Experiment 22

Out[5]:

Design a Python script to generate statistical reports (Minimum, Maximum, Count, Average, Sum etc) on public datasets.

```
In [1]:
          import pandas as pd
          data=pd.read_csv("ebola_2014_2016_clean.csv")
In [2]:
          data.info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 2485 entries, 0 to 2484
        Data columns (total 4 columns):
              Column
                                                                              Non-Null Count Dtype
          #
              -----
                                                                              -----
          0
              Country
                                                                              2485 non-null
                                                                                               object
                                                                              2485 non-null
                                                                                               object
          1
              Date
              Cumulative no. of confirmed, probable and suspected cases
                                                                              2477 non-null
                                                                                               float64
              Cumulative no. of confirmed, probable and suspected deaths 2485 non-null
                                                                                               float64
         dtypes: float64(2), object(2)
        memory usage: 77.8+ KB
In [3]:
          data.dtypes
Out[3]: Country
                                                                            object
                                                                            object
        Date
        Cumulative no. of confirmed, probable and suspected cases
                                                                           float64
        Cumulative no. of confirmed, probable and suspected deaths
                                                                           float64
        dtype: object
In [4]:
          data.head()
Out[4]:
                             Cumulative no. of confirmed, probable
                                                                  Cumulative no. of confirmed, probable
            Country
                      Date
                                             and suspected cases
                                                                                 and suspected deaths
                      2014-
             Guinea
                                                          648.0
                                                                                               430.0
                      08-29
                      2014-
             Nigeria
                                                           19.0
                                                                                                 7.0
         1
                      08-29
                      2014-
              Sierra
         2
                                                         1026.0
                                                                                               422.0
              Leone
                     08-29
                      2014-
             Liberia
         3
                                                         1378.0
                                                                                               694.0
                      08-29
                      2014-
              Sierra
                                                         1261.0
                                                                                               491.0
              Leone
                     09-05
In [5]:
          data.tail()
```

2480 Liberia 2016- 03-23 1.0 2481 Italy 2016- 03-23 5.0 2482 Liberia 2016- 03-23 20.0 2483 Nigeria 2016- 03-23 20.0 2484 United States of America 2014-08-29 1 Nigeria 2014-08-29 2 Sierra Leone 2014-08-29 1 Nigeria 2014-08-29 2 Sierra Leone 2014-08-29 3 Liberia 2014-08-29 4 Sierra Leone 2014-09-05 2480 Liberia 2016-03-23 2481 Tally 2016-03-23 2482 Liberia 2016-03-23 2483 Nigeria 2016-03-23 2484 United States of America 2016-03-23 2485 Liberia 2016-03-23 2486 Liberia 2016-03-23 2487 Tally 2016-03-23 2488 Liberia 2016-03-23 2489 Liberia 2016-03-23 2480 Liberia 2016-03-23 2481 Tally 2016-03-23 2482 Liberia 2016-03-23 2483 Nigeria 2016-03-23 2484 United States of America 2016-03-23 2484 United States of America 2016-03-23 2485 Liberia 2016-03-23 2486 Liberia 2016-03-23 2487 Liberia 2016-03-23 2488 Liberia 2016-03-23 2489 Liberia 2016-03-25 2480 Liberia 2016-03-25 2480 Liberia 2016-03-25 2481 Liberia 2016-03-23 2482 Liberia 2016-03-23 2483 Nigeria 2016-03-23 2484 United States of America 2016-03-23 2488 Liberia 2016-03-23 2489 Liberia 2016-03-24 2481 Liberia 2016-03-24 2481 Liberia 2016-03-25 2482 Liberia 2016-03-25 2483 Liberia 2016-03-23 2484 United States of America 2016-03-23 2484 United States of America 2016-03-23 2485 Liberia 2016-03-23 2486 Liberia 2016-03-23 2487 Liberia 2016-03-23 2488 Liberia 2016-03-23 2489 Liberia 2016-03-23 2480 Liberia 2016-03		Country	Date	Cumulative no. of conf probable and suspected		Cumulative no. probable and sus	
Tally 03-23 1.0	480	Liberia		1	0666.0		4806.0
2482 Liberia 03-23 5.0 2483 Nigeria 03-23 2016-03-23 20.0 2484 United States of America 03-23 4.0 data.columns Index(['Country', 'Date', 'Cumulative no. of confirmed, probable and suspected cases', 'Cumulative no. of confirmed, probable and suspected deaths'], dtype='object') data.items cbound method DataFrame.items of Country Date 0 Ginea 2014-08-29 1 Nigeria 2014-08-29 2 Sierra Leone 2014-08-29 3 Liberia 2014-08-29 4 Sierra Leone 2014-08-29 4 Sierra Leone 2014-08-29 4 Sierra Leone 2014-08-29 5 Liberia 2016-03-23 1 Liberia 2016-03-23 1 Liberia 2016-03-23 2481 Taly 2016-03-23 2482 Liberia 2016-03-23 2483 Nigeria 2016-03-23 2484 United States of America 2016-03-23 Cumulative no. of confirmed, probable and suspected cases \ 0 648.0 19.0 2 1 1026.0 1378.0 1378.0 1261.0 2480 10666.0 1778.0 10666.0 2481 1.0 2683 20.0 2484 2482 5.0 20.0 2484 2484 2483 20.0 0	481	Italy			1.0		0.0
United States of America O3-23 O	482	Liberia			5.0		4.0
data.columns Index(['Country', 'Date',	483	Nigeria			20.0		8.0
Index(['Country', 'Date',	484				4.0		1.0
'Cumulative no. of confirmed, probable and suspected cases', 'Cumulative no. of confirmed, probable and suspected deaths'], data.items cbound method DataFrame.items of	data.cc	olumns					
<pre>cbound method DataFrame.items of</pre>		'Cumulative	e no. of cone e no. of cone				
0	data.it	tems					
0 648.0 1 19.0 2 1026.0 3 1378.0 4 1261.0 2480 10666.0 2481 1.0 2482 5.0 2483 20.0 2484 4.0	••• 480 481 482 483 484 Ur	nited State	Guine Nigeri Sierra Leon Liberi Sierra Leon Liberi Ital Liberi Nigeri es of Americ	a 2014-08-29 a 2014-08-29 e 2014-08-29 a 2014-08-29 e 2014-09-05 a 2016-03-23 y 2016-03-23 a 2016-03-23 a 2016-03-23 a 2016-03-23			te \
0 430.0 1 7.0 2 422.0 3 694.0 4 491.0	 480 481 482 483 484				648.0 19.0 1026.0 1378.0 1261.0 10666.0 1.0 5.0 20.0 4.0 suspected of 430.0 7.0 422.0 694.0 491.0		

```
2482
                                                                   4.0
          2483
                                                                   8.0
          2484
                                                                   1.0
          [2485 rows x 4 columns]>
 In [8]:
           data.describe()
 Out[8]:
                     Cumulative no. of confirmed, probable and
                                                                 Cumulative no. of confirmed, probable and
                                                                                        suspected deaths
                                             suspected cases
          count
                                                2477.000000
                                                                                             2485.000000
                                                2553.678644
                                                                                             1028.347686
          mean
            std
                                                4427.118148
                                                                                             1656.064372
            min
                                                   0.000000
                                                                                                0.000000
           25%
                                                   1.000000
                                                                                                0.000000
           50%
                                                   8.000000
                                                                                                6.000000
           75%
                                                3657.000000
                                                                                             2386.000000
            max
                                               14122.000000
                                                                                             4806.000000
 In [9]:
           data.count(axis=0)
 Out[9]: Country
                                                                              2485
                                                                               2485
          Cumulative no. of confirmed, probable and suspected cases
                                                                              2477
          Cumulative no. of confirmed, probable and suspected deaths
                                                                              2485
          dtype: int64
In [10]:
           data.count(axis=1)
                   4
Out[10]:
          1
                   4
          2
                   4
          3
                   4
          4
                   4
          2480
                   4
          2481
                   4
          2482
                   4
          2483
                   4
          2484
          Length: 2485, dtype: int64
In [11]:
           data.sum(axis=0)
                                                                              GuineaNigeriaSierra LeoneL
          Country
Out[11]:
          iberiaSierra LeoneNi...
                                                                              2014-08-292014-08-292014-0
          8-292014-08-292014-0...
          Cumulative no. of confirmed, probable and suspected cases
```

0.0

2481

6.32546e+06

Cumulative no. of confirmed, probable and suspected deaths 2.55544e+06

dtype: object

In [12]: data['Cumulative no. of confirmed, probable and suspected cases'].sum()

Out[12]: 6325462.0

In [13]: data.median()

Out[13]: Cumulative no. of confirmed, probable and suspected cases 8.0 Cumulative no. of confirmed, probable and suspected deaths 6.0 dtype: float64

In [14]: data.mean()

Out[14]: Cumulative no. of confirmed, probable and suspected cases
Cumulative no. of confirmed, probable and suspected deaths
dtype: float64

Cumulative no. of confirmed, probable and suspected deaths
dtype: float64

In [15]: data['Cumulative no. of confirmed, probable and suspected deaths'].mean()

Out[15]: 1028.3476861167003

In [16]: data.sum()

Out[16]: Country GuineaNigeriaSierra LeoneL iberiaSierra LeoneNi...
Date 2014-08-292014-08-292014-0

8-292014-08-292014-0...

Cumulative no. of confirmed, probable and suspected cases

6.32546e+06

Cumulative no. of confirmed, probable and suspected deaths

2.55544e+06 dtype: object

In [17]: data['Cumulative no. of confirmed, probable and suspected deaths'].sum()

Out[17]: 2555444.0

In [18]: data.cummin()

Out[18]:		Country	Date	Cumulative no. of confirmed, probable and suspected cases	Cumulative no. of confirmed, probable and suspected deaths
	0	Guinea	2014- 08-29	648.0	430.0
	1	Guinea	2014- 08-29	19.0	7.0
	2	Guinea	2014- 08-29	19.0	7.0

		Country	Date	Cumulative no. of confirmed, probable and suspected cases	Cumulative no. of confirmed, probable and suspected deaths
	3	Guinea	2014- 08-29	19.0	7.0
	4	Guinea	2014- 08-29	19.0	7.0
	•••	•••	•••		***
	2480	Guinea	2014- 08-29	0.0	0.0
	2481	Guinea	2014- 08-29	0.0	0.0
	2482	Guinea	2014- 08-29	0.0	0.0
	2483	Guinea	2014- 08-29	0.0	0.0
	2484	Guinea	2014- 08-29	0.0	0.0

2485 rows × 4 columns

```
In [19]:
```

data['Cumulative no. of confirmed, probable and suspected cases'].cumprod()

Out[19]: 0 6.480000e+02 1.231200e+04 1 1.263211e+07 1.740705e+10 2.195029e+13 2480 NaN 2481 NaN 2482 NaN 2483 NaN 2484 NaN

Name: Cumulative no. of confirmed, probable and suspected cases, Length: 2485, dtype: fl oat64

In [20]:

data.cummin()

Out[20]:

	Country	Date	Cumulative no. of confirmed, probable and suspected cases	Cumulative no. of confirmed, probable and suspected deaths
0	Guinea	2014- 08-29	648.0	430.0
1	Guinea	2014- 08-29	19.0	7.0
2	Guinea	2014- 08-29	19.0	7.0
3	Guinea	2014- 08-29	19.0	7.0

		Country	Date	Cumulative no. of confirmed, probable and suspected cases	Cumulative no. of confirmed, probable and suspected deaths
	4	Guinea	2014- 08-29	19.0	7.0
	•••	•••	•••		•••
	2480	Guinea	2014- 08-29	0.0	0.0
	2481	Guinea	2014- 08-29	0.0	0.0
	2482	Guinea	2014- 08-29	0.0	0.0
	2483	Guinea	2014- 08-29	0.0	0.0
	2484	Guinea	2014- 08-29	0.0	0.0

2485 rows × 4 columns

In [21]: data.cummax()

Out[21]:

	Country	Date	Cumulative no. of confirmed, probable and suspected cases	Cumulative no. of confirmed, probable and suspected deaths
0	Guinea	2014- 08-29	648.0	430.0
1	Nigeria	2014- 08-29	648.0	430.0
2	Sierra Leone	2014- 08-29	1026.0	430.0
3	Sierra Leone	2014- 08-29	1378.0	694.0
4	Sierra Leone	2014- 09-05	1378.0	694.0
•••	•••			
2480	United States of America	2016- 03-23	14122.0	4806.0
2481	United States of America	2016- 03-23	14122.0	4806.0
2482	United States of America	2016- 03-23	14122.0	4806.0
2483	United States of America	2016- 03-23	14122.0	4806.0
2484	United States of America	2016- 03-23	14122.0	4806.0

```
In [22]:
          data['Cumulative no. of confirmed, probable and suspected cases'].pct_change()
Out[22]: 0
                         NaN
                  -0.970679
         2
                   53.000000
         3
                   0.343080
         4
                   -0.084906
         2480
               1184.111111
                  -0.999906
         2481
         2482
                    4.000000
         2483
                    3.000000
                   -0.800000
         2484
         Name: Cumulative no. of confirmed, probable and suspected cases, Length: 2485, dtype: fl
```

By

Harsha Praneeth D

122010407003

ECE VLSI