# IST 652 Final Project Code

December 10, 2021

## 0.1 COVID 19 State Wise Analysis

### 0.1.1 Team 1

1. Kirat Saran - 550246249
2. Pranav Addipalli - 482561915
3. Vinit Horakeri - 537701299

Submission Date: 12/10/2021

### 0.1.2 Importing Libraries

[1]:
```python
import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
#!pip install plotly
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots
import numpy as np
import datetime as dt
from datetime import timedelta
from sklearn.model_selection import GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score,silhouette_samples
from sklearn.linear_model import LinearRegression,Ridge,Lasso
from sklearn.svm import SVR
from sklearn.metrics import mean_squared_error,r2_score, mean_absolute_error
#import statsmodels.api as sm
#from statsmodels.tsa.api import Holt,SimpleExpSmoothing,ExponentialSmoothing
#from fbprophet import Prophet
from sklearn.preprocessing import PolynomialFeatures
#from statsmodels.tsa.stattools import adfuller
#import statsmodels.api as sm
#from statsmodels.tsa.api import Holt,SimpleExpSmoothing,ExponentialSmoothing
#from fbprophet import Prophet
```

```
from sklearn.preprocessing import PolynomialFeatures
#from statsmodels.tsa.stattools import adfuller
#std=StandardScaler()
```

[2]:

```
# setting up few restrictions
np.set_printoptions(precision=4)
pd.options.display.max_rows = 20
```

[3]:

```
#reading data
df = pd.read_csv("all-states-history.csv")
df
```

[3]:

```
             date state    death  deathConfirmed  deathIncrease  \
0      2021-03-07    AK    305.0             NaN              0
1      2021-03-07    AL  10148.0          7963.0             -1
2      2021-03-07    AR   5319.0          4308.0             22
3      2021-03-07    AS      0.0             NaN              0
4      2021-03-07    AZ  16328.0         14403.0              5
…           …   …      …             …            …
20775  2020-01-17    WA      NaN             NaN              0
20776  2020-01-16    WA      NaN             NaN              0
20777  2020-01-15    WA      NaN             NaN              0
20778  2020-01-14    WA      NaN             NaN              0
20779  2020-01-13    WA      NaN             NaN              0

       deathProbable  hospitalized  hospitalizedCumulative  \
0                NaN        1293.0                  1293.0
1             2185.0       45976.0                 45976.0
2             1011.0       14926.0                 14926.0
3                NaN           NaN                     NaN
4             1925.0       57907.0                 57907.0
…              …           …                     …
20775            NaN           NaN                     NaN
20776            NaN           NaN                     NaN
20777            NaN           NaN                     NaN
20778            NaN           NaN                     NaN
20779            NaN           NaN                     NaN

       hospitalizedCurrently  hospitalizedIncrease  …  totalTestResults  \
0                       33.0                     0  …         1731628.0
1                      494.0                     0  …         2323788.0
2                      335.0                    11  …         2736442.0
3                        NaN                     0  …            2140.0
4                      963.0                    44  …         7908105.0
…                      …                   …  …               …
20775                    NaN                     0  …               NaN
20776                    NaN                     0  …               NaN
```

```
20777                    NaN                     0  …               NaN
20778                    NaN                     0  …               NaN
20779                    NaN                     0  …               NaN

       totalTestResultsIncrease  totalTestsAntibody  totalTestsAntigen  \
0                             0                 NaN                NaN
1                          2347                 NaN                NaN
2                          3380                 NaN                NaN
3                             0                 NaN                NaN
4                         45110            580569.0                NaN
…                         …                 …                …
20775                         0                 NaN                NaN
20776                         0                 NaN                NaN
20777                         0                 NaN                NaN
20778                         0                 NaN                NaN
20779                         0                 NaN                NaN

       totalTestsPeopleAntibody  totalTestsPeopleAntigen  \
0                           NaN                      NaN
1                      119757.0                      NaN
2                           NaN                 481311.0
3                           NaN                      NaN
4                      444089.0                      NaN
…                         …                      …
20775                       NaN                      NaN
20776                       NaN                      NaN
20777                       NaN                      NaN
20778                       NaN                      NaN
20779                       NaN                      NaN

       totalTestsPeopleViral  totalTestsPeopleViralIncrease  totalTestsViral  \
0                        NaN                              0        1731628.0
1                  2323788.0                           2347              NaN
2                        NaN                              0        2736442.0
3                        NaN                              0           2140.0
4                  3842945.0                          14856        7908105.0
…                      …                            …              …
20775                    NaN                              0              NaN
20776                    NaN                              0              NaN
20777                    NaN                              0              NaN
20778                    NaN                              0              NaN
20779                    NaN                              0              NaN

       totalTestsViralIncrease
0                            0
1                            0
2                         3380
```

```
3                            0
4                        45110
…                        …
20775                        0
20776                        0
20777                        0
20778                        0
20779                        0

[20780 rows x 41 columns]
```

[4]:

```
# checking the datset schema
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20780 entries, 0 to 20779
Data columns (total 41 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   date                              20780 non-null  object
 1   state                             20780 non-null  object
 2   death                             19930 non-null  float64
 3   deathConfirmed                    9422 non-null   float64
 4   deathIncrease                     20780 non-null  int64
 5   deathProbable                     7593 non-null   float64
 6   hospitalized                      12382 non-null  float64
 7   hospitalizedCumulative            12382 non-null  float64
 8   hospitalizedCurrently             17339 non-null  float64
 9   hospitalizedIncrease              20780 non-null  int64
 10  inIcuCumulative                   3789 non-null   float64
 11  inIcuCurrently                    11636 non-null  float64
 12  negative                          13290 non-null  float64
 13  negativeIncrease                  20780 non-null  int64
 14  negativeTestsAntibody             1458 non-null   float64
 15  negativeTestsPeopleAntibody       972 non-null    float64
 16  negativeTestsViral                5024 non-null   float64
 17  onVentilatorCumulative            1290 non-null   float64
 18  onVentilatorCurrently             9126 non-null   float64
 19  positive                          20592 non-null  float64
 20  positiveCasesViral                14246 non-null  float64
 21  positiveIncrease                  20780 non-null  int64
 22  positiveScore                     20780 non-null  int64
 23  positiveTestsAntibody             3346 non-null   float64
 24  positiveTestsAntigen              2233 non-null   float64
 25  positiveTestsPeopleAntibody       1094 non-null   float64
 26  positiveTestsPeopleAntigen        633 non-null    float64
 27  positiveTestsViral                8958 non-null   float64
 28  recovered                         12003 non-null  float64
 29  totalTestEncountersViral          5231 non-null   float64
 30  totalTestEncountersViralIncrease  20780 non-null  int64
 31  totalTestResults                  20614 non-null  float64
 32  totalTestResultsIncrease          20780 non-null  int64
 33  totalTestsAntibody                4789 non-null   float64
 34  totalTestsAntigen                 3421 non-null   float64
 35  totalTestsPeopleAntibody          2200 non-null   float64
 36  totalTestsPeopleAntigen           999 non-null    float64
 37  totalTestsPeopleViral             9197 non-null   float64
 38  totalTestsPeopleViralIncrease     20780 non-null  int64
 39  totalTestsViral                   14516 non-null  float64
 40  totalTestsViralIncrease           20780 non-null  int64
dtypes: float64(30), int64(9), object(2)
memory usage: 6.5+ MB
```

[5]:

```
#checking for count of null values
df.isnull().sum()
```

[5]:

```
date                                0
state                               0
death                             850
deathConfirmed                  11358
deathIncrease                       0
                                ...
totalTestsPeopleAntigen         19781
totalTestsPeopleViral           11583
totalTestsPeopleViralIncrease       0
totalTestsViral                  6264
totalTestsViralIncrease             0
Length: 41, dtype: int64
```

### 0.1.3 Data Cleaning

[6]:

```
# Replacing Null Values
df = df.replace(np.nan,0)
df.isnull().sum()
```

[6]:

```
date                            0
state                           0
death                           0
deathConfirmed                  0
deathIncrease                   0
                               ..
totalTestsPeopleAntigen         0
totalTestsPeopleViral           0
totalTestsPeopleViralIncrease   0
```

```
totalTestsViral                    0
totalTestsViralIncrease            0
Length: 41, dtype: int64
```

Since all the values in this dataset are factual values, we cannot impute them with mean, median etc. Hence, we have replcaed them all with 0s.

```
[7]: # Setting the index to date which will help in implementing time-series anlysis
     df_new = df.set_index(['date'], append = True)
     df_new
```

```
[7]:                  state    death  deathConfirmed  deathIncrease  deathProbable  \
           date
     0     2021-03-07    AK    305.0             0.0              0            0.0
     1     2021-03-07    AL  10148.0          7963.0             -1         2185.0
     2     2021-03-07    AR   5319.0          4308.0             22         1011.0
     3     2021-03-07    AS      0.0             0.0              0            0.0
     4     2021-03-07    AZ  16328.0         14403.0              5         1925.0
     …                 …      …             …              …             …
     20775 2020-01-17    WA      0.0             0.0              0            0.0
     20776 2020-01-16    WA      0.0             0.0              0            0.0
     20777 2020-01-15    WA      0.0             0.0              0            0.0
     20778 2020-01-14    WA      0.0             0.0              0            0.0
     20779 2020-01-13    WA      0.0             0.0              0            0.0

                        hospitalized  hospitalizedCumulative  hospitalizedCurrently  \
           date
     0     2021-03-07         1293.0                  1293.0                   33.0
     1     2021-03-07        45976.0                 45976.0                  494.0
     2     2021-03-07        14926.0                 14926.0                  335.0
     3     2021-03-07            0.0                     0.0                    0.0
     4     2021-03-07        57907.0                 57907.0                  963.0
     …                            …                       …                     …
     20775 2020-01-17            0.0                     0.0                    0.0
     20776 2020-01-16            0.0                     0.0                    0.0
     20777 2020-01-15            0.0                     0.0                    0.0
     20778 2020-01-14            0.0                     0.0                    0.0
     20779 2020-01-13            0.0                     0.0                    0.0

                        hospitalizedIncrease  inIcuCumulative  …  \
           date                                                …
     0     2021-03-07                      0              0.0  …
     1     2021-03-07                      0           2676.0  …
     2     2021-03-07                     11              0.0  …
     3     2021-03-07                      0              0.0  …
     4     2021-03-07                     44              0.0  …
     …                                  …              …   …
```

```
20775 2020-01-17                     0              0.0  …
20776 2020-01-16                     0              0.0  …
20777 2020-01-15                     0              0.0  …
20778 2020-01-14                     0              0.0  …
20779 2020-01-13                     0              0.0  …

                  totalTestResults  totalTestResultsIncrease  \
      date
0     2021-03-07         1731628.0                         0
1     2021-03-07         2323788.0                      2347
2     2021-03-07         2736442.0                      3380
3     2021-03-07            2140.0                         0
4     2021-03-07         7908105.0                     45110
…                            …                       …
20775 2020-01-17               0.0                         0
20776 2020-01-16               0.0                         0
20777 2020-01-15               0.0                         0
20778 2020-01-14               0.0                         0
20779 2020-01-13               0.0                         0

                  totalTestsAntibody  totalTestsAntigen  \
      date
0     2021-03-07                 0.0                0.0
1     2021-03-07                 0.0                0.0
2     2021-03-07                 0.0                0.0
3     2021-03-07                 0.0                0.0
4     2021-03-07            580569.0                0.0
…                              …                …
20775 2020-01-17                 0.0                0.0
20776 2020-01-16                 0.0                0.0
20777 2020-01-15                 0.0                0.0
20778 2020-01-14                 0.0                0.0
20779 2020-01-13                 0.0                0.0

                  totalTestsPeopleAntibody  totalTestsPeopleAntigen  \
      date
0     2021-03-07                       0.0                      0.0
1     2021-03-07                  119757.0                      0.0
2     2021-03-07                       0.0                 481311.0
3     2021-03-07                       0.0                      0.0
4     2021-03-07                  444089.0                      0.0
…                                    …                      …
20775 2020-01-17                       0.0                      0.0
20776 2020-01-16                       0.0                      0.0
20777 2020-01-15                       0.0                      0.0
20778 2020-01-14                       0.0                      0.0
20779 2020-01-13                       0.0                      0.0
```

```
                  totalTestsPeopleViral  totalTestsPeopleViralIncrease  \
      date
0     2021-03-07                    0.0                              0
1     2021-03-07              2323788.0                           2347
2     2021-03-07                    0.0                              0
3     2021-03-07                    0.0                              0
4     2021-03-07              3842945.0                          14856
…                                 …                             …
20775 2020-01-17                    0.0                              0
20776 2020-01-16                    0.0                              0
20777 2020-01-15                    0.0                              0
20778 2020-01-14                    0.0                              0
20779 2020-01-13                    0.0                              0

                  totalTestsViral  totalTestsViralIncrease
      date
0     2021-03-07        1731628.0                        0
1     2021-03-07              0.0                        0
2     2021-03-07        2736442.0                     3380
3     2021-03-07           2140.0                        0
4     2021-03-07        7908105.0                    45110
…                           …                      …
20775 2020-01-17              0.0                        0
20776 2020-01-16              0.0                        0
20777 2020-01-15              0.0                        0
20778 2020-01-14              0.0                        0
20779 2020-01-13              0.0                        0

[20780 rows x 40 columns]
```

[8]:

```
# checking for duplicates
duplicate_rows_df = df[df.duplicated()]
print("number of duplicate rows: ", duplicate_rows_df.shape)
```

```
number of duplicate rows:  (0, 41)
```

[9]:

```
#Dropping the duplicate values
df = df.drop_duplicates()
df
```

[9]:

```
             date state    death  deathConfirmed  deathIncrease  \
0      2021-03-07    AK    305.0             0.0              0
1      2021-03-07    AL  10148.0          7963.0             -1
2      2021-03-07    AR   5319.0          4308.0             22
3      2021-03-07    AS      0.0             0.0              0
4      2021-03-07    AZ  16328.0         14403.0              5
```

```
…           …   …      …             …            …
20775  2020-01-17    WA      0.0             0.0              0
20776  2020-01-16    WA      0.0             0.0              0
20777  2020-01-15    WA      0.0             0.0              0
20778  2020-01-14    WA      0.0             0.0              0
20779  2020-01-13    WA      0.0             0.0              0

       deathProbable  hospitalized  hospitalizedCumulative  \
0                0.0        1293.0                  1293.0
1             2185.0       45976.0                 45976.0
2             1011.0       14926.0                 14926.0
3                0.0           0.0                     0.0
4             1925.0       57907.0                 57907.0
…              …           …                     …
20775            0.0           0.0                     0.0
20776            0.0           0.0                     0.0
20777            0.0           0.0                     0.0
20778            0.0           0.0                     0.0
20779            0.0           0.0                     0.0

       hospitalizedCurrently  hospitalizedIncrease  …  totalTestResults  \
0                       33.0                     0  …         1731628.0
1                      494.0                     0  …         2323788.0
2                      335.0                    11  …         2736442.0
3                        0.0                     0  …            2140.0
4                      963.0                    44  …         7908105.0
…                      …                   …  …               …
20775                    0.0                     0  …               0.0
20776                    0.0                     0  …               0.0
20777                    0.0                     0  …               0.0
20778                    0.0                     0  …               0.0
20779                    0.0                     0  …               0.0

       totalTestResultsIncrease  totalTestsAntibody  totalTestsAntigen  \
0                             0                 0.0                0.0
1                          2347                 0.0                0.0
2                          3380                 0.0                0.0
3                             0                 0.0                0.0
4                         45110            580569.0                0.0
…                         …                 …                …
20775                         0                 0.0                0.0
20776                         0                 0.0                0.0
20777                         0                 0.0                0.0
20778                         0                 0.0                0.0
20779                         0                 0.0                0.0

       totalTestsPeopleAntibody  totalTestsPeopleAntigen  \
0                           0.0                      0.0
1                      119757.0                      0.0
2                           0.0                 481311.0
3                           0.0                      0.0
4                      444089.0                      0.0
…                         …                      …
20775                       0.0                      0.0
20776                       0.0                      0.0
20777                       0.0                      0.0
20778                       0.0                      0.0
20779                       0.0                      0.0

       totalTestsPeopleViral  totalTestsPeopleViralIncrease  totalTestsViral  \
0                        0.0                              0        1731628.0
1                  2323788.0                           2347              0.0
2                        0.0                              0        2736442.0
3                        0.0                              0           2140.0
4                  3842945.0                          14856        7908105.0
…                      …                            …              …
20775                    0.0                              0              0.0
20776                    0.0                              0              0.0
20777                    0.0                              0              0.0
20778                    0.0                              0              0.0
20779                    0.0                              0              0.0

       totalTestsViralIncrease
0                            0
1                            0
2                         3380
3                            0
4                        45110
…                        …
20775                        0
20776                        0
20777                        0
20778                        0
20779                        0

[20780 rows x 41 columns]
```

### 0.1.4 Vizualizations

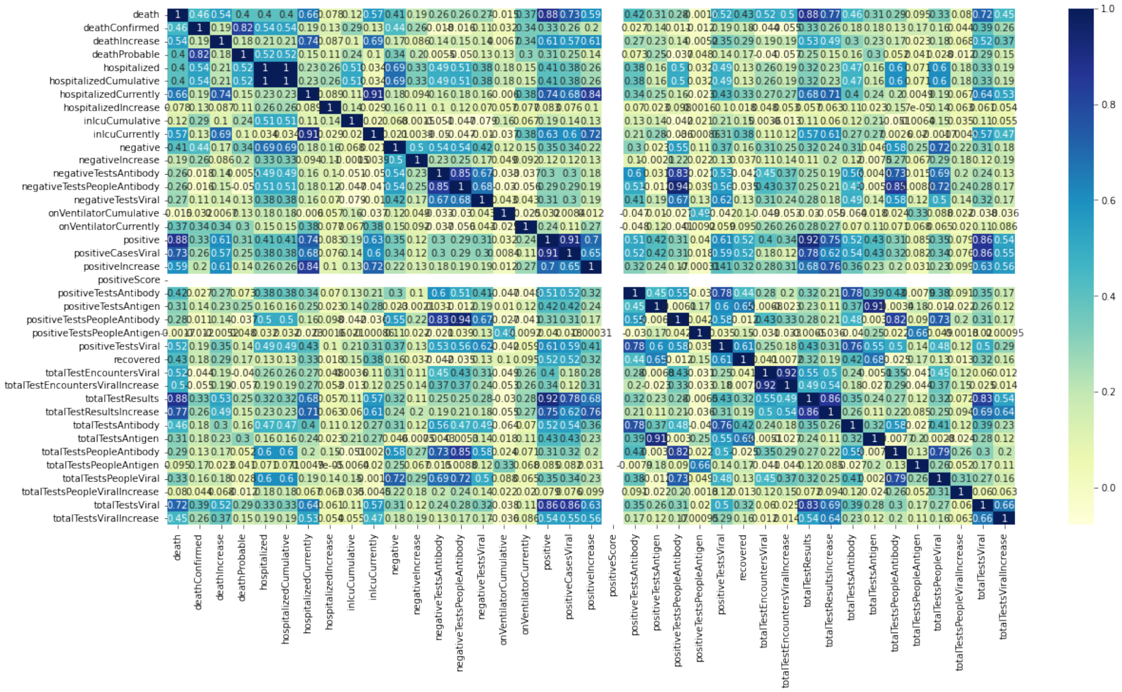
```
[10]: #Plotting a heatmap
plt.figure(figsize=(20,10))  #Define the size of the heat map
heat_map = df.corr()    #Use the correlation function
sns.heatmap(heat_map,cmap = "YlGnBu",annot=True)   #Define the colors of the␣
 ↪heatmap
```

```
print(heat_map)    #Print the heat map
```

```
                                death  deathConfirmed  deathIncrease  \
death                        1.000000        0.458202       0.544996   
deathConfirmed               0.458202        1.000000       0.190040   
deathIncrease                0.544996        0.190040       1.000000   
deathProbable                0.401842        0.822783       0.178782   
hospitalized                 0.399603        0.535771       0.211280   
…                               …             …              …      
totalTestsPeopleAntigen      0.095040        0.166877       0.022706   
totalTestsPeopleViral        0.328475        0.160163       0.184259   
totalTestsPeopleViralIncrease 0.079824        0.044079       0.067781   
totalTestsViral              0.715859        0.392120       0.522430   
totalTestsViralIncrease      0.452903        0.255140       0.370743   

                             deathProbable  hospitalized  \
death                             0.401842      0.399603   
deathConfirmed                    0.822783      0.535771   
deathIncrease                     0.178782      0.211280   
deathProbable                     1.000000      0.522451   
hospitalized                      0.522451      1.000000   
…                                    …            …      
totalTestsPeopleAntigen           0.040742      0.071220   
totalTestsPeopleViral             0.028018      0.604871   
totalTestsPeopleViralIncrease    -0.011589      0.175071   
totalTestsViral                   0.290240      0.328153   
totalTestsViralIncrease           0.150921      0.187719   

                             hospitalizedCumulative  hospitalizedCurrently  \
death                                      0.399603               0.660668   
deathConfirmed                             0.535771               0.190038   
deathIncrease                              0.211280               0.744475   
deathProbable                              0.522451               0.150118   
hospitalized                               1.000000               0.226965   
…                                             …                      …      
totalTestsPeopleAntigen                    0.071220               0.004858   
totalTestsPeopleViral                      0.604871               0.185062   
totalTestsPeopleViralIncrease              0.175071               0.067353   
totalTestsViral                            0.328153               0.644951   
totalTestsViralIncrease                    0.187719               0.528704   

                             hospitalizedIncrease  inIcuCumulative  \
death                                    0.077612         0.118222   
deathConfirmed                           0.127143         0.285404   
deathIncrease                            0.087028         0.103279   
deathProbable                            0.111791         0.239080   
hospitalized                             0.256983         0.508123   
```

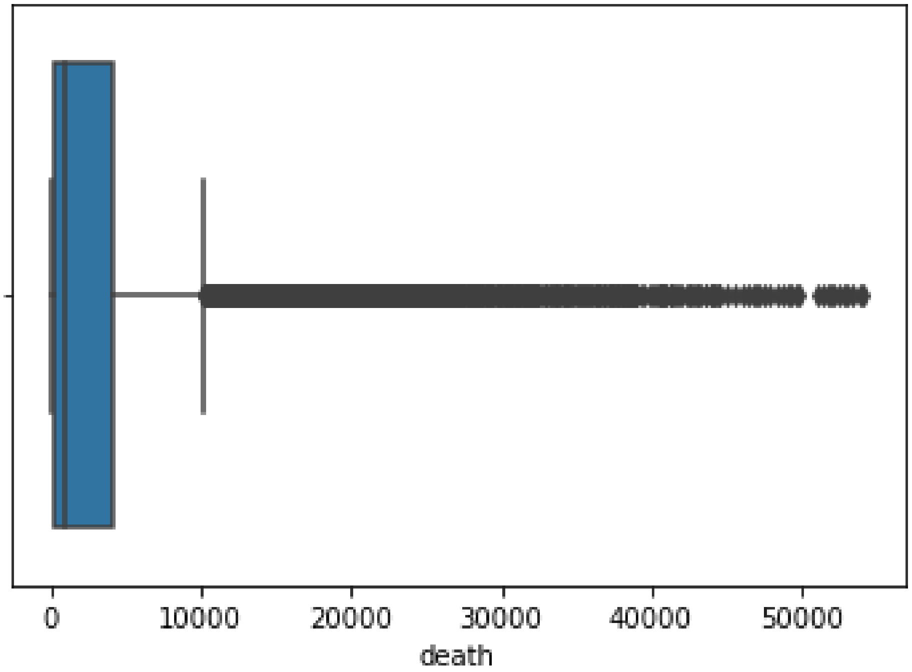
```
…                                             …              …
totalTestsPeopleAntigen                 0.000070      -0.006350
totalTestsPeopleViral                   0.139861       0.150128
totalTestsPeopleViralIncrease           0.062758       0.035433
totalTestsViral                         0.061483       0.105095
totalTestsViralIncrease                 0.053820       0.054545

                               inIcuCurrently  …  totalTestResults  \
death                                0.567410  …          0.879331
deathConfirmed                       0.133216  …          0.330150
deathIncrease                        0.693506  …          0.531354
deathProbable                        0.099909  …          0.247814
hospitalized                         0.033632  …          0.318375
…                                        …  …               …
totalTestsPeopleAntigen              0.019933  …          0.120531
totalTestsPeopleViral               -0.001711  …          0.322604
totalTestsPeopleViralIncrease       -0.004461  …          0.072154
totalTestsViral                      0.566572  …          0.830146
totalTestsViralIncrease              0.469967  …          0.539653

                               totalTestResultsIncrease  totalTestsAntibody  \
death                                          0.766859            0.460725
deathConfirmed                                 0.258907            0.181075
deathIncrease                                  0.493253            0.302596
deathProbable                                  0.148138            0.156670
hospitalized                                   0.227365            0.468539
…                                                  …                 …
totalTestsPeopleAntigen                        0.085138           -0.027398
totalTestsPeopleViral                          0.254955            0.414889
totalTestsPeopleViralIncrease                  0.094097            0.118399
totalTestsViral                                0.694387            0.390429
totalTestsViralIncrease                        0.643758            0.225608

                               totalTestsAntigen  totalTestsPeopleAntibody  \
death                                   0.311406                  0.294211
deathConfirmed                          0.177717                  0.127631
deathIncrease                           0.227180                  0.173444
deathProbable                           0.300462                  0.052314
hospitalized                            0.164995                  0.596687
…                                          …                        …
totalTestsPeopleAntigen                 0.201083                  0.133025
totalTestsPeopleViral                   0.002836                  0.788615
totalTestsPeopleViralIncrease          -0.023865                  0.258874
totalTestsViral                         0.276267                  0.304605
totalTestsViralIncrease                 0.117404                  0.196094

                               totalTestsPeopleAntigen  totalTestsPeopleViral  \
death                                         0.095040               0.328475
```

```
deathConfirmed                                0.166877               0.160163
deathIncrease                                 0.022706               0.184259
deathProbable                                 0.040742               0.028018
hospitalized                                  0.071220               0.604871
…                                                …                    …
totalTestsPeopleAntigen                       1.000000               0.262697
totalTestsPeopleViral                         0.262697               1.000000
totalTestsPeopleViralIncrease                 0.052360               0.309263
totalTestsViral                               0.166669               0.266132
totalTestsViralIncrease                       0.106756               0.160747

                               totalTestsPeopleViralIncrease  totalTestsViral  \
death                                               0.079824         0.715859
deathConfirmed                                      0.044079         0.392120
deathIncrease                                       0.067781         0.522430
deathProbable                                      -0.011589         0.290240
hospitalized                                        0.175071         0.328153
…                                                      …              …
totalTestsPeopleAntigen                             0.052360         0.166669
totalTestsPeopleViral                               0.309263         0.266132
totalTestsPeopleViralIncrease                       1.000000         0.059796
totalTestsViral                                     0.059796         1.000000
totalTestsViralIncrease                             0.062927         0.663808

                               totalTestsViralIncrease
death                                         0.452903
deathConfirmed                                0.255140
deathIncrease                                 0.370743
deathProbable                                 0.150921
hospitalized                                  0.187719
…                                                …
totalTestsPeopleAntigen                       0.106756
totalTestsPeopleViral                         0.160747
totalTestsPeopleViralIncrease                 0.062927
totalTestsViral                               0.663808
totalTestsViralIncrease                       1.000000

[39 rows x 39 columns]
```



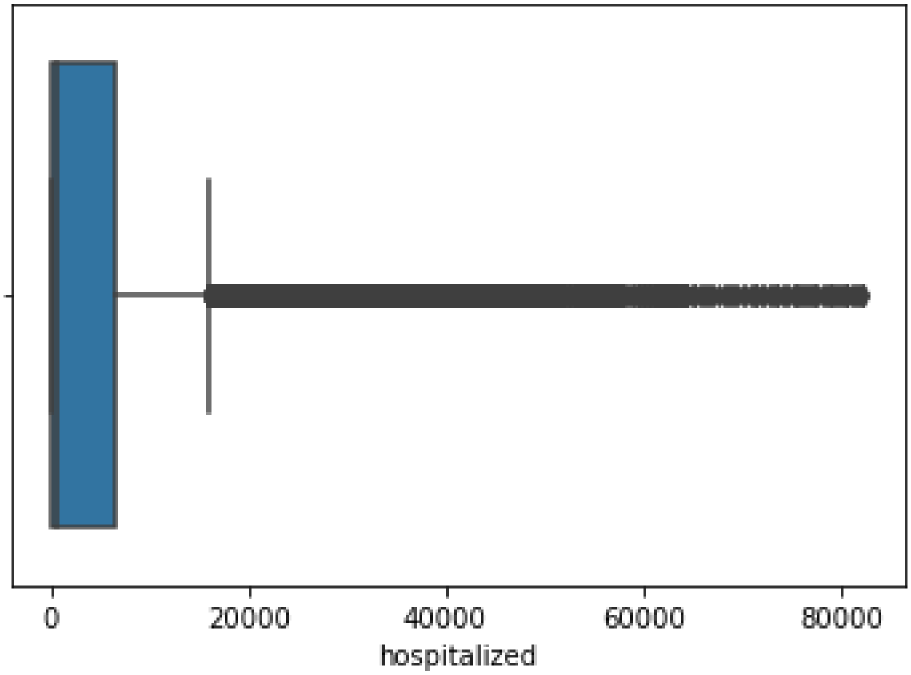
```
[11]: sns.boxplot(x = df['death'])                    #Check for outliers in 'death'␣
      ↪column using boxplot
```

```
[11]: <AxesSubplot:xlabel='death'>
```



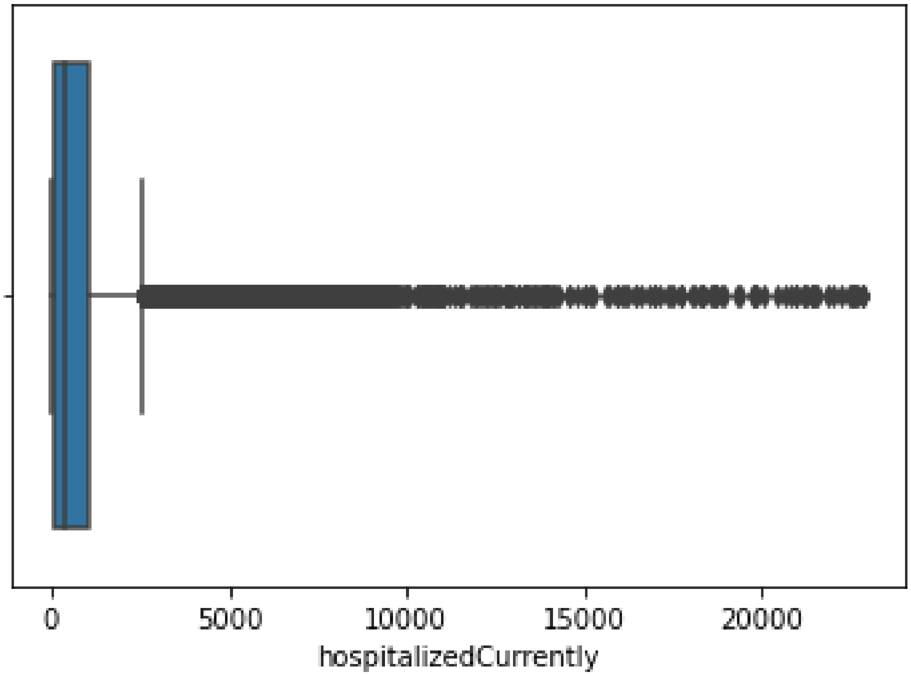
```
[12]: sns.boxplot(x = df['hospitalized'])            #Check for outliers in␣
       ↪'hospitalized' column using boxplot
```

```
[12]: <AxesSubplot:xlabel='hospitalized'>
```



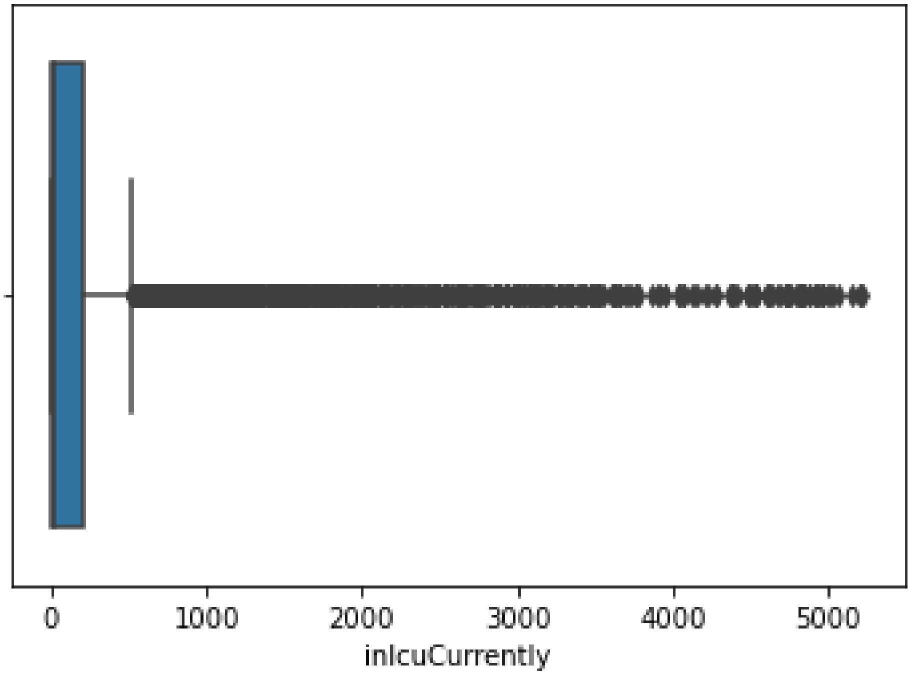
```
[13]: sns.boxplot(x = df['hospitalizedCurrently'])       #Check for outliers in␣
       ↪'hospitalizedCurrently' column using boxplot
```

```
[13]: <AxesSubplot:xlabel='hospitalizedCurrently'>
```



```
[14]: sns.boxplot(x = df['inIcuCurrently'])          #Check for outliers in␣
      ↪'inIcuCurrently' column using boxplot
```

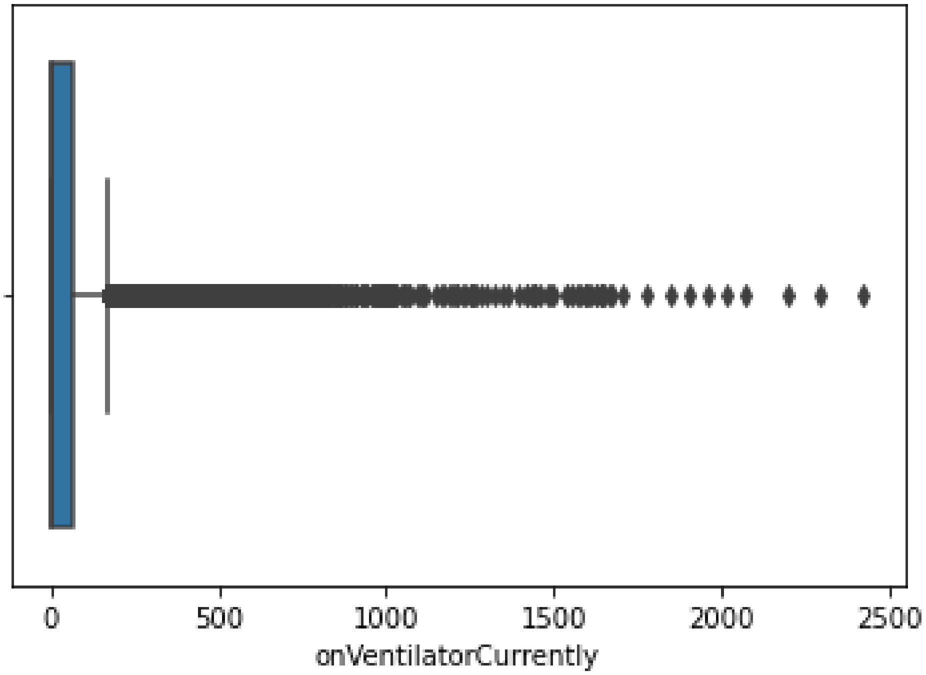
```
[14]: <AxesSubplot:xlabel='inIcuCurrently'>
```



```
[ ]:
```

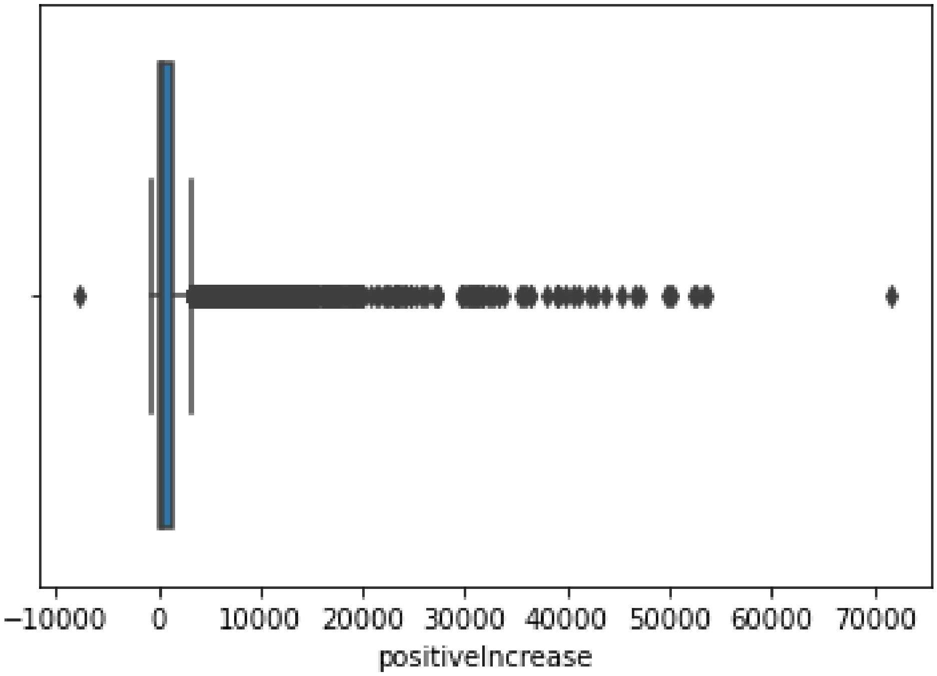
```
[15]: sns.boxplot(x = df['onVentilatorCurrently'])        #Check for outliers in␣
       ↪'onVentilatorCurrently' column using boxplot
```

```
[15]: <AxesSubplot:xlabel='onVentilatorCurrently'>
```



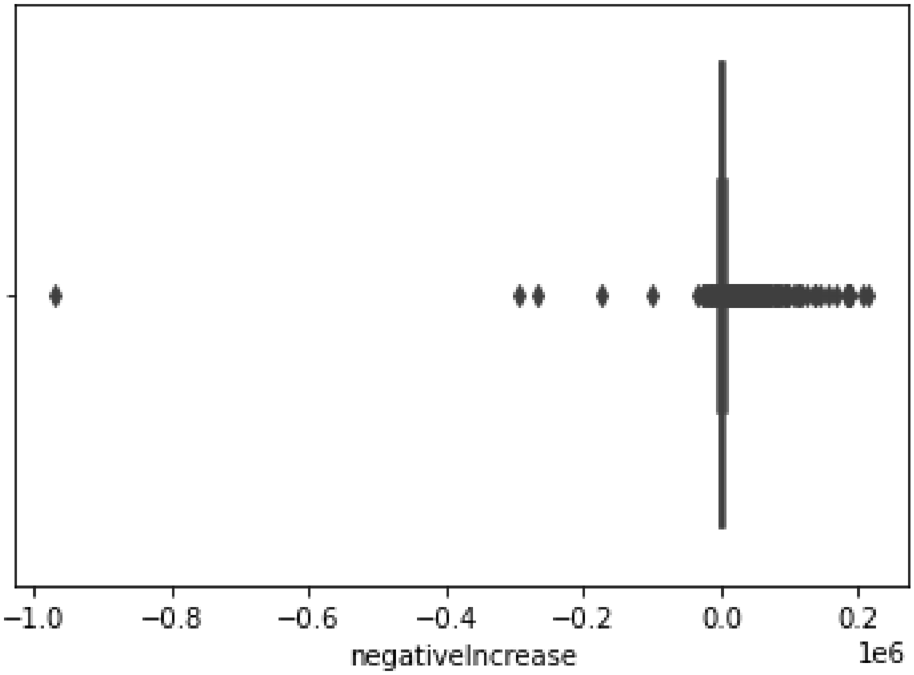
```
[16]: sns.boxplot(x = df['positiveIncrease'])        #Check for outliers in␣
       ↪'positiveIncreased' column using boxplot
```

```
[16]: <AxesSubplot:xlabel='positiveIncrease'>
```



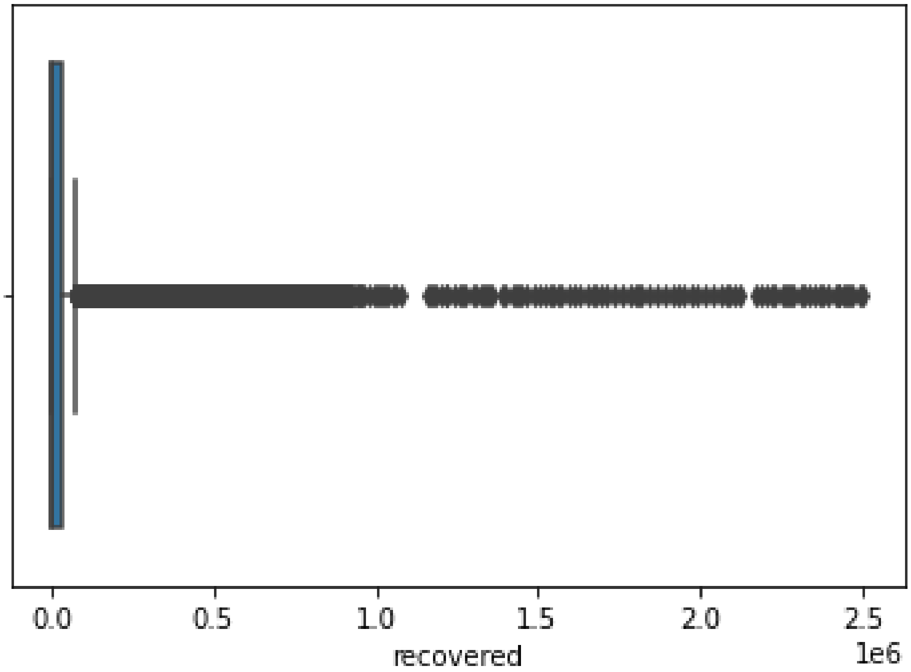
```
[17]: sns.boxplot(x = df['negativeIncrease'])      #Check for outliers in␣
       ↪'negativeIncreased' column using boxplot
```

```
[17]: <AxesSubplot:xlabel='negativeIncrease'>
```



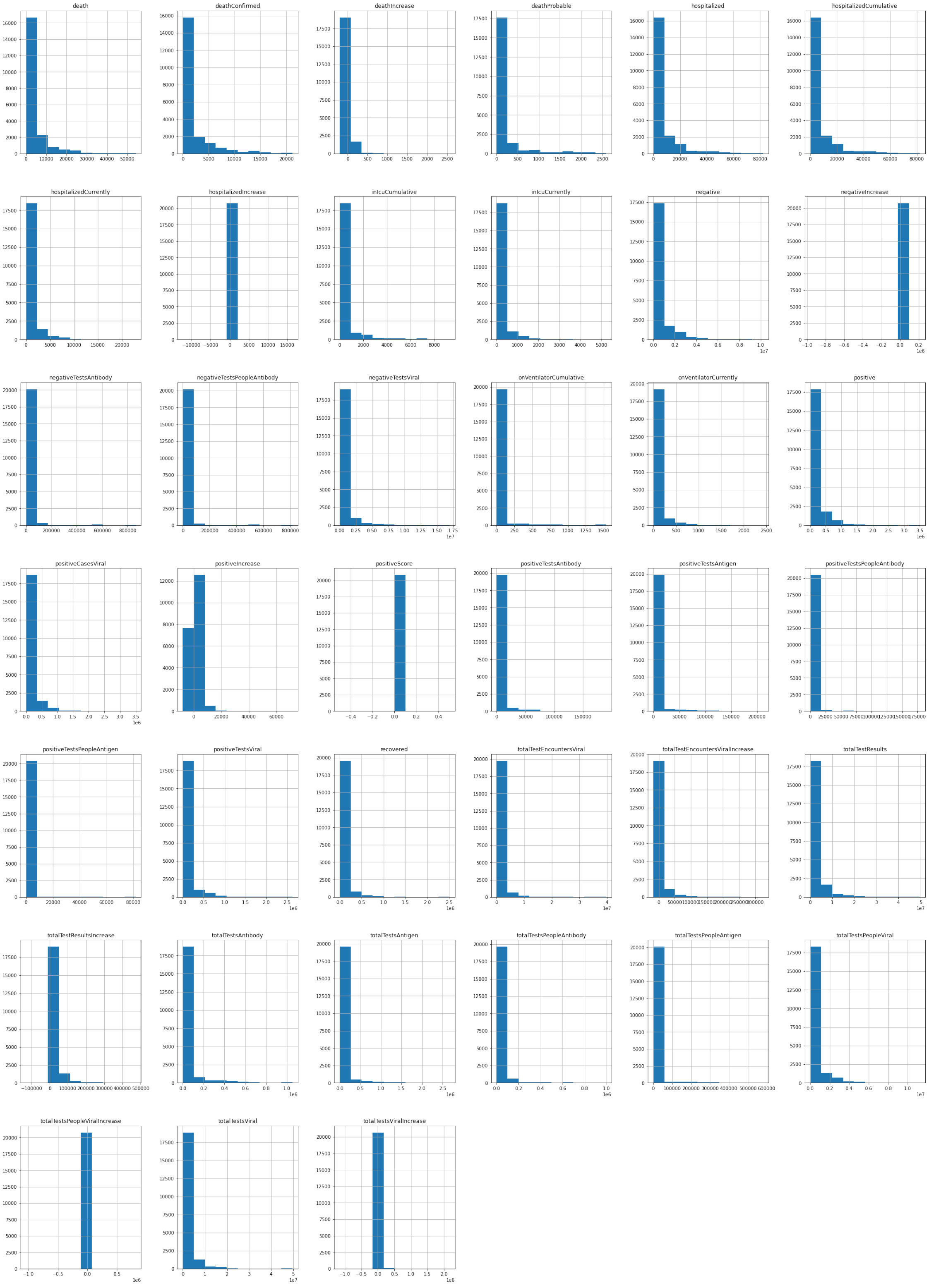
[18]:
```
sns.boxplot(x = df['recovered'])   #Check for outliers in 'recovered' column␣
 ↪using boxplot
```

[18]:
```
<AxesSubplot:xlabel='recovered'>
```



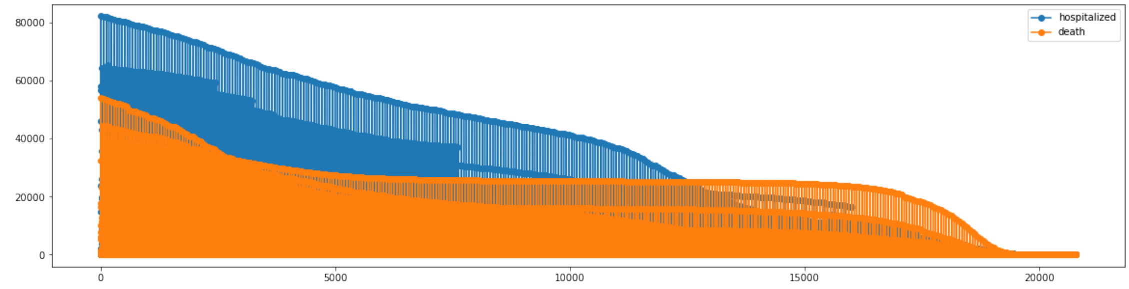
[19]:
```
# Plotting bar graphs for all the variables
fig = plt.figure(figsize = (35,50))
ax = fig.gca()
df.hist(ax = ax)
```

[19]:
```
array([[<AxesSubplot:title={'center':'death'}>,
        <AxesSubplot:title={'center':'deathConfirmed'}>,
        <AxesSubplot:title={'center':'deathIncrease'}>,
        <AxesSubplot:title={'center':'deathProbable'}>,
        <AxesSubplot:title={'center':'hospitalized'}>,
        <AxesSubplot:title={'center':'hospitalizedCumulative'}>],
       [<AxesSubplot:title={'center':'hospitalizedCurrently'}>,
        <AxesSubplot:title={'center':'hospitalizedIncrease'}>,
        <AxesSubplot:title={'center':'inIcuCumulative'}>,
        <AxesSubplot:title={'center':'inIcuCurrently'}>,
        <AxesSubplot:title={'center':'negative'}>,
        <AxesSubplot:title={'center':'negativeIncrease'}>],
       [<AxesSubplot:title={'center':'negativeTestsAntibody'}>,
        <AxesSubplot:title={'center':'negativeTestsPeopleAntibody'}>,
        <AxesSubplot:title={'center':'negativeTestsViral'}>,
        <AxesSubplot:title={'center':'onVentilatorCumulative'}>,
        <AxesSubplot:title={'center':'onVentilatorCurrently'}>,
        <AxesSubplot:title={'center':'positive'}>],
       [<AxesSubplot:title={'center':'positiveCasesViral'}>,
        <AxesSubplot:title={'center':'positiveIncrease'}>,
        <AxesSubplot:title={'center':'positiveScore'}>,
        <AxesSubplot:title={'center':'positiveTestsAntibody'}>,
        <AxesSubplot:title={'center':'positiveTestsAntigen'}>,
        <AxesSubplot:title={'center':'positiveTestsPeopleAntibody'}>],
       [<AxesSubplot:title={'center':'positiveTestsPeopleAntigen'}>,
        <AxesSubplot:title={'center':'positiveTestsViral'}>,
        <AxesSubplot:title={'center':'recovered'}>,
        <AxesSubplot:title={'center':'totalTestEncountersViral'}>,
        <AxesSubplot:title={'center':'totalTestEncountersViralIncrease'}>,
        <AxesSubplot:title={'center':'totalTestResults'}>],
       [<AxesSubplot:title={'center':'totalTestResultsIncrease'}>,
        <AxesSubplot:title={'center':'totalTestsAntibody'}>,
        <AxesSubplot:title={'center':'totalTestsAntigen'}>,
        <AxesSubplot:title={'center':'totalTestsPeopleAntibody'}>,
        <AxesSubplot:title={'center':'totalTestsPeopleAntigen'}>,
        <AxesSubplot:title={'center':'totalTestsPeopleViral'}>],
       [<AxesSubplot:title={'center':'totalTestsPeopleViralIncrease'}>,
        <AxesSubplot:title={'center':'totalTestsViral'}>,
        <AxesSubplot:title={'center':'totalTestsViralIncrease'}>,
        <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>]], dtype=object)
```


death
0 10000 20000 30000 40000 50000
deathConfirmed
0 5000 10000 15000 20000
deathIncrease
0 500 1000 1500 2000 2500
deathProbable
0 500 1000 1500 2000 2500
hospitalized
0 20000 40000 60000 80000
hospitalizedCumulative
0 20000 40000 60000 80000
hospitalizedCurrently
0 5000 10000 15000 20000
hospitalizedIncrease
−10000 −5000 0 5000 10000 15000
inIcuCumulative
0 2000 4000 6000 8000
inIcuCurrently
0 1000 2000 3000 4000 5000
negative
0.0 0.2 0.4 0.6 0.8 1.0
1e7
negativeIncrease
−1.0 −0.8 −0.6 −0.4 −0.2 0.0 0.2
1e6
negativeTestsAntibody
0 200000 400000 600000 800000
negativeTestsPeopleAntibody
0 200000 400000 600000 800000
negativeTestsViral
0.00 0.25 0.50 0.75 1.00 1.25 1.50 1.75
1e7
onVentilatorCumulative
0 250 500 750 1000 1250 1500
onVentilatorCurrently
0 500 1000 1500 2000 2500
positive
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
1e6
positiveCasesViral
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
1e6
positiveIncrease
0 20000 40000 60000
positiveScore
−0.4 −0.2 0.0 0.2 0.4
positiveTestsAntibody
0 50000 100000 150000
positiveTestsAntigen
0 50000 100000 150000 200000
positiveTestsPeopleAntibody
0 25000 50000 75000 100000 125000 150000 175000
positiveTestsPeopleAntigen
0 20000 40000 60000 80000
positiveTestsViral
0.0 0.5 1.0 1.5 2.0 2.5
1e6
recovered
0.0 0.5 1.0 1.5 2.0 2.5
1e6
totalTestEncountersViral
0 1 2 3 4
1e7
totalTestEncountersViralIncrease
0 50000 100000 150000 200000 250000 300000
totalTestResults
0 1 2 3 4 5
1e7
totalTestResultsIncrease
−100000 0 100000 200000 300000 400000 500000
totalTestsAntibody
0.0 0.2 0.4 0.6 0.8 1.0
1e6
totalTestsAntigen
0.0 0.5 1.0 1.5 2.0 2.5
1e6
totalTestsPeopleAntibody
0.0 0.2 0.4 0.6 0.8 1.0
1e6
totalTestsPeopleAntigen
0 100000 200000 300000 400000 500000 600000
totalTestsPeopleViral
0.0 0.2 0.4 0.6 0.8 1.0
1e7
totalTestsPeopleViralIncrease
−1.0 −0.5 0.0 0.5
1e6
totalTestsViral
0 1 2 3 4 5
1e7
totalTestsViralIncrease
−1.0 −0.5 0.0 0.5 1.0 1.5 2.0
1e6

**Most of the columns are positively skewed, here the mean of positively skewed data will be greater than the median.**

[21]:
```
# Creating a plot to visualize deaths vs hospitalized patients
df.plot(y=['hospitalized', 'death',], figsize=(20,5), marker='o');
```

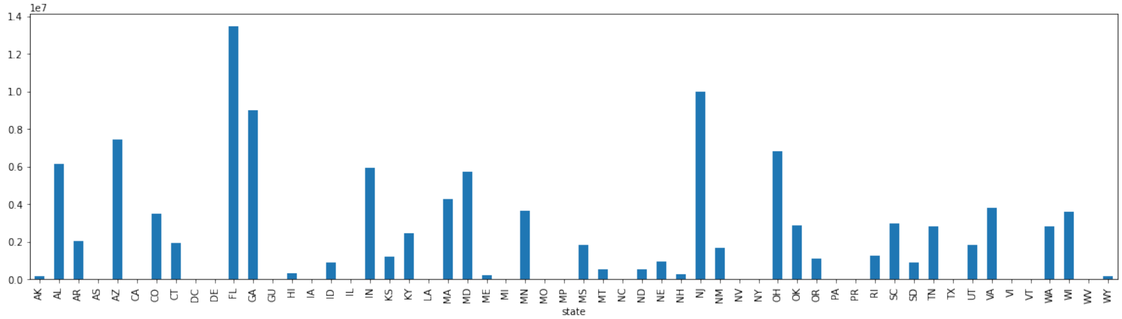


**This helps understand the mortality rate of the population due to COVID-19 virus.**

**Even though the hospitalizations seem to be decreasing the deaths are constant for a long period of time.**

[23]:
```
#Plot for state-wise hospitalizations
df.groupby(['state'])['hospitalized'].sum().plot(kind = 'bar',  figsize=(20,5))
```

[23]:
```
<AxesSubplot:xlabel='state'>
```



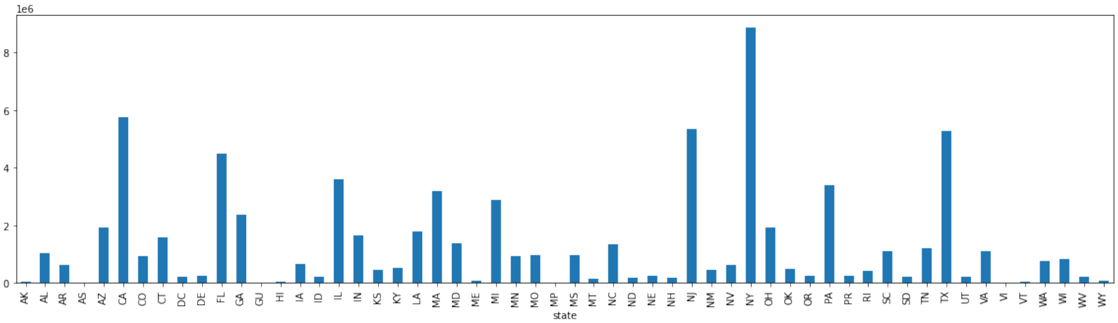
[25]:
```
# Finding maximum deaths ever recorded per day
max_deaths = df.loc[df['death'] == max(df['death'])]
max_deaths
```

[25]:
```
         date state    death  deathConfirmed  deathIncrease  deathProbable  \
5  2021-03-07    CA  54124.0             0.0            258            0.0

   hospitalized  hospitalizedCumulative  hospitalizedCurrently  \
5           0.0                     0.0                 4291.0
```

```
   hospitalizedIncrease  …  totalTestResults  totalTestResultsIncrease  \
5                     0  …        49646014.0                    133186

   totalTestsAntibody  totalTestsAntigen  totalTestsPeopleAntibody  \
5                 0.0                0.0                       0.0

   totalTestsPeopleAntigen  totalTestsPeopleViral  \
5                      0.0                    0.0

   totalTestsPeopleViralIncrease  totalTestsViral  totalTestsViralIncrease
5                              0       49646014.0                   133186

[1 rows x 41 columns]
```

[26]:
```
#Plot for state-wise deaths
df.groupby(['state'])['death'].sum().plot(kind = 'bar',  figsize=(20,5))
```

[26]:
```
<AxesSubplot:xlabel='state'>
```



**Although the hospitalizations in Florida is the highest, the maximum number of deaths are in New York**

[28]:
```
df.dtypes
```

[28]:
```
date                              object
state                             object
death                            float64
deathConfirmed                   float64
deathIncrease                      int64
                                   …
totalTestsPeopleAntigen          float64
totalTestsPeopleViral            float64
totalTestsPeopleViralIncrease      int64
totalTestsViral                  float64
```

```
totalTestsViralIncrease           int64
Length: 41, dtype: object
```

[29]:

```
# setting df index as date
df["date"]=pd.to_datetime(df["date"])
```

[30]:

```
#Grouping different types of cases as per the date
datewise=df.groupby(["date"]).agg({"positive":'sum',"recovered":'sum',"death":
 ↪'sum'})
datewise["Days Since"]=datewise.index-datewise.index[0]
```

[31]:

```
datewise
```

[31]:

```
              positive   recovered     death Days Since
date
2020-01-13         0.0         0.0       0.0     0 days
2020-01-14         0.0         0.0       0.0     1 days
2020-01-15         0.0         0.0       0.0     2 days
2020-01-16         0.0         0.0       0.0     3 days
2020-01-17         0.0         0.0       0.0     4 days
…                …         …       …        …
2021-03-03  28520365.0  10809992.0  508665.0   415 days
2021-03-04  28585852.0  10855515.0  510408.0   416 days
2021-03-05  28654639.0  10882166.0  512629.0   417 days
2021-03-06  28714654.0  10912208.0  514309.0   418 days
2021-03-07  28756489.0  10933942.0  515151.0   419 days

[420 rows x 4 columns]
```

[32]:

```
# Plotting Date-wise Recovered Cases
fig=px.bar(x=datewise.index,y=datewise["recovered"]+datewise["death"])
fig.update_layout(title="Distribution of Number of Recovered Cases",
                  xaxis_title="Date",yaxis_title="Number of Cases")
fig.show()
```

**This plot depicts that the recovered cases steeply started rising from Jan onwards.**

[33]:

```
# Plotting Weekly Growth of types of cases
datewise["WeekOfYear"]=datewise.index.weekofyear

week_num=[]
weekwise_confirmed=[]
weekwise_recovered=[]
weekwise_deaths=[]
w=1
for i in list(datewise["WeekOfYear"].unique()):
    weekwise_confirmed.append(datewise[datewise["WeekOfYear"]==i]["positive"].
 ↪iloc[-1])
    weekwise_recovered.append(datewise[datewise["WeekOfYear"]==i]["recovered"].
 ↪iloc[-1])
    weekwise_deaths.append(datewise[datewise["WeekOfYear"]==i]["death"].
 ↪iloc[-1])
    week_num.append(w)
    w=w+1

fig=go.Figure()
fig.add_trace(go.Scatter(x=week_num, y=weekwise_confirmed,
                    mode='lines+markers',
                    name='Weekly Growth of Confirmed Cases'))
fig.add_trace(go.Scatter(x=week_num, y=weekwise_recovered,
                    mode='lines+markers',
                    name='Weekly Growth of Recovered Cases'))
fig.add_trace(go.Scatter(x=week_num, y=weekwise_deaths,
                    mode='lines+markers',
                    name='Weekly Growth of Death Cases'))
fig.update_layout(title="Weekly Growth of different types of Cases",
                 xaxis_title="Week Number",yaxis_title="Number of␣
 ↪Cases",legend=dict(x=0,y=1,traceorder="normal"))
fig.show()
```

**Through Week 25, the confirmed cases started rising exponentially while the recovery rate is increasing gradually.**

[34]:

```
#Plotting Growth of different types of cases
fig=go.Figure()
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["positive"],
                    mode='lines+markers',
                    name='Confirmed Cases'))
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["recovered"],
                    mode='lines+markers',
                    name='Recovered Cases'))
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["death"],
                    mode='lines+markers',
                    name='Death Cases'))
fig.update_layout(title="Growth of different types of cases",
                 xaxis_title="Date",yaxis_title="Number of␣
 ↪Cases",legend=dict(x=0,y=1,traceorder="normal"))
fig.show()
```

**Confirmed cases started increasing with huge spike from Nov 2020**

[35]:

```
#Calculating the Mortality Rate and Recovery Rate
datewise["Mortality Rate"]=(datewise["death"]/datewise["positive"])*100
```

```
datewise["Recovery Rate"]=(datewise["recovered"]/datewise["positive"])*100
datewise["Active␣
 ↪Cases"]=datewise["positive"]-datewise["recovered"]-datewise["death"]
datewise["Closed Cases"]=datewise["recovered"]+datewise["death"]

print("Average Mortality Rate",datewise["Mortality Rate"].mean())
print("Median Mortality Rate",datewise["Mortality Rate"].median())
print("Average Recovery Rate",datewise["Recovery Rate"].mean())
print("Median Recovery Rate",datewise["Recovery Rate"].median())

#Plotting Mortality and Recovery Rate
fig = make_subplots(rows=2, cols=1,
                   subplot_titles=("Recovery Rate", "Mortatlity Rate"))
fig.add_trace(
    go.Scatter(x=datewise.index, y=(datewise["recovered"]/
 ↪datewise["positive"])*100,name="Recovery Rate"),
    row=1, col=1
)
fig.add_trace(
    go.Scatter(x=datewise.index, y=(datewise["death"]/
 ↪datewise["positive"])*100,name="Mortality Rate"),
    row=2, col=1
)
fig.update_layout(height=1000,legend=dict(x=-0.1,y=1.2,traceorder="normal"))
fig.update_xaxes(title_text="Date", row=1, col=1)
fig.update_yaxes(title_text="Recovery Rate", row=1, col=1)
fig.update_xaxes(title_text="Date", row=1, col=2)
fig.update_yaxes(title_text="Mortality Rate", row=1, col=2)
fig.show()
```

```
Average Mortality Rate 3.0811072348595387
Median Mortality Rate 2.6626517254495714
Average Recovery Rate 22.09855525602634
Median Recovery Rate 27.668506986117173
```

**Mortality rate = (Number of Death Cases / Number of Confirmed Cases) x 100**

**Recovery Rate= (Number of Recoverd Cases / Number of Confirmed Cases) x 100**

**Mortality rate is showing a considerable for a pretty long time, which is positive sign**

**Recovery Rate has started to pick up again which is a good sign, another supportive reason to why number of Closed Cases are increasing**

[36]:

```
#Plotting daily increase in cases, by type of case
print("Average increase in number of Confirmed Cases every day: ",np.
 ↪round(datewise["positive"].diff().fillna(0).mean()))
```

```
print("Average increase in number of Recovered Cases every day: ",np.
 ↪round(datewise["recovered"].diff().fillna(0).mean()))
print("Average increase in number of Deaths Cases every day: ",np.
 ↪round(datewise["death"].diff().fillna(0).mean()))

fig=go.Figure()
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["positive"].diff().
 ↪fillna(0),mode='lines+markers',
                    name='Confirmed Cases'))
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["recovered"].diff().
 ↪fillna(0),mode='lines+markers',
                    name='Recovered Cases'))
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["death"].diff().
 ↪fillna(0),mode='lines+markers',
                    name='Death Cases'))
fig.update_layout(title="Daily increase in different types of Cases",
                 xaxis_title="Date",yaxis_title="Number of␣
 ↪Cases",legend=dict(x=0,y=1,traceorder="normal"))
fig.show()
```

```
Average increase in number of Confirmed Cases every day:  68468.0
Average increase in number of Recovered Cases every day:  26033.0
Average increase in number of Deaths Cases every day:  1227.0
```

**Although the deaths were not increasing at a huge rate as it did during May 2020, the confirmed cases have never been below 100k after Nov 2020**

[37]:

```
# plotting different cases based on 7 days rolling mean
fig=go.Figure()
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["positive"].diff().
 ↪rolling(window=7).mean(),mode='lines+markers',
                    name='Confirmed Cases'))
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["recovered"].diff().
 ↪rolling(window=7).mean(),mode='lines+markers',
                    name='Recovered Cases'))
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["death"].diff().
 ↪rolling(window=7).mean(),mode='lines+markers',
                    name='Death Cases'))
fig.update_layout(title="7 Days Rolling Mean of Daily Increase of Confirmed,␣
 ↪Recovered and Death Cases",
                 xaxis_title="Date",yaxis_title="Number of␣
 ↪Cases",legend=dict(x=0,y=1,traceorder="normal"))
fig.show()
```

### 0.1.5 Exploratory Analysis (Derived Plots)

**Growth Factor** Growth factor is the factor by which a quantity multiplies itself over time. The formula used is:

**Formula: Every day's new (Confirmed,Recovered,Deaths) / new (Confirmed,Recovered,Deaths) on the previous day.**

A growth factor **above 1 indicates an increase correspoding cases**.

A growth factor **above 1 but trending downward** is a positive sign, whereas a **growth factor constantly above 1 is the sign of exponential growth**.

A growth factor **constant at 1 indicates there is no change in any kind of cases**.

[38]:

```
# PLotting datewise growth factor of different types of cases
fig=go.Figure()
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["positive"]/
 ↪datewise["positive"].shift(),
                    mode='lines',
                    name='Growth Factor of Confirmed Cases'))
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["recovered"]/
 ↪datewise["recovered"].shift(),
                    mode='lines',
                    name='Growth Factor of Recovered Cases'))
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["death"]/
 ↪datewise["death"].shift(),
                    mode='lines',
                    name='Growth Factor of Death Cases'))
fig.update_layout(title="Datewise Growth Factor of different types of cases",
                 xaxis_title="Date",yaxis_title="Growth Factor",
                 legend=dict(x=0,y=-0.4,traceorder="normal"))
fig.show()
```

**Growth Factor for Active and Closed Cases** Growth factor is the factor by which a quantity multiplies itself over time. The formula used is:

**Formula: Every day's new (Active and Closed Cases) / new (Active and Closed Cases) on the previous day.**

A growth factor **above 1 indicates an increase correspoding cases.**

A growth factor **above 1 but trending downward is a positive sign.**

A growth factor **constant at 1 indicates there is no change in any kind of cases.**

A growth factor **below 1 indicates real positive sign implying more patients are getting recovered or dying as compared to the Confirmed Cases.**

[39]:

```
# PLotting datewise growth factor of different types of active and closed cases
fig=go.Figure()
fig.add_trace(go.Scatter(x=datewise.index,
                        ␣
 ↪y=(datewise["positive"]-datewise["recovered"]-datewise["death"])/
 ↪(datewise["positive"]-datewise["recovered"]-datewise["death"]).shift(),
                    mode='lines',
                    name='Growth Factor of Active Cases'))
fig.add_trace(go.Scatter(x=datewise.index,␣
 ↪y=(datewise["recovered"]+datewise["death"])/
 ↪(datewise["recovered"]+datewise["death"]).shift(),
                    mode='lines',
                    name='Growth Factor of Closed Cases'))
fig.update_layout(title="Datewise Growth Factor of Active and Closed Cases",
                 xaxis_title="Date",yaxis_title="Growth Factor",
                 legend=dict(x=0,y=-0.4,traceorder="normal"))
fig.show()
```

**Growth Factor constantly above 1 is an clear indication of Exponential increase in all form of cases.**

### 0.1.6 Rate of Doubling for Confirmed Cases

[40]:
```
# setting the doubling rate cases count to 500
d=500
double_days=[]
C=[]
while(1):
    double_days.append(datewise[datewise["positive"]<=d].iloc[[-1]]["Days␣
 ↪Since"][0])
    C.append(d)
    d=d*2
    if(d<datewise["positive"].max()):
        continue
    else:
        break
```

[41]:
```
# checking the doubling count
doubling_rate=pd.DataFrame(list(zip(C,double_days)),columns=["No. of␣
 ↪cases","Days since first Case"])
doubling_rate["Number of days for doubling"]=doubling_rate["Days since first␣
 ↪Case"].diff().fillna(doubling_rate["Days since first Case"])
doubling_rate
```

[41]:
```
    No. of cases Days since first Case Number of days for doubling
0            500               53 days                     53 days
1           1000               55 days                      2 days
2           2000               58 days                      3 days
3           4000               60 days                      2 days
```

```
4          8000        63 days          3 days
5         16000        65 days          2 days
6         32000        68 days          3 days
7         64000        71 days          3 days
8        128000        74 days          3 days
9        256000        80 days          6 days
10       512000        88 days          8 days
11      1024000       106 days         18 days
12      2048000       151 days         45 days
13      4096000       193 days         42 days
14      8192000       280 days         87 days
15     16384000       335 days         55 days
```

**Doubling Rate is fluctuating very much, which ideally supposed to increase if we are successfully faltening the curve.**

# 1 Prediction

### 1.0.1 Linear Regression for Prediction of Confirmed Cases

[42]:
```
datewise["Days Since"]=datewise.index-datewise.index[0]
datewise["Days Since"]=datewise["Days Since"].dt.days
```

[45]:
```
train_ml=datewise.iloc[:int(datewise.shape[0]*0.95)]
valid_ml=datewise.iloc[int(datewise.shape[0]*0.95):]
model_scores=[]
```

[46]:
```
lin_reg=LinearRegression(normalize=True)
```

[47]:
```
lin_reg.fit(np.array(train_ml["Days Since"]).reshape(-1,1),np.
 ↪array(train_ml["positive"]).reshape(-1,1))
```

[47]:
```
LinearRegression(normalize=True)
```

[48]:
```
prediction_valid_linreg=lin_reg.predict(np.array(valid_ml["Days Since"]).
 ↪reshape(-1,1))
```

[49]:
```
model_scores.append(np.
 ↪sqrt(mean_absolute_error(valid_ml["positive"],prediction_valid_linreg)))
print("Mean Absolute Error for Linear Regression: ",np.
 ↪sqrt(mean_absolute_error(valid_ml["positive"],prediction_valid_linreg)))
```

```
Mean Absolute Error for Linear Regression:  2843.2559842886844
```

[50]:
```
plt.figure(figsize=(11,6))
```

```
prediction_linreg=lin_reg.predict(np.array(datewise["Days Since"]).
 ↪reshape(-1,1))
linreg_output=[]
for i in range(prediction_linreg.shape[0]):
    linreg_output.append(prediction_linreg[i][0])

fig=go.Figure()
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["positive"],
                    mode='lines+markers',name="Train Data for Confirmed Cases"))
fig.add_trace(go.Scatter(x=datewise.index, y=linreg_output,
                    mode='lines',name="Linear Regression Best Fit Line",
                    line=dict(color='black', dash='dot')))
fig.update_layout(title="Confirmed Cases Linear Regression Prediction",
                 xaxis_title="Date",yaxis_title="Confirmed␣
 ↪Cases",legend=dict(x=0,y=1,traceorder="normal"))
fig.show()
```

```
<Figure size 792x432 with 0 Axes>
```

**The Linear Regression Model is absolutely falling aprat. As it is clearly visible that the trend of Confirmed Cases in absolutely not Linear.**

### 1.0.2 Support Vector Machine ModelRegressor for Prediction of Confirmed Cases

[51]:
```
train_ml=datewise.iloc[:int(datewise.shape[0]*0.95)]
valid_ml=datewise.iloc[int(datewise.shape[0]*0.95):]
```

[52]:
```
#Intializing SVR Model
svm=SVR(C=1,degree=6,kernel='poly',epsilon=0.01)
```

[53]:
```
#Fitting model on the training data
svm.fit(np.array(train_ml["Days Since"]).reshape(-1,1),np.
 ↪array(train_ml["positive"]).reshape(-1,1))
```

[53]:
```
SVR(C=1, degree=6, epsilon=0.01, kernel='poly')
```

[54]:
```
prediction_valid_svm=svm.predict(np.array(valid_ml["Days Since"]).reshape(-1,1))
```

[55]:
```
model_scores.append(np.
 ↪sqrt(mean_absolute_error(valid_ml["positive"],prediction_valid_svm)))
print("Mean Absolute Error for Support Vectore Machine: ",np.
 ↪sqrt(mean_absolute_error(valid_ml["positive"],prediction_valid_svm)))
```

```
Mean Absolute Error for Support Vectore Machine:  3593.8969962012893
```

[56]:
```
plt.figure(figsize=(11,6))
prediction_svm=svm.predict(np.array(datewise["Days Since"]).reshape(-1,1))
fig=go.Figure()
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["positive"],
                    mode='lines+markers',name="Train Data for Confirmed Cases"))
fig.add_trace(go.Scatter(x=datewise.index, y=prediction_svm,
                    mode='lines',name="Support Vector Machine Best fit Kernel",
                    line=dict(color='black', dash='dot')))
fig.update_layout(title="Confirmed Cases Support Vectore Machine Regressor␣
 ↪Prediction",
                 xaxis_title="Date",yaxis_title="Confirmed␣
 ↪Cases",legend=dict(x=0,y=1,traceorder="normal"))
fig.show()
```

```
<Figure size 792x432 with 0 Axes>
```

**Support Vector Machine model isn't providing great results now, the predictions are either overshooting or really lower than what's expected.**

### 1.0.3 Prediction Deaths

[57]:
```
datewise["Days Since"]=datewise.index-datewise.index[0]
datewise["Days Since"]=datewise["Days Since"].dt.days
```

[58]:
```
train_ml=datewise.iloc[:int(datewise.shape[0]*0.95)]
valid_ml=datewise.iloc[int(datewise.shape[0]*0.95):]
model_scores=[]
```

[59]:
```
lin_reg=LinearRegression(normalize=True)
```

[60]:
```
lin_reg.fit(np.array(train_ml["Days Since"]).reshape(-1,1),np.
 ↪array(train_ml["death"]).reshape(-1,1))
```

[60]:
```
LinearRegression(normalize=True)
```

[61]:
```
prediction_valid_linreg=lin_reg.predict(np.array(valid_ml["Days Since"]).
 ↪reshape(-1,1))
```

[62]:
```
model_scores.append(np.
 ↪sqrt(mean_absolute_error(valid_ml["death"],prediction_valid_linreg)))
print("Mean Absolute Error for Linear Regression: ",np.
 ↪sqrt(mean_absolute_error(valid_ml["death"],prediction_valid_linreg)))

### HUGE RMS, bad MODEL
```

```
Mean Absolute Error for Linear Regression:  334.94448077283846
```

[63]:

```
plt.figure(figsize=(11,6))
prediction_linreg=lin_reg.predict(np.array(datewise["Days Since"]).
 ↪reshape(-1,1))
linreg_output=[]
for i in range(prediction_linreg.shape[0]):
    linreg_output.append(prediction_linreg[i][0])

fig=go.Figure()
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["death"],
                    mode='lines+markers',name="Train Data for Deaths"))
fig.add_trace(go.Scatter(x=datewise.index, y=linreg_output,
                    mode='lines',name="Linear Regression Best Fit Line",
                    line=dict(color='black', dash='dot')))
fig.update_layout(title="Deaths Linear Regression Prediction",
                 xaxis_title="Date",yaxis_title="Confirmed␣
 ↪Cases",legend=dict(x=0,y=1,traceorder="normal"))
fig.show()
```

```
<Figure size 792x432 with 0 Axes>
```

### 1.0.4 SVM Deaths

[64]:

```
train_ml=datewise.iloc[:int(datewise.shape[0]*0.95)]
valid_ml=datewise.iloc[int(datewise.shape[0]*0.95):]
```

[65]:

```
#Intializing SVR Model
svm=SVR(C=1,degree=6,kernel='poly',epsilon=0.01)
```

[66]:

```
#Fitting model on the training data
svm.fit(np.array(train_ml["Days Since"]).reshape(-1,1),np.
 ↪array(train_ml["death"]).reshape(-1,1))
```

[66]:

```
SVR(C=1, degree=6, epsilon=0.01, kernel='poly')
```

[67]:

```
prediction_valid_svm=svm.predict(np.array(valid_ml["Days Since"]).reshape(-1,1))
```

[68]:

```
model_scores.append(np.
 ↪sqrt(mean_absolute_error(valid_ml["death"],prediction_valid_svm)))
print("Mean Absolute Error for Support Vectore Machine: ",np.
 ↪sqrt(mean_absolute_error(valid_ml["death"],prediction_valid_svm)))
```

```
Mean Absolute Error for Support Vectore Machine:  372.1425302380849
```

[69]:

```
plt.figure(figsize=(11,6))
prediction_svm=svm.predict(np.array(datewise["Days Since"]).reshape(-1,1))
fig=go.Figure()
```

```
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["death"],
                    mode='lines+markers',name="Train Data for Confirmed Cases"))
fig.add_trace(go.Scatter(x=datewise.index, y=prediction_svm,
                    mode='lines',name="Support Vector Machine Best fit Kernel",
                    line=dict(color='black', dash='dot')))
fig.update_layout(title="Deaths Support Vectore Machine Regressor Prediction",
                 xaxis_title="Date",yaxis_title="Confirmed␣
 ↪Cases",legend=dict(x=0,y=1,traceorder="normal"))
fig.show()
```

```
<Figure size 792x432 with 0 Axes>
```

### 1.0.5 Linear regression for Recovered

[70]:

```
datewise["Days Since"]=datewise.index-datewise.index[0]
datewise["Days Since"]=datewise["Days Since"].dt.days
```

[71]:

```
train_ml=datewise.iloc[:int(datewise.shape[0]*0.95)]
valid_ml=datewise.iloc[int(datewise.shape[0]*0.95):]
model_scores=[]
```

[72]:

```
lin_reg=LinearRegression(normalize=True)
```

[73]:

```
lin_reg.fit(np.array(train_ml["Days Since"]).reshape(-1,1),np.
 ↪array(train_ml["recovered"]).reshape(-1,1))
```

[73]:

```
LinearRegression(normalize=True)
```

[74]:

```
prediction_valid_linreg=lin_reg.predict(np.array(valid_ml["Days Since"]).
 ↪reshape(-1,1))
```

[75]:

```
model_scores.append(np.
 ↪sqrt(mean_absolute_error(valid_ml["recovered"],prediction_valid_linreg)))
print("Mean Absolute Error for Linear Regression: ",np.
 ↪sqrt(mean_absolute_error(valid_ml["recovered"],prediction_valid_linreg)))

### HUGE RMS, bad MODEL
```

```
Mean Absolute Error for Linear Regression:  1941.720716154996
```

[76]:

```
plt.figure(figsize=(11,6))
prediction_linreg=lin_reg.predict(np.array(datewise["Days Since"]).
 ↪reshape(-1,1))
linreg_output=[]
for i in range(prediction_linreg.shape[0]):
    linreg_output.append(prediction_linreg[i][0])
```

```
fig=go.Figure()
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["recovered"],
                    mode='lines+markers',name="Train Data for Recovered"))
fig.add_trace(go.Scatter(x=datewise.index, y=linreg_output,
                    mode='lines',name="Linear Regression Best Fit Line",
                    line=dict(color='black', dash='dot')))
fig.update_layout(title="Recovered Linear Regression Prediction",
                 xaxis_title="Date",yaxis_title="Confirmed␣
 ↪Cases",legend=dict(x=0,y=1,traceorder="normal"))
fig.show()
```

```
<Figure size 792x432 with 0 Axes>
```

### 1.0.6 SVM for recovered

[77]:
```
train_ml=datewise.iloc[:int(datewise.shape[0]*0.95)]
valid_ml=datewise.iloc[int(datewise.shape[0]*0.95):]
```

[78]:
```
#Intializing SVR Model
svm=SVR(C=1,degree=6,kernel='poly',epsilon=0.01)
```

[79]:
```
#Fitting model on the training data
svm.fit(np.array(train_ml["Days Since"]).reshape(-1,1),np.
 ↪array(train_ml["recovered"]).reshape(-1,1))
```

[79]:
```
SVR(C=1, degree=6, epsilon=0.01, kernel='poly')
```

[80]:
```
prediction_valid_svm=svm.predict(np.array(valid_ml["Days Since"]).reshape(-1,1))
```

[81]:
```
model_scores.append(np.
 ↪sqrt(mean_absolute_error(valid_ml["recovered"],prediction_valid_svm)))
print("Mean Absolute Error for Support Vectore Machine: ",np.
 ↪sqrt(mean_absolute_error(valid_ml["recovered"],prediction_valid_svm)))
```

```
Mean Absolute Error for Support Vectore Machine:  2074.5004112041993
```

[82]:
```
plt.figure(figsize=(11,6))
prediction_svm=svm.predict(np.array(datewise["Days Since"]).reshape(-1,1))
fig=go.Figure()
fig.add_trace(go.Scatter(x=datewise.index, y=datewise["recovered"],
                    mode='lines+markers',name="Train Data for Confirmed Cases"))
fig.add_trace(go.Scatter(x=datewise.index, y=prediction_svm,
                    mode='lines',name="Support Vector Machine Best fit Kernel",
                    line=dict(color='black', dash='dot')))
fig.update_layout(title="Deaths Support Vectore Machine Regressor Prediction",
                 xaxis_title="Date",yaxis_title="Confirmed␣
↪Cases",legend=dict(x=0,y=1,traceorder="normal"))
fig.show()
```

```
<Figure size 792x432 with 0 Axes>
```

### 1.0.7 Insights and Conclusion

**We can see a common trend where the case count in high from Nov 2020. This might be because of many factors such as it being the Flu season, Vaccine rollout.**

**The recovery and doubling rate increased as time passed by because of antibodies tend to increase once a person contracts the virus**