KUMARAGURU COLLEGE OF TECHNOLOGY

DATA SCIENCE VISUALIZATION

PROJECT REPORT

Team : 18

COURSE CODE : 24ADI204

Subject : Data Science Visualization

Team : 18

1. Akhil C – 24BAD007
2. Kaviya S - 24BAD059
3. Hamphiritha BS – 24BAD403

Submitted to
Faculty Name: Shriaarthy E

Problem Selection Summary

We selected the problem statement **"Olympic History & Geopolitics"** because the Olympic Games represent more than just sports—they reflect global power, political influence, and the evolution of athletic performance over time. The availability of a rich dataset covering **120 years of Olympic history** provides an excellent opportunity to apply data science visualization techniques to uncover long-term trends, patterns, and insights across countries and sports. This problem allows us to combine historical data with analytical storytelling, making it both data-intensive and socially meaningful.

Purpose of the Project

The primary purpose of this project is to **visualize and analyze country dominance in the Olympics over time** and to understand how the **physical attributes (height and weight) of gold medalists vary across different sports**. By doing so, the project aims to show how athletic excellence has evolved and how certain body types are optimized for specific events. Through effective visualizations, we tell the story **"Faster, Higher, Stronger: The Evolution of the Ultimate Athlete,"** highlighting how training, selection, and global competition have shaped modern Olympic champions. This project also demonstrates the power of data visualization in transforming complex historical data into clear, insightful narratives.

WEEK 1 PROGRESS REPORT :

1. Dataset Selection

- **Dataset Name:** 120 Years of Olympic History – Athletes and Results
- **Source:**

  https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results
- **Format:** CSV
- **Domain:** Sports and Entertainment

Reason for Selection

We selected this dataset because it contains long-term Olympic records across countries and sports. It supports visualization of country dominance trends and analysis of athlete physical attributes. The dataset is large and real-world, which fits the project requirement for data storytelling and visualization.

2. Tools and Environment Setup

The following tools and libraries were installed and configured:

- Python 3.x
- Jupyter Notebook
- Pandas
- NumPy
- Matplotlib
- Seaborn
- Scikit-learn (installed for future use)

Environment setup was verified by running sample data loading scripts.

3. GitHub Repository Setup

A GitHub repository was created for version control and collaboration.

Repository Structure Created

- Raw dataset uploaded
- Initial README file added
- Team access configured
- Version control workflow started

Repository Link:

https://github.com/Akhil-coderr/DSV-Olympic-History-Geopolitics-TEAM---18.git

4. Initial Project Planning

The team discussed the project story theme:

"Faster, Higher, Stronger – The Evolution of the Ultimate Athlete."

Planned key analyses:

- Country medal dominance over time
- Sport-wise medal patterns
- Height and weight analysis of gold medalists
- Visual storytelling dashboard

Week 1 Outcome

- Team formed and roles assigned
- Dataset selected and downloaded
- Tools installed and verified
- GitHub repository created
- Project folder structure organized
- Initial project direction defined