MGT 6203 Group Project Proposal Template

**Please edit the following template to record your responses and provide details on your project plan.**

**TEAM INFORMATION (1 point)**

**Team #:**

**Team Members:**

1. Team Member 1 Name; GT Id (OMSA) or EdX username (MM)

[Insert background information: Name, professional background, education background, previous analytics related projects you have worked on]

1. Team Member 2 Name; GT Id or EdX username
2. Team Member 3 Name; GT Id or EdX username
3. Team Member 4 Name; GT Id or EdX username
4. Team Member 5 Name; GT Id or EdX username

**OBJECTIVE/PROBLEM (5 points)**

**Project Title:**

**Background Information on chosen project topic:**

**Problem Statement (clear and concise statement explaining purpose of your analysis and investigation):**

**State your Primary Research Question (RQ):**

**Add some possible Supporting Research Questions (2-4 RQs that support problem statement):**

**Business Justification:** **(Why is this problem interesting to solve from a business viewpoint? Try to quantify the financial, marketing or operational aspects and implications of this problem, as if you were running a company, non-profit organization, city or government that is encountering this problem.)**

**DATASET/PLAN FOR DATA (4 points)**

**Data Sources (links, attachments, etc.):**

The data for our projected are sourced from the website http://ufcstats.com/, which is an exhaustive repository of statistics for the Ultimate Fighting Championship (UFC) covering a broad spectrum of information related to UFC events, fights, and fighter statistics.

To systematically extract and organize this corpus of data into structured CSV files, the "https://github.com/Greco1899/scrape\_ufc\_stats" GitHub repository was employed. This repository houses a Python-based web scraping tool tailored for navigating "ufcstats.com". It uses web scraping techniques, with libraries such as BeautifulSoup for HTML parsing and Pandas for data structuring, to collect and compile the data into a format suitable for analysis. This process transforms the rich, web-based data into accessible datasets for a variety of analytical purposes.

**Data Description (describe each of your data sources, include screenshots of a few rows of data):**

There are five main datasets that are derived from this scraping effort.

1. **UFC Event Details Dataset.** This dataset aggregates information about UFC events, serving as a comprehensive event catalog. It encompasses essential metadata for each event, providing a macro view of UFC activities, including the scheduling and geographical distribution of events.
2. **UFC Fight Details Dataset.** Offering a drill-down from events to individual fights, this dataset itemizes each bout within an event. It links fights to their detailed statistics, facilitating in-depth analyses of matchups, fight dynamics, and historical trends.
3. **UFC Fight Results Dataset**. Focused on the outcomes of individual bouts, this dataset captures the essence of competition results. It's crucial for analyzing fight outcomes, understanding victory methods, and studying fighters' performance trajectories.
4. **UFC Fight Stats Dataset**. This dataset stands out by providing round-by-round performance metrics for fighters in each bout. It's rich in statistical data, ideal for nuanced analyses of combat strategies, efficacy, and in-fight behavior.
5. **UFC Fighter Details Dataset**. As a directory of fighters, this dataset connects names to comprehensive profiles. It's the starting point for exploring fighters' careers and achievements, enabling comparative analyses and profiling.
6. **UFC Fighter Totals (TOTT) Dataset.** This dataset enriches fighter profiles with physical attributes and background information. It supports analyses that correlate physical characteristics with fight outcomes, contributing to a deeper understanding of the sport's physical demands.

**Key Variables: (which ones will be considered independent and dependent? Are you going to create new variables?** **What variables do you hypothesize beforehand to be most important?)**

The dependent variable for our model is a simplified version of the fight outcome, which will only include “Win” and “Lose”, removing the complexity of handling rarely occurring outcomes such as “Draw” or “No Contest”.

The independent variables will be drawn from the 6 datasets and can be organized in the following taxonomy.

1. Fighter Background
   * Physical Attributes: Information such as height, weight, and reach derived from the ufc\_fighter\_details.csv and ufc\_fighter\_tott.csv datasets.
   * Age: Calculated from the date of birth in the ufc\_fighter\_tott.csv dataset and the date of the fight from the ufc\_event\_details.csv dataset.
2. Historical Fight Performance
   * Win-Loss Record: Aggregated from the ufc\_fight\_results.csv dataset to calculate each fighter's career win-loss ratio before the fight.
   * Recent Form: Performance in recent fights (e.g., last 3-5 fights), including win/loss streaks, derived from the ufc\_fight\_results.csv dataset.
3. In-Fight Metrics (Aggregated)
   * Striking and Grappling Stats: Aggregate statistics like significant strikes landed/attempted, takedown attempts/successes, and submission attempts from the ufc\_fight\_stats.csv dataset.
   * Fighter Efficiency: Ratios such as significant strike accuracy and takedown accuracy, calculated from the ufc\_fight\_stats.csv dataset.
4. Comparative Metrics
   * Physical Differentials: Differences in height, reach, and weight between fighters, calculated using the ufc\_fighter\_tott.csv dataset.
   * Experience Differential: Difference in the number of total fights or years active, derived from the fighters' records in the ufc\_fight\_results.csv and ufc\_fighter\_details.csv datasets.
   * Style Matchup: Analysis of fighters' styles (striker vs. grappler, counter-fighter vs. pressure fighter) and stance compatibility.
5. Fight Context
   * Event Type and Location: Information from the ufc\_event\_details.csv dataset, which might influence fighter performance due to factors like travel distance or fighting in familiar versus foreign environments.
6. Advanced Statistical Features
   * Performance Trends: Moving averages and weighted metrics for recent performance indicators.
   * Variability & Consistency: Standard deviations in key metrics across fights, fight-to-fight performance deltas.
   * Round-by-Round Analysis: Performance metrics broken down by round to identify fighters' pacing and adaptations.
   * Composite Scores: Weighted combinations of selected metrics to create overall performance scores.
   * Pressure and Adversity Metrics: Performance under high-pressure conditions and in situations of adversity (e.g., after a loss or knockdown).

**APPROACH/METHODOLOGY (8 points)**

**Planned Approach (In paragraph(s), describe the approach you will take and what are the models you will try to use? Mention any data transformations that would need to happen. How do you plan to compare your models? How do you plan to train and optimize your model hyper-parameters?))**

**Anticipated Conclusions/Hypothesis (what results do you expect, how will you approach lead you to determining the final conclusion of your analysis) Note: At the end of the project, you do not have to be correct or have acceptable accuracy, the purpose is to walk us through an analysis that gives the reader insight into the conclusion regarding your objective/problem statement**

**What business decisions will be impacted by the results of your analysis? What could be some benefits?**

**PROJECT TIMELINE/PLANNING (2 points)**

**Project Timeline/Mention key dates you hope to achieve certain milestones by:**

**Appendix (any preliminary figures or charts that you would like to include):**