Assignment

Step 1 :

Create new folder called final\_transformation\_data in the same storage account

Step 2 :

Load remaining 3 files in the raw data

Step 3 :

Use that 3 files in the Databricks notebook and undserstand the schema

Step 4 :

Once you have loaded the files then try to solve this problem

**Task 1**: Count the number of teams each country has and store the result.

**Hint**: Use groupBy and count functions to aggregate data based on the country.

**Task 1**: Create a dataframe listing all coaches for each sport and store the result.

**Hint**: Group data by the sport and collect list of coaches using groupBy and agg with collect\_list.

**Task 3**: Calculate the gender distribution (male and female count) for each team and store the result.

**Hint**: Join EntriesGender with Teams on team ID, then group by team and count genders.

**Task 4**: Identify teams that have more than one coach and store the result.

**Hint**: Group data by team and filter where the count of coaches is greater than one.

**Task 5**: Find teams that do not have both male and female participants and store the result.

**Hint**: After joining EntriesGender with Teams, group by team and check if only one gender is present.