

Open the file

My BU ID is U64501194 Therefore, I am dealing with the silver medals

1. load the "country medals" csv as a list of lines using Python and construct a sublist for you group

```
In [35]: # -*- coding: utf-8 -*-
        """
        Created on Mon Nov 5 14:37:29 2018

        @author: epinsky
        this scripts reads your ticker file (e.g. MSFT.csv) and
        constructs a list of lines
        """
        import os

        ticker='Country_Medals'
        input_dir = os.getcwd()
        ticker_file = os.path.join(input_dir, ticker + '.csv')

        try:
            with open(ticker_file) as f:
                lines = f.read().splitlines()
                print('opened file for ticker:', ticker)
                """ your code for assignment 1 goes here """

        except Exception as e:
            print(e)
            print('failed to read stock data for ticker: ', ticker)
```

opened file for ticker: Country_Medals

```
In [36]: lines[0:5]

Out[36]: ['Year,Country_Code,Country_Name,Host_city,Host_country,Gold,Silver,Bronze',
          '1932,(USA),United States,Los Angeles,United States,41,32,30',
          '1932,(ITA),Italy,Los Angeles,United States,12,12,12',
          '1932,(FRA),France,Los Angeles,United States,10,5,4',
          '1932,(SWE),Sweden,Los Angeles,United States,9,5,9']
```

Extract the data

```
In [37]: # Class a data type (year, countryName, silverMetal)
        class DataNode:
            def __init__(self, year, countryName, silver):
                self.year = year
                self.countryName = countryName
                self.silver = silver
            def show(self):
                print("{:5} {:15} {:3}".format(self.year, self.countryName, self.silver))

In [38]: # Split the raw data into list
        table = []
        for line in lines:
            temp = line.split(',')
            table.append(temp)

        # Extract the raw data
        extractData = []
        for row in table:
            y = row[0]
            cn = row[2]
            s = row[6]
            dataNode = DataNode(y, cn, s)
            extractData.append(dataNode)

        # Remove the title from the extractData
        extractData = extractData[1:]

In [39]: # Show the first 20 row of the extracted data
        for i in range(20):
            extractData[i].show()

1932 United States 32
1932 Italy 12
1932 France 5
1932 Sweden 5
1932 Japan 7
1932 Hungary 4
1932 Finland 8
1932 Great Britain 7
1932 Germany 12
1932 Australia 1
1932 Argentina 1
1932 Canada 5
1932 Netherlands 5
1932 Poland 1
1932 South Africa 0
1932 Ireland 0
1932 Czechoslovakia 2
1932 Austria 1
1932 India 0
1932 Denmark 3
```

Analysis

2. how many entries are there?

```
In [40]: # How many entries
        entries = len(extractData)
        print("we have {} entries of data and 1 row of title".format(entries))

we have 1344 entries of data and 1 row of title
```

Calculate the average medals for each country

```
In [41]: # Sort the list with the key 'countryName'
        sortData = sorted(extractData, key=lambda x:x.countryName)
        # Show first 20 data
        for node in sortData[0:20]:
            node.show()

2012 Afghanistan 0
2008 Afghanistan 0
1984 Algeria 0
2016 Algeria 2
2012 Algeria 0
2008 Algeria 1
2000 Algeria 1
1996 Algeria 0
1992 Algeria 0
1932 Argentina 1
1928 Argentina 3
1924 Argentina 3
1972 Argentina 1
1968 Argentina 0
1964 Argentina 1
1960 Argentina 1
2020 Argentina 1
2016 Argentina 1
2012 Argentina 1
2008 Argentina 0
```

```
In [42]: # Extract the country from data
        temp = []
        for node in sortData:
            c = node.countryName
            temp.append(c)

        # Processing the data into the form (countryname, list(a series of medal wins)
        countrySet = sorted(set(temp))
        medalList = [[] for i in range(len(countrySet))]

        for node in sortData:
            n = node.countryName
            s = node.silver
            index = countrySet.index(n)
            medalList[index].append(int(s))

        countryMedalData = list(zip(countrySet, medalList))

In [43]: # Print the first 20 countryMedalData
        for data in countryMedalData[0:20]:
            print(data)

('Afghanistan', [0, 0])
('Algeria', [0, 2, 0, 1, 1, 0, 0])
('Argentina', [1, 3, 3, 1, 0, 1, 1, 1, 1, 1, 0, 0, 2, 2, 0, 1, 1, 2, 3, 2])
('Armenia', [2, 3, 1, 1, 0, 1])
('Australasia', [2, 2])
('Australia', [1, 2, 1, 2, 3, 0, 0, 8, 2, 1, 7, 7, 2, 8, 7, 11, 15, 15, 16, 25, 9, 9, 6, 8, 2, 6, 0])
('Austria', [1, 0, 3, 2, 0, 1, 3, 1, 1, 2, 0, 1, 2, 1, 1, 0, 1, 4, 1, 1, 2, 0, 0, 1, 0, 6])
('Azerbaijan', [3, 7, 2, 1, 0, 0, 1])
('Bahamas', [0, 0, 0, 0, 1, 0, 0, 1, 0, 0])
('Bahrain', [1, 1, 0])
('Barbados', [0])
('Belarus', [3, 4, 5, 4, 5, 3, 6])
('Belgium', [1, 7, 12, 1, 5, 7, 1, 0, 3, 2, 1, 0, 2, 1, 2, 1, 0, 0, 2, 2, 1, 0, 2, 2, 0])
('Bermuda', [0, 0])
('Bohemia', [0, 1])
('Botswana', [0, 1])
('Brazil', [1, 5, 0, 0, 0, 1, 0, 0, 6, 6, 5, 4, 2, 6, 3, 1, 2, 0, 0, 0])
('British West Indies', [0])
('Bulgaria', [16, 9, 10, 4, 5, 3, 1, 1, 2, 1, 1, 6, 7, 12, 3, 0])
('Burkina Faso', [0])
```

3. compute the average numbers of medals per country and write this (in decreasing order) to a le "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

```
In [44]: # Average medal for each country
        averageMedal = []
        for data in countryMedalData:
            name = data[0]
            medal = data[1]
            aveMedal = sum(data[1]) // len(data[1])
            averageMedal.append((name, aveMedal))

In [45]: # Sorting the averageMedal in decreasing ordert
        temp = sorted(averageMedal, key=lambda x:x[1], reverse = True)
        print(temp)

[('Unified Team', 38), ('Soviet Union', 35), ('United States', 29), ('ROC', 28), ('East Germany', 25), ('Russia', 20), ('China', 19), ('United Team of Germany', 18), ('West Germany', 13), ('Germany', 12), ('Great Britain', 11), ('France', 8), ('Australia', 6), ('Italy', 6), ('Japan', 6), ('Sweden', 6), ('Bulgaria', 5), ('Ukraine', 5), ('Hungary', 5), ('South Korea', 5), ('Ukraine', 5), ('Belarus', 4), ('Canada', 4), ('Cuba', 4), ('Romania', 4), ('Netherlands', 4), ('Poland', 4), ('Romania', 4), ('Czech Republic', 3), ('Czechoslovakia', 3), ('Finland', 3), ('Kazakhstan', 3), ('Kenya', 3), ('Spain', 3), ('Australasia', 2), ('Azerbaijan', 2), ('Belgium', 2), ('Brazil', 2), ('Denmark', 2), ('Greece', 2), ('Jamaica', 2), ('Mixed team', 2), ('Norway', 2), ('Russian Empire', 2), ('Serbia and Montenegro', 2), ('Slovakia', 2), ('Switzerland', 2), ('Tanzania', 2), ('Argentina', 1), ('Armenia', 1), ('Austria', 1), ('Ceylon', 1), ('Chile', 1), ('Chinese Taipei', 1), ('Colombia', 1), ('Croatia', 1), ('Cyprus', 1), ('Egypt', 1), ('Gabon', 1), ('Georgia', 1), ('Guatemala', 1), ('Independent Olympic Participants', 1), ('Indonesia', 1), ('Iran', 1), ('Kyrgyzstan', 1), ('Latvia', 1), ('Malaysia', 1), ('Mexico', 1), ('Montenegro', 1), ('Namibia', 1), ('Netherlands Antilles', 1), ('New Zealand', 1), ('Nigeria', 1), ('North Korea', 1), ('North Macedonia', 1), ('Paraguay', 1), ('Samoa', 1), ('San Marino', 1), ('Senegal', 1), ('Serbia', 1), ('Slovenia', 1), ('South Africa', 1), ('Sri Lanka', 1), ('Sudan', 1), ('Tonga', 1), ('Turkey', 1), ('Turkmenistan', 1), ('Virgin Islands', 1), ('Yugoslavia', 1), ('Zimbabwe', 1), ('Zimbabwe', 1), ('Zimbabwe', 1), ('Afghanistan', 0), ('Algeria', 0), ('Bahamas', 0), ('Bahrain', 0), ('Barbados', 0), ('Bermuda', 0), ('Bohemia', 0), ('Botswana', 0), ('British West Indies', 0), ('Burkina Faso', 0), ('Burundi', 0), ('Cameroon', 0), ('Costa Rica', 0), ('Djibouti', 0), ('Dominican Republic', 0), ('Ecuador', 0), ('Eritrea', 0), ('Estonia', 0), ('Ethiopia', 0), ('Fiji', 0), ('Ghana', 0), ('Grenada', 0), ('Guyana', 0), ('Haiti', 0), ('Hong Kong', 0), ('Iceland', 0), ('Independent Olympic Athletes', 0), ('India', 0), ('Iraq', 0), ('Ireland', 0), ('Israel', 0), ('Ivory Coast', 0), ('Jordan', 0), ('Kosovo', 0), ('Kuwait', 0), ('Lebanon', 0), ('Lithuania', 0), ('Luxembourg', 0), ('Macedonia', 0), ('Mauritius', 0), ('Moldova', 0), ('Mongolia', 0), ('Morocco', 0), ('Mozambique', 0), ('Niger', 0), ('Pakistan', 0), ('Panama', 0), ('Peru', 0), ('Philippines', 0), ('Portugal', 0), ('Puerto Rico', 0), ('Republic of China', 0), ('Saudi Arabia', 0), ('Singapore', 0), ('Suriname', 0), ('Syria', 0), ('Tajikistan', 0), ('Thailand', 0), ('Togo', 0), ('Trinidad and Tobago', 0), ('Tunisia', 0), ('Uganda', 0), ('United Arab Emirates', 0), ('Uruguay', 0), ('Uzbekistan', 0), ('Venezuela', 0), ('Vietnam', 0), ('Zambia', 0)]
```

```
In [46]: print("\n\n\n{} has won the most averaged silver medal".format(temp[0][0]), )
        print("Unified Team only once participated the olympic game in 1992, and the team won 38 silver metals")

Unified Team has won the most averaged silver medal
Unified Team only once participated the olympic game in 1992, and the team won 38 silver metals
```

```
In [47]: # Wirte the average data into a new cvs name ""average medals per country.csv"
        import csv
        try:
            with open('average_medals_per_country.csv', 'w', newline = '') as file:
                writer = csv.writer(file)
                writer.writerow(['Country Name', 'Average Silver Medal'])
                for data in temp:
                    name = data[0]
                    medal = data[1]
                    writer.writerow([name, medal])
                print('\nOutput csv SUCCESSFUL\n')
        except:
            print("Something wrong when output the csv file")

Output csv SUCCESSFUL
```

```
In [48]: # List top 10 countries by (average) number of medals
        for data in temp[0:10]:
            print("{:25} {}".format(data[0], data[1]))

Unified Team 38
Soviet Union 35
United States 29
ROC 28
East Germany 25
Russia 20
China 19
United Team of Germany 18
West Germany 13
Germany 12
```

6. compute the median number of medals per country and write this (in decreasing order) to a le "median medals per country.csv" for your group

```
In [49]: def medianNum(inputList):
        inputList = sorted(inputList)
        l = len(inputList)
        if l % 2 == 0:
            index = l // 2
            return (inputList[index] + inputList[index - 1]) // 2

        else:
            index = l // 2
            return inputList[index]
```

```
In [50]: medMedal = []
        for data in countryMedalData:
            name = data[0]
            medal = data[1]
            l = len(medal)
            median = medianNum(medal)
            medMedal.append((name, median))

        medMedal = sorted(medMedal, key=lambda d:d[1], reverse=True)
        print(medMedal)

[('Unified Team', 38), ('Soviet Union', 31), ('ROC', 28), ('United States', 27), ('East Germany', 25), ('China', 20), ('Russia', 20), ('United Team of Germany', 19), ('West Germany', 12), ('Germany', 11), ('Great Britain', 9), ('Italy', 8), ('Japan', 7), ('Australia', 6), ('France', 6), ('Hungary', 5), ('Sweden', 5), ('Ukraine', 5), ('Belarus', 4), ('Bulgaria', 4), ('Czech Republic', 4), ('Kenya', 4), ('Poland', 4), ('Romania', 4), ('South Korea', 4), ('Canada', 3), ('Cuba', 3), ('Czech Republic', 3), ('Czechoslovakia', 3), ('Netherlands', 3), ('Australasia', 2), ('Denmark', 2), ('Finland', 2), ('Georgia', 2), ('Jamaica', 2), ('Namibia', 2), ('Norway', 2), ('Russian Empire', 2), ('Serbia and Montenegro', 2), ('Slovakia', 2), ('Spain', 2), ('Switzerland', 2), ('Tanzania', 2), ('Argentina', 1), ('Armenia', 1), ('Austria', 1), ('Azerbaijan', 1), ('Bahrain', 1), ('Belgium', 1), ('Brazil', 1), ('Ceylon', 1), ('Chile', 1), ('Chinese Taipei', 1), ('Colombia', 1), ('Croatia', 1), ('Cyprus', 1), ('Ecuador', 1), ('Egypt', 1), ('Estonia', 1), ('Ethiopia', 1), ('Gabon', 1), ('Greece', 1), ('Guatemala', 1), ('Independent Olympic Participants', 1), ('Indonesia', 1), ('Iran', 1), ('Ireland', 1), ('Kyrgyzstan', 1), ('Latvia', 1), ('Lebanon', 1), ('Malaysia', 1), ('Mexico', 1), ('Mixed team', 1), ('Mongolia', 1), ('Montenegro', 1), ('Netherlands Antilles', 1), ('New Zealand', 1), ('Nigeria', 1), ('North Korea', 1), ('North Macedonia', 1), ('Paraguay', 1), ('Peru', 1), ('Samoa', 1), ('San Marino', 1), ('Saudi Arabia', 1), ('Senegal', 1), ('Serbia', 1), ('Slovenia', 1), ('South Africa', 1), ('Sri Lanka', 1), ('Sudan', 1), ('Tonga', 1), ('Turkey', 1), ('Turkmenistan', 1), ('Uganda', 1), ('Uzbekistan', 1), ('Vietnam', 1), ('Virgin Islands', 1), ('Yugoslavia', 1), ('Zimbabwe', 1), ('Zimbabwe', 1), ('Zimbabwe', 1), ('Afghanistan', 0), ('Algeria', 0), ('Bermuda', 0), ('Bohemia', 0), ('Botswana', 0), ('British West Indies', 0), ('Burkina Faso', 0), ('Burundi', 0), ('Cameroon', 0), ('Costa Rica', 0), ('Djibouti', 0), ('Dominican Republic', 0), ('Eritrea', 0), ('Fiji', 0), ('Ghana', 0), ('Grenada', 0), ('Guyana', 0), ('Haiti', 0), ('Hong Kong', 0), ('Iceland', 0), ('Independent Olympic Athletes', 0), ('India', 0), ('Iraq', 0), ('Israel', 0), ('Ivory Coast', 0), ('Jordan', 0), ('Kosovo', 0), ('Kuwait', 0), ('Lebanon', 0), ('Lithuania', 0), ('Luxembourg', 0), ('Macedonia', 0), ('Mauritius', 0), ('Moldova', 0), ('Morocco', 0), ('Mozambique', 0), ('Niger', 0), ('Pakistan', 0), ('Panama', 0), ('Philippines', 0), ('Portugal', 0), ('Puerto Rico', 0), ('Republic of China', 0), ('Saudi Arabia', 0), ('Singapore', 0), ('Suriname', 0), ('Syria', 0), ('Tajikistan', 0), ('Thailand', 0), ('Togo', 0), ('Trinidad and Tobago', 0), ('Tunisia', 0), ('Uganda', 0), ('United Arab Emirates', 0), ('Uruguay', 0), ('Uzbekistan', 0), ('Venezuela', 0), ('Vietnam', 0), ('Zambia', 0)]
```

```
In [51]: # Output csv file
        try:
            with open('median_medals_per_country.csv', 'w', newline = '') as file:
                writer = csv.writer(file)
                writer.writerow(['Country Name', 'median of Silver Medal'])
                for data in medMedal:
                    name = data[0]
                    m = data[1]
                    writer.writerow([name, m])
                print('\nOutput csv SUCCESSFUL\n')
        except:
            print("Something wrong when output the csv file")

Output csv SUCCESSFUL
```

7. which country has the highest median number of medals?

```
In [52]: print("{} has the highest median number of silver medals, which is {}".format(medMedal[0][0], medMedal[0][1]))

Unified Team has the highest median number of silver medals, which is 38
```

8. list top 10 countries by median number of medals

```
In [53]: for data in medMedal[0:10]:
        country = data[0]
        m = data[1]
        print("{:25} {}".format(country, m))

Unified Team 38
Soviet Union 31
ROC 28
United States 27
East Germany 25
China 20
Russia 20
United Team of Germany 19
West Germany 12
Germany 11
```