Caitlin Sisilli

Homework 3

Question 1)

With choosing between dot notation and indexing operator, the preferred notation is dot notation, even though indexing operator will always work. Dot notation is more preferred because of multiple reason like, it is easier to remember to do since most programming with libraries we become use to using dot notation. As well as using dot notation, I know it is faster to write, as well as dot notation limits the number of brackets that can be used in code, so it become efficient to use less.

Question 2)

The object data type is used for columns that contain the data is: The answer is: A and B which is Strings and Numbers (Int and float).

Question 3)

A)

import numpy as np

import pandas as pd

df = pd.DataFrame(index=["BOP", "POP"], columns=["Foo", "Boo"])

with pd.ExcelWriter("file.xlsx") as writer:

df.to\_excel(writer)

B)

```
import numpy as np
import pandas as pd
df = pd.DataFrame(np.random.randn(5, 4),
           columns=list('ABCD'),
           index=pd.date_range('20130101', periods=5))
df
print(df.index)
Output:
DatetimeIndex(['2013-01-01', '2013-01-02', '2013-01-03', '2013-01-04',
        '2013-01-05'],
        dtype='datetime64[ns]', freq='D')
C)
import numpy as np
import pandas as pd
df = pd.DataFrame(np.random.randn(5, 4),
           columns=list('ABCD'),
           index=pd.date_range('20130101', periods=5))
df
```

```
print(df.loc['20130101'])
Output:
A 2.724066
B -0.521298
C -2.290498
D 0.544622
Name: 2013-01-01 00:00:00, dtype: float64
      D)
import numpy as np
import pandas as pd
df =
pd.DataFrame({'o':[70001,np.nan,70002,70004,np.nan,70005,np.nan,70010,70003,70012,np.
nan,70013]})
df['o'].fillna(df['o'].median(),inplace=True)
print(df)
Output:
     0
0 70001.0
1 70004.5
```

2 70002.0
3 70004.0
4 70004.5
5 70005.0
6 70004.5
7 70010.0
8 70003.0
9 70012.0
10 70004.5
11 70013.0
11 70013.0 E)
<b>E</b> )
E) import numpy as np
E) import numpy as np import pandas as pd
E) import numpy as np import pandas as pd $df = pd.DataFrame(\{"A": [1, 2, 3], "B": [4, 5, 6]\})$
E)  import numpy as np  import pandas as pd  df = pd.DataFrame({"A": [1, 2, 3], "B": [4, 5, 6]})  bf=df.rename(columns={"A": "a", "B": "c"})

```
Original
 A B
0 1 4
1 2 5
2 3 6
Rename
 a c
0 1 4
1 2 5
2 3 6
      F)
import numpy as np
import pandas as pd
df = pd.DataFrame(np.random.randn(5, 4),
          columns=list('ABCD'),
          index=pd.date_range('20130101', periods=5))
df
print(df.size)
```

```
Output:
20
                  Question 4)
A)
import numpy as np
import pandas as pd
frame= pd.read_table('subset-covid-data.csv',sep=',')
print(frame)
df=frame[["country", "continent"]]
print(df)
Output:

        country continent
        date day month
        year cases described

        Afghanistan
        Asia 2020-04-12 12 4 2020

        Albania
        Europe 2020-04-12 12 4 2020

        Algeria
        Africa 2020-04-12 12 4 2020

        Andorra
        Europe 2020-04-12 12 4 2020

        Angola
        Africa 2020-04-12 12 4 2020

        Ol Venezuela
        America 2020-04-12 12 4 2020

        O2 Vietnam
        Asia 2020-04-12 12 4 2020

        O3 Yemen
        Asia 2020-04-12 12 4 2020

        O4 Zambia
        Africa 2020-04-12 12 4 2020

        O5 Zimbabwe
        Africa 2020-04-12 12 4 2020

                                                                                                                                                                                   deaths \
                                                                                                                                                                                             34
                                                                                                                                                                                                                        3
0
                                                                                                                                                                                                                        0
                                                                                                                                                                                            17
1
2
                                                                                                                                                                                            64
                                                                                                                                                                                                                     19
3
                                                                                                                                                                                            21
4
                                                                                                                                                                                               0
                                                                                                                                                                                                                        0
                                                                                                                                                                                         ...
                                                                                                                                                                                          0
201
                                                                                                                                                                                                                        0
                                                                                                                                                                                             4
                                                                                                                                                                                                                        0
202
203
                                                                                                                                                                                           0
                                                                                                                                                                                                                        0
                                                                                                                                                                                           0
                                                                                                                                                                                                                        0
204
205
                                                                                                                                                                                               3
                                                                                                                                                                                                                        0
            country code population
0
                                        AFG 37172386.0
```

ALB 2866376.0

DZA 42228429.0

AGO 30809762.0

VEN 28870195.0

77006.0

AND

. . .

Albania Europe
Algeria Africa
Andorra Europe
Angola Africa
Angola Africa

Nenezuela America
Venezuela America
Vietnam Asia
Vemen Asia
Africa
Zoda Zambia Africa
Zota Zimbabwe Africa

[206 rows x 2 columns]

#### B)

import numpy as np

import pandas as pd

frame= pd.read\_table('subset-covid-data.csv',sep=',',index\_col='country')

print(frame)

df=frame.loc['Algeria','population']

print(df)

	continent	date	day	month	year	cases	deaths	\
country								
Afghanistan	Asia	2020-04-12	12	4	2020	34	3	
Albania	Europe	2020-04-12	12	4	2020	17	0	
Algeria	Africa	2020-04-12	12	4	2020	64	19	

Andorra	Europe Africa	2020-04-12 2020-04-12	12 12	4	2020 2020	21	2
Angola	Allica	2020-04-12	12	4	2020	U	U
• • •							
Venezuela	America	2020-04-12	12	4	2020	0	0
Vietnam	Asia	2020-04-12	12	4	2020	4	0
Yemen	Asia	2020-04-12	12	4	2020	0	0
Zambia	Africa	2020-04-12	12	4	2020	0	0
Zimbabwe	Africa	2020-04-12	12	4	2020	3	0

 country\_code
 population

 country
 Afghanistan
 AFG
 37172386.0

 Albania
 ALB
 2866376.0
 2866376.0

 Algeria
 DZA
 42228429.0
 42228429.0

 Andorra
 AND
 77006.0
 30809762.0

 Angola
 AGO
 30809762.0
 ...

 Venezuela
 VEN
 28870195.0
 Vietnam

 Vietnam
 VNM
 95540395.0
 Yemen
 YEM
 28498687.0

 Zambia
 ZMB
 17351822.0
 Zimbabwe
 ZWE
 14439018.0

[206 rows x 9 columns] 42228429.0

### C)

I wasn't sure if you wanted the index in country but I did both 3 and 2 using the index just in case

import numpy as np

import pandas as pd

frame= pd.read\_table('subset-covid-data.csv',sep=',',index\_col='country')

print(frame)

df=frame.iloc[50,3]

# print(df)

pf=frame.iloc[50,2]

print(pf)

# Output:

(	continent		date	day	month	year	cases	deaths	\
country									
Afghanistan	Asia		-04-12		4		34		
Albania	Europe		-04-12					_	
Algeria	Africa		-04-12		_		_		
Andorra	Europe		-04-12		4				
Angola	Africa	2020	-04-12	12	4	2020	0	0	
• • •									
Venezuela	America		-04-12		4			0	
Vietnam	Asia		-04-12	12	4	2020	4	0	
Yemen	Asia		-04-12		4		0	0	
Zambia	Africa	2020	-04-12			2020	0	0	
Zimbabwe	Africa	2020	-04-12	12	4	2020	3	0	
country Afghanistan Albania Algeria Andorra Angola Venezuela Vietnam Yemen Zambia Zimbabwe	A D A A V V V Y Z.	FG 3 LB ZA 4 ND GO 3 EN 2 NM 9 EM 2 MB 1	717238 286637 222842 7700 080976	6.0 6.0 9.0 6.0 2.0  5.0 7.0					
[206 rows x 4 12	9 columns]								

D)

import numpy as np

import pandas as pd

```
frame= pd.read_table('subset-covid-data.csv',sep=',',index_col='country')
```

## print(frame)

df=frame.iloc[203:206,7:9]

### print(df)

## Output:

continent	Ċ	late	day	mc	onth	year	С	ases	deaths	\	
country											
Afghanistan	Asia	202	0-04-	12	12		4	2020	34		3
Albania	Europe	202	0-04-	12	12		4	2020	17		0
Algeria	Africa	202	0-04-	12	12		4	2020	64		19
Andorra	Europe	202	0-04-	12	12		4	2020	21		2
Angola	Africa	202	0-04-	12	12		4	2020	0		0
						•					
Venezuela	America	202	0-04-	12	12		4	2020	0		0
Vietnam	Asia	202	0-04-	12	12		4	2020	4		0
Yemen	Asia	202	0-04-	12	12		4	2020	0		0
Zambia	Africa	202	0-04-	12	12		4	2020	0		0
Zimbabwe	Africa	202	0-04-	12	12		4	2020	3		0

	country_code	population
country		
Afghanistan	AFG	37172386.0
Albania	ALB	2866376.0
Algeria	DZA	42228429.0
Andorra	AND	77006.0
Angola	AGO	30809762.0
Venezuela	VEN	28870195.0
Vietnam	VNM	95540395.0
Yemen	YEM	28498687.0
Zambia	ZMB	17351822.0
Zimbabwe	ZWE	14439018.0

[206 rows x 9 columns]

country\_code population

country

Yemen YEM 28498687.0

```
Zambia ZMB 17351822.0 Zimbabwe ZWE 14439018.0 \mathbf{E}
```

import numpy as np

import pandas as pd

frame= pd.read\_table('subset-covid-data.csv',sep=',')

print(frame)

df=frame[frame['population'] >2500000, frame['cases']>3000]

print(df)

	country	continent	date	day	month	year	cases	deaths	\
0	Afghanistan	Asia	2020-04-12	12	4	2020	34	3	
1	Albania	Europe	2020-04-12	12	4	2020	17	0	
2	Algeria	Africa	2020-04-12	12	4	2020	64	19	
3	Andorra	Europe	2020-04-12	12	4	2020	21	2	
4	Angola	Africa	2020-04-12	12	4	2020	0	0	
201	Venezuela	America	2020-04-12	12	4	2020	0	0	
202	Vietnam	Asia	2020-04-12	12	4	2020	4	0	
203	Yemen	Asia	2020-04-12	12	4	2020	0	0	
204	Zambia	Africa	2020-04-12	12	4	2020	0	0	
205	Zimbabwe	Africa	2020-04-12	12	4	2020	3	0	
	gountry godo	nonulatio	n						

	country_code	populacion
0	AFG	37172386.0
1	ALB	2866376.0
2	DZA	42228429.0
3	AND	77006.0
4	AGO	30809762.0
201	VEN	28870195.0
202	VNM	95540395.0
203	YEM	28498687.0
204	ZMB	17351822.0
205	ZWE	14439018.0

```
[206 rows x 10 columns]
'(0
         True
        True
1
2
       True
3
       False
       True
       . . .
201
       True
202
       True
203
        True
204
        True
205
        True
Name: population, Length: 206, dtype: bool, 0 False
       False
2
       False
3
      False
4
      False
       . . .
201
     False
202
     False
203
      False
     False
204
205
     False
Name: cases, Length: 206, dtype: bool)' is an invalid key
      Question 5)
A and B)
import numpy as np
import pandas as pd
frame= pd.read_table('subset-covid-data.csv',sep=',')
print(frame)
cols=frame.columns
```

frames=frame.pop('country\_code',)

# print(frame)

frame=frame.join(frames).reindex(columns=cols)

print(frame)

	country	continent	date	day	month	year	cases	deaths	\
0	Afghanistan	Asia	2020-04-12	12	4	2020	34	3	
1	Albania	Europe	2020-04-12	12	4	2020	17	0	
2	Algeria	Africa	2020-04-12	12	4	2020	64	19	
3	Andorra	Europe	2020-04-12	12	4	2020	21	2	
4	Angola	Africa	2020-04-12	12	4	2020	0	0	
	• • •		• • •						
201			2020-04-12	12	4	2020	0	0	
202		Asia	2020-04-12	12	4	2020	4	0	
203	3 Yemen	Asia	2020-04-12	12	4	2020	0	0	
204	4 Zambia	Africa	2020-04-12	12	4	2020	0	0	
20	5 Zimbabwe	Africa	2020-04-12	12	4	2020	3	0	
	country_code	populatio	n						
0	AFG	37172386.	0						
1	ALB 2866376.0								
2	DZA 42228429.0								
3	AND	AND 77006.0							
4	AGO	30809762.	0						
			•						
201	l VEN	28870195.	0						
202	2 VNM	95540395.	0						
203	3 YEM	28498687.	0						
204	4 ZMB	17351822.	0						
20	5 ZWE	14439018.	0						
[20	06 rows x 10 co	olumns]							
	country	continent	date	day	month	year	cases	deaths	\
0	Afghanistan	Asia	2020-04-12	12	4	2020	34	3	
1	Albania	Europe	2020-04-12	12	4	2020	17	0	
2	Algeria	Africa	2020-04-12	12	4	2020	64	19	
3	Andorra	Europe	2020-04-12	12	4	2020	21	2	
4	Angola	Africa	2020-04-12	12	4	2020	0	0	

201	Venezuela	America	2020-04-12	12	4	2020	0	0	
202	Vietnam	Asia	2020-04-12	12	4	2020	4	0	
203	Yemen		2020-04-12	12	4	2020	0	0	
204	Zambia		2020-04-12	12	4	2020	0	0	
205	Zimbabwe		2020-04-12	12	4	2020	3	0	
	population								
0	37172386.0								
1	2866376.0								
2	42228429.0								
3	77006.0								
4	30809762.0								
201	28870195.0								
202	95540395.0								
203	28498687.0								
204	17351822.0								
205	14439018.0								
[206	rows x 9 col	Lumns]							
	country	continent	date	day	month	year	cases	deaths	\
0	Afghanistan	Asia	2020-04-12	12	4	2020	34	3	
1	Albania	Europe	2020-04-12	12	4	2020	17	0	
	Albania	паторс							
2	Algeria		2020-04-12	12	4	2020	64	19	
2		Africa		12 12	4	2020 2020	64 21	19 2	
	Algeria	Africa Europe	2020-04-12						
3	Algeria Andorra Angola	Africa Europe Africa	2020-04-12 2020-04-12 2020-04-12	12 12	4	2020 2020 	21	2	
3 4  201	Algeria Andorra Angola	Africa Europe Africa America	2020-04-12 2020-04-12 2020-04-12  2020-04-12	12 12  12	4	2020 2020  2020	21	2	
3 4	Algeria Andorra Angola	Africa Europe Africa America Asia	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12	12 12  12 12	4 4 ••• 4 4	2020 2020 	21 0	2 0	
3 4  201	Algeria Andorra Angola  Venezuela	Africa Europe Africa America Asia Asia	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12 2020-04-12	12 12  12 12	4 4 ••• 4 4	2020 2020  2020	21 0 	2 0 ···	
3 4  201 202	Algeria Andorra Angola  Venezuela Vietnam	Africa Europe Africa America Asia Asia	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12	12 12  12 12	4  4 4 4	2020 2020  2020 2020 2020 2020	21 0  0 4	2 0 0	
3 4  201 202 203	Algeria Andorra Angola Venezuela Vietnam Yemen Zambia	Africa Europe Africa America Asia Asia Africa	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12 2020-04-12	12 12  12 12	4  4 4 4	2020 2020  2020 2020 2020	21 0  0 4 0	2 0  0 0	
3 4  201 202 203 204	Algeria Andorra Angola Venezuela Vietnam Yemen Zambia Zimbabwe	Africa Europe Africa America Asia Asia Africa Africa	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12	12 12  12 12 12 12	4  4 4 4	2020 2020  2020 2020 2020 2020	21 0  0 4 0	2 0  0 0 0	
3 4  201 202 203 204 205	Algeria Andorra Angola Venezuela Vietnam Yemen Zambia Zimbabwe country_code	Africa Europe Africa America Asia Asia Africa Africa populatio	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12 2020-04-12 2020-04-12	12 12  12 12 12 12	4  4 4 4	2020 2020  2020 2020 2020 2020	21 0  0 4 0	2 0  0 0 0	
3 4  201 202 203 204	Algeria Andorra Angola Venezuela Vietnam Yemen Zambia Zimbabwe country_code AFG	Africa Europe Africa America Asia Asia Africa Africa 7172386.	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12 2020-04-12 2020-04-12	12 12  12 12 12 12	4  4 4 4	2020 2020  2020 2020 2020 2020	21 0  0 4 0	2 0  0 0 0	
3 4  201 202 203 204 205	Algeria Andorra Angola Venezuela Vietnam Yemen Zambia Zimbabwe country_code	Africa Europe Africa America Asia Asia Africa Africa 2866376.	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12 2020-04-12 2020-04-12	12 12  12 12 12 12	4  4 4 4	2020 2020  2020 2020 2020 2020	21 0  0 4 0	2 0  0 0 0	
3 4  201 202 203 204 205	Algeria Andorra Angola Venezuela Vietnam Yemen Zambia Zimbabwe country_code AFG	Africa Europe Africa America Asia Asia Africa Africa 2866376. 42228429.	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12	12 12  12 12 12 12	4  4 4 4	2020 2020  2020 2020 2020 2020	21 0  0 4 0	2 0  0 0 0	
3 4  201 202 203 204 205	Algeria Andorra Angola Venezuela Vietnam Yemen Zambia Zimbabwe  country_code AFG ALB	Africa Europe Africa America Asia Asia Africa Africa 2866376. 42228429. 77006.	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12	12 12  12 12 12 12	4  4 4 4	2020 2020  2020 2020 2020 2020	21 0  0 4 0	2 0  0 0 0	
3 4  201 202 203 204 205	Algeria Andorra Angola Venezuela Vietnam Yemen Zambia Zimbabwe  country_code AFG ALB DZA	Africa Europe Africa America Asia Asia Africa Africa 2866376. 42228429.	2020-04-12 2020-04-12 2020-04-12  2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12	12 12  12 12 12 12	4  4 4 4	2020 2020  2020 2020 2020 2020	21 0  0 4 0	2 0  0 0 0	
3 4  201 202 203 204 205	Algeria Andorra Angola Venezuela Vietnam Yemen Zambia Zimbabwe  country_code AFG ALB DZA AND	Africa Europe Africa America Asia Asia Africa Africa 286376. 42228429. 77006. 30809762.	2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12	12 12  12 12 12 12	4  4 4 4	2020 2020  2020 2020 2020 2020	21 0  0 4 0	2 0  0 0 0	
3 4  201 202 203 204 205	Algeria Andorra Angola Venezuela Vietnam Yemen Zambia Zimbabwe  country_code AFG ALB DZA AND AGO	Africa Europe Africa America Asia Asia Africa Africa 286376. 42228429. 77006. 30809762.	2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12 2020-04-12	12 12  12 12 12 12	4  4 4 4	2020 2020  2020 2020 2020 2020	21 0  0 4 0	2 0  0 0 0	

203 YEM 28498687.0

```
204 ZMB 17351822.0
205 ZWE 14439018.0
[206 rows x 10 columns]
```

C and D) Need to do D

import numpy as np

import pandas as pd

frame= pd.read\_table('subset-covid-data.csv',sep=',')

print(frame)

frames= frame.drop(frame.index[[0,1,2]])

print(frames)

	country conti	inent	date day	month	yea	r cases	deaths	\
0	Afghanistan	Asia	2020-04-12	12	4	2020	34	3
1	Albania	Europe	2020-04-12	12	4	2020	17	0
2	Algeria	Africa	2020-04-12	12	4	2020	64	19
3	Andorra	Europe	2020-04-12	12	4	2020	21	2
4	Angola	Africa	2020-04-12	12	4	2020	0	0
201	Venezuela	America	2020-04-12	12	4	2020	0	0
202	Vietnam	Asia	2020-04-12	12	4	2020	4	0
203	Yemen	Asia	2020-04-12	12	4	2020	0	0
204	Zambia	Africa	2020-04-12	12	4	2020	0	0
205	Zimbabwe	Africa	2020-04-12	12	4	2020	3	0

```
country_code population
0 AFG 37172386.0
1 ALB 2866376.0
2 DZA 42228429.0
3 AND 77006.0
4 AGO 30809762.0
```

```
VEN 28870195.0
201
202
           VNM 95540395.0
203
           YEM 28498687.0
204
           ZMB
               17351822.0
205
               14439018.0
           ZWE
[206 rows x 10 columns]
               country continent date day month year cases
               Andorra Europe 2020-04-12 12
                                               4 2020
                                                              21
4
                Angola
                        Africa 2020-04-12 12
                                                   4 2020
                                                              0
5
              Anguilla America 2020-04-12 12
                                                   4 2020
                                                              0
    Antigua_and_Barbuda America 2020-04-12 12
Argentina America 2020-04-12 12
6
                                                   4 2020
                                                              0
7
                                                   4 2020
                                                             162
                 ...
                                                 . . .
                                          . . .
201
             Venezuela America 2020-04-12
                                           12
                                                  4 2020
                                                              0
202
                        Asia 2020-04-12 12
                                                   4 2020
               Vietnam
                          Asia 2020-04-12 12
                                                   4 2020
                                                              0
203
                 Yemen
                                                              0
204
                Zambia Africa 2020-04-12 12
                                                  4 2020
                                                              3
205
              Zimbabwe Africa 2020-04-12 12
                                                  4 2020
    deaths country code population
3
       2
                  AND 77006.0
4
        0
                  AGO 30809762.0
5
        0
                 NaN
                             NaN
                 ATG
6
         0
                          96286.0
7
        7
                 ARG 44494502.0
                  . . .
. .
       . . .
201
                 VEN 28870195.0
       0
202
        0
                  VNM 95540395.0
        0
203
                  YEM 28498687.0
204
                  ZMB 17351822.0
205
        0
                  ZWE 14439018.0
```

[203 rows x 10 columns]