

I have taken the covid dataset which is happend in india.where the data contains of states ,total cases in that particular state,active cases,discharged ,number of deaths and percentages of the active,discharged and death ratio.

```
In [1]: #importing the libraries
import os
import pandas as pd
import numpy as np
```

```
In [2]: #reading the dataset
path = "covid.csv"
data = pd.read_csv(path)
print(data)
```

| | State | Total Cases | Active | Discharged | \ |
|----|--|-------------|--------|------------|---|
| 0 | Maharashtra | 6122893 | 117869 | 5881167 | |
| 1 | Kerala | 3011694 | 108400 | 2889186 | |
| 2 | Karnataka | 2862338 | 39626 | 2787111 | |
| 3 | Tamil Nadu | 2506848 | 34076 | 2439576 | |
| 4 | Andhra Pradesh | 1911231 | 32356 | 1865956 | |
| 5 | Uttar Pradesh | 1706934 | 1947 | 1682321 | |
| 6 | West Bengal | 1508223 | 16655 | 1473718 | |
| 7 | Delhi | 1434780 | 858 | 1408917 | |
| 8 | Chhattisgarh | 996689 | 5017 | 978208 | |
| 9 | Rajasthan | 952887 | 935 | 943010 | |
| 10 | Odisha | 929788 | 25148 | 900282 | |
| 11 | Gujarat | 824029 | 1969 | 811988 | |
| 12 | Madhya Pradesh | 780870 | 441 | 780610 | |
| 13 | Haryana | 769093 | 1066 | 758512 | |
| 14 | Bihar | 722850 | 1183 | 712055 | |
| 15 | Telangana | 629054 | 11472 | 613872 | |
| 16 | Punjab | 596970 | 1964 | 578865 | |
| 17 | Assam | 524556 | 23371 | 496442 | |
| 18 | Jharkhand | 346113 | 532 | 340463 | |
| 19 | Uttarakhand | 340959 | 1506 | 332115 | |
| 20 | Jammu and Kashmir | 317761 | 34391 | 309973 | |
| 21 | Himachal Pradesh | 293117 | 1338 | 198293 | |
| 22 | Goa | 168015 | 1959 | 162983 | |
| 23 | Puducherry | 118416 | 1753 | 114898 | |
| 24 | Manipur | 74341 | 6534 | 66578 | |
| 25 | Tripura | 69019 | 4122 | 64194 | |
| 26 | Chandigarh | 61780 | 99 | 60872 | |
| 27 | Meghalaya | 52712 | 4229 | 47597 | |
| 28 | Arunachal Pradesh | 38283 | 3363 | 34736 | |
| 29 | Nagaland | 25663 | 1070 | 24088 | |
| 30 | Mizoram | 22703 | 3960 | 18643 | |
| 31 | Sikkim | 21573 | 1975 | 19287 | |
| 32 | Ladakh | 20143 | 183 | 19756 | |
| 33 | Dadra and Nagar Haveli and Daman and Diu | 10576 | 38 | 10534 | |
| 34 | Lakshadweep | 9965 | 246 | 9670 | |
| 35 | Andaman and Nicobar | 7491 | 14 | 7349 | |

| | Deaths | Active Ratio (%) | Discharge Ratio (%) | Death Ratio (%) |
|----|--------|------------------|---------------------|-----------------|
| 0 | 123857 | 1.93 | 96.05 | 2.02 |
| 1 | 14108 | 3.60 | 95.93 | 0.47 |
| 2 | 35601 | 1.38 | 97.37 | 1.24 |
| 3 | 33196 | 1.36 | 97.32 | 1.32 |
| 4 | 12919 | 1.69 | 97.63 | 1.68 |
| 5 | 22666 | 0.11 | 98.56 | 1.33 |
| 6 | 17850 | 1.10 | 97.71 | 1.18 |
| 7 | 25005 | 0.06 | 98.20 | 1.74 |
| 8 | 13464 | 0.50 | 98.15 | 1.35 |
| 9 | 8942 | 0.10 | 98.96 | 0.94 |
| 10 | 4358 | 2.70 | 96.83 | 0.47 |
| 11 | 10072 | 0.24 | 98.54 | 1.22 |
| 12 | 9019 | 0.06 | 98.80 | 1.14 |
| 13 | 9515 | 0.14 | 98.62 | 1.24 |
| 14 | 9612 | 0.16 | 98.51 | 1.33 |
| 15 | 3710 | 1.82 | 97.59 | 0.59 |
| 16 | 16141 | 0.33 | 96.97 | 2.70 |
| 17 | 4743 | 4.46 | 94.64 | 0.90 |
| 18 | 5118 | 0.15 | 98.37 | 1.48 |
| 19 | 7338 | 0.44 | 97.41 | 2.15 |
| 20 | 4349 | 1.08 | 97.55 | 1.37 |
| 21 | 3486 | 0.66 | 97.63 | 1.72 |
| 22 | 3082 | 1.16 | 97.01 | 1.83 |
| 23 | 1765 | 1.48 | 97.03 | 1.49 |
| 24 | 1229 | 8.79 | 89.56 | 1.65 |
| 25 | 703 | 5.97 | 93.01 | 1.02 |
| 26 | 809 | 0.16 | 98.53 | 1.31 |
| 27 | 886 | 0.02 | 90.30 | 1.68 |
| 28 | 184 | 8.78 | 90.73 | 0.48 |
| 29 | 505 | 4.17 | 93.86 | 1.97 |
| 30 | 100 | 17.44 | 82.12 | 0.44 |
| 31 | 311 | 9.15 | 89.40 | 1.44 |
| 32 | 204 | 0.91 | 98.08 | 1.01 |
| 33 | 4 | 0.36 | 99.60 | 0.04 |
| 34 | 49 | 2.47 | 97.04 | 0.49 |
| 35 | 128 | 0.19 | 98.10 | 1.71 |

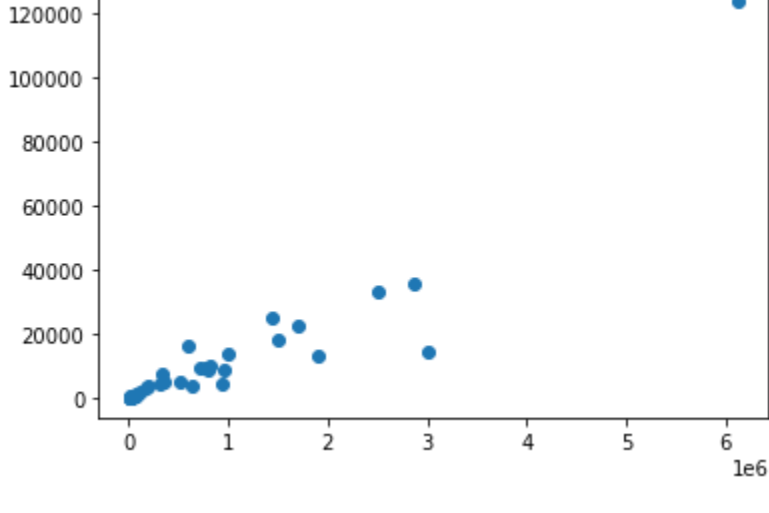
```
In [3]: data.head()
```

| | State | Total Cases | Active | Discharged | Deaths | Active Ratio (%) | Discharge Ratio (%) | Death Ratio (%) |
|---|----------------|-------------|--------|------------|--------|------------------|---------------------|-----------------|
| 0 | Maharashtra | 6122893 | 117869 | 5881167 | 123857 | 1.93 | 96.05 | 2.02 |
| 1 | Kerala | 3011694 | 108400 | 2889186 | 14108 | 3.60 | 95.93 | 0.47 |
| 2 | Karnataka | 2862338 | 39626 | 2787111 | 35601 | 1.38 | 97.37 | 1.24 |
| 3 | Tamil Nadu | 2506848 | 34076 | 2439576 | 33196 | 1.36 | 97.32 | 1.32 |
| 4 | Andhra Pradesh | 1911231 | 32356 | 1865956 | 12919 | 1.69 | 97.63 | 1.68 |

```
In [4]: print(data.shape)
print(data.columns)
print(data.dtypes)

(36, 8)
Index(['State', 'Total Cases', 'Active', 'Discharged', 'Deaths',
       'Active Ratio (%)', 'Discharge Ratio (%)', 'Death Ratio (%)'],
      dtype='object')
State          object
Total Cases    int64
Active         int64
Discharged     int64
Deaths         int64
Active Ratio (%) float64
Discharge Ratio (%) float64
Death Ratio (%) float64
dtype: object
```

```
In [5]: import matplotlib.pyplot as plt
x=data['Total Cases']
y=data['Deaths']
plt.scatter(x,y)
plt.show()
```



```
In [6]: data['Deaths'].sort_values()
```

| | |
|----------------------------|--------|
| Out[6]: | 4 |
| 34 | 49 |
| 30 | 100 |
| 35 | 128 |
| 28 | 184 |
| 32 | 204 |
| 31 | 311 |
| 29 | 505 |
| 25 | 703 |
| 26 | 809 |
| 27 | 886 |
| 24 | 1229 |
| 23 | 1765 |
| 22 | 3082 |
| 21 | 3486 |
| 15 | 3710 |
| 20 | 4349 |
| 10 | 4358 |
| 17 | 4743 |
| 18 | 5118 |
| 19 | 7338 |
| 9 | 8942 |
| 12 | 9019 |
| 13 | 9515 |
| 14 | 9612 |
| 11 | 10072 |
| 4 | 12919 |
| 8 | 13464 |
| 1 | 14108 |
| 16 | 16141 |
| 6 | 17850 |
| 5 | 22666 |
| 7 | 25005 |
| 3 | 33196 |
| 2 | 35601 |
| 0 | 123857 |
| Name: Deaths, dtype: int64 | |

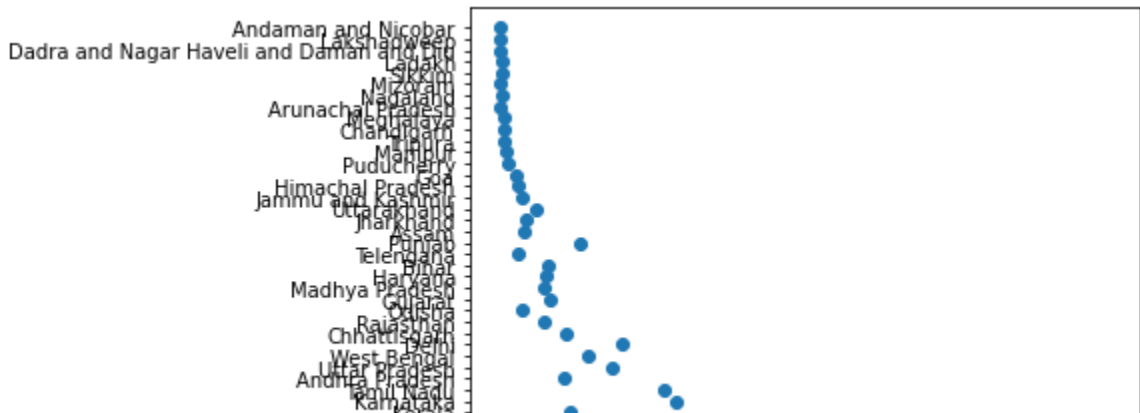
```
In [7]: data.Deaths.value_counts()
```

| | |
|----------------------------|---|
| Out[7]: | 1 |
| 703 | 1 |
| 3710 | 1 |
| 3486 | 1 |
| 14108 | 1 |
| 3082 | 1 |
| 13464 | 1 |
| 311 | 1 |
| 123857 | 1 |
| 16141 | 1 |
| 9612 | 1 |
| 22666 | 1 |
| 5118 | 1 |
| 4743 | 1 |
| 4358 | 1 |
| 204 | 1 |
| 4 | 1 |
| 10072 | 1 |
| 100 | 1 |
| 1765 | 1 |
| 809 | 1 |
| 12919 | 1 |
| 4349 | 1 |
| 35601 | 1 |
| 9019 | 1 |
| 17850 | 1 |
| 505 | 1 |
| 184 | 1 |
| 886 | 1 |
| 7338 | 1 |
| 49 | 1 |
| 1229 | 1 |
| 8942 | 1 |
| 25005 | 1 |
| 33196 | 1 |
| 9515 | 1 |
| 128 | 1 |
| Name: Deaths, dtype: int64 | |

```
In [8]: data['State'].sort_values()
```

| | |
|----------------------------|--|
| Out[8]: | Andaman and Nicobar |
| 4 | Andhra Pradesh |
| 28 | Arunachal Pradesh |
| 17 | Assam |
| 14 | Bihar |
| 26 | Chandigarh |
| 8 | Chhattisgarh |
| 33 | Dadra and Nagar Haveli and Daman and Diu |
| 7 | Delhi |
| 22 | Goa |
| 11 | Gujarat |
| 13 | Haryana |
| 21 | Himachal Pradesh |
| 20 | Jammu and Kashmir |
| 18 | Jharkhand |
| 2 | Karnataka |
| 1 | Kerala |
| 32 | Ladakh |
| 34 | Lakshadweep |
| 12 | Madhya Pradesh |
| 0 | Maharashtra |
| 24 | Manipur |
| 27 | Meghalaya |
| 30 | Mizoram |
| 29 | Nagaland |
| 10 | Odisha |
| 23 | Puducherry |
| 16 | Punjab |
| 9 | Rajasthan |
| 31 | Sikkim |
| 3 | Tamil Nadu |
| 15 | Telangana |
| 25 | Tripura |
| 5 | Uttar Pradesh |
| 19 | Uttarakhand |
| 6 | West Bengal |
| Name: State, dtype: object | |

```
In [9]: y=data['State']
x = data['Deaths']
plt.scatter(x,y)
plt.show()
```



The one Which is in the rigth bottom is called outlier.

```
In [11]: data.groupby('Deaths').median()
```

| | Total Cases | Active | Discharged | Active Ratio (%) | Discharge Ratio (%) | Death Ratio (%) |
|--------|-------------|--------|------------|------------------|---------------------|-----------------|
| Deaths | | | | | | |
| 4 | 10576 | 38 | 10534 | 0.36 | 99.60 | 0.04 |
| 49 | 9965 | 246 | 9670 | 2.47 | 97.04 | 0.49 |
| 100 | 22703 | 3960 | 18643 | 17.44 | 82.12 | 0.44 |
| 128 | 7491 | 14 | 7349 | 0.19 | 98.10 | 1.71 |
| 184 | 38283 | 3363 | 34736 | 8.78 | 90.73 | 0.48 |
| 204 | 20143 | 183 | 19756 | 0.91 | 98.08 | 1.01 |
| 311 | 21573 | 1975 | 19287 | 9.15 | 89.40 | 1.44 |
| 505 | 25663 | 1070 | 24088 | 4.17 | 93.86 | 1.97 |
| 703 | 69019 | 4122 | 64194 | 5.97 | 93.01 | 1.02 |
| 809 | 61780 | 99 | 60872 | 0.16 | 98.53 | 1.31 |
| 886 | 52712 | 4229 | 47597 | 8.02 | 90.30 | 1.68 |
| 1229 | 74341 | 6534 | 66578 | 8.79 | 89.56 | 1.65 |
| 1765 | 118416 | 1753 | 114898 | 1.48 | 97.03 | 1.49 |
| 3082 | 168015 | 1950 | 162983 | 1.16 | 97.01 | 1.83 |
| 3486 | 203117 | 1338 | 198293 | 0.66 | 97.63 | 1.72 |
| 3710 | 629054 | 11472 | 613872 | 1.82 | 97.59 | 0.59 |
| 4349 | 317761 | 34391 | 309973 | 1.08 | 97.55 | 1.37 |
| 4358 | 929788 | 25148 | 900282 | 2.70 | 96.83 | 0.47 |
| 4743 | 524556 | 23371 | 496442 | 4.46 | 94.64 | 0.90 |
| 5118 | 346113 | 532 | 340463 | 0.15 | 98.37 | 1.48 |
| 7338 | 340959 | 1506 | 332115 | 0.44 | 97.41 | 2.15 |
| 8942 | 952887 | 935 | 943010 | 0.10 | 98.96 | 0.94 |
| 9019 | 790070 | 441 | 780610 | 0.06 | 98.80 | 1.14 |
| 9515 | 769093 | 1066 | 758512 | 0.14 | 98.62 | 1.24 |
| 9612 | 722850 | 1183 | 712055 | 0.16 | 98.51 | 1.33 |
| 10072 | 824029 | 1969 | 811988 | 0.24 | 98.54 | 1.22 |
| 12919 | 1911231 | 32356 | 1865956 | 1.69 | 97.63 | 0.68 |
| 13464 | 996689 | 5017 | 978208 | 0.50 | 98.15 | 1.35 |
| 14108 | 3011694 | 108400 | 2889186 | 3.60 | 95.93 | 0.47 |
| 16141 | 596970 | 1964 | 578865 | 0.33 | 96.97 | 2.70 |
| 17850 | 1508223 | 16655 | 1473718 | 1.10 | 97.71 | 1.18 |
| 22666 | 1706934 | 1947 | 1682321 | 0.11 | 98.56 | 1.33 |
| 25005 | 1434780 | 858 | 1408917 | 0.06 | 98.20 | 1.74 |
| 33196 | 2506848 | 34076 | 2439576 | 1.36 | 97.32 | 1.32 |
| 35601 | 2862338 | 39626 | 2787111 | 1.38 | 97.37 | 1.24 |
| 123857 | 6122893 | 117869 | 5881167 | 1.93 | 96.05 | 2.02 |

```
In [12]: from scipy import stats
import math
```

```
In [13]: from scipy.stats import binom
prob = 1 - binom.cdf(15,36,0.5)
print(str(round(prob*100, 1))+"%")

79.7%
```

```
In [14]: from scipy.stats import binom
print(binom.ppf(0.5,100,0.5)+1)
```

51.0

```
In [ ]:
```