Part1: Summary Report

Exploration & Code: The script connects to a MySQL database, loads athlete performance data, and cleans inconsistent metric names and formatting issues found during initial exploration. The dataset originally had multiple columns with varied formats and inconsistent metric naming, which were standardized for clarity. It then explores the data by counting unique athletes and teams from the **6,617,426** rows, checking the date range of records, identifying the most active data sources, and analyzing the top metrics from Hawkins, Kinexon, and Vald. The result is a clean, structured dataset ready for deeper analysis and insights.

Key Data: Athletes: **1,287** Sports/Teams: **92** Date Range: **2018-10-15 19:27:41 to 2025-10-21 12:24:21** Kinexon: **4,073,754 (Max)** Hawkins: **2,492,372** Vald: **51,300** Athletes with Missing/Invalid Names: **0** Athletes with data from Multiple Sources: **541** Unique metrics: **548** Count of Unknown in "team" column: **5904** Count of rows with "0" in value column: **1520074**

Top 3 metrics/source: **Hawkins:** System Weight (N) (**32,373** records**)**, Avg. Propulsive Power (W), Peak Propulsive Force (N), etc. (**32,123** records) **Kinexon:** Accel_Load_Accum, Distance_Total, Event_Count_Exertion_Category7, etc. (Top 10 metrics each have **40,803** records) **Vald:** Leftavgforce, Leftcalibration, Leftimpulse, Rightavgforce, etc. (Top 10 metrics each have **4,275** records). Metric data range **2018 - 2025**. **Total records count for selected sports:** Football **1411664** Basketball Men **771335** Women **658730**

Research: Reporting on women's cohorts is consistently limited across all metrics. Metrics are often siloed by sport, and further separated by neuromuscular aspects like sprint and jump velocity. Thresholds for performance decline and asymmetry vary considerably across studies.: Variability among sports positional analysis is insufficiently explored. The absence of integrated frameworks diminishes methodological rigor.

Metric Selection: Peak Propulsive Power, Jump Height (m), Peak Velocity (m/s), Speed Max and Distance Total were selected to explore further based on performance concepts of "Movement-Efficiency/Gait-Complexity" and "Force/Strength metrics". Jump Height is a key Vald metric, but SBU seems to track the CMJ metric via Hawkins device.The 5 metrics (Hawkins and Kinexon) were selected for 2 reasons: they have the maximum number of records, and to determine if these can be used to create a meaningful Derived Metric.

Part1 underscores the necessity for reproducible, threshold-based monitoring across multiple metrics. It highlights that an integrated approach can potentially improve fatigue detection and

enhance operational relevance in collegiate sports. Gender gaps and incorporating positional subgroup analysis are stark gaps seen from the Lit review (later confirmed in the dataset). All this led to the use of RFD and ME&GC to explore gender based gaps by applying standardized thresholds across 5 top metrics in Women's and Men's Basketball teams at SBU.