured in a data storytelling format that covers:
o Context
o Analytical approach
o Insights and analysis
o Recommendations for business action
• Presentation Slides: A set of narrative slides summarizing the analysis, visualizations, and key findings, along with a demonstration of the final dashboard.