SportsStats Executive Summary *(High-Level Overview of Milestone 1 Work)*

Purpose of the Project:
SportsStats *engaged the analytics team to provide a historical, data-driven perspective on how the modern Olympic Games have evolved over the last **120 years**. The goal is to identify long-term trends in athlete demographics, national participation, and competitive outcomes that can support media storytelling, performance analysis, and strategic planning within sports organizations.*

Data Sources & Scope:
The analysis is based on two historical datasets covering the Olympics from 1896 to 2016

- ***Athlete Events Dataset*** *(269k+ records): Contains athlete-level information across all events, including age, gender, physical attributes, sport, and medals.*
- ***NOC Regions Dataset****: Provides geopolitical and regional information for each National Olympic Committee (NOC).*

These datasets were cleaned, validated, and integrated using a structured SQL-based pipeline to ensure accuracy and analytical readiness.

Data Quality Assessment:
A comprehensive data quality audit was conducted to evaluate

- *Accuracy and real-world validity*
- *Completeness and missingness patterns*
- *Consistency of naming standards and categories*
- *Historical correctness (verified through Olympedia and IOC sources)*
- *Duplicate identification and removal*
- *Region mapping and geopolitical alignment*

Key findings included inconsistent NOC codes, missing athlete attributes, and discrepancies in team names, all of which were corrected or documented. A total of **1,455 duplicate records were removed**, country codes were standardized, and critical demographic information was validated and supplemented when possible.

Data Cleaning Outcomes
- *Cleaned datasets were created and exported for long-term use.*
- *Region mappings were fully corrected and standardized.*
- *Medal information was harmonized to replace nulls with meaningful values.*
- *Demographic completeness was improved using authoritative external sources.*
- *All column relationships were validated for structural integrity (e.g., Year–Games alignment).*

The dataset is now fully prepared for exploratory, statistical, and advanced analytic phases.

Initial Exploration & Key Observations
*Initial exploration of the cleaned Olympic dataset confirms strong completeness and reliability across 120 years of competition. The primary dataset includes **269,661 records** representing **135,571 athletes**, spanning **51 Olympic Games**, with clear era-based trends in athlete participation and event growth. Demographic indicators such as age, height, and weight follow expected athletic patterns, with median age at **24**, and physical attributes well distributed despite partial missingness typical of early Olympic years. Geographic representation remains extensive, covering **230 NOCs** linked to **209 regions**, with historically active nations (France, USA, UK) appearing most frequently across team associations. The Games encompass **66 sports and 765 events**, including several discontinued disciplines, with Shooting and Athletics showing the highest event counts.*

Analytical Focus Going Forward:
The next phase of analysis will center on three core questions:

1. *How have athlete demographics evolved over time?*
2. *Which regions demonstrate the strongest long-term performance patterns?*
3. *How do competitive characteristics differ across sports and eras?*

The project will use statistical measures, trend analysis, and comparative metrics to test hypotheses regarding demographic standardization, regional dominance, and sport-specific athlete profiles.

Expected Business Impact:
The insights derived from this analysis will help SportsStats:

- *Strengthen data-driven storytelling for media clients*
- *Identify long-term competitive indicators across nations*
- *Support athlete performance research with historical context*
- *Improve predictive capabilities for future Olympic trends*

These findings will directly enhance SportsStats's value proposition to its journalism, performance, and sports management partners.