JGI JAIN DEEMED-TO-BE UNIVERSITY | FACULTY OF ENGINEERING AND TECHNOLOGY

School of Computer Science and Engineering

(Computer Science & Engineering)

Faculty of Engineering & Technology

Jain Global Campus, Kanakapura Taluk - 562112
Ramanagara District, Karnataka, India

# Project Outline

**Project Title:** Olympics Data Analysis

**Project Goal:** The goal of this data analysis project is to find patterns and information in the Olympics dataset. Through an examination of participation rates, medal distribution, and athlete characteristics in several Olympic Games, the project aims to illuminate factors influencing success, identify areas of growth, and potentially reveal social impacts associated with this global sporting event.

**Team Name:** Strategy Stars

**Team Members:**

- Hitha Choudhary G – 22BTRAD015
- K Shreeshanth Gouda – 22BTRAD017
- Lakshya Sharma – 22BTRAD021

**Resources:**

The dataset was collected from a public repository which was a zip file("Olympic Dataset") that included two csv files "athlete_events.csv" and "noc_regions.csv"

- athlete_events.csv contains a total of 271116 rows and 15 columns and the att ributes are as follows:
  ID, Name, Sex, Age, Height, Weight, Team, NOC, Games, Year, Season, City , Sport, Event, Medal
- Noc_regions.csv contains a total of 230 rows and 3 columns and the attributes are as follows:
  NOC, region, notes

**Project Stages:**

- Stage 1: Data Acquisition and Exploration
  Objective: Acquire dataset and understand its structure and summary statistics.

- Stage 2: Data Cleaning and Preparation
  Objective: Handle missing values, inconsistencies and errors within the data and also to format data appropriately (data types), create new features if needed and ensure the data is ready for analysis.

- Stage 3: Data Analysis and Visualization
  Objective: To identify key trends, patterns, and relationships in the data and producing clear and informative visualizations to effectively communicate through our findings and insights.

- Stage 4: Conclusion
  Objective: Outlining the important conclusions and understandings we obtained from each analysis to conclude our project.

**Tasks with Onus:**

- Task: Athlete Demographics and Participation, Medal Distribution
  - Objective: The goal of this work is to examine participant patterns and athlete statistics such as age, sex, and nationality. An examination of medal distribution among nations (NOCs), sports, and potentially athlete demographics is required for this work.
  - Onus: Hitha Choudhary G

- Task: Trends in Sports, Summer vs Winter Olympics
  - Objective: Exploring medal distribution, and performance trends within different sports over time. Conduct a comparative analysis of athlete participation, medal distribution, and trends between Summer and Winter Games.
  - Onus: K Shreeshanth Gouda

- Task: Social and Economic Factors, Predictive Modeling
  - Objective: Investigate potential relationships between a country's social and economic factors (e.g., GDP, gender equality) and their Olympic performance. Develop models to forecast future outcomes, such as medal winners or participation trends, leveraging historical data from the Olympics dataset.
  - Onus: Lakshya Sharma

These core areas will be tackled with the collaboration of all the team members to answer any potential project questions posed. This analysis provide valuable insights into the who, what, why, and how of the Olympic Games by uncovering the above mentioned tasks.

**Constraints:**

- Data Availability and Quality: The limited data quality and availability can restrict the types of questions that can be answered and potentially introduce bias into your analysis.
- Data Complexity and Size: The complexity and size of the data can impact the feasibility of certain analysis methods and potentially lengthen the processing time.
- Technical Challenges: Beyond data limitations, technical hurdles can arise. Choosing the most effective analysis methods for your specific data and research questions requires careful consideration.

**Deliverables:**

- Analysis Methodology:
  Document outlining the approach adopted for data acquisition, cleaning, analysis, visualization, and inference derivation.
- Report on Data Analysis:
  Comprehensive report showcasing significant findings derived from the analysis, including descriptive statistics, patterns, trends, and relationships identified within the dataset.
- Key Findings and Insights:
  Summary document highlighting the most important conclusions and practical insights gained from the project, accompanied by concise justifications and relevant images or visuals.

**Project Deadline:**

10/04/2024 – Wednesday