**CAPSTONE PROJECT**

# **A. Bhuvan Deep**

**OVERVIEW**

The dataset here is a comprehensive collection of information about the Olympic Games, offering a detailed view of various aspects related to this prestigious international sporting event. It encompasses data from multiple interconnected tables, providing valuable insights into the history, sports, participants, and locations of the Olympic Games.

**THE PROCESS**

1. Data Acquisition from GitHub:

Obtain the requisite dataset from a designated GitHub repository, containing essential information on university rankings, encompassing various countries and their performance across distinct ranking systems.

2. Data Transformation and Enhancement:

If necessary, execute data transformation procedures to ensure data quality and consistency. Additionally, consider augmenting the dataset with new problem statements to enrich the analysis potential.

3. Connecting with Tools:

Establish connections between the dataset and various analytical tools. Interface the dataset with Power BI, Excel, and MySQL Workbench, facilitating seamless data integration and processing.

4. Problem Statement Solution in Power BI:

Utilize Power BI to delve into the specified problem statements. Employ its robust features for data visualization, exploration, and analysis, effectively deriving insights and solutions.

5. Exploratory Data Analysis (EDA):

Perform exploratory data analysis using either Excel or SQL Workbench, depending on the complexity of the analysis. Extract meaningful patterns, relationships, and trends from the data to inform subsequent decision-making.

6. Creation of Visual and Insightful PowerPoint:

Develop a comprehensive PowerPoint presentation that encapsulates the project's objectives, methodologies, problem statement solutions, and key visualizations. Each problem statement should be accompanied by a dedicated section with pertinent conclusions and insights.

7. Detailed Documentation:

Compile a detailed report that meticulously documents the entire project lifecycle. Include sections on data collection, transformation, problem statement formulation, tools integration, Power BI solutions, EDA insights, and PowerPoint visualizations.

**Project Objective**

* The objective of this project is to explore and analyze historical Olympic Games data to uncover meaningful trends and insights. The focus is on understanding how the Games have evolved across years in terms of sports disciplines, athlete involvement, medal distributions, and worldwide participation. This helps identify long-term patterns and transformations in the Olympic movement.

**Analysis Scope**

* The scope includes examining various Olympic editions, host nations, and event structures to study their development. It also analyzes participation trends in individual sports, athlete demographics, and medal achievements. Key areas include sport-wise popularity, country-level dominance, and notable athlete performances. The structure of the underlying database is also explored through an ER diagram that maps entities like Games, Athletes, Events, and Medals.

**Project Goal**

* The goal is to provide a comprehensive, data-driven overview of Olympic history. It aims to highlight trends such as shifts in host city choices, changes in sports prominence, and consistent top performers by country and region. Using these insights, the project aims to recommend strategies for future Olympic planning, sport inclusion, and enhancing global participation. The final output will be a detailed report and presentation supported by visual analytics.

**DATA DICTIONARY**

### **Table: City**

* **Fields:**
  + id: Unique identifier for each city.
  + city\_name: Name of the city (e.g., Barcelona, Tokyo).

### **Table: Sport**

* **Fields:**
  + id: Unique identifier for each sport.
  + sport\_name: Name of the sport (e.g., Swimming, Athletics).

### **Table: Event**

* **Fields:**
  + id: Unique identifier for each event.
  + sport\_id: Foreign key referencing the id field in the Sport table.
  + event\_name: Name of the specific Olympic event (e.g., Men's 100m, Women's Basketball).

### **Table: Games**

* **Fields:**
  + id: Unique identifier for each Olympic Games edition.
  + games\_year: The year in which the Olympics were held.
  + games\_name: Official name of the Olympic Games (e.g., "1992 Summer").
  + season: Season of the games — Summer or Winter.

### **Table: Games\_City**

* **Fields:**
  + games\_id: Foreign key referencing the id field in the Games table.
  + city\_id: Foreign key referencing the id field in the City table.

### **Table: NOC\_Region**

* **Fields:**
  + id: Unique identifier for each NOC entry.
  + noc: National Olympic Committee code (e.g., IND for India).
  + region\_name: Full name of the country or region (e.g., India, Australia).

### **Table: Person**

* **Fields:**
  + id: Unique identifier for each athlete.
  + full\_name: Full name of the athlete.
  + gender: Gender of the athlete (M/F).
  + height: Athlete’s height in centimeters.
  + weight: Athlete’s weight in kilograms.

### **Table: Person\_Region**

* **Fields:**
  + person\_id: Foreign key referencing the id field in the Person table.
  + region\_id: Foreign key referencing the id field in the NOC\_Region table.

### **Table: Games\_Competitor**

* **Fields:**
  + id: Unique identifier for each participation record.
  + games\_id: Foreign key referencing the id field in the Games table.
  + person\_id: Foreign key referencing the id field in the Person table.
  + age: Age of the athlete during that Olympic edition.

### **Table: Medal**

* **Fields:**
  + id: Unique identifier for each medal type.
  + medal\_name: Name of the medal (Gold, Silver, Bronze, or N/A).

### **Table: Competitor\_Event**

* **Fields:**
  + event\_id: Foreign key referencing the id field in the Event table.
  + competitor\_id: Foreign key referencing the id field in the Games\_Competitor table.
  + medal\_id: Foreign key referencing the id field in the Medal table.

















