**Top Olympic Medal Winner Countries**

This project is about finding the top 5 TOTAL Olympic medals winning countries for the summer games, winter games and both summer and winter games combined from 1896 to 2018 (124 years).

The Summer Olympic Games, also known as the Games of the Olympiad, are a major international multi-sport event normally held once every four years. The Games were first held in 1896 in Athens, Greece, and were most recently the 2020 Summer Olympics held in 2021 in Tokyo, Japan.

The International Olympic Committee (IOC) organizes the Games and oversees the host city's preparations. In each Olympic event, gold medals are awarded for first place, silver medals are awarded for second place, and bronze medals are awarded for third place; this tradition began in 1904.

This project is built to meet the following requirements of “Python Dara Analysis 1” of Code Louisville January 2021 session:

1. Read data from an external file, such as text, JSON, CSV, etc. and use that data in your application

2. Visualize data in a graph, chart, or other visual representation of data

3. Create a dictionary or list, populate it with several values, retrieve at least one value, and use it in your program

4. Create and call at least 3 functions or methods, at least one of which must return a value that is used somewhere else in your code. To clarify, at least one function should be called in your code, that function should calculate, retrieve, or otherwise set the value of a variable or data structure, return a value to where it was called, and use that value somewhere else in your code. For example, you could create a function that reads how many items there are in a text file, returns that value, and later uses that value to execute a loop a certain number of times.

5. Use pandas, matplotlib, and/or numpy to perform a data analysis project. Ingest 2 or more pieces of data, analyze that data in some manner, and display a new result to a graph, chart, or other display

More information about the three datasets:

The three datasets were downloaded from two different pages on Kaggle website (<https://www.kaggle.com/code/arpitsolanki14/olympic-games-data-analysis/data> “summer.csv” and “winter.csv”

and

<https://www.kaggle.com/datasets/mysarahmadbhat/120-years-of-olympic-history?select=country_definitions.csv> “country\_definitions.csv).

This project was written in the Jupyter notebook. The following packages were imported in this project:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

How to run this project:

1. Start with .ipynb file "medals.ipynb".

2. You must install jupyter labs to run the code in the jupyter notebook (visit https://jupyterlabs.org/install) or run jupyter notebooks in a browser.

3. Once in the jupyter notebook, navigate to the directory where you downloaded the project files. (You should see the .ipynb file, and "data" folder needed to run this project).

4. Open the .ipynb file.

5. You may need to install the above-mentioned 4 packages. Open terminal and run the following commands to install each package:

Pip install pandas

Pip install numpy

pip install matplotlib.pyplot

pip install seaborn

6. Go back to the jupyter notebooks. You see we are now importing these packages. So you should be able to run the code now.

7. Select "Kernel" (top ribbon) and "Restart and run all cells."

8. The script should read in the two csv files, from the "data" folder ("annual\_aqi\_by\_cbsa\_2020.csv" and "annual\_aqi\_by\_cbsa\_2019.csv")

9. See comments in each code block (notebook cell) for further details about the project. You will see the features listed in the comments such as the dictionary and data visualization.

For visual studio (VS) Code users on Windows, make sure you get the python and Jupyter extensions.