US covid tracking project

February 9, 2021

```python
[1]: import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
```

```python
[2]: import io
     import requests
     url = "https://api.covidtracking.com/v1/states/daily.csv" # Access date and time␣
     ↪: Feb 8, 2021, 10.30 A.M.
     s=requests.get(url).content
```

```python
[3]: df = pd.read_csv(io.StringIO(s.decode('utf-8')))
```

```python
[4]: df.to_csv('../Dataset/dataset.csv', index=False, header=True) # Saving the␣
     ↪dataset for reuse in future time
```

```python
[5]: pd.set_option('display.max_columns', None) # To display all columns
     df.head()
```

```
[5]:        date state  positive  probableCases    negative  pending  \
     0  20210207    AK   53279.0            NaN         NaN      NaN
     1  20210207    AL  472423.0       101367.0   1816273.0      NaN
     2  20210207    AR  306736.0        62862.0   2285451.0      NaN
     3  20210207    AS       0.0            NaN      2140.0      NaN
     4  20210207    AZ  780637.0        50509.0   2818265.0      NaN

       totalTestResultsSource  totalTestResults  hospitalizedCurrently  \
     0           totalTestsViral         1536911.0                   44.0
     1     totalTestsPeopleViral         2187329.0                 1513.0
     2           totalTestsViral         2529325.0                  781.0
     3           totalTestsViral            2140.0                    NaN
     4           totalTestsViral         6982148.0                 2910.0

       hospitalizedCumulative  inIcuCurrently  inIcuCumulative  \
     0                 1219.0             NaN              NaN
     1                43005.0             NaN           2576.0
     2                14066.0           270.0              NaN
```

1

```
3                NaN             NaN             NaN
4            54657.0           838.0             NaN


   onVentilatorCurrently  onVentilatorCumulative  recovered dataQualityGrade  \
0                   11.0                     NaN        NaN                A
1                    NaN                  1460.0   252880.0                A
2                  126.0                  1458.0   285306.0               A+
3                    NaN                     NaN        NaN              NaN
4                  561.0                     NaN   107979.0               A+


       lastUpdateEt          dateModified checkTimeEt    death  hospitalized  \
0   2/5/2021 03:59  2021-02-05T03:59:00Z  02/04 22:59    279.0        1219.0
1   2/7/2021 11:00  2021-02-07T11:00:00Z  02/07 06:00   8515.0       43005.0
2   2/6/2021 00:00  2021-02-06T00:00:00Z  02/05 19:00   5076.0       14066.0
3  12/1/2020 00:00  2020-12-01T00:00:00Z  11/30 19:00      0.0           NaN
4   2/7/2021 00:00  2021-02-07T00:00:00Z  02/06 19:00  14048.0       54657.0


           dateChecked  totalTestsViral  positiveTestsViral  \
0  2021-02-05T03:59:00Z        1536911.0             64404.0
1  2021-02-07T11:00:00Z              NaN                 NaN
2  2021-02-06T00:00:00Z        2529325.0                 NaN
3  2020-12-01T00:00:00Z           2140.0                 NaN
4  2021-02-07T00:00:00Z        6982148.0                 NaN


   negativeTestsViral  positiveCasesViral  deathConfirmed  deathProbable  \
0           1470760.0                 NaN             NaN            NaN
1                 NaN            371056.0          6747.0         1768.0
2           2285451.0            243874.0          4054.0         1022.0
3                 NaN                 0.0             NaN            NaN
4                 NaN            730128.0         12480.0         1568.0


   totalTestEncountersViral  totalTestsPeopleViral  totalTestsAntibody  \
0                       NaN                    NaN                 NaN
1                       NaN              2187329.0                 NaN
2                       NaN                    NaN                 NaN
3                       NaN                    NaN                 NaN
4                       NaN              3548393.0            423601.0


   positiveTestsAntibody  negativeTestsAntibody  totalTestsPeopleAntibody  \
0                    NaN                    NaN                       NaN
1                    NaN                    NaN                  109260.0
2                    NaN                    NaN                       NaN
3                    NaN                    NaN                       NaN
4                    NaN                    NaN                       NaN


   positiveTestsPeopleAntibody  negativeTestsPeopleAntibody  \
0                          NaN                          NaN
```

1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	totalTestsPeopleAntigen	positiveTestsPeopleAntigen	totalTestsAntigen \
0	NaN	NaN	NaN
1	NaN	NaN	NaN
2	389367.0	73756.0	NaN
3	NaN	NaN	NaN
4	NaN	NaN	NaN

	positiveTestsAntigen	fips	positiveIncrease	negativeIncrease	total \
0	NaN	2	0	0	53279
1	NaN	1	1112	4462	2288696
2	NaN	5	672	8180	2592187
3	NaN	60	0	0	2140
4	NaN	4	1544	16776	3598902

	totalTestResultsIncrease	posNeg	deathIncrease	hospitalizedIncrease \
0	0	53279	0	0
1	5308	2288696	2	0
2	8840	2592187	15	17
3	0	2140	0	0
4	59968	3598902	37	150

	hash	commercialScore \
0	07a5d43f958541e9cdabb5ea34c8fb481835e130	0
1	bde38ab9d426d29691fb40de1edeb285b1674fdc	0
2	c2f88044a45b0669c5355b2cddbaca15de43c7f3	0
3	80aa91d3878a87a2e94edb3586244a5df450d0bb	0
4	380468038ad1b440ada7da318650aefc8e8a772d	0

	negativeRegularScore	negativeScore	positiveScore	score	grade
0	0	0	0	0	NaN
1	0	0	0	0	NaN
2	0	0	0	0	NaN
3	0	0	0	0	NaN
4	0	0	0	0	NaN

[6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19205 entries, 0 to 19204
Data columns (total 55 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
```

```
0    date                          19205 non-null   int64
1    state                         19205 non-null   object
2    positive                      19016 non-null   float64
3    probableCases                 8489 non-null    float64
4    negative                      15289 non-null   float64
5    pending                       1999 non-null    float64
6    totalTestResultsSource        19205 non-null   object
7    totalTestResults              19103 non-null   float64
8    hospitalizedCurrently         15854 non-null   float64
9    hospitalizedCumulative        11827 non-null   float64
10   inIcuCurrently                10358 non-null   float64
11   inIcuCumulative               3456 non-null    float64
12   onVentilatorCurrently         8226 non-null    float64
13   onVentilatorCumulative        1181 non-null    float64
14   recovered                     13772 non-null   float64
15   dataQualityGrade              17873 non-null   object
16   lastUpdateEt                  18624 non-null   object
17   dateModified                  18624 non-null   object
18   checkTimeEt                   18624 non-null   object
19   death                         18334 non-null   float64
20   hospitalized                  11827 non-null   float64
21   dateChecked                   18624 non-null   object
22   totalTestsViral               12701 non-null   float64
23   positiveTestsViral            7372 non-null    float64
24   negativeTestsViral            4413 non-null    float64
25   positiveCasesViral            13731 non-null   float64
26   deathConfirmed                8927 non-null    float64
27   deathProbable                 6817 non-null    float64
28   totalTestEncountersViral      4839 non-null    float64
29   totalTestsPeopleViral         8472 non-null    float64
30   totalTestsAntibody            4332 non-null    float64
31   positiveTestsAntibody         3328 non-null    float64
32   negativeTestsAntibody         1373 non-null    float64
33   totalTestsPeopleAntibody      1699 non-null    float64
34   positiveTestsPeopleAntibody   982 non-null     float64
35   negativeTestsPeopleAntibody   888 non-null     float64
36   totalTestsPeopleAntigen       859 non-null     float64
37   positiveTestsPeopleAntigen    549 non-null     float64
38   totalTestsAntigen             2830 non-null    float64
39   positiveTestsAntigen          1862 non-null    float64
40   fips                          19205 non-null   int64
41   positiveIncrease              19205 non-null   int64
42   negativeIncrease              19205 non-null   int64
43   total                         19205 non-null   int64
44   totalTestResultsIncrease      19205 non-null   int64
45   posNeg                        19205 non-null   int64
46   deathIncrease                 19205 non-null   int64
47   hospitalizedIncrease          19205 non-null   int64
```

```
48   hash                          19205 non-null   object
49   commercialScore               19205 non-null   int64
50   negativeRegularScore          19205 non-null   int64
51   negativeScore                 19205 non-null   int64
52   positiveScore                 19205 non-null   int64
53   score                         19205 non-null   int64
54   grade                         0 non-null       float64
dtypes: float64(33), int64(14), object(8)
memory usage: 8.1+ MB
```

0.1 Metadata information

1. **checkTimeEt** *Field type:string* Deprecated

2. **commercialScore** *Field type:integer* Deprecated *Returns : null, if no data is available*

3. **dataQualityGrade** Field type:string Data Quality Grade The COVID Tracking Project grade of the completeness of the data reporting by a state.

4. **date** *Field type:integer* Date on which data was collected by The COVID Tracking Project.

5. **dateChecked** Field type:string Deprecated

6. **dateModified** *Field type:string* Deprecated, use lastUpdateEt instead

7. **death** *Field type:integer* Deaths (confirmed and probable) Total fatalities with confirmed OR probable COVID-19 case diagnosis (per the expanded CSTE case definition of April 5th, 2020 approved by the CDC). In some states, these individuals must also have COVID-19 listed on the death certificate to count as a COVID-19 death. When states post multiple numbers for fatalities, the metric includes only deaths with COVID-19 listed on the death certificate, unless deaths among cases is a more reliable metric in the state. *Returns : null. if no data is available*

8. **deathConfirmed** *Field type:integer* Deaths (confirmed) Total fatalities with confirmed COVID-19 case diagnosis (per the expanded CSTE case definition of April 5th, 2020 approved by the CDC). In some states, these individuals must also have COVID-19 listed on the death certificate to count as a COVID-19 death. When states post multiple numbers for confirmed fatalities, the metric includes only lab-confirmed deaths with COVID-19 listed on the death certificate, unless deaths among confirmed cases is a more reliable metric in the state. *Returns : null, if no data is available*

9. **deathIncrease** *Field type:integer* New deaths Daily increase in death, calculated from the previous day's value. *Returns: null, if no data is available*

10. **deathProbable** *Field type:integer* Deaths (probable) Total fatalities with probable COVID-19 case diagnosis (per the expanded CSTE case definition of April 5th, 2020 approved by the CDC). In some states, these individuals must also have COVID-19 listed on the death certificate to count as a COVID-19 death. When states post multiple numbers for probable fatalities, the metric includes only probable fatalities with COVID-19 listed on the death certificate, unless deaths among probable cases is a more reliable metric in the state. *Returns : null, if no data is available*

11. **fips** *Field type:string* FIPS code State FIPS code

12. **grade** *Field type:string* Deprecated

13. **hash** *Field type:string* Deprecated - A hash of the current record.

14. **hospitalized** *Field type:integer* Deprecated *Returns : null, if no data is available*

15. **hospitalizedCumulative** *Field type:integer* Cumulative hospitalized/Ever hospitalized Total number of individuals who have ever been hospitalized with COVID-19. Definitions vary by state / territory, and it is not always clear whether pediatric patients are included in this metric. Where possible, we report patients hospitalized with confirmed or suspected COVID-19 cases. *Returns : null, if no data is available*

16. **hospitalizedCurrently** *Field type:integer* Currently hospitalized/Now hospitalized Individuals who are currently hospitalized with COVID-19. Definitions vary by state / territory, and it is not always clear whether pediatric patients are included in this metric. Where possible, we report patients hospitalized with confirmed or suspected COVID-19 cases. *Returns : null, if no data is available*

17. **hospitalizedIncrease** *Field type:integer* New total hospitalizations Daily increase in hospitalizedCumulative, calculated from the previous day's value. *Returns : null, if no data is available*

18. **inIcuCumulative** *Field type:integer* Cumulative in ICU/Ever in ICU Total number of individuals who have ever been hospitalized in the Intensive Care Unit with COVID-19. Definitions vary by state / territory, and it is not always clear whether pediatric patients are included in this metric. Where possible, we report patients in the ICU with confirmed or suspected COVID-19 cases. _Returns : null, if no data is available

19. **inIcuCurrently** *Field type:integer* Currently in ICU/Now in ICU Individuals who are currently hospitalized in the Intensive Care Unit with COVID-19. Definitions vary by state / territory, and it is not always clear whether pediatric patients are included in this metric. Where possible, we report patients in the ICU with confirmed or suspected COVID-19 cases. *Returns : null, if no data is available*

20. **lastUpdateEt** *Field type:string* Last Update (ET) Date and time in Eastern time the state or territory last updated the data.

21. **negative** *Field type:integer* Negative PCR tests (people) Total number of unique people with a completed PCR test that returns negative. For states / territories that do not report this number directly, we compute it using one of several methods, depending on which data points the state provides. Due to complex reporting procedures, this number might be mixing units and therefore, at best, it should only be considered an estimate of the number of people with a completed PCR test that return negative. *Returns : null, if no data is available*

22. **negativeIncrease** *Field type:integer* Deprecated *Returns : null, if no data is available*

23. **negativeRegularScore** *Field type:integer* Deprecated *Returns : null, if no data is available*

24. **negativeScore** *Field type:integer* Deprecated *Returns : null, if no data is available*

25. **negativeTestsAntibody** *Field type:integer* Negative antibody tests (specimens) The total number of completed antibody tests that return negative as reported by the state or territory. *Returns : null, if no data is available*

26. **negativeTestsPeopleAntibody** *Field type:integer* Negative antibody tests (people) The total number of unique people with completed antibody tests that return negative as reported by

the state or territory. *Returns : null, if no data is available*

27. **negativeTestsViral** *Field type:integer* Negative PCR tests (specimens) Total number of completed PCR tests (or specimens tested) that return negative as reported by the state or territory. For states/territories that do not report this number directly, we compute it using one of several methods, depending on which data points the state provides. If we discover that a jurisdiction is including antigen tests in this metric, we will annotate that state or territory's data accordingly. *Returns : null, if no data is available*

28. **onVentilatorCumulative** *Field type:integer* Cumulative on ventilator/Ever on ventilator Total number of individuals who have ever been hospitalized under advanced ventilation with COVID-19. Definitions vary by state / territory, and it is not always clear whether pediatric patients are included in this metric. Where possible, we report patients on ventilation with confirmed or suspected COVID-19 cases. *Returns : null, if no data is available*

29. **onVentilatorCurrently** *Field type:integer* Currently on ventilator/Now on ventilator Individuals who are currently hospitalized under advanced ventilation with COVID-19. Definitions vary by state / territory, and it is not always clear whether pediatric patients are included in this metric. Where possible, we report patients on ventilation with confirmed or suspected COVID-19 cases. *Returns : null, if no data is available*

30. **pending** *Field type:integer* Pending Total number of viral tests that have not been completed as reported by the state or territory. *Returns : null, if no data is available*

31. **posNeg** *Field type:integer* Deprecated *Returns : null, if no data is available*

32. **positive** *Field type:integer* Cases (confirmed plus probable) Total number of confirmed plus probable cases of COVID-19 reported by the state or territory, ideally per the August 5, 2020 CSTE case definition. Some states are following the older April 5th, 2020 CSTE case definition or using their own custom definitions. Not all states and territories report probable cases. If a state is not reporting probable cases, this field will just represent confirmed cases. *Returns : null, if no data is available*

33. **positiveCasesViral** *Field type:integer* Confirmed Cases Total number of unique people with a positive PCR or other approved nucleic acid amplification test (NAAT), as reported by the state or territory. This is equivalent to a confirmed case as per the CSTE case definitions of August 5th, 2020 and April 5th, 2020. If we discover a jurisdiction is labeling cases as confirmed using other evidence (e.g. positive antigen tests), we will annotate that state or territory's data accordingly. *Returns : null, if no data is available*

34. **positiveIncrease** *Field type:integer* New cases The daily increase in API field positive, which measures Cases (confirmed plus probable) calculated based on the previous day's value. *Returns : null, if no data is available*

35. **positiveScore** *Field type:integer* Deprecated *Returns : null, if no data is available*

36. **positiveTestsAntibody** *Field type:integer* Positive antibody tests (specimens) Total number of completed antibody tests that return positive as reported by the state or territory. *Returns : null, if no data is available*

37. **positiveTestsAntigen** *Field type:integer* Positive antigen tests (specimens) Total number of completed antigen tests that return positive as reported by the state or territory. *Returns : null, if no data is available*

38. **positiveTestsPeopleAntibody** *Field type:integer* Positive antibody tests (people) The total number of unique people with completed antibody tests that return positive as reported by the state or territory. *Returns : null, if no data is available*

39. **positiveTestsPeopleAntigen** *Field type:integer* Positive antigen tests (people) Total number of unique people with a completed antigen test that returned positive as reported by the state or territory. *Returns : null, if no data is available*

40. **positiveTestsViral** *Field type:integer* Positive PCR tests (specimens) Total number of completed PCR tests (or specimens tested) that return positive as reported by the state or territory. If we discover that a jurisdiction is including antigen tests in this metric, we will annotate that state or territory's data accordingly. *Returns : null, if no data is available*

41. **probableCases** *Field type:integer* Probable Cases Total number of probable cases of COVID-19 as reported by the state or territory, ideally per the August 5, 2020 CSTE case definition. By this definition, a probable case is someone who tests positive via antigen without a positive PCR or other approved nucleic acid amplification test (NAAT), someone with clinical and epidemiological evidence of COVID-19 infection with no confirmatory laboratory testing performed for SARS-CoV-2, or someone with COVID-19 listed on their death certificate with no confirmatory laboratory testing performed for SARS-CoV-2. Some states are following the older April 5th, 2020 CSTE case definition or using their own custom definitions. *Returns : null, if no data is available*

42. **recovered** *Field type:integer* Recovered Total number of people that are identified as recovered from COVID-19. States provide very disparate definitions on what constitutes a "recovered" COVID-19 case. Types of "recovered" cases include those who are discharged from hospitals, released from isolation after meeting CDC guidance on symptoms cessation, or those who have not been identified as fatalities after a number of days (30 or more) post disease onset. Specifics vary for each state or territory. *Returns : null, if no data is available*

43. **score** *Field type:integer* Deprecated *Returns : null, if no data is available*

44. **state** *Field type:string* State (or territory) Two-letter abbreviation for the state or territory.

45. **total** *Field type:integer* Deprecated *Returns : null, if no data is available*

46. **totalTestEncountersViral** *Field type:integer* Total PCR tests (test encounters) Total number of people tested per day via PCR testing as reported by the state or territory. The count for this metric is incremented up by one for each day on which an individual person is tested, no matter how many specimens are collected from that person on that day. If an individual person is tested twice a day on three different days, this count will increment up by three. If we discover that a jurisdiction is including antigen tests in this metric, we will annotate that state or territory's data accordingly. *Returns : null, if no data is available*

47. **totalTestResults** *Field type:integer* Total test results At the national level, this metric is a summary statistic which, because of the variation in test reporting methods, is at best an estimate of US viral (PCR) testing. Some states/territories report tests in units of test encounters, some report tests in units of specimens, and some report tests in units of unique people. Moreover, some jurisdictions include antigen tests in their total test counts without reporting a separate total of viral (PCR) tests. Therefore, this value is an aggregate calculation of heterogeneous figures. Please consult each state or territory's individual data page to see whether that jurisdiction lumps antigen and PCR tests together and to see what units

that jurisdiction uses for test reporting. In most states, the totalTestResults field is currently computed by adding positive and negative values because, historically, some states do not report totals, and to work around different reporting cadences for cases and tests. In Colorado, Delaware, the District of Columbia, Florida, Hawaii, Minnesota, Nevada, New York, North Dakota, Pennsylvania, Rhode Island, Virginia, Washington, and Wisconsin, where reliable testing encounters figures are available with a complete time series, we directly report those figures in this field. In Alaska, America Samoa, Arizona, Arkansas, California, Georgia, Indiana, Kentucky, Maryland, Massachusetts, Missouri, Nebraska, New Hampshire, Ohio, Oregon, Texas, Utah, Vermont, and Wyoming, where reliable specimens figures are available with a complete time series, we directly report those figures in this field. In Alabama, Idaho, and South Dakota, where reliable unique people figures are available with a complete time series, we directly report those figures in this field. We are in the process of switching all states over to use directly reported total figures, using a policy of preferring testing encounters, specimens, and people, in that order. *Returns : null, if no data is available*

48. **totalTestResultsIncrease** *Field type:integer* New tests Daily increase in totalTestResults, calculated from the previous day's value. This calculation includes all the caveats associated with Total tests/totalTestResults, and we recommend against using it at the state/territory level. *Returns : null, if no data is available*

49. **totalTestResultsSource** *Field type:string* Indicates which field is being used for total test results. If it is posNeg, then it is calculated by adding all positive and negative values.

50. **totalTestsAntibody** *Field type:integer* Total antibody tests (specimens) Total number of completed antibody tests as reported by the state or territory. *Returns : null, if no data is available*

51. **totalTestsAntigen** *Field type:integer* Total antigen tests (specimens) Total number of completed antigen tests, as reported by the state or territory. *Returns : null, if no data is available*

52. **totalTestsPeopleAntibody** *Field type:integer* Total antibody tests (people) The total number of unique people who have been tested at least once via antibody testing as reported by the state or territory. *Returns : null, if no data is available*

53. **totalTestsPeopleAntigen** *Field type:integer* Total antigen tests (people) Total number of unique people who have been tested at least once via antigen testing, as reported by the state or territory. *Returns : null, if no data is available*

54. **totalTestsPeopleViral** *Field type:integer* Total PCR tests (people) Total number of unique people tested at least once via PCR testing, as reported by the state or territory. The count for this metric is incremented up only the first time an individual person is tested and their result is reported. Future tests of the same person will not be added to this count. In the case where the state only provides negative cases, this field is calculated as the summation of people who tested positive ("Positive Cases (People") and the number of people who tested negative ("Negative (People or Cases)"). If we discover that a jurisdiction is including antigen tests in this metric, we will annotate that state or territory's data accordingly. *Returns : null, if no data is available*

55. **totalTestsViral** *Field type:integer* Total PCR tests (specimens) Total number of PCR tests (or specimens tested) as reported by the state or territory. The count for this metric is incremented up each time a specimen is tested and the result is reported. If we discover that a jurisdiction is including antigen tests in this metric, we will annotate that state or territory's data accordingly. For states with ambiguous annotations, we have assigned their total

tests to this category; these states and territories are identified in the new API field covid-TrackingProjectPreferredTotalTestUnits as having "Unclear units." *Returns : null, if no data is available*

```
[7]: df.corr() # Checking the correlation values between the different features
```

[7]:

	date	positive	probableCases	negative \
date	1.000000	0.403350	0.471940	0.356075
positive	0.403350	1.000000	0.901683	0.861454
probableCases	0.471940	0.901683	1.000000	0.620788
negative	0.356075	0.861454	0.620788	1.000000
pending	0.050661	0.226807	0.708051	0.253257
totalTestResults	0.343073	0.918836	0.760209	0.916254
hospitalizedCurrently	0.182523	0.788657	0.801676	0.612512
hospitalizedCumulative	0.236209	0.810323	0.717151	0.856888
inIcuCurrently	0.109469	0.738103	0.791933	0.466253
inIcuCumulative	0.376537	0.869864	0.751898	0.855113
onVentilatorCurrently	0.130084	0.545098	0.473090	0.322949
onVentilatorCumulative	0.463882	0.809749	0.893915	0.913105
recovered	0.363205	0.868807	0.867334	0.691471
death	0.270351	0.845571	0.707644	0.824111
hospitalized	0.236209	0.810323	0.717151	0.856888
totalTestsViral	0.333125	0.938391	0.765405	0.925259
positiveTestsViral	0.382847	0.985520	0.905840	0.867252
negativeTestsViral	0.360361	0.959161	0.826980	0.952165
positiveCasesViral	0.365034	0.996971	0.873507	0.848898
deathConfirmed	0.320239	0.773162	0.434814	0.785706
deathProbable	0.382624	0.587166	0.502899	0.539372
totalTestEncountersViral	0.376650	0.906785	0.860444	0.983941
totalTestsPeopleViral	0.317357	0.954197	0.771963	0.998758
totalTestsAntibody	0.228938	0.902632	0.733022	0.865302
positiveTestsAntibody	0.329251	0.873263	0.768252	0.743162
negativeTestsAntibody	0.161242	0.913019	0.699025	0.920001
totalTestsPeopleAntibody	0.154359	0.910430	0.621145	0.938031
positiveTestsPeopleAntibody	0.285720	0.952671	0.716976	0.944991
negativeTestsPeopleAntibody	0.228077	0.932197	0.763865	0.971803
totalTestsPeopleAntigen	0.711857	0.841856	0.576764	0.820655
positiveTestsPeopleAntigen	0.820228	0.916889	0.999773	0.779542
totalTestsAntigen	0.534225	0.791388	0.925628	0.393414
positiveTestsAntigen	0.599870	0.806190	0.900656	0.603656
fips	0.001356	-0.140387	-0.084427	-0.197303
positiveIncrease	0.235618	0.802033	0.704384	0.682545
negativeIncrease	0.094668	0.220117	0.108135	0.714452
total	0.344687	0.564814	0.458565	0.998705
totalTestResultsIncrease	0.196783	0.784191	0.507602	0.737069
posNeg	0.344719	0.564790	0.458481	0.998714
deathIncrease	0.207830	0.635918	0.642568	0.451441
hospitalizedIncrease	0.040072	0.095773	0.107298	0.106330

```
commercialScore                      NaN        NaN          NaN      NaN
negativeRegularScore                 NaN        NaN          NaN      NaN
negativeScore                        NaN        NaN          NaN      NaN
positiveScore                        NaN        NaN          NaN      NaN
score                                NaN        NaN          NaN      NaN
grade                                NaN        NaN          NaN      NaN

                               pending  totalTestResults  \
date                          0.050661          0.343073
positive                      0.226807          0.918836
probableCases                 0.708051          0.760209
negative                      0.253257          0.916254
pending                       1.000000          0.222137
totalTestResults              0.222137          1.000000
hospitalizedCurrently         0.180884          0.711930
hospitalizedCumulative        0.552573          0.793775
inIcuCurrently                0.260953          0.620174
inIcuCumulative              -0.478638          0.862973
onVentilatorCurrently         0.428136          0.421197
onVentilatorCumulative             NaN          0.814084
recovered                     0.226951          0.594212
death                         0.207302          0.852370
hospitalized                  0.552573          0.793775
totalTestsViral               0.101123          0.999024
positiveTestsViral            0.498398          0.936751
negativeTestsViral            0.346680          0.994728
positiveCasesViral            0.688307          0.947366
deathConfirmed                0.502042          0.812353
deathProbable                 0.244726          0.492943
totalTestEncountersViral      0.874735          1.000000
totalTestsPeopleViral         0.934190          0.971421
totalTestsAntibody            0.425722          0.923967
positiveTestsAntibody         0.404907          0.852600
negativeTestsAntibody         0.895392          0.927599
totalTestsPeopleAntibody      0.926843          0.835316
positiveTestsPeopleAntibody   0.952406          0.853410
negativeTestsPeopleAntibody   0.946156          0.926972
totalTestsPeopleAntigen            NaN          0.751520
positiveTestsPeopleAntigen         NaN          0.746389
totalTestsAntigen            -0.221305          0.718710
positiveTestsAntigen         -0.210559          0.754916
fips                         -0.319006         -0.145488
positiveIncrease              0.213856          0.765520
negativeIncrease              0.229540          0.232615
total                         0.272606          0.550488
totalTestResultsIncrease      0.202865          0.886305
posNeg                        0.270488          0.550484
```

```
deathIncrease                     0.196535           0.546314
hospitalizedIncrease              0.182067           0.057354
commercialScore                        NaN                NaN
negativeRegularScore                   NaN                NaN
negativeScore                          NaN                NaN
positiveScore                          NaN                NaN
score                                  NaN                NaN
grade                                  NaN                NaN


                            hospitalizedCurrently  hospitalizedCumulative  \
date                                     0.182523                0.236209
positive                                 0.788657                0.810323
probableCases                            0.801676                0.717151
negative                                 0.612512                0.856888
pending                                  0.180884                0.552573
totalTestResults                         0.711930                0.793775
hospitalizedCurrently                    1.000000                0.622000
hospitalizedCumulative                   0.622000                1.000000
inIcuCurrently                           0.979078                0.476767
inIcuCumulative                          0.829367                0.924660
onVentilatorCurrently                    0.912319                0.551642
onVentilatorCumulative                   0.802693                0.878012
recovered                                0.610649                0.337116
death                                    0.657336                0.941707
hospitalized                             0.622000                1.000000
totalTestsViral                          0.809487                0.869269
positiveTestsViral                       0.815944                0.917417
negativeTestsViral                       0.770702                0.960339
positiveCasesViral                       0.856058                0.919290
deathConfirmed                           0.564088                0.825622
deathProbable                            0.451712                0.656862
totalTestEncountersViral                 0.535554                0.778397
totalTestsPeopleViral                    0.691258                0.919764
totalTestsAntibody                       0.787010                0.961843
positiveTestsAntibody                    0.708139                0.966267
negativeTestsAntibody                    0.798613                0.972595
totalTestsPeopleAntibody                 0.751585                0.894796
positiveTestsPeopleAntibody              0.678598                0.979221
negativeTestsPeopleAntibody              0.776259                0.934963
totalTestsPeopleAntigen                  0.453008                0.896059
positiveTestsPeopleAntigen               0.599423                0.930071
totalTestsAntigen                        0.560587                0.766302
positiveTestsAntigen                     0.559463                0.597775
fips                                    -0.116283               -0.091445
positiveIncrease                         0.840328                0.584614
negativeIncrease                         0.123113                0.155386
total                                    0.318290                0.471580
```

```
totalTestResultsIncrease                    0.700612                0.740902
posNeg                                      0.318255                0.471566
deathIncrease                               0.795999                0.452838
hospitalizedIncrease                        0.212689                0.179962
commercialScore                                  NaN                     NaN
negativeRegularScore                             NaN                     NaN
negativeScore                                    NaN                     NaN
positiveScore                                    NaN                     NaN
score                                            NaN                     NaN
grade                                            NaN                     NaN

                              inIcuCurrently  inIcuCumulative  \
date                                0.109469         0.376537
positive                            0.738103         0.869864
probableCases                       0.791933         0.751898
negative                            0.466253         0.855113
pending                             0.260953        -0.478638
totalTestResults                    0.620174         0.862973
hospitalizedCurrently               0.979078         0.829367
hospitalizedCumulative              0.476767         0.924660
inIcuCurrently                      1.000000         0.766121
inIcuCumulative                     0.766121         1.000000
onVentilatorCurrently               0.971572         0.741309
onVentilatorCumulative              0.862205         0.873577
recovered                           0.595909         0.777482
death                               0.599082         0.957400
hospitalized                        0.476767         0.924660
totalTestsViral                     0.764186         0.896704
positiveTestsViral                  0.844731         0.899474
negativeTestsViral                  0.601912         0.995796
positiveCasesViral                  0.843887         0.861831
deathConfirmed                      0.472970         0.959488
deathProbable                       0.417315         0.733858
totalTestEncountersViral            0.304105         0.823227
totalTestsPeopleViral               0.493814         0.861893
totalTestsAntibody                  0.809407         0.978717
positiveTestsAntibody               0.725652         0.963656
negativeTestsAntibody               0.915512         0.988691
totalTestsPeopleAntibody            0.358473         0.937153
positiveTestsPeopleAntibody         0.692965              NaN
negativeTestsPeopleAntibody         0.701513              NaN
totalTestsPeopleAntigen             0.386949         0.942685
positiveTestsPeopleAntigen          0.499357         0.934141
totalTestsAntigen                   0.553266         0.737923
positiveTestsAntigen                0.544574         0.716944
fips                               -0.129240        -0.176316
positiveIncrease                    0.765221         0.657874
```

negativeIncrease	0.046053	0.014449
total	0.192413	0.423836
totalTestResultsIncrease	0.618542	0.678791
posNeg	0.192387	0.423854
deathIncrease	0.813762	0.464835
hospitalizedIncrease	0.205549	0.415252
commercialScore	NaN	NaN
negativeRegularScore	NaN	NaN
negativeScore	NaN	NaN
positiveScore	NaN	NaN
score	NaN	NaN
grade	NaN	NaN

	onVentilatorCurrently	onVentilatorCumulative \
date	0.130084	0.463882
positive	0.545098	0.809749
probableCases	0.473090	0.893915
negative	0.322949	0.913105
pending	0.428136	NaN
totalTestResults	0.421197	0.814084
hospitalizedCurrently	0.912319	0.802693
hospitalizedCumulative	0.551642	0.878012
inIcuCurrently	0.971572	0.862205
inIcuCumulative	0.741309	0.873577
onVentilatorCurrently	1.000000	0.888654
onVentilatorCumulative	0.888654	1.000000
recovered	0.378104	0.786621
death	0.456948	0.886804
hospitalized	0.551642	0.878012
totalTestsViral	0.518983	0.807190
positiveTestsViral	0.611297	0.968977
negativeTestsViral	0.459612	0.788034
positiveCasesViral	0.533924	0.842570
deathConfirmed	0.379337	0.929325
deathProbable	0.364788	0.780981
totalTestEncountersViral	0.494518	NaN
totalTestsPeopleViral	0.364073	0.858928
totalTestsAntibody	0.625848	NaN
positiveTestsAntibody	0.113211	NaN
negativeTestsAntibody	0.887897	NaN
totalTestsPeopleAntibody	0.349926	0.834913
positiveTestsPeopleAntibody	0.494253	NaN
negativeTestsPeopleAntibody	-0.019264	NaN
totalTestsPeopleAntigen	0.367665	0.823230
positiveTestsPeopleAntigen	0.439335	0.841165
totalTestsAntigen	0.529593	0.786556
positiveTestsAntigen	0.509959	0.797217

fips	-0.058111	-0.391002
positiveIncrease	0.597193	0.518707
negativeIncrease	0.165617	0.358224
total	0.284886	0.911902
totalTestResultsIncrease	0.387825	0.356866
posNeg	0.284882	0.911902
deathIncrease	0.701856	0.449584
hospitalizedIncrease	0.128445	0.325341
commercialScore	NaN	NaN
negativeRegularScore	NaN	NaN
negativeScore	NaN	NaN
positiveScore	NaN	NaN
score	NaN	NaN
grade	NaN	NaN

	recovered	death	hospitalized \
date	0.363205	0.270351	0.236209
positive	0.868807	0.845571	0.810323
probableCases	0.867334	0.707644	0.717151
negative	0.691471	0.824111	0.856888
pending	0.226951	0.207302	0.552573
totalTestResults	0.594212	0.852370	0.793775
hospitalizedCurrently	0.610649	0.657336	0.622000
hospitalizedCumulative	0.337116	0.941707	1.000000
inIcuCurrently	0.595909	0.599082	0.476767
inIcuCumulative	0.777482	0.957400	0.924660
onVentilatorCurrently	0.378104	0.456948	0.551642
onVentilatorCumulative	0.786621	0.886804	0.878012
recovered	1.000000	0.560364	0.337116
death	0.560364	1.000000	0.941707
hospitalized	0.337116	0.941707	1.000000
totalTestsViral	0.793388	0.866362	0.869269
positiveTestsViral	0.957705	0.935220	0.917417
negativeTestsViral	0.813287	0.952401	0.960339
positiveCasesViral	0.922332	0.873448	0.919290
deathConfirmed	0.469037	0.997826	0.825622
deathProbable	0.226439	0.803305	0.656862
totalTestEncountersViral	0.387618	0.854868	0.778397
totalTestsPeopleViral	0.773410	0.919917	0.919764
totalTestsAntibody	0.779116	0.947354	0.961843
positiveTestsAntibody	0.802504	0.874202	0.966267
negativeTestsAntibody	0.042416	0.942509	0.972595
totalTestsPeopleAntibody	0.332518	0.872524	0.894796
positiveTestsPeopleAntibody	0.106602	0.926676	0.979221
negativeTestsPeopleAntibody	-0.216030	0.960307	0.934963
totalTestsPeopleAntigen	0.920482	0.709359	0.896059
positiveTestsPeopleAntigen	0.948567	0.868110	0.930071

totalTestsAntigen	0.747069	0.710912	0.766302
positiveTestsAntigen	0.758075	0.712723	0.597775
fips	0.063993	-0.148035	-0.091445
positiveIncrease	0.667410	0.626802	0.584614
negativeIncrease	0.151039	0.269465	0.155386
total	0.462639	0.555719	0.471580
totalTestResultsIncrease	0.438014	0.771587	0.740902
posNeg	0.462630	0.555724	0.471566
deathIncrease	0.478043	0.547131	0.452838
hospitalizedIncrease	0.030771	0.103984	0.179962
commercialScore	NaN	NaN	NaN
negativeRegularScore	NaN	NaN	NaN
negativeScore	NaN	NaN	NaN
positiveScore	NaN	NaN	NaN
score	NaN	NaN	NaN
grade	NaN	NaN	NaN

	totalTestsViral	positiveTestsViral \
date	0.333125	0.382847
positive	0.938391	0.985520
probableCases	0.765405	0.905840
negative	0.925259	0.867252
pending	0.101123	0.498398
totalTestResults	0.999024	0.936751
hospitalizedCurrently	0.809487	0.815944
hospitalizedCumulative	0.869269	0.917417
inIcuCurrently	0.764186	0.844731
inIcuCumulative	0.896704	0.899474
onVentilatorCurrently	0.518983	0.611297
onVentilatorCumulative	0.807190	0.968977
recovered	0.793388	0.957705
death	0.866362	0.935220
hospitalized	0.869269	0.917417
totalTestsViral	1.000000	0.929717
positiveTestsViral	0.929717	1.000000
negativeTestsViral	0.999440	0.961108
positiveCasesViral	0.945970	0.984962
deathConfirmed	0.807765	0.822507
deathProbable	0.479289	0.826974
totalTestEncountersViral	0.999086	0.993593
totalTestsPeopleViral	0.965439	0.964925
totalTestsAntibody	0.934753	0.913355
positiveTestsAntibody	0.854592	0.839378
negativeTestsAntibody	0.950931	0.941644
totalTestsPeopleAntibody	0.800595	0.920822
positiveTestsPeopleAntibody	0.799619	0.952810
negativeTestsPeopleAntibody	0.915151	0.930037

```
totalTestsPeopleAntigen            0.765026            0.857356
positiveTestsPeopleAntigen         0.904188            0.957355
totalTestsAntigen                  0.745217            0.744009
positiveTestsAntigen               0.763658            0.770396
fips                              -0.251865           -0.063936
positiveIncrease                   0.791239            0.775952
negativeIncrease                   0.258855            0.214689
total                              0.585156            0.672787
totalTestResultsIncrease           0.869193            0.750348
posNeg                             0.585155            0.672722
deathIncrease                      0.686104            0.664946
hospitalizedIncrease               0.072906            0.313925
commercialScore                         NaN                 NaN
negativeRegularScore                    NaN                 NaN
negativeScore                           NaN                 NaN
positiveScore                           NaN                 NaN
score                                   NaN                 NaN
grade                                   NaN                 NaN

                          negativeTestsViral  positiveCasesViral  \
date                                0.360361            0.365034
positive                           0.959161            0.996971
probableCases                      0.826980            0.873507
negative                           0.952165            0.848898
pending                            0.346680            0.688307
totalTestResults                   0.994728            0.947366
hospitalizedCurrently              0.770702            0.856058
hospitalizedCumulative             0.960339            0.919290
inIcuCurrently                     0.601912            0.843887
inIcuCumulative                    0.995796            0.861831
onVentilatorCurrently              0.459612            0.533924
onVentilatorCumulative             0.788034            0.842570
recovered                          0.813287            0.922332
death                              0.952401            0.873448
hospitalized                       0.960339            0.919290
totalTestsViral                    0.999440            0.945970
positiveTestsViral                 0.961108            0.984962
negativeTestsViral                 1.000000            0.959359
positiveCasesViral                 0.959359            1.000000
deathConfirmed                     0.887400            0.729639
deathProbable                      0.715895            0.538521
totalTestEncountersViral           0.998908            0.988379
totalTestsPeopleViral              0.991373            0.957867
totalTestsAntibody                 0.945647            0.913355
positiveTestsAntibody              0.738316            0.876718
negativeTestsAntibody              0.946685            0.924777
totalTestsPeopleAntibody           0.934672            0.932429
```

```
positiveTestsPeopleAntibody             0.973849            0.962656
negativeTestsPeopleAntibody             0.938882            0.962868
totalTestsPeopleAntigen                 0.653248            0.800682
positiveTestsPeopleAntigen              0.906130            0.889109
totalTestsAntigen                       0.627906            0.769372
positiveTestsAntigen                    0.844436            0.784298
fips                                   -0.327026           -0.196756
positiveIncrease                        0.750879            0.810812
negativeIncrease                        0.399919            0.148130
total                                   0.953926            0.519196
totalTestResultsIncrease                0.776801            0.791675
posNeg                                  0.953864            0.519162
deathIncrease                           0.652171            0.730500
hospitalizedIncrease                    0.528901            0.097555
commercialScore                              NaN                 NaN
negativeRegularScore                         NaN                 NaN
negativeScore                                NaN                 NaN
positiveScore                                NaN                 NaN
score                                        NaN                 NaN
grade                                        NaN                 NaN

                            deathConfirmed  deathProbable  \
date                              0.320239       0.382624
positive                         0.773162       0.587166
probableCases                    0.434814       0.502899
negative                         0.785706       0.539372
pending                          0.502042       0.244726
totalTestResults                 0.812353       0.492943
hospitalizedCurrently            0.564088       0.451712
hospitalizedCumulative           0.825622       0.656862
inIcuCurrently                   0.472970       0.417315
inIcuCumulative                  0.959488       0.733858
onVentilatorCurrently            0.379337       0.364788
onVentilatorCumulative           0.929325       0.780981
recovered                        0.469037       0.226439
death                            0.997826       0.803305
hospitalized                     0.825622       0.656862
totalTestsViral                  0.807765       0.479289
positiveTestsViral               0.822507       0.826974
negativeTestsViral               0.887400       0.715895
positiveCasesViral               0.729639       0.538521
deathConfirmed                   1.000000       0.759771
deathProbable                    0.759771       1.000000
totalTestEncountersViral         0.926593       0.599288
totalTestsPeopleViral            0.847268       0.361371
totalTestsAntibody               0.894220       0.673467
positiveTestsAntibody            0.513122       0.463622
```

```
negativeTestsAntibody            0.693940         0.893359
totalTestsPeopleAntibody         0.133053        -0.094329
positiveTestsPeopleAntibody      0.216874         0.022572
negativeTestsPeopleAntibody      0.563326         0.936954
totalTestsPeopleAntigen          0.593997         0.506991
positiveTestsPeopleAntigen       0.990677         0.968930
totalTestsAntigen                0.720865         0.711874
positiveTestsAntigen             0.707393         0.648409
fips                            -0.163917        -0.072153
positiveIncrease                 0.511701         0.323800
negativeIncrease                 0.523335         0.339596
total                            0.758052         0.521578
totalTestResultsIncrease         0.629699         0.309864
posNeg                           0.758051         0.521577
deathIncrease                    0.430019         0.349139
hospitalizedIncrease             0.124616         0.116476
commercialScore                       NaN              NaN
negativeRegularScore                  NaN              NaN
negativeScore                         NaN              NaN
positiveScore                         NaN              NaN
score                                 NaN              NaN
grade                                 NaN              NaN

                         totalTestEncountersViral  totalTestsPeopleViral  \
date                                     0.376650               0.317357
positive                                 0.906785               0.954197
probableCases                            0.860444               0.771963
negative                                 0.983941               0.998758
pending                                  0.874735               0.934190
totalTestResults                         1.000000               0.971421
hospitalizedCurrently                    0.535554               0.691258
hospitalizedCumulative                   0.778397               0.919764
inIcuCurrently                           0.304105               0.493814
inIcuCumulative                          0.823227               0.861893
onVentilatorCurrently                    0.494518               0.364073
onVentilatorCumulative                        NaN               0.858928
recovered                                0.387618               0.773410
death                                    0.854868               0.919917
hospitalized                             0.778397               0.919764
totalTestsViral                          0.999086               0.965439
positiveTestsViral                       0.993593               0.964925
negativeTestsViral                       0.998908               0.991373
positiveCasesViral                       0.988379               0.957867
deathConfirmed                           0.926593               0.847268
deathProbable                            0.599288               0.361371
totalTestEncountersViral                 1.000000               0.987577
totalTestsPeopleViral                    0.987577               1.000000
```

```
totalTestsAntibody                       0.938612              0.921408
positiveTestsAntibody                    0.962995              0.958315
negativeTestsAntibody                    0.931448              0.951507
totalTestsPeopleAntibody                 0.921510              0.940166
positiveTestsPeopleAntibody              0.945595              0.949888
negativeTestsPeopleAntibody              0.913326              0.967442
totalTestsPeopleAntigen                       NaN              0.861804
positiveTestsPeopleAntigen                    NaN              0.806807
totalTestsAntigen                        0.716568              0.394367
positiveTestsAntigen                     0.867685              0.715981
fips                                     0.025212             -0.250267
positiveIncrease                         0.745160              0.762786
negativeIncrease                         0.256736              0.388698
total                                    0.582777              0.998343
totalTestResultsIncrease                 0.913752              0.841698
posNeg                                   0.582807              0.998339
deathIncrease                            0.357743              0.492643
hospitalizedIncrease                     0.067795              0.376734
commercialScore                               NaN                   NaN
negativeRegularScore                          NaN                   NaN
negativeScore                                 NaN                   NaN
positiveScore                                 NaN                   NaN
score                                         NaN                   NaN
grade                                         NaN                   NaN

                             totalTestsAntibody  positiveTestsAntibody  \
date                                   0.228938               0.329251
positive                               0.902632               0.873263
probableCases                          0.733022               0.768252
negative                               0.865302               0.743162
pending                                0.425722               0.404907
totalTestResults                       0.923967               0.852600
hospitalizedCurrently                  0.787010               0.708139
hospitalizedCumulative                 0.961843               0.966267
inIcuCurrently                         0.809407               0.725652
inIcuCumulative                        0.978717               0.963656
onVentilatorCurrently                  0.625848               0.113211
onVentilatorCumulative                      NaN                    NaN
recovered                              0.779116               0.802504
death                                  0.947354               0.874202
hospitalized                           0.961843               0.966267
totalTestsViral                        0.934753               0.854592
positiveTestsViral                     0.913355               0.839378
negativeTestsViral                     0.945647               0.738316
positiveCasesViral                     0.913355               0.876718
deathConfirmed                         0.894220               0.513122
deathProbable                          0.673467               0.463622
```

totalTestEncountersViral	0.938612	0.962995
totalTestsPeopleViral	0.921408	0.958315
totalTestsAntibody	1.000000	0.974044
positiveTestsAntibody	0.974044	1.000000
negativeTestsAntibody	0.999679	0.972845
totalTestsPeopleAntibody	0.999921	0.967632
positiveTestsPeopleAntibody	0.999785	0.999963
negativeTestsPeopleAntibody	0.999975	0.999757
totalTestsPeopleAntigen	0.996730	0.999015
positiveTestsPeopleAntigen	NaN	NaN
totalTestsAntigen	0.767948	0.829444
positiveTestsAntigen	0.723795	0.822314
fips	-0.351115	-0.104808
positiveIncrease	0.724485	0.690727
negativeIncrease	0.343802	0.208922
total	0.657973	0.559548
totalTestResultsIncrease	0.794665	0.711357
posNeg	0.657996	0.559668
deathIncrease	0.668739	0.620899
hospitalizedIncrease	0.299655	0.311521
commercialScore	NaN	NaN
negativeRegularScore	NaN	NaN
negativeScore	NaN	NaN
positiveScore	NaN	NaN
score	NaN	NaN
grade	NaN	NaN

	negativeTestsAntibody	totalTestsPeopleAntibody \
date	0.161242	0.154359
positive	0.913019	0.910430
probableCases	0.699025	0.621145
negative	0.920001	0.938031
pending	0.895392	0.926843
totalTestResults	0.927599	0.835316
hospitalizedCurrently	0.798613	0.751585
hospitalizedCumulative	0.972595	0.894796
inIcuCurrently	0.915512	0.358473
inIcuCumulative	0.988691	0.937153
onVentilatorCurrently	0.887897	0.349926
onVentilatorCumulative	NaN	0.834913
recovered	0.042416	0.332518
death	0.942509	0.872524
hospitalized	0.972595	0.894796
totalTestsViral	0.950931	0.800595
positiveTestsViral	0.941644	0.920822
negativeTestsViral	0.946685	0.934672
positiveCasesViral	0.924777	0.932429

```
deathConfirmed                       0.693940               0.133053
deathProbable                        0.893359              -0.094329
totalTestEncountersViral             0.931448               0.921510
totalTestsPeopleViral                0.951507               0.940166
totalTestsAntibody                   0.999679               0.999921
positiveTestsAntibody                0.972845               0.967632
negativeTestsAntibody                1.000000               0.999781
totalTestsPeopleAntibody             0.999781               1.000000
positiveTestsPeopleAntibody          0.999703               0.971365
negativeTestsPeopleAntibody          0.999990               0.999594
totalTestsPeopleAntigen              0.995065               0.820393
positiveTestsPeopleAntigen                NaN               0.973903
totalTestsAntigen                    0.728787               0.554146
positiveTestsAntigen                 0.722569               0.832472
fips                                -0.551163              -0.424297
positiveIncrease                     0.756795               0.730633
negativeIncrease                     0.650835               0.718550
total                                0.922172               0.940385
totalTestResultsIncrease             0.851442               0.721153
posNeg                               0.922116               0.940274
deathIncrease                        0.668649               0.629507
hospitalizedIncrease                 0.686985               0.565465
commercialScore                           NaN                    NaN
negativeRegularScore                      NaN                    NaN
negativeScore                             NaN                    NaN
positiveScore                             NaN                    NaN
score                                     NaN                    NaN
grade                                     NaN                    NaN

                         positiveTestsPeopleAntibody  \
date                                        0.285720
positive                                    0.952671
probableCases                               0.716976
negative                                    0.944991
pending                                     0.952406
totalTestResults                            0.853410
hospitalizedCurrently                       0.678598
hospitalizedCumulative                      0.979221
inIcuCurrently                              0.692965
inIcuCumulative                                  NaN
onVentilatorCurrently                       0.494253
onVentilatorCumulative                           NaN
recovered                                   0.106602
death                                       0.926676
hospitalized                                0.979221
totalTestsViral                             0.799619
positiveTestsViral                          0.952810
```

```
negativeTestsViral                          0.973849
positiveCasesViral                          0.962656
deathConfirmed                              0.216874
deathProbable                               0.022572
totalTestEncountersViral                    0.945595
totalTestsPeopleViral                       0.949888
totalTestsAntibody                          0.999785
positiveTestsAntibody                       0.999963
negativeTestsAntibody                       0.999703
totalTestsPeopleAntibody                    0.971365
positiveTestsPeopleAntibody                 1.000000
negativeTestsPeopleAntibody                 0.964205
totalTestsPeopleAntigen                     0.892081
positiveTestsPeopleAntigen                  0.924561
totalTestsAntigen                           0.991828
positiveTestsAntigen                        0.978641
fips                                       -0.647935
positiveIncrease                            0.688476
negativeIncrease                            0.634676
total                                       0.950720
totalTestResultsIncrease                    0.685351
posNeg                                      0.950643
deathIncrease                               0.640799
hospitalizedIncrease                        0.594803
commercialScore                                  NaN
negativeRegularScore                             NaN
negativeScore                                    NaN
positiveScore                                    NaN
score                                            NaN
grade                                            NaN


                            negativeTestsPeopleAntibody  \
date                                        0.228077
positive                                    0.932197
probableCases                               0.763865
negative                                    0.971803
pending                                     0.946156
totalTestResults                            0.926972
hospitalizedCurrently                       0.776259
hospitalizedCumulative                      0.934963
inIcuCurrently                              0.701513
inIcuCumulative                                  NaN
onVentilatorCurrently                      -0.019264
onVentilatorCumulative                           NaN
recovered                                  -0.216030
death                                       0.960307
hospitalized                                0.934963
```

```
totalTestsViral                      0.915151
positiveTestsViral                   0.930037
negativeTestsViral                   0.938882
positiveCasesViral                   0.962868
deathConfirmed                       0.563326
deathProbable                        0.936954
totalTestEncountersViral             0.913326
totalTestsPeopleViral                0.967442
totalTestsAntibody                   0.999975
positiveTestsAntibody                0.999757
negativeTestsAntibody                0.999990
totalTestsPeopleAntibody             0.999594
positiveTestsPeopleAntibody          0.964205
negativeTestsPeopleAntibody          1.000000
totalTestsPeopleAntigen              0.988565
positiveTestsPeopleAntigen           0.964157
totalTestsAntigen                    0.974740
positiveTestsAntigen                 0.954946
fips                                -0.742274
positiveIncrease                     0.758061
negativeIncrease                     0.750473
total                                0.968643
totalTestResultsIncrease             0.836180
posNeg                               0.968602
deathIncrease                        0.727735
hospitalizedIncrease                 0.733249
commercialScore                           NaN
negativeRegularScore                      NaN
negativeScore                             NaN
positiveScore                             NaN
score                                     NaN
grade                                     NaN

                          totalTestsPeopleAntigen  \
date                                     0.711857
positive                                 0.841856
probableCases                            0.576764
negative                                 0.820655
pending                                       NaN
totalTestResults                         0.751520
hospitalizedCurrently                    0.453008
hospitalizedCumulative                   0.896059
inIcuCurrently                           0.386949
inIcuCumulative                          0.942685
onVentilatorCurrently                    0.367665
onVentilatorCumulative                   0.823230
recovered                                0.920482
```

```
death                           0.709359
hospitalized                    0.896059
totalTestsViral                 0.765026
positiveTestsViral              0.857356
negativeTestsViral              0.653248
positiveCasesViral              0.800682
deathConfirmed                  0.593997
deathProbable                   0.506991
totalTestEncountersViral             NaN
totalTestsPeopleViral           0.861804
totalTestsAntibody              0.996730
positiveTestsAntibody           0.999015
negativeTestsAntibody           0.995065
totalTestsPeopleAntibody        0.820393
positiveTestsPeopleAntibody     0.892081
negativeTestsPeopleAntibody     0.988565
totalTestsPeopleAntigen         1.000000
positiveTestsPeopleAntigen      0.983404
totalTestsAntigen               0.881709
positiveTestsAntigen            0.969383
fips                           -0.273796
positiveIncrease                0.460288
negativeIncrease                0.248027
total                           0.853695
totalTestResultsIncrease        0.516910
posNeg                          0.853695
deathIncrease                   0.407718
hospitalizedIncrease            0.177073
commercialScore                      NaN
negativeRegularScore                 NaN
negativeScore                        NaN
positiveScore                        NaN
score                                NaN
grade                                NaN

                          positiveTestsPeopleAntigen  totalTestsAntigen  \
date                                        0.820228           0.534225
positive                                    0.916889           0.791388
probableCases                               0.999773           0.925628
negative                                    0.779542           0.393414
pending                                          NaN          -0.221305
totalTestResults                            0.746389           0.718710
hospitalizedCurrently                       0.599423           0.560587
hospitalizedCumulative                      0.930071           0.766302
inIcuCurrently                              0.499357           0.553266
inIcuCumulative                             0.934141           0.737923
onVentilatorCurrently                       0.439335           0.529593
```

25

onVentilatorCumulative	0.841165	0.786556
recovered	0.948567	0.747069
death	0.868110	0.710912
hospitalized	0.930071	0.766302
totalTestsViral	0.904188	0.745217
positiveTestsViral	0.957355	0.744009
negativeTestsViral	0.906130	0.627906
positiveCasesViral	0.889109	0.769372
deathConfirmed	0.990677	0.720865
deathProbable	0.968930	0.711874
totalTestEncountersViral	NaN	0.716568
totalTestsPeopleViral	0.806807	0.394367
totalTestsAntibody	NaN	0.767948
positiveTestsAntibody	NaN	0.829444
negativeTestsAntibody	NaN	0.728787
totalTestsPeopleAntibody	0.973903	0.554146
positiveTestsPeopleAntibody	0.924561	0.991828
negativeTestsPeopleAntibody	0.964157	0.974740
totalTestsPeopleAntigen	0.983404	0.881709
positiveTestsPeopleAntigen	1.000000	0.972424
totalTestsAntigen	0.972424	1.000000
positiveTestsAntigen	0.999902	0.930921
fips	-0.363611	0.050712
positiveIncrease	0.311321	0.546100
negativeIncrease	0.083110	0.018802
total	0.834063	0.280844
totalTestResultsIncrease	0.081945	0.417288
posNeg	0.834063	0.280843
deathIncrease	0.469934	0.530992
hospitalizedIncrease	0.188699	0.193458
commercialScore	NaN	NaN
negativeRegularScore	NaN	NaN
negativeScore	NaN	NaN
positiveScore	NaN	NaN
score	NaN	NaN
grade	NaN	NaN

	positiveTestsAntigen	fips	positiveIncrease \
date	0.599870	0.001356	0.235618
positive	0.806190	-0.140387	0.802033
probableCases	0.900656	-0.084427	0.704384
negative	0.603656	-0.197303	0.682545
pending	-0.210559	-0.319006	0.213856
totalTestResults	0.754916	-0.145488	0.765520
hospitalizedCurrently	0.559463	-0.116283	0.840328
hospitalizedCumulative	0.597775	-0.091445	0.584614
inIcuCurrently	0.544574	-0.129240	0.765221

```
inIcuCumulative                    0.716944 -0.176316         0.657874
onVentilatorCurrently              0.509959 -0.058111         0.597193
onVentilatorCumulative             0.797217 -0.391002         0.518707
recovered                          0.758075  0.063993         0.667410
death                              0.712723 -0.148035         0.626802
hospitalized                       0.597775 -0.091445         0.584614
totalTestsViral                    0.763658 -0.251865         0.791239
positiveTestsViral                 0.770396 -0.063936         0.775952
negativeTestsViral                 0.844436 -0.327026         0.750879
positiveCasesViral                 0.784298 -0.196756         0.810812
deathConfirmed                     0.707393 -0.163917         0.511701
deathProbable                      0.648409 -0.072153         0.323800
totalTestEncountersViral           0.867685  0.025212         0.745160
totalTestsPeopleViral              0.715981 -0.250267         0.762786
totalTestsAntibody                 0.723795 -0.351115         0.724485
positiveTestsAntibody              0.822314 -0.104808         0.690727
negativeTestsAntibody              0.722569 -0.551163         0.756795
totalTestsPeopleAntibody           0.832472 -0.424297         0.730633
positiveTestsPeopleAntibody        0.978641 -0.647935         0.688476
negativeTestsPeopleAntibody        0.954946 -0.742274         0.758061
totalTestsPeopleAntigen            0.969383 -0.273796         0.460288
positiveTestsPeopleAntigen         0.999902 -0.363611         0.311321
totalTestsAntigen                  0.930921  0.050712         0.546100
positiveTestsAntigen               1.000000  0.008309         0.558915
fips                               0.008309  1.000000        -0.125490
positiveIncrease                   0.558915 -0.125490         1.000000
negativeIncrease                   0.063571 -0.119969         0.224184
total                              0.473288 -0.157948         0.405251
totalTestResultsIncrease           0.455054 -0.161360         0.784863
posNeg                             0.473284 -0.157913         0.405231
deathIncrease                      0.555051 -0.120435         0.650633
hospitalizedIncrease               0.151090 -0.065061         0.141312
commercialScore                         NaN       NaN              NaN
negativeRegularScore                    NaN       NaN              NaN
negativeScore                           NaN       NaN              NaN
positiveScore                           NaN       NaN              NaN
score                                   NaN       NaN              NaN
grade                                   NaN       NaN              NaN

                       negativeIncrease     total  \
date                           0.094668  0.344687
positive                       0.220117  0.564814
probableCases                  0.108135  0.458565
negative                       0.714452  0.998705
pending                        0.229540  0.272606
totalTestResults               0.232615  0.550488
hospitalizedCurrently          0.123113  0.318290
```

hospitalizedCumulative	0.155386	0.471580
inIcuCurrently	0.046053	0.192413
inIcuCumulative	0.014449	0.423836
onVentilatorCurrently	0.165617	0.284886
onVentilatorCumulative	0.358224	0.911902
recovered	0.151039	0.462639
death	0.269465	0.555719
hospitalized	0.155386	0.471580
totalTestsViral	0.258855	0.585156
positiveTestsViral	0.214689	0.672787
negativeTestsViral	0.399919	0.953926
positiveCasesViral	0.148130	0.519196
deathConfirmed	0.523335	0.758052
deathProbable	0.339596	0.521578
totalTestEncountersViral	0.256736	0.582777
totalTestsPeopleViral	0.388698	0.998343
totalTestsAntibody	0.343802	0.657973
positiveTestsAntibody	0.208922	0.559548
negativeTestsAntibody	0.650835	0.922172
totalTestsPeopleAntibody	0.718550	0.940385
positiveTestsPeopleAntibody	0.634676	0.950720
negativeTestsPeopleAntibody	0.750473	0.968643
totalTestsPeopleAntigen	0.248027	0.853695
positiveTestsPeopleAntigen	0.083110	0.834063
totalTestsAntigen	0.018802	0.280844
positiveTestsAntigen	0.063571	0.473288
fips	-0.119969	-0.157948
positiveIncrease	0.224184	0.405251
negativeIncrease	1.000000	0.624508
total	0.624508	1.000000
totalTestResultsIncrease	0.343153	0.439504
posNeg	0.624522	1.000000
deathIncrease	0.159763	0.323062
hospitalizedIncrease	0.046284	0.087038
commercialScore	NaN	NaN
negativeRegularScore	NaN	NaN
negativeScore	NaN	NaN
positiveScore	NaN	NaN
score	NaN	NaN
grade	NaN	NaN

	totalTestResultsIncrease	posNeg \
date	0.196783	0.344719
positive	0.784191	0.564790
probableCases	0.507602	0.458481
negative	0.737069	0.998714
pending	0.202865	0.270488

```
totalTestResults                    0.886305  0.550484
hospitalizedCurrently               0.700612  0.318255
hospitalizedCumulative              0.740902  0.471566
inIcuCurrently                      0.618542  0.192387
inIcuCumulative                     0.678791  0.423854
onVentilatorCurrently               0.387825  0.284882
onVentilatorCumulative              0.356866  0.911902
recovered                           0.438014  0.462630
death                               0.771587  0.555724
hospitalized                        0.740902  0.471566
totalTestsViral                     0.869193  0.585155
positiveTestsViral                  0.750348  0.672722
negativeTestsViral                  0.776801  0.953864
positiveCasesViral                  0.791675  0.519162
deathConfirmed                      0.629699  0.758051
deathProbable                       0.309864  0.521577
totalTestEncountersViral            0.913752  0.582807
totalTestsPeopleViral               0.841698  0.998339
totalTestsAntibody                  0.794665  0.657996
positiveTestsAntibody               0.711357  0.559668
negativeTestsAntibody               0.851442  0.922116
totalTestsPeopleAntibody            0.721153  0.940274
positiveTestsPeopleAntibody         0.685351  0.950643
negativeTestsPeopleAntibody         0.836180  0.968602
totalTestsPeopleAntigen             0.516910  0.853695
positiveTestsPeopleAntigen          0.081945  0.834063
totalTestsAntigen                   0.417288  0.280843
positiveTestsAntigen                0.455054  0.473284
fips                               -0.161360 -0.157913
positiveIncrease                    0.784863  0.405231
negativeIncrease                    0.343153  0.624522
total                               0.439504  1.000000
totalTestResultsIncrease            1.000000  0.439501
posNeg                              0.439501  1.000000
deathIncrease                       0.511043  0.323049
hospitalizedIncrease                0.068900  0.087009
commercialScore                          NaN       NaN
negativeRegularScore                     NaN       NaN
negativeScore                            NaN       NaN
positiveScore                            NaN       NaN
score                                    NaN       NaN
grade                                    NaN       NaN

                     deathIncrease  hospitalizedIncrease  \
date                      0.207830              0.040072
positive                  0.635918              0.095773
probableCases             0.642568              0.107298
```

29

	deathIncrease	hospitalizedIncrease
negative	0.451441	0.106330
pending	0.196535	0.182067
totalTestResults	0.546314	0.057354
hospitalizedCurrently	0.795999	0.212689
hospitalizedCumulative	0.452838	0.179962
inIcuCurrently	0.813762	0.205549
inIcuCumulative	0.464835	0.415252
onVentilatorCurrently	0.701856	0.128445
onVentilatorCumulative	0.449584	0.325341
recovered	0.478043	0.030771
death	0.547131	0.103984
hospitalized	0.452838	0.179962
totalTestsViral	0.686104	0.072906
positiveTestsViral	0.664946	0.313925
negativeTestsViral	0.652171	0.528901
positiveCasesViral	0.730500	0.097555
deathConfirmed	0.430019	0.124616
deathProbable	0.349139	0.116476
totalTestEncountersViral	0.357743	0.067795
totalTestsPeopleViral	0.492643	0.376734
totalTestsAntibody	0.668739	0.299655
positiveTestsAntibody	0.620899	0.311521
negativeTestsAntibody	0.668649	0.686985
totalTestsPeopleAntibody	0.629507	0.565465
positiveTestsPeopleAntibody	0.640799	0.594803
negativeTestsPeopleAntibody	0.727735	0.733249
totalTestsPeopleAntigen	0.407718	0.177073
positiveTestsPeopleAntigen	0.469934	0.188699
totalTestsAntigen	0.530992	0.193458
positiveTestsAntigen	0.555051	0.151090
fips	-0.120435	-0.065061
positiveIncrease	0.650633	0.141312
negativeIncrease	0.159763	0.046284
total	0.323062	0.087038
totalTestResultsIncrease	0.511043	0.068900
posNeg	0.323049	0.087009
deathIncrease	1.000000	0.260957
hospitalizedIncrease	0.260957	1.000000
commercialScore	NaN	NaN
negativeRegularScore	NaN	NaN
negativeScore	NaN	NaN
positiveScore	NaN	NaN
score	NaN	NaN
grade	NaN	NaN

	commercialScore	negativeRegularScore \
date	NaN	NaN

positive	NaN	NaN
probableCases	NaN	NaN
negative	NaN	NaN
pending	NaN	NaN
totalTestResults	NaN	NaN
hospitalizedCurrently	NaN	NaN
hospitalizedCumulative	NaN	NaN
inIcuCurrently	NaN	NaN
inIcuCumulative	NaN	NaN
onVentilatorCurrently	NaN	NaN
onVentilatorCumulative	NaN	NaN
recovered	NaN	NaN
death	NaN	NaN
hospitalized	NaN	NaN
totalTestsViral	NaN	NaN
positiveTestsViral	NaN	NaN
negativeTestsViral	NaN	NaN
positiveCasesViral	NaN	NaN
deathConfirmed	NaN	NaN
deathProbable	NaN	NaN
totalTestEncountersViral	NaN	NaN
totalTestsPeopleViral	NaN	NaN
totalTestsAntibody	NaN	NaN
positiveTestsAntibody	NaN	NaN
negativeTestsAntibody	NaN	NaN
totalTestsPeopleAntibody	NaN	NaN
positiveTestsPeopleAntibody	NaN	NaN
negativeTestsPeopleAntibody	NaN	NaN
totalTestsPeopleAntigen	NaN	NaN
positiveTestsPeopleAntigen	NaN	NaN
totalTestsAntigen	NaN	NaN
positiveTestsAntigen	NaN	NaN
fips	NaN	NaN
positiveIncrease	NaN	NaN
negativeIncrease	NaN	NaN
total	NaN	NaN
totalTestResultsIncrease	NaN	NaN
posNeg	NaN	NaN
deathIncrease	NaN	NaN
hospitalizedIncrease	NaN	NaN
commercialScore	NaN	NaN
negativeRegularScore	NaN	NaN
negativeScore	NaN	NaN
positiveScore	NaN	NaN
score	NaN	NaN
grade	NaN	NaN

	negativeScore	positiveScore	score	grade
date	NaN	NaN	NaN	NaN
positive	NaN	NaN	NaN	NaN
probableCases	NaN	NaN	NaN	NaN
negative	NaN	NaN	NaN	NaN
pending	NaN	NaN	NaN	NaN
totalTestResults	NaN	NaN	NaN	NaN
hospitalizedCurrently	NaN	NaN	NaN	NaN
hospitalizedCumulative	NaN	NaN	NaN	NaN
inIcuCurrently	NaN	NaN	NaN	NaN
inIcuCumulative	NaN	NaN	NaN	NaN
onVentilatorCurrently	NaN	NaN	NaN	NaN
onVentilatorCumulative	NaN	NaN	NaN	NaN
recovered	NaN	NaN	NaN	NaN
death	NaN	NaN	NaN	NaN
hospitalized	NaN	NaN	NaN	NaN
totalTestsViral	NaN	NaN	NaN	NaN
positiveTestsViral	NaN	NaN	NaN	NaN
negativeTestsViral	NaN	NaN	NaN	NaN
positiveCasesViral	NaN	NaN	NaN	NaN
deathConfirmed	NaN	NaN	NaN	NaN
deathProbable	NaN	NaN	NaN	NaN
totalTestEncountersViral	NaN	NaN	NaN	NaN
totalTestsPeopleViral	NaN	NaN	NaN	NaN
totalTestsAntibody	NaN	NaN	NaN	NaN
positiveTestsAntibody	NaN	NaN	NaN	NaN
negativeTestsAntibody	NaN	NaN	NaN	NaN
totalTestsPeopleAntibody	NaN	NaN	NaN	NaN
positiveTestsPeopleAntibody	NaN	NaN	NaN	NaN
negativeTestsPeopleAntibody	NaN	NaN	NaN	NaN
totalTestsPeopleAntigen	NaN	NaN	NaN	NaN
positiveTestsPeopleAntigen	NaN	NaN	NaN	NaN
totalTestsAntigen	NaN	NaN	NaN	NaN
positiveTestsAntigen	NaN	NaN	NaN	NaN
fips	NaN	NaN	NaN	NaN
positiveIncrease	NaN	NaN	NaN	NaN
negativeIncrease	NaN	NaN	NaN	NaN
total	NaN	NaN	NaN	NaN
totalTestResultsIncrease	NaN	NaN	NaN	NaN
posNeg	NaN	NaN	NaN	NaN
deathIncrease	NaN	NaN	NaN	NaN
hospitalizedIncrease	NaN	NaN	NaN	NaN
commercialScore	NaN	NaN	NaN	NaN
negativeRegularScore	NaN	NaN	NaN	NaN
negativeScore	NaN	NaN	NaN	NaN
positiveScore	NaN	NaN	NaN	NaN
score	NaN	NaN	NaN	NaN

```
grade                                NaN          NaN    NaN    NaN
```

[9]: `# Checking the correlation between the death vs all the features only`
`df.corr()['death']`

```
[9]: date                              0.270351
     positive                          0.845571
     probableCases                     0.707644
     negative                          0.824111
     pending                           0.207302
     totalTestResults                  0.852370
     hospitalizedCurrently             0.657336
     hospitalizedCumulative            0.941707
     inIcuCurrently                    0.599082
     inIcuCumulative                   0.957400
     onVentilatorCurrently             0.456948
     onVentilatorCumulative            0.886804
     recovered                         0.560364
     death                             1.000000
     hospitalized                      0.941707
     totalTestsViral                   0.866362
     positiveTestsViral                0.935220
     negativeTestsViral                0.952401
     positiveCasesViral                0.873448
     deathConfirmed                    0.997826
     deathProbable                     0.803305
     totalTestEncountersViral          0.854868
     totalTestsPeopleViral             0.919917
     totalTestsAntibody                0.947354
     positiveTestsAntibody             0.874202
     negativeTestsAntibody             0.942509
     totalTestsPeopleAntibody          0.872524
     positiveTestsPeopleAntibody       0.926676
     negativeTestsPeopleAntibody       0.960307
     totalTestsPeopleAntigen           0.709359
     positiveTestsPeopleAntigen        0.868110
     totalTestsAntigen                 0.710912
     positiveTestsAntigen              0.712723
     fips                             -0.148035
     positiveIncrease                  0.626802
     negativeIncrease                  0.269465
     total                             0.555719
     totalTestResultsIncrease          0.771587
     posNeg                            0.555724
     deathIncrease                     0.547131
     hospitalizedIncrease              0.103984
     commercialScore                        NaN
```

```
negativeRegularScore              NaN
negativeScore                     NaN
positiveScore                     NaN
score                             NaN
grade                             NaN
Name: death, dtype: float64
```

[10]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19205 entries, 0 to 19204
Data columns (total 55 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   date                    19205 non-null  int64
 1   state                   19205 non-null  object
 2   positive                19016 non-null  float64
 3   probableCases           8489 non-null   float64
 4   negative                15289 non-null  float64
 5   pending                 1999 non-null   float64
 6   totalTestResultsSource  19205 non-null  object
 7   totalTestResults        19103 non-null  float64
 8   hospitalizedCurrently   15854 non-null  float64
 9   hospitalizedCumulative  11827 non-null  float64
 10  inIcuCurrently          10358 non-null  float64
 11  inIcuCumulative         3456 non-null   float64
 12  onVentilatorCurrently   8226 non-null   float64
 13  onVentilatorCumulative  1181 non-null   float64
 14  recovered               13772 non-null  float64
 15  dataQualityGrade        17873 non-null  object
 16  lastUpdateEt            18624 non-null  object
 17  dateModified            18624 non-null  object
 18  checkTimeEt             18624 non-null  object
 19  death                   18334 non-null  float64
 20  hospitalized            11827 non-null  float64
 21  dateChecked             18624 non-null  object
 22  totalTestsViral         12701 non-null  float64
 23  positiveTestsViral      7372 non-null   float64
 24  negativeTestsViral      4413 non-null   float64
 25  positiveCasesViral      13731 non-null  float64
 26  deathConfirmed          8927 non-null   float64
 27  deathProbable           6817 non-null   float64
 28  totalTestEncountersViral 4839 non-null  float64
 29  totalTestsPeopleViral   8472 non-null   float64
 30  totalTestsAntibody      4332 non-null   float64
 31  positiveTestsAntibody   3328 non-null   float64
 32  negativeTestsAntibody   1373 non-null   float64
```

```
33  totalTestsPeopleAntibody    1699 non-null   float64
34  positiveTestsPeopleAntibody 982 non-null    float64
35  negativeTestsPeopleAntibody 888 non-null    float64
36  totalTestsPeopleAntigen     859 non-null    float64
37  positiveTestsPeopleAntigen  549 non-null    float64
38  totalTestsAntigen           2830 non-null   float64
39  positiveTestsAntigen        1862 non-null   float64
40  fips                        19205 non-null  int64
41  positiveIncrease            19205 non-null  int64
42  negativeIncrease            19205 non-null  int64
43  total                       19205 non-null  int64
44  totalTestResultsIncrease    19205 non-null  int64
45  posNeg                      19205 non-null  int64
46  deathIncrease               19205 non-null  int64
47  hospitalizedIncrease        19205 non-null  int64
48  hash                        19205 non-null  object
49  commercialScore             19205 non-null  int64
50  negativeRegularScore        19205 non-null  int64
51  negativeScore               19205 non-null  int64
52  positiveScore               19205 non-null  int64
53  score                       19205 non-null  int64
54  grade                       0 non-null      float64
dtypes: float64(33), int64(14), object(8)
memory usage: 8.1+ MB
```

[11]:
```python
# Seems the total tests done = Positive + Negative
# However there are NAN values in negative and positive columns.
# positive = 19016
# negative = 15289
# total = 19205
# The data is lagging some positive cases, many negative cases
print(df['positive'].isna().sum())
print(df['negative'].isna().sum())
print(df['total'].isna().sum())
```

```
189
3916
0
```

[12]:
```python
# Some analysis on NAN values in negative columns
df[df['negative'].isna()][['positive', 'negative', 'total']]

# It shows that, if the negative value is nan then, the total equals positive
```

[12]:
```
     positive   negative    total
0     53279.0        NaN    53279
5   3335926.0        NaN  3335926
```

```
8         38035.0       NaN      38035
11       940991.0       NaN     940991
13        27163.0       NaN      27163
...           ...       ...        ...
19200         0.0       NaN          0
19201         0.0       NaN          0
19202         0.0       NaN          0
19203         0.0       NaN          0
19204         NaN       NaN          0

[3916 rows x 3 columns]
```

```python
[13]: # Some analysis on NAN values in positive columns
      df[df['positive'].isna()][['positive', 'negative', 'total']]

      # It shows that, if the negative value is nan then, the total equals positive
```

```
[13]:        positive  negative  total
      17587       NaN       NaN      0
      17643       NaN       NaN      0
      17699       NaN       NaN      0
      17755       NaN       NaN      0
      17811       NaN       NaN      0
      ...         ...       ...    ...
      19188       NaN       NaN      0
      19190       NaN       NaN      0
      19192       NaN       NaN      0
      19194       NaN       NaN      0
      19204       NaN       NaN      0

      [189 rows x 3 columns]
```

```python
[14]: # Dropping all the columns with total = 0, which means the data is not available
      df = df[df['total']!=0]
```

```python
[15]: # Based on above filtering, checking out Nan values in positive columns
      df[df['positive'].isna()][['positive', 'negative', 'total']]
```

```
[15]:        positive  negative  total
      18577       NaN       NaN     14
      18628       NaN       NaN     14
      18679       NaN       NaN      9
      18730       NaN       NaN      9
      18781       NaN       NaN      6
      18832       NaN       NaN      2
      18883       NaN       NaN      1
```

```
[16]:   # Based on above filtering, checking out Nan values in negative columns
        df[df['negative'].isna()][['positive', 'negative', 'total']]

[16]:          positive  negative    total
        0         53279.0      NaN    53279
        5       3335926.0      NaN  3335926
        8         38035.0      NaN    38035
        11       940991.0      NaN   940991
        13        27163.0      NaN    27163
        ...           ...      ...      ...
        19193         2.0      NaN        2
        19195         2.0      NaN        2
        19196         2.0      NaN        2
        19197         1.0      NaN        1
        19198         1.0      NaN        1

        [3669 rows x 3 columns]

[17]:   # Looking at the 3669 scenarios, all the positives = total when negative is nan,
        #thus replacing all the nan of negative with 0
        df['negative'].fillna(0, inplace=True)

[18]:   df[df['negative'].isna()][['positive', 'negative', 'total']]

[18]:   Empty DataFrame
        Columns: [positive, negative, total]
        Index: []

[19]:   # Placing all the values of total to positive, when the positive is nan
        df['positive'] = df.apply(lambda row: row['total'] if np.isnan(row['positive'])⎵
         ↪else row['positive'], axis=1)

[20]:   df[df['positive'].isna()][['positive', 'negative', 'total']]

[20]:   Empty DataFrame
        Columns: [positive, negative, total]
        Index: []

[21]:   df.info()

        <class 'pandas.core.frame.DataFrame'>
        Int64Index: 18953 entries, 0 to 19198
        Data columns (total 55 columns):
         #   Column                    Non-Null Count  Dtype
        ---  ------                    --------------  -----
         0   date                      18953 non-null  int64
         1   state                     18953 non-null  object
```

```
2    positive                    18953 non-null   float64
3    probableCases               8489 non-null    float64
4    negative                    18953 non-null   float64
5    pending                     1997 non-null    float64
6    totalTestResultsSource      18953 non-null   object
7    totalTestResults            18892 non-null   float64
8    hospitalizedCurrently       15854 non-null   float64
9    hospitalizedCumulative      11823 non-null   float64
10   inIcuCurrently              10358 non-null   float64
11   inIcuCumulative             3456 non-null    float64
12   onVentilatorCurrently       8226 non-null    float64
13   onVentilatorCumulative      1181 non-null    float64
14   recovered                   13772 non-null   float64
15   dataQualityGrade            17873 non-null   object
16   lastUpdateEt                18583 non-null   object
17   dateModified                18583 non-null   object
18   checkTimeEt                 18583 non-null   object
19   death                       18283 non-null   float64
20   hospitalized                11823 non-null   float64
21   dateChecked                 18583 non-null   object
22   totalTestsViral             12585 non-null   float64
23   positiveTestsViral          7271 non-null    float64
24   negativeTestsViral          4382 non-null    float64
25   positiveCasesViral          13703 non-null   float64
26   deathConfirmed              8927 non-null    float64
27   deathProbable               6793 non-null    float64
28   totalTestEncountersViral    4790 non-null    float64
29   totalTestsPeopleViral       8468 non-null    float64
30   totalTestsAntibody          4332 non-null    float64
31   positiveTestsAntibody       3328 non-null    float64
32   negativeTestsAntibody       1373 non-null    float64
33   totalTestsPeopleAntibody    1699 non-null    float64
34   positiveTestsPeopleAntibody 982 non-null     float64
35   negativeTestsPeopleAntibody 888 non-null     float64
36   totalTestsPeopleAntigen     859 non-null     float64
37   positiveTestsPeopleAntigen  549 non-null     float64
38   totalTestsAntigen           2792 non-null    float64
39   positiveTestsAntigen        1824 non-null    float64
40   fips                        18953 non-null   int64
41   positiveIncrease            18953 non-null   int64
42   negativeIncrease            18953 non-null   int64
43   total                       18953 non-null   int64
44   totalTestResultsIncrease    18953 non-null   int64
45   posNeg                      18953 non-null   int64
46   deathIncrease               18953 non-null   int64
47   hospitalizedIncrease        18953 non-null   int64
48   hash                        18953 non-null   object
49   commercialScore             18953 non-null   int64
```

```
50   negativeRegularScore        18953 non-null  int64
51   negativeScore               18953 non-null  int64
52   positiveScore               18953 non-null  int64
53   score                       18953 non-null  int64
54   grade                       0 non-null      float64
dtypes: float64(33), int64(14), object(8)
memory usage: 8.1+ MB
```

0.1.1 Lets confirm, that the model, we want should be able to predict the death cases given the total tests done, and several other relevant features

0.1.2 The relevant features that should be used for prediction are as follows:

1. Date, just to divide the test and train data
2. state, its a categorical value, need to do one hot encoding on it, because based on the state info, the results could vary.
3. positive
4. negative
5. recovered
6. hospitalized
7. hospitalizedCumulative
8. hospitalizedCurrently
9. hospitalizedIncrease
10. inIcuCurrently
11. negativeIncrease
12. onVentilatorCumulative
13. onVentilatorCurrently
14. positiveCasesViral
15. positiveIncrease

Most of the features are dropped based on the number of non-null counts, since, the total number of the values in such features are very less, thus appending zeros or other values in such features will deviate the performance and the actual data relevancy, thus will try predicting on the basis of highly available feature columns and the relevant feature columns only.

Target is calculating the total **deaths** based on the above features.

```
[22]: # Selecting only the above mentioned features for further processing
      selected_columns = ["date", "state", "positive", "negative", "recovered",
       →"hospitalized",
                          "hospitalizedCumulative", "hospitalizedCurrently",
       →"hospitalizedIncrease",
                          "inIcuCurrently", "negativeIncrease",
       →"onVentilatorCurrently",
                          "positiveCasesViral", "positiveIncrease", "total", "death"]

      new_df = df[selected_columns]
```

```
new_df.head()
```

[22]:
```
     date state  positive    negative  recovered  hospitalized  \
0  20210207    AK   53279.0         0.0        NaN        1219.0
1  20210207    AL  472423.0   1816273.0   252880.0       43005.0
2  20210207    AR  306736.0   2285451.0   285306.0       14066.0
3  20210207    AS       0.0      2140.0        NaN          NaN
4  20210207    AZ  780637.0   2818265.0   107979.0       54657.0

   hospitalizedCumulative  hospitalizedCurrently  hospitalizedIncrease  \
0                  1219.0                   44.0                     0
1                 43005.0                 1513.0                     0
2                 14066.0                  781.0                    17
3                     NaN                    NaN                     0
4                 54657.0                 2910.0                   150

   inIcuCurrently  negativeIncrease  onVentilatorCurrently  \
0             NaN                 0                   11.0
1             NaN              4462                    NaN
2           270.0              8180                  126.0
3             NaN                 0                    NaN
4           838.0             16776                  561.0

   positiveCasesViral  positiveIncrease     total    death
0                 NaN                 0     53279    279.0
1            371056.0              1112   2288696   8515.0
2            243874.0               672   2592187   5076.0
3                 0.0                 0      2140      0.0
4            730128.0              1544   3598902  14048.0
```

[23]:
```
new_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 18953 entries, 0 to 19198
Data columns (total 16 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   date                    18953 non-null  int64
 1   state                   18953 non-null  object
 2   positive                18953 non-null  float64
 3   negative                18953 non-null  float64
 4   recovered               13772 non-null  float64
 5   hospitalized            11823 non-null  float64
 6   hospitalizedCumulative  11823 non-null  float64
 7   hospitalizedCurrently   15854 non-null  float64
 8   hospitalizedIncrease    18953 non-null  int64
 9   inIcuCurrently          10358 non-null  float64
```

```
10   negativeIncrease      18953 non-null   int64
11   onVentilatorCurrently   8226 non-null   float64
12   positiveCasesViral    13703 non-null   float64
13   positiveIncrease      18953 non-null   int64
14   total                 18953 non-null   int64
15   death                 18283 non-null   float64
dtypes: float64(10), int64(5), object(1)
memory usage: 2.5+ MB
```

[24]: `new_df.corr()['hospitalizedCumulative'] # hospitalized and` `↪hospitalizedCumulative are redundant features`

```
[24]: date                   0.236124
      positive               0.810348
      negative               0.391955
      recovered              0.337116
      hospitalized           1.000000
      hospitalizedCumulative 1.000000
      hospitalizedCurrently  0.622000
      hospitalizedIncrease   0.179922
      inIcuCurrently         0.476767
      negativeIncrease       0.155319
      onVentilatorCurrently  0.551642
      positiveCasesViral     0.919290
      positiveIncrease       0.584562
      total                  0.471509
      death                  0.941700
      Name: hospitalizedCumulative, dtype: float64
```

[25]:
```
# dropping redundant features
new_df.drop('hospitalizedCumulative', inplace=True, axis=1)
```

```
/Users/adarsh/Desktop/Machine Learning/Projects/Assignment 1 Covid
Dataset/.venv/lib/python3.8/site-packages/pandas/core/frame.py:4305:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  return super().drop(
```

[26]: `new_df.head()`

```
[26]:       date state  positive    negative   recovered  hospitalized  \
      0  20210207    AK   53279.0         0.0        NaN        1219.0
      1  20210207    AL  472423.0   1816273.0   252880.0       43005.0
      2  20210207    AR  306736.0   2285451.0   285306.0       14066.0
      3  20210207    AS       0.0      2140.0        NaN          NaN
```

```
4   20210207     AZ   780637.0  2818265.0    107979.0          54657.0

     hospitalizedCurrently  hospitalizedIncrease  inIcuCurrently  \
0                     44.0                     0             NaN
1                   1513.0                     0             NaN
2                    781.0                    17           270.0
3                      NaN                     0             NaN
4                   2910.0                   150           838.0

     negativeIncrease  onVentilatorCurrently  positiveCasesViral  \
0                   0                   11.0                 NaN
1                4462                    NaN            371056.0
2                8180                  126.0            243874.0
3                   0                    NaN                 0.0
4               16776                  561.0            730128.0

     positiveIncrease    total     death
0                   0    53279     279.0
1                1112  2288696    8515.0
2                 672  2592187    5076.0
3                   0     2140       0.0
4                1544  3598902   14048.0
```

[27]: ```python
new_df['date'] = pd.to_datetime(new_df['date'], format='%Y%m%d')
```

```
<ipython-input-27-214f810a4fa9>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  new_df['date'] = pd.to_datetime(new_df['date'], format='%Y%m%d')
```

[28]: ```python
new_df.head()
```

[28]:
```
        date state   positive    negative   recovered  hospitalized  \
0 2021-02-07    AK    53279.0         0.0         NaN        1219.0
1 2021-02-07    AL   472423.0   1816273.0    252880.0       43005.0
2 2021-02-07    AR   306736.0   2285451.0    285306.0       14066.0
3 2021-02-07    AS        0.0      2140.0         NaN           NaN
4 2021-02-07    AZ   780637.0   2818265.0    107979.0       54657.0

     hospitalizedCurrently  hospitalizedIncrease  inIcuCurrently  \
0                     44.0                     0             NaN
1                   1513.0                     0             NaN
2                    781.0                    17           270.0
3                      NaN                     0             NaN
```

```
4                            2910.0                  150          838.0

   negativeIncrease  onVentilatorCurrently  positiveCasesViral  \
0                 0                   11.0                 NaN
1              4462                    NaN            371056.0
2              8180                  126.0            243874.0
3                 0                    NaN                 0.0
4             16776                  561.0            730128.0

   positiveIncrease     total      death
0                 0     53279      279.0
1              1112   2288696     8515.0
2               672   2592187     5076.0
3                 0      2140        0.0
4              1544   3598902    14048.0
```

[29]: `new_df["state"].unique()`

[29]:
```
array(['AK', 'AL', 'AR', 'AS', 'AZ', 'CA', 'CO', 'CT', 'DC', 'DE', 'FL',
       'GA', 'GU', 'HI', 'IA', 'ID', 'IL', 'IN', 'KS', 'KY', 'LA', 'MA',
       'MD', 'ME', 'MI', 'MN', 'MO', 'MP', 'MS', 'MT', 'NC', 'ND', 'NE',
       'NH', 'NJ', 'NM', 'NV', 'NY', 'OH', 'OK', 'OR', 'PA', 'PR', 'RI',
       'SC', 'SD', 'TN', 'TX', 'UT', 'VA', 'VI', 'VT', 'WA', 'WI', 'WV',
       'WY'], dtype=object)
```

0.1.3 One hot encoding the states column

[30]:
```python
from sklearn.preprocessing import OneHotEncoder
# creating instance of one-hot-encoder
enc = OneHotEncoder(handle_unknown='ignore')
# passing state column (label encoded values of states)
enc_df = pd.DataFrame(enc.fit_transform(new_df[['state']]).toarray())
# merge with new_df on key values
new_df = new_df.join(enc_df)
new_df.head()
# enc_df
```

[30]:
```
        date state   positive     negative   recovered  hospitalized  \
0 2021-02-07    AK    53279.0          0.0         NaN        1219.0
1 2021-02-07    AL   472423.0    1816273.0    252880.0       43005.0
2 2021-02-07    AR   306736.0    2285451.0    285306.0       14066.0
3 2021-02-07    AS        0.0       2140.0         NaN          NaN
4 2021-02-07    AZ   780637.0    2818265.0    107979.0       54657.0

   hospitalizedCurrently  hospitalizedIncrease  inIcuCurrently  \
0                   44.0                     0             NaN
```

1	1513.0	0	NaN
2	781.0	17	270.0
3	NaN	0	NaN
4	2910.0	150	838.0

	negativeIncrease	onVentilatorCurrently	positiveCasesViral \
0	0	11.0	NaN
1	4462	NaN	371056.0
2	8180	126.0	243874.0
3	0	NaN	0.0
4	16776	561.0	730128.0

	positiveIncrease	total	death	0	1	2	3	4	5	6	7 \
0	0	53279	279.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1112	2288696	8515.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
2	672	2592187	5076.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
3	0	2140	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
4	1544	3598902	14048.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0

	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22 \
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37 \
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52 \
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

	53	54	55
0	0.0	0.0	0.0
1	0.0	0.0	0.0
2	0.0	0.0	0.0
3	0.0	0.0	0.0
4	0.0	0.0	0.0

```python
[31]: # Lets drop the state column now, since its already one hot encoded
      new_df.drop('state', inplace=True, axis=1)
```

```python
[32]: new_df.head()
```

```
[32]:          date    positive    negative   recovered  hospitalized  \
      0  2021-02-07    53279.0         0.0         NaN        1219.0
      1  2021-02-07   472423.0   1816273.0    252880.0       43005.0
      2  2021-02-07   306736.0   2285451.0    285306.0       14066.0
      3  2021-02-07        0.0      2140.0         NaN          NaN
      4  2021-02-07   780637.0   2818265.0    107979.0       54657.0

         hospitalizedCurrently  hospitalizedIncrease  inIcuCurrently  \
      0                   44.0                     0             NaN
      1                 1513.0                     0             NaN
      2                  781.0                    17           270.0
      3                    NaN                     0             NaN
      4                 2910.0                   150           838.0

         negativeIncrease  onVentilatorCurrently  positiveCasesViral  \
      0                 0                   11.0                 NaN
      1              4462                    NaN            371056.0
      2              8180                  126.0            243874.0
      3                 0                    NaN                 0.0
      4             16776                  561.0            730128.0

         positiveIncrease     total     death    0    1    2    3    4    5    6    7  \
      0                 0     53279     279.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
      1              1112   2288696    8515.0  0.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
      2               672   2592187    5076.0  0.0  0.0  1.0  0.0  0.0  0.0  0.0  0.0
      3                 0      2140       0.0  0.0  0.0  0.0  1.0  0.0  0.0  0.0  0.0
      4              1544   3598902   14048.0  0.0  0.0  0.0  0.0  1.0  0.0  0.0  0.0

           8    9   10   11   12   13   14   15   16   17   18   19   20   21   22  \
      0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
      1  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
      2  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
      3  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
      4  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0

          23   24   25   26   27   28   29   30   31   32   33   34   35   36   37  \
      0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
      1  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
      2  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
      3  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
      4  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
```

```
       38    39    40    41    42    43    44    45    46    47    48    49    50    51    52  \
0     0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
1     0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
2     0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
3     0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
4     0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

       53    54    55
0     0.0   0.0   0.0
1     0.0   0.0   0.0
2     0.0   0.0   0.0
3     0.0   0.0   0.0
4     0.0   0.0   0.0
```

[33]: ```python
# Lets refill all the Nans with zeros
new_df.fillna(0, inplace=True)
```

[34]: ```python
new_df.head()
```

[34]:
```
         date   positive   negative   recovered   hospitalized  \
0   2021-02-07    53279.0        0.0         0.0         1219.0
1   2021-02-07   472423.0  1816273.0    252880.0        43005.0
2   2021-02-07   306736.0  2285451.0    285306.0        14066.0
3   2021-02-07        0.0     2140.0         0.0            0.0
4   2021-02-07   780637.0  2818265.0    107979.0        54657.0

    hospitalizedCurrently   hospitalizedIncrease   inIcuCurrently  \
0                    44.0                      0              0.0
1                  1513.0                      0              0.0
2                   781.0                     17            270.0
3                     0.0                      0              0.0
4                  2910.0                    150            838.0

    negativeIncrease   onVentilatorCurrently   positiveCasesViral  \
0                  0                    11.0                  0.0
1               4462                     0.0             371056.0
2               8180                   126.0             243874.0
3                  0                     0.0                  0.0
4              16776                   561.0             730128.0

    positiveIncrease     total      death     0    1    2    3    4    5    6    7  \
0                  0     53279      279.0   1.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
1               1112   2288696     8515.0   0.0  1.0  0.0  0.0  0.0  0.0  0.0  0.0
2                672   2592187     5076.0   0.0  0.0  1.0  0.0  0.0  0.0  0.0  0.0
3                  0      2140        0.0   0.0  0.0  0.0  1.0  0.0  0.0  0.0  0.0
4               1544   3598902    14048.0   0.0  0.0  0.0  0.0  1.0  0.0  0.0  0.0
```

```
       8     9    10    11    12    13    14    15    16    17    18    19    20    21    22  \
0    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
1    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
2    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
3    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
4    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

      23    24    25    26    27    28    29    30    31    32    33    34    35    36    37  \
0    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
1    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
2    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
3    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
4    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

      38    39    40    41    42    43    44    45    46    47    48    49    50    51    52  \
0    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
1    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
2    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
3    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
4    0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

      53    54    55
0    0.0   0.0   0.0
1    0.0   0.0   0.0
2    0.0   0.0   0.0
3    0.0   0.0   0.0
4    0.0   0.0   0.0
```

0.1.4 Dividing data into train, and test sets

```python
[35]: # Since the dataset is ordered in chronological manner,
      # lets put 70% of past data as training data and latest 30% as test data
      mask = round(0.3*len(new_df))
      print(mask)
      test = new_df[:mask]
      train = new_df[mask:]
```

```
5686
```

```python
[36]: len(train)
```

```
[36]: 13267
```

```python
[37]: len(test)
```

```
[37]: 5686
```

```
[38]: train.drop("date", inplace=True, axis=1)
      test.drop('date', inplace=True, axis=1)
```

/Users/adarsh/Desktop/Machine Learning/Projects/Assignment 1 Covid
Dataset/.venv/lib/python3.8/site-packages/pandas/core/frame.py:4305:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  return super().drop(

```
[39]: # resetting the index
      train.reset_index(inplace=True, drop=True)
```

```
[40]: train_features = list(set(train.columns) - set(['death']))
      # print(train_features)
      X_train = train[train_features]
      y_train = train["death"]
      X_test = test[train_features]
      y_test = test["death"]
```

0.1.5 Simple Linear regression without any feature manipulations

```
[41]: from sklearn.linear_model import LinearRegression
      from sklearn.metrics import mean_squared_error
```

```
[42]: simple_regressor = LinearRegression()
      simple_regressor.fit(X_train, y_train)
```

```
[42]: LinearRegression()
```

```
[43]: predictions = simple_regressor.predict(X_train)
```

```
[44]: print('RMSE for Simple Linear Regression : ',np.sqrt(mean_squared_error(y_train,␣
      →predictions)))
```

RMSE for Simple Linear Regression : 781.457450314224

Standardization of feature columns
```
[49]: from sklearn.preprocessing import StandardScaler
```

```
[50]: scaler = StandardScaler()
```

```
[51]: scaler.fit(X_train)
```

```
[51]: StandardScaler()
```

```
[52]: scaler.transform(X_train)
```

```
[52]: array([[-0.13370237, -0.13341003, -0.13457585, ...,  1.89592534,
               -0.33048884,  4.21786569],
              [-0.13370237, -0.13341003, -0.13457585, ..., -0.0849822 ,
               -0.33048884, -0.2120831 ],
              [-0.13370237, -0.13341003, -0.13457585, ...,  0.14745542,
               -0.33048884,  0.1298092 ],
              ...,
              [-0.13370237, -0.13341003, -0.13457585, ..., -0.45924014,
               -0.33048884, -0.53158853],
              [-0.13370237, -0.13341003, -0.13457585, ..., -0.45924925,
               -0.33048884, -0.53158853],
              [-0.13370237, -0.13341003, -0.13457585, ..., -0.45924925,
               -0.33048884, -0.53158853]])
```

```
[53]: standardized_regressor = LinearRegression()
      standardized_regressor.fit(scaler.transform(X_train), y_train)
```

```
[53]: LinearRegression()
```

```
[54]: standardized_predictions = standardized_regressor.predict(scaler.
      ↪transform(X_train))
```

```
[55]: print('RMSE for Standardized Linear Regression : ',np.
      ↪sqrt(mean_squared_error(y_train, standardized_predictions)))
```

```
RMSE for Standardized Linear Regression :  781.4574503142238
```

Clearly, the above result of RMSE is very bad, saying that the model is underfitting

---

Trying out polynomial features for fitting the data

```
[56]: from sklearn.preprocessing import PolynomialFeatures
```

```
[59]: poly_2 = PolynomialFeatures(degree = 2)
      X_poly_2 = poly_2.fit_transform(X_train)

      poly_scaler_2 = StandardScaler()
      poly_scaler_2.fit(X_poly_2)

      poly_reg_2 = LinearRegression()
      poly_reg_2.fit(poly_scaler_2.transform(X_poly_2), y_train)
      pred_2 = poly_reg_2.predict(poly_scaler_2.transform(X_poly_2))
```

```
print('RMSE for polynomial features with degree 2 Regression :', np.
 ↪sqrt(mean_squared_error(y_train, pred_2)))
```

```
RMSE for polynomial features with degree 2 Regression : 80.83447281126186
```

Degree 2 for the polynomial feature has reduced the RMSE value by almost 10 times, the model is obviously not underfitting.

```
[60]: poly_3 = PolynomialFeatures(degree = 3)
      X_poly_3 = poly_3.fit_transform(X_train)

      poly_scaler_3 = StandardScaler()
      poly_scaler_3.fit(X_poly_3)

      poly_reg_3 = LinearRegression()
      poly_reg_3.fit(poly_scaler_3.transform(X_poly_3), y_train)
      pred_3 = poly_reg_3.predict(poly_scaler_3.transform(X_poly_3))

      print('RMSE for polynomial features with degree 3 Regression :', np.
       ↪sqrt(mean_squared_error(y_train, pred_3)))
```

```
RMSE for polynomial features with degree 3 Regression : 17.07924163175371
```

0.1.6 Result analysis

The linear regression with polynomial feature has decreased the RMSE error significantly. The degree 3 polynomial feature has the lowest RMSE in training data, that is 17.079.

```
[79]: # Lets check if the model is overfitting.
      print("RMSE for test data with polynomial regression of degree 3 : ", np.
       ↪sqrt(mean_squared_error(y_test,

                                                                              ␣
       ↪                poly_reg_3.predict(poly_scaler_3.transform(poly_3.
       ↪transform(X_test))))))
```

```
RMSE for test data with polynomial regression of degree 3 :
7.659619838501448e+16
```

It can be clearly seen from the RMSE score that the polynomial regression model of degree 3 is overfitting alot

```
[81]: # Lets check if the model is overfitting in polynomial regression with degree 2.
      test_poly_2 = PolynomialFeatures(degree = 2)
      X_test_2 = test_poly_2.fit_transform(X_test)
      print("RMSE for test data with polynomial regression of degree 2 : ", np.
       ↪sqrt(mean_squared_error(y_test,
```

```
↪                poly_reg_2.predict(poly_scaler_2.transform(poly_2.
↪transform(X_test))))))
```

RMSE for test data with polynomial regression of degree 2 : 106669159937314.16

Polynomial regression model of degree 2 is also very overfitting

```
[82]: print("RMSE for test data with standardized regressor : ", np.
      ↪sqrt(mean_squared_error(y_test, standardized_regressor.predict(scaler.
      ↪transform(X_test)))))
```

RMSE for test data with standardized regressor : 3025.8240038989434

0.1.7 Applying regularization

L2 regularization

L2 regularization on linear regression

```
[161]: from sklearn.linear_model import Ridge
       ridge_standardized = Ridge(alpha=100.0)
```

```
[162]: ridge_standardized.fit(scaler.transform(X_train), y_train)
```

```
[162]: Ridge(alpha=100.0)
```

```
[163]: print('RMSE on training data for Standardized Ridge Regression : ',np.
       ↪sqrt(mean_squared_error(y_train, ridge_standardized.predict(scaler.
       ↪transform(X_train)))))
```

RMSE on training data for Standardized Ridge Regression : 790.8509437430997

```
[164]: print("RMSE for test data with ridge regressor : ", np.
       ↪sqrt(mean_squared_error(y_test, ridge_standardized.predict(scaler.
       ↪transform(X_test)))))
```

RMSE for test data with ridge regressor : 2819.4099616242806

As from the result of regularization on standard linear regression, it can be seen that the model performance on test data has improved slightly as compared to the without regularization result.

L2 regularization on polynomial regression of degree 2

```
[165]: ridge_poly_2 = Ridge(alpha=100.0)
```

```
[166]: ridge_poly_2.fit(poly_scaler_2.transform(poly_2.transform(X_train)), y_train)
```

```
[166]: Ridge(alpha=100.0)
```

```
[167]: print('RMSE on training data for Standardized Ridge Regression with degree 2 :␣
       ↪',np.sqrt(mean_squared_error(y_train, ridge_poly_2.predict(poly_scaler_2.
       ↪transform(poly_2.transform(X_train))))))
```

    RMSE on training data for Standardized Ridge Regression : 156.11568778186938

```
[168]: print('RMSE on test data for Standardized Ridge Regression with degree 2 : ',np.
       ↪sqrt(mean_squared_error(y_test, ridge_poly_2.predict(poly_scaler_2.
       ↪transform(poly_2.transform(X_test))))))
```

    RMSE on training data for Standardized Ridge Regression : 4272.891237210047

By adding L2 regularization on polynomial regression of order 2, the model performance on test data improved by 1.06*10^15 to 4272. The addition of L2 regularization on the model has scaled the test error from factor of 10^15 to 1K, which is a huge improvement interms of the model generalizability.

L2 regularization on polynomial regression of degree 3

```
[165]: ridge_poly_3 = Ridge(alpha=100.0)
```

```
[169]: ridge_poly_3.fit(poly_scaler_3.transform(poly_3.transform(X_train)), y_train)
```

```
[169]: Ridge()
```

```
[170]: print('RMSE on training data for Standardized Ridge Regression with degree 3 :␣
       ↪',np.sqrt(mean_squared_error(y_train, ridge_poly_3.predict(poly_scaler_3.
       ↪transform(poly_3.transform(X_train))))))
```

    RMSE on training data for Standardized Ridge Regression with degree 3 :
    28.74896331312586

```
[171]: print('RMSE on test data for Standardized Ridge Regression with degree 3 : ',np.
       ↪sqrt(mean_squared_error(y_test, ridge_poly_3.predict(poly_scaler_3.
       ↪transform(poly_3.transform(X_test))))))
```

    RMSE on test data for Standardized Ridge Regression with degree 3 :
    17804.929700532004

By adding L2 regularization on polynomial regression of order 3, the model performance on test data improved by 7.6*10^16 to 17804. The generalizability of the model has improved very highly. From the result of normal polynomial regression with degree 3 the training error was only 17 however the test error was in 10^16 which is a very high error, and that was showing that the model was overfitting. But by adding regularization the training error increased by a little however the test error reduced from factor of 10^16 to 10^3. Eventhough, the model performance is not good, but the adding of penalty and its effect on model generalizibility can be clearly seen.

L1 regularization

L1 regularization on linear regression

```
[180]: from sklearn.linear_model import Lasso
       lasso_standardized = Lasso(alpha=100.0)
```

```
[181]: lasso_standardized.fit(scaler.transform(X_train), y_train)
```

```
[181]: Lasso(alpha=100.0)
```

```
[182]: print('RMSE on training data for Standardized lasso Regression : ',np.
       ↪sqrt(mean_squared_error(y_train, lasso_standardized.predict(scaler.
       ↪transform(X_train)))))
```

```
RMSE on training data for Standardized Ridge Regression :   953.3144018862494
```

```
[183]: print("RMSE for test data with lasso regressor : ", np.
       ↪sqrt(mean_squared_error(y_test, lasso_standardized.predict(scaler.
       ↪transform(X_test)))))
```

```
RMSE for test data with ridge regressor :   3070.182596553438
```

As from the result of regularization on standard linear regression, it can be seen that the model performance on test data has improved slightly as compared to the without regularization result. In addition, the model performance with L1 or L2 regression are quiet comparable

L1 regularization on polynomial regression of degree 2

```
[184]: lasso_poly_2 = Lasso(alpha=100.0)
```

```
[185]: lasso_poly_2.fit(poly_scaler_2.transform(poly_2.transform(X_train)), y_train)
```

```
[185]: Lasso(alpha=100.0)
```

```
[186]: print('RMSE on training data for Standardized Lasso Regression with degree 2 :␣
       ↪',np.sqrt(mean_squared_error(y_train, lasso_poly_2.predict(poly_scaler_2.
       ↪transform(poly_2.transform(X_train))))))
```

```
RMSE on training data for Standardized Lasso Regression with degree 2 :
607.7821862627763
```

```
[187]: print('RMSE on test data for Standardized Lasso Regression with degree 2 : ',np.
       ↪sqrt(mean_squared_error(y_test, lasso_poly_2.predict(poly_scaler_2.
       ↪transform(poly_2.transform(X_test))))))
```

```
RMSE on test data for Standardized Lasso Regression with degree 2 :
4599.883232362514
```

By adding L1 regularization on polynomial regression of order 2, the model performance on test data improved by 1.06*10^15 to 4599. The addition of L1 regularization on the model has scaled the test error from factor of 10^15 to 1K, which is a huge improvement interms of the model generalizability.

L1 regularization on polynomial regression of degree 3

```
[188]: lasso_poly_3 = Lasso(alpha=100.0)
```

```
[189]: lasso_poly_3.fit(poly_scaler_3.transform(poly_3.transform(X_train)), y_train)
```

```
[189]: Lasso(alpha=100.0)
```

```
[190]: print('RMSE on training data for Standardized Lasso Regression with degree 3 :␣
       ↪',np.sqrt(mean_squared_error(y_train, lasso_poly_3.predict(poly_scaler_3.
       ↪transform(poly_3.transform(X_train))))))
```

```
RMSE on training data for Standardized Lasso Regression with degree 3 :
598.5867956501127
```

```
[191]: print('RMSE on test data for Standardized Lasso Regression with degree 3 : ',np.
       ↪sqrt(mean_squared_error(y_test, lasso_poly_3.predict(poly_scaler_3.
       ↪transform(poly_3.transform(X_test))))))
```

```
RMSE on test data for Standardized Lasso Regression with degree 3 :
5682.515308429499
```

By adding L1 regularization on polynomial regression of order 3, the model performance on test data improved by 7.6*10^16 to 5682. The generalizability of the model has improved very highly. From the result of normal polynomial regression with degree 3 the training error was only 17 however the test error was in 10^16 which is a very high error, and that was showing that the model was overfitting. But by adding regularization the training error increased by a little however the test error reduced from factor of 10^16 to 10^3. The L1 regularization has shown better generalizibility than the L2 regularization for the polynomial regression of order 3. Eventhough, the model performance is not good, but the adding of penalty and its effect on model generalizibility can be clearly seen.

L1 and L2 regularization

L1/L2 regularization on linear regression

```
[192]: from sklearn.linear_model import ElasticNet
```

```
[200]: elastic_standardized = ElasticNet(alpha=1.0)
```

```
[201]: elastic_standardized.fit(scaler.transform(X_train), y_train)
```

```
[201]: ElasticNet()
```

```
[202]: print('RMSE on training data for Standardized lasso Regression : ',np.
       →sqrt(mean_squared_error(y_train, elastic_standardized.predict(scaler.
       →transform(X_train)))))
```

```
RMSE on training data for Standardized lasso Regression :   1191.3545334885946
```

```
[203]: print("RMSE for test data with lasso regressor : ", np.
       →sqrt(mean_squared_error(y_test, elastic_standardized.predict(scaler.
       →transform(X_test)))))
```

```
RMSE for test data with lasso regressor :   2419.857338129787
```

As from the result of regularization on standard linear regression, it can be seen that the model performance on test data has improved slightly as compared to the without regularization result. In addition, the model performance with L1, L2, regularization are quiet comparable. However, the L1/L2 regularization with very less penalty has very good performance.

L1/L2 regularization on polynomial regression of degree 2

```
[212]: elastic_poly_2 = ElasticNet(alpha=1.0)
```

```
[213]: elastic_poly_2.fit(poly_scaler_2.transform(poly_2.transform(X_train)), y_train)
```

```
[213]: ElasticNet()
```

```
[216]: print('RMSE on training data for Standardized Elastic Regression with degree 2 :␣
       →',np.sqrt(mean_squared_error(y_train, elastic_poly_2.predict(poly_scaler_2.
       →transform(poly_2.transform(X_train))))))
```

```
RMSE on training data for Standardized Elastic Regression with degree 2 :
498.7737480143985
```

```
[217]: print('RMSE on test data for Standardized Elastic Regression with degree 2 :␣
       →',np.sqrt(mean_squared_error(y_test, elastic_poly_2.predict(poly_scaler_2.
       →transform(poly_2.transform(X_test))))))
```

```
RMSE on test data for Standardized Elastic Regression with degree 2 :
9373.648415816955
```

By adding L1/L2 regularization on polynomial regression of order 2, the model performance on test data improved by $1.06*10^{15}$ to 9373. The addition of L1/L2 regularization on the model has scaled the test error from factor of 10^{15} to 10K, which is a huge improvement in terms of the model generalizability. However, the model performance is not par as compared to the L1 only and L2 only regularization.

L1/L2 regularization on polynomial regression of degree 3

```
[218]: elastic_poly_3 = ElasticNet(alpha=1.0)
```

```
[219]: elastic_poly_3.fit(poly_scaler_3.transform(poly_3.transform(X_train)), y_train)
```

/Users/adarsh/Desktop/Machine Learning/Projects/Assignment 1 Covid
Dataset/.venv/lib/python3.8/site-
packages/sklearn/linear_model/_coordinate_descent.py:530: ConvergenceWarning:
Objective did not converge. You might want to increase the number of iterations.
Duality gap: 2448670537.5238323, tolerance: 22177518.87666834
    model = cd_fast.enet_coordinate_descent(

```
[219]: ElasticNet()
```

```
[220]: print('RMSE on training data for Standardized Elastic Regression with degree 3 :␣
       →',np.sqrt(mean_squared_error(y_train, elastic_poly_3.predict(poly_scaler_3.
       →transform(poly_3.transform(X_train))))))
```

RMSE on training data for Standardized Elastic Regression with degree 3 :
346.99643131125737

```
[221]: print('RMSE on test data for Standardized Elastic Regression with degree 3 :␣
       →',np.sqrt(mean_squared_error(y_test, elastic_poly_3.predict(poly_scaler_3.
       →transform(poly_3.transform(X_test))))))
```

RMSE on test data for Standardized Elastic Regression with degree 3 :
18681.878235384993

By adding L1/L2 regularization on polynomial regression of order 3, the model performance on test data improved by 7.6*10^16 to 18681. The generalizability of the model has improved very highly. From the result of normal polynomial regression with degree 3 the training error was only 17 however the test error was in 10^16 which is a very high error, and that was showing that the model was overfitting. But by adding regularization the training error increased by a little however the test error reduced from factor of 10^16 to 10^4. Eventhough, the model performance is not good, but the adding of penalty and its effect on model generalizibility can be clearly seen.

Lastly, the addition of the regularization has improved the model generalazibility, plus improved the training time by alot. However, the model performance is not as expected, but impact of regularization can be clearly seen from the results above. The reason behind the bad model performance is because of the data. The NAN value filled with zeros is one of the reason why the model was not able to perform well. Since, introducing zeros introduces a noise in the data, which means the data behavior is completely changed due to it. Secondly, the model was not performing good enough because the polynomial features that has been introduced as also not able to capture the trend and relation in data, thus a deeper data analysis is required such that more significant and more meaningful features can be taken into consideration for the development of the model.

```
[ ]:
```