from pandas import DataFrame, Series

#################

# Syntax Reminder:

#

# The following code would create a two-column pandas DataFrame

# named df with columns labeled 'name' and 'age':

#

# people = ['Sarah', 'Mike', 'Chrisna']

# ages = [28, 32, 25]

# df = DataFrame({'name' : Series(people),

# 'age' : Series(ages)})

def create\_dataframe():

'''

Create a pandas dataframe called 'olympic\_medal\_counts\_df' containing

the data from the table of 2014 Sochi winter olympics medal counts.

The columns for this dataframe should be called

'country\_name', 'gold', 'silver', and 'bronze'.

There is no need to specify row indexes for this dataframe

(in this case, the rows will automatically be assigned numbered indexes).

You do not need to call the function in your code when running it in the

browser - the grader will do that automatically when you submit or test it.

'''

countries = ['Russian Fed.', 'Norway', 'Canada', 'United States',

'Netherlands', 'Germany', 'Switzerland', 'Belarus',

'Austria', 'France', 'Poland', 'China', 'Korea',

'Sweden', 'Czech Republic', 'Slovenia', 'Japan',

'Finland', 'Great Britain', 'Ukraine', 'Slovakia',

'Italy', 'Latvia', 'Australia', 'Croatia', 'Kazakhstan']

gold = [13, 11, 10, 9, 8, 8, 6, 5, 4, 4, 4, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0]

silver = [11, 5, 10, 7, 7, 6, 3, 0, 8, 4, 1, 4, 3, 7, 4, 2, 4, 3, 1, 0, 0, 2, 2, 2, 1, 0]

bronze = [9, 10, 5, 12, 9, 5, 2, 1, 5, 7, 1, 2, 2, 6, 2, 4, 3, 1, 2, 1, 0, 6, 2, 1, 0, 1]

# your code here

olympic\_medal\_counts\_df = DataFrame ({'country\_name': Series(countries),

'gold': Series(gold),

'silver': Series(silver),

'bronze': Series(bronze)})

return olympic\_medal\_counts\_df

import pandas as pd

'''

Also a couple pointers:

1) Selecting a single column from the DataFrame will return a Series

2) Selecting multiple columns from the DataFrame will return a DataFrame

# Change False to True to see Series indexing in action

if True:

data = {'year': [2010, 2011, 2012, 2011, 2012, 2010, 2011, 2012],

'team': ['Bears', 'Bears', 'Bears', 'Packers', 'Packers', 'Lions',

'Lions', 'Lions'],

'wins': [11, 8, 10, 15, 11, 6, 10, 4],

'losses': [5, 8, 6, 1, 5, 10, 6, 12]}

football = pd.DataFrame(data)

print football['year']

print ''

print football.year # shorthand for football['year']

print ''

print football[['year', 'wins', 'losses']]

'''

Row selection can be done through multiple ways.

Some of the basic and common methods are:

1) Slicing

2) An individual index (through the functions iloc or loc)

3) Boolean indexing

You can also combine multiple selection requirements through boolean

operators like & (and) or | (or)

'''

# Change False to True to see boolean indexing in action

if True:

data = {'year': [2010, 2011, 2012, 2011, 2012, 2010, 2011, 2012],

'team': ['Bears', 'Bears', 'Bears', 'Packers', 'Packers', 'Lions',

'Lions', 'Lions'],

'wins': [11, 8, 10, 15, 11, 6, 10, 4],

'losses': [5, 8, 6, 1, 5, 10, 6, 12]}

football = pd.DataFrame(data)

print football.iloc[[0]]

print ""

print football.loc[[0]]

print ""

print football[3:5]

print ""

print football[football.wins > 10]

print ""

print football[(football.wins > 10) & (football.team == "Packers")]

0 2010

1 2011

2 2012

3 2011

4 2012

5 2010

6 2011

7 2012

Name: year, dtype: int64

0 2010

1 2011

2 2012

3 2011

4 2012

5 2010

6 2011

7 2012

Name: year, dtype: int64

year wins losses

0 2010 11 5

1 2011 8 8

2 2012 10 6

3 2011 15 1

4 2012 11 5

5 2010 6 10

6 2011 10 6

7 2012 4 12

losses team wins year

0 5 Bears 11 2010

losses team wins year

0 5 Bears 11 2010

losses team wins year

3 1 Packers 15 2011

4 5 Packers 11 2012

losses team wins year

0 5 Bears 11 2010

3 1 Packers 15 2011

4 5 Packers 11 2012

losses team wins year

3 1 Packers 15 2011

4 5 Packers 11 2012

import numpy

from pandas import DataFrame, Series

def avg\_medal\_count():

'''

Using the dataframe's apply method, create a new Series called

avg\_medal\_count that indicates the average number of gold, silver,

and bronze medals earned amongst countries who earned at

least one medal of any kind at the 2014 Sochi olympics. Note that

the countries list already only includes countries that have earned

at least one medal. No additional filtering is necessary.

'''

countries = ['Russian Fed.', 'Norway', 'Canada', 'United States',

'Netherlands', 'Germany', 'Switzerland', 'Belarus',

'Austria', 'France', 'Poland', 'China', 'Korea',

'Sweden', 'Czech Republic', 'Slovenia', 'Japan',

'Finland', 'Great Britain', 'Ukraine', 'Slovakia',

'Italy', 'Latvia', 'Australia', 'Croatia', 'Kazakhstan']

gold = [13, 11, 10, 9, 8, 8, 6, 5, 4, 4, 4, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0]

silver = [11, 5, 10, 7, 7, 6, 3, 0, 8, 4, 1, 4, 3, 7, 4, 2, 4, 3, 1, 0, 0, 2, 2, 2, 1, 0]

bronze = [9, 10, 5, 12, 9, 5, 2, 1, 5, 7, 1, 2, 2, 6, 2, 4, 3, 1, 2, 1, 0, 6, 2, 1, 0, 1]

olympic\_medal\_counts = {'country\_name':countries,

'gold': Series(gold),

'silver': Series(silver),

'bronze': Series(bronze)}

olympic\_medal\_counts\_df = DataFrame(olympic\_medal\_counts)

# YOUR CODE HERE

avg\_medal\_count = olympic\_medal\_counts\_df[['gold', 'silver','bronze']].apply(numpy.mean)

return avg\_medal\_count

import numpy

from pandas import DataFrame, Series

def points():

'''

Imagine a point system in which each country is awarded 4 points for each

gold medal, 2 points for each silver medal, and one point for each

bronze medal.

Using the numpy.dot function, create a new dataframe called

'olympic\_points\_df' that includes:

a) a column called 'country\_name' with the country name

b) a column called 'points' with the total number of points the country

earned at the Sochi olympics.

'''

countries = ['Russian Fed.', 'Norway', 'Canada', 'United States',

'Netherlands', 'Germany', 'Switzerland', 'Belarus',

'Austria', 'France', 'Poland', 'China', 'Korea',

'Sweden', 'Czech Republic', 'Slovenia', 'Japan',

'Finland', 'Great Britain', 'Ukraine', 'Slovakia',

'Italy', 'Latvia', 'Australia', 'Croatia', 'Kazakhstan']

gold = [13, 11, 10, 9, 8, 8, 6, 5, 4, 4, 4, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0]

silver = [11, 5, 10, 7, 7, 6, 3, 0, 8, 4, 1, 4, 3, 7, 4, 2, 4, 3, 1, 0, 0, 2, 2, 2, 1, 0]

bronze = [9, 10, 5, 12, 9, 5, 2, 1, 5, 7, 1, 2, 2, 6, 2, 4, 3, 1, 2, 1, 0, 6, 2, 1, 0, 1]

medals = [gold,silver,bronze]

df = DataFrame({"country\_name":countries,

"gold":gold,

"silver":silver,

"bronze":bronze})

df['points'] = df[['gold','silver','bronze']].dot([4, 2, 1])

olympic\_points\_df = df[['country\_name','points']]

return olympic\_points\_df