Exploring the Data

- World Cup performance: This can be measured in different ways, such as:
  - Final placement (1st, 2nd, 3rd, etc.)
  - Knockout stage reached (Group Stage, Quarterfinals, Semifinals, etc.)
  - Win percentage in the World Cup
  - Number of points per match in World Cup games

Data Wrangling and Transformation

- Filtering data: Keeping only matches from 1999 onward.
- Merging datasets: We will assess both the World Cup datasets and non-World Cup datasets.
- Creating new metrics (e.g., rolling averages, Elo ratings).
- Creating new columns: For example, you might want to create a column that indicates whether a match was a knockout stage match or not.
- Create a few new columns which may include:
- Transforming categorical variables (e.g., converting team names into numerical indicators or factors).
- Remove unnecessary columns if needed (e.g., if stadium or city is not part of your analysis).

```
# Create a table of terms and their definitions
import pandas as pd
terms = {
    "Term": [
        "World Cup performance",
        "Final placement",
        "Knockout stage reached",
        "Win percentage",
        "Points per match",

    ],
    "Definition": [
        "A measure of a team's success in the World Cup.",
        "The final position a team finishes in the tournament.",
        "The stage of the tournament a team reaches (e.g., Group Stage, Quarterfinals).",
        "The percentage of matches won by a team in the World Cup.",
        "The average number of points scored by a team per match in the World Cup.",

    ]
}
terms_df = pd.DataFrame(terms)
print(terms_df)
```

```
##                      Term                                         Definition
## 0   World Cup performance    A measure of a team's success in the World Cup.
## 1         Final placement  The final position a team finishes in the tour...
## 2  Knockout stage reached  The stage of the tournament a team reaches (e....
## 3          Win percentage  The percentage of matches won by a team in the...
## 4        Points per match  The average number of points scored by a team ...
```