

# Assignment

In this assignment, we will construct simple predictive models for olympic medals using a dataset from Kaggle. You will complete this assignment using only (generic) Python. No additional Python modules like Pandas, Numpy, SciPy are allowed. You will get 0 for this assignment if you use any of these modules.

For the dataset, we use olympic medals dataset from Kaggle: <https://www.kaggle.com/ramontanoeiro/summer-olympic-medals-1986-2020>

**Dataset Description:** From the Kaggle website: "This dataset contains all medals awarded for all participating countries from 1896 up to 2020(1). Data was scraped from Wikipedia pages containing the tables of medals. This dataset was used to create this animated bar chart race:

<https://www.youtube.com/watch?v=81ZJL07GL4s>

A data frame with 1345 rows (including headers) and 8 variables"

1. **Year:** date of games (1896-2020) no games in 1916 (World War I) and 1940, 1944 (World War II)
2. **Country\_Code:** 3 Letters Code used to identify the country

3. **Country\_Name**: quality of the cut (Fair, Good, Very Good, Premium, Ideal)
4. **Host\_city**: City of the Edition
5. **Host\_country**: a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
6. **Gold**: Number of Gold medals awarded
7. **Silver**: Number of Silver medals awarded
8. **Bronze**: Number of Bronze medals awarded

You will use a subset of this dataset depending on the last digit of your BU ID as follows:

1. gold medals if your BU ID ends on 0, 1, 2, 3 (Group 1)
2. silver medals if your BU ID ends on 4, 5, 6 (Group 2)
3. bronze medals if your BU ID ends on 7, 8, 9 (Group 3)

**Question 1:**

1. load the "country\_medals" csv file as a list of lines using Python and construct a sublist for you group
2. how many entries are there?
3. compute the average numbers of medals per country and write this (in decreasing order) to a file "average\_medals\_per\_country.csv" for your group
4. which country has the highest (average) number of medals?
5. list top 10 countries by (averaged) number of medals
6. compute the median number of medals per country and write this (in decreasing order) to a file "median\_medals\_per\_country.csv" for your group
7. which country has the highest median number of medals?
8. list top 10 countries by median number of medals