**CSE-564 - Visualization**

**Lab 1 – Assignment: Analysis of FIFA World Cup Matches Dataset**

**Introduction**

The FIFA World Cup is one of the most prestigious and widely viewed sports events in the world, bringing together nations in a celebration of football (soccer). This report delves into the dataset available on Kaggle, which offers a comprehensive look at World Cup matches over the years. The dataset is curated by a Kaggle user, shivan118, and can be accessed through the following URL:

[FIFA World Cup Matches Dataset](https://www.kaggle.com/code/shivan118/fifa-world-cup-data-analysis/input?select=WorldCupMatches.csv)

**Source of Dataset**

The dataset is hosted on Kaggle, a prominent platform for sharing and exploring data. It is part of a broader project titled "FIFA World Cup Data Analysis" by the user shivan118. The direct link to the dataset is here.

**Attributes Description**

The dataset encompasses several attributes that provide a detailed account of World Cup matches. Key attributes include:

* **Year:** The year the match was played.
* **Datetime:** The date and time of the match.
* **Stage:** The stage of the competition (e.g., Group Stage, Quarter-finals).
* **Stadium:** The stadium where the match was held.
* **City:** The city in which the match was played.
* **Home Team Name:** The name of the home team.
* **Away Team Name:** The name of the away team.
* **Home Team Goals:** The number of goals scored by the home team.
* **Away Team Goals:** The number of goals scored by the away team.
* **Win conditions:** Conditions under which the game was won (e.g., extra time, penalties).
* **Attendance:** The number of spectators present at the match.

**Data Fuse Operation:**

**Identify Common Attributes: The first step is to identify attributes that can serve as keys for merging the datasets. In this case, RoundID and MatchID are present in both datasets and uniquely identify each match, making them suitable for this purpose.**

**Prepare the Datasets: Before fusing, ensure that the datasets are clean and the key attributes (RoundID, MatchID) are in a consistent format across both datasets. This might involve data cleaning steps like removing duplicates, handling missing values, or standardizing formats.**

**Merge Operation: We used Javascript to merge datasets and libraries like “csv-parser” and “csv-writer fs”. I have attached the JavaScript code file for reference.**

**Select and Rename Attributes: After merging, select the relevant attributes from both datasets to be included in the final dataset. The final set of attributes will be:**

**From Dataset-1 (WorldCupMatches.csv): Year, Datetime, Stage, Stadium, City, Home Team Name, Home Team Goals, Away Team Goals, Away Team Name, Win conditions, Attendance, Half-time Home Goals, Half-time Away Goals, RoundID, MatchID**

**From Dataset-2 (FIFAReferee.csv): Home Team Initials, Away Team Initials, Referee, Assistant 1, Assistant 2 (Note that RoundID and MatchID are used for joining and are not counted again).**

**The Fused Dataset**

**The fused dataset, “fusedDataset.xlsx”, will contain the following attributes:**



**This sample data row shows a merged record from the two datasets, including all the specified attributes.**

**Interest in the Data**

This dataset is particularly interesting for several reasons:

**Historical Insight:** It offers a comprehensive overview of the World Cup matches, allowing for analysis of trends and patterns over the years.

**Sporting Analysis:** Enables detailed statistical analysis of team performances, game outcomes, and the impact of different stages or locations on match results.

**Cultural Impact:** The attendance and global interest in matches reflect the cultural significance and global reach of the World Cup.

**Noteworthy Implementation Aspects**

When working with this dataset, certain aspects stand out:

**Data Cleaning**: Given the historical range, ensuring data consistency and accuracy across years is crucial.

**Analytical Potential:** The dataset supports a wide range of analyses, from predictive modeling of match outcomes to sociocultural studies on the popularity of football.

**Visualization Opportunities:** The data lends itself well to visual representations, such as heat maps of goal distributions or timelines of team performances.

**Conclusion**

The FIFA World Cup Matches dataset provides a rich source of information for analyzing the dynamics of one of the world's most popular sporting events. Its comprehensive attributes and the potential for deep analytical insights make it a valuable resource for both sports enthusiasts and researchers.