CSE 544 Project, Spring 2021

Group members :

1. Nikhil Sira (113273807)
2. Mohit Khandelia (113786286)
3. Shreyash Hisariya (113216219)
4. Saurabh Verma (113222104)

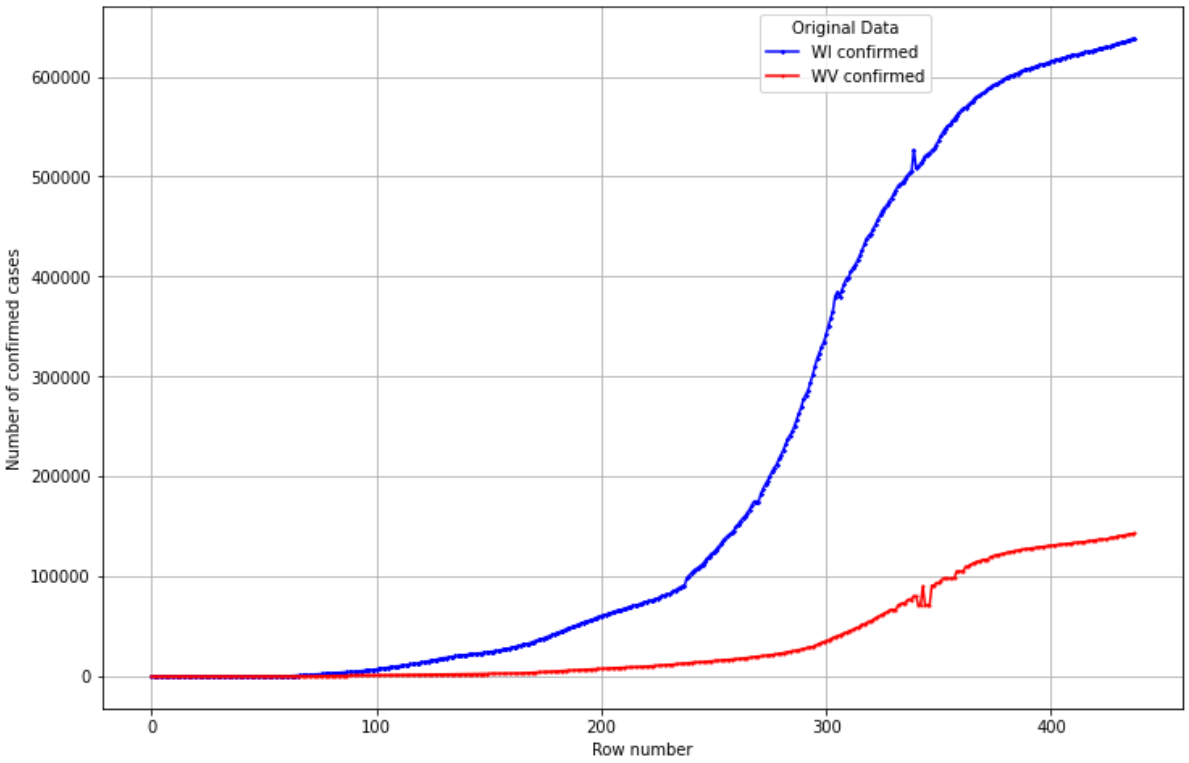
Question 1 :

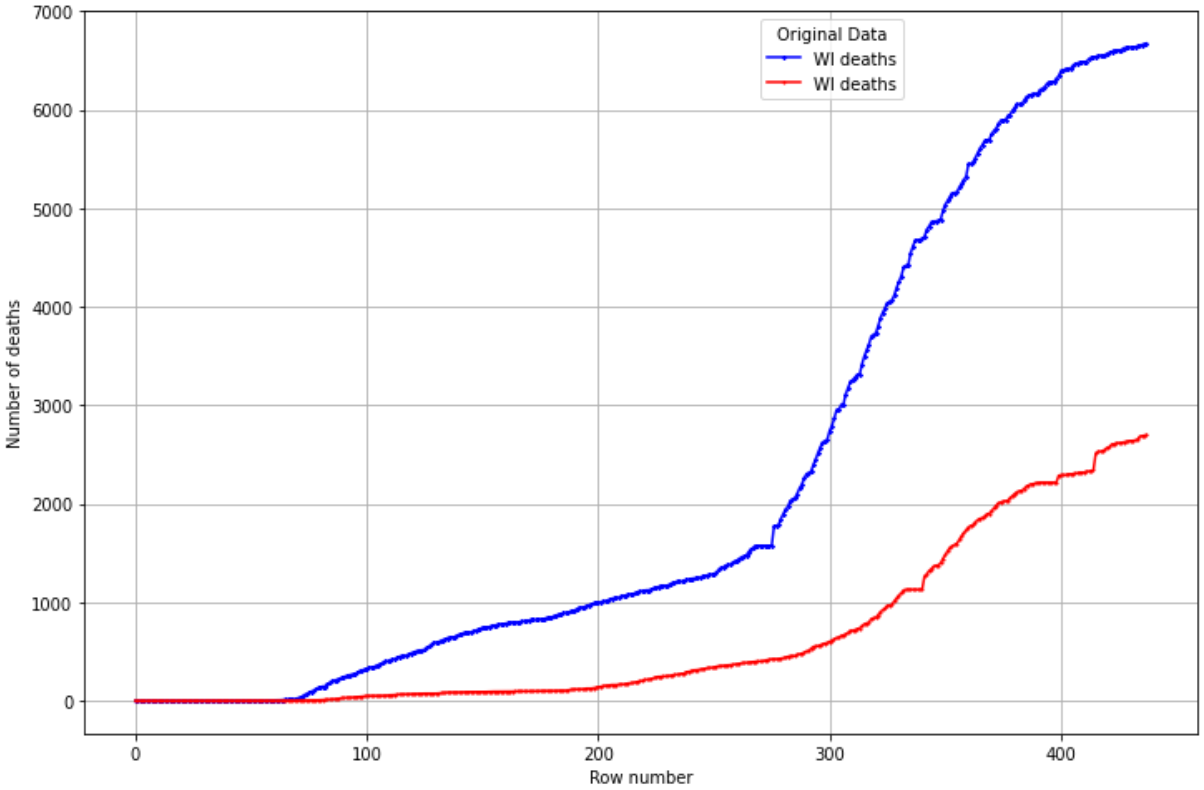
#################### Data Cleaning (Ans 1)####################

Steps Followed:

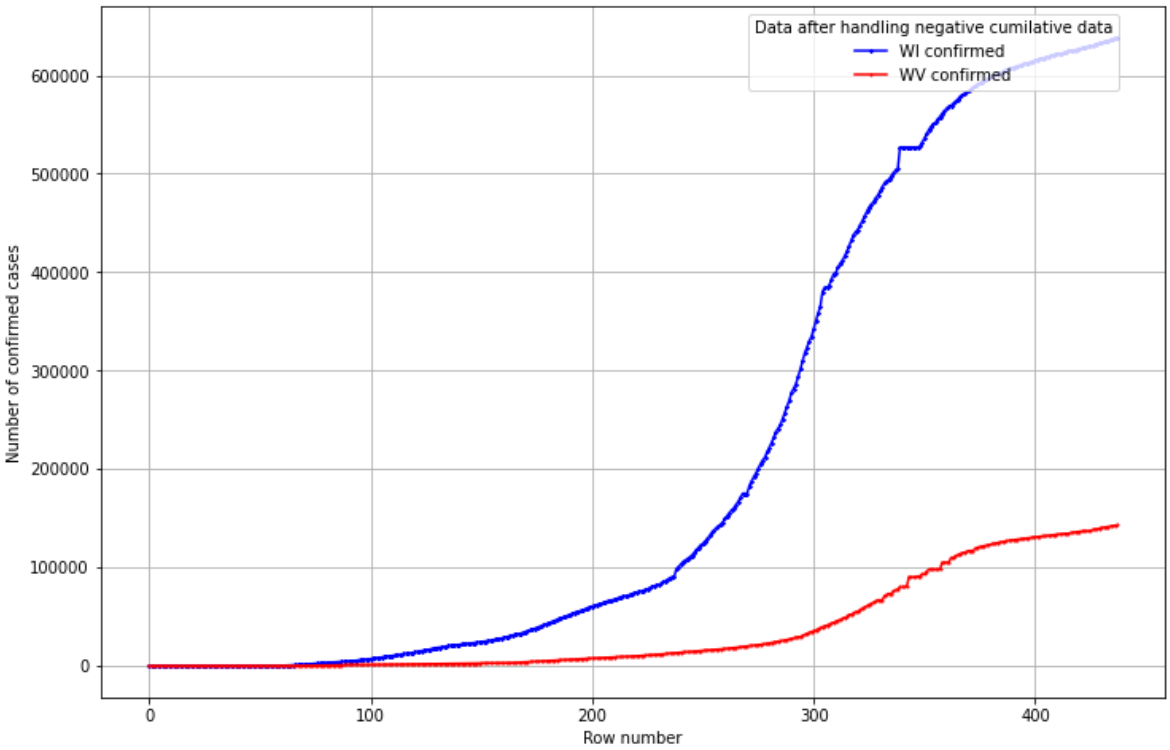
1. We first checked if there are any Null values in the entire dataset and we observed that the given data did not have any NULL values.
2. We plotted the original data and observed that there were instances when cumulative data had inappropriate values(i.e, the value at a particular place was smaller than the previous rows). For this, we modified the inappropriate values with the values present at the previous row in the dataset to make the cumulative data proper. We again plotted the data to make sure if the cumulative data was proper.
3. Now, we converted the cumulative data to instantaneous data(daily data) for all the states data. We also plotted the data to make sure if cumulative data is converted to daily data.
4. We then applied Tukey Rule for Outlier detection. We also made sure that the outliers did not correspond to the zero values. We got several outliers for each column, So we took the union of outliers indexes and then removed those rows from the original data.

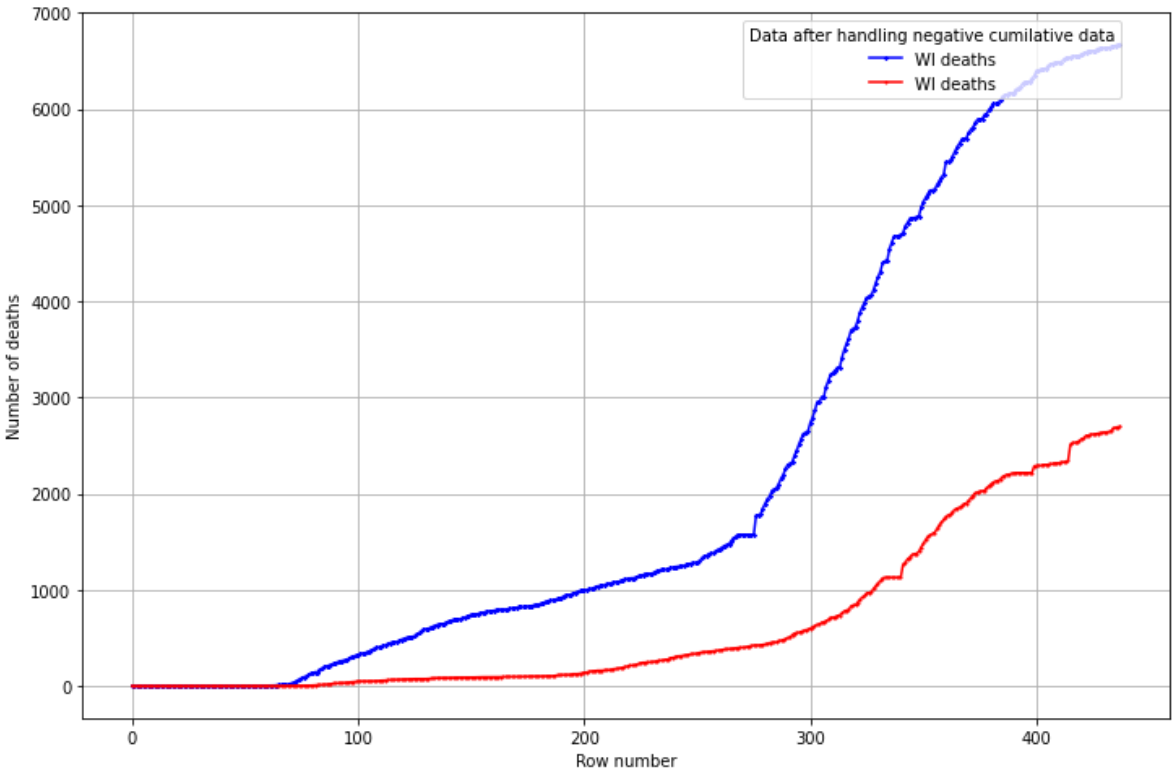
```
Plotting the original data
```



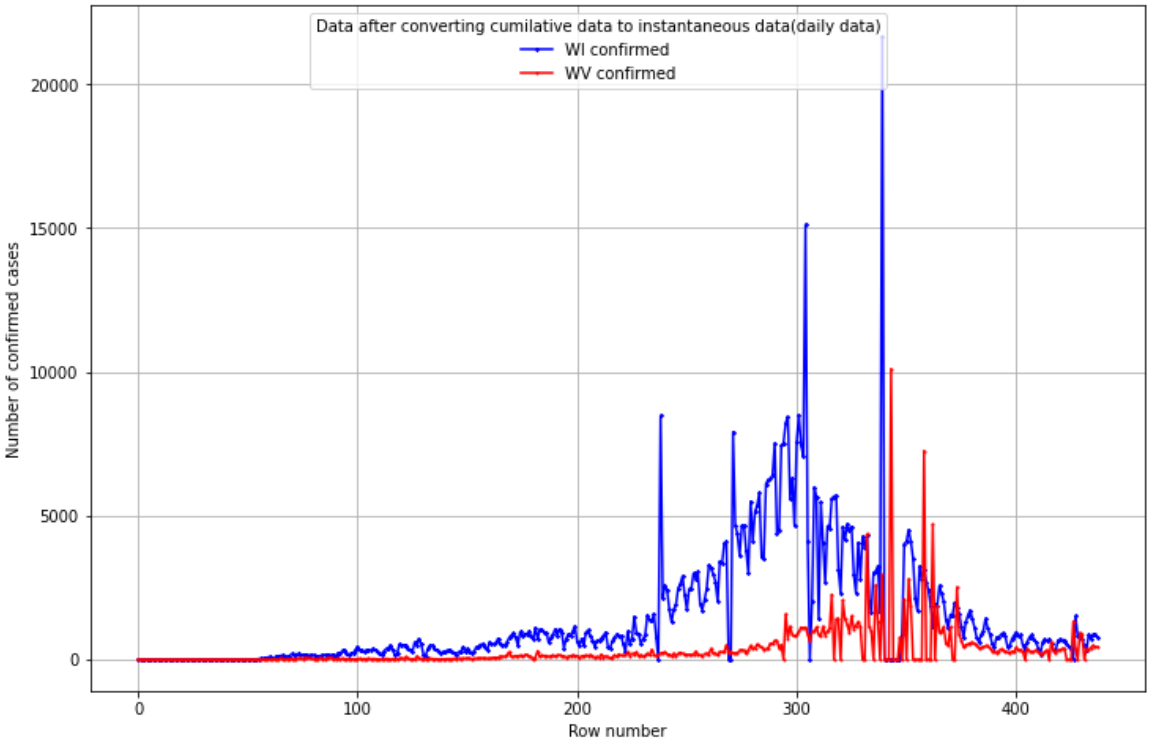


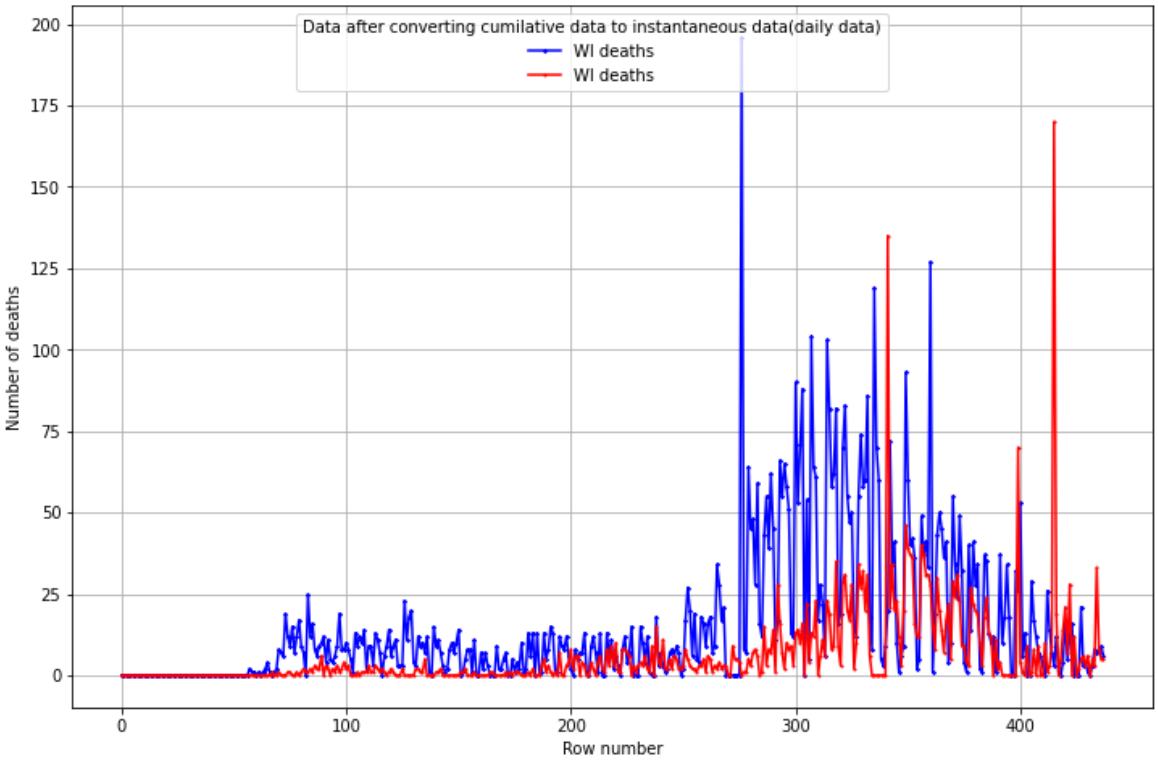
```
#   Plotting the data after handling negative cumulative data
```






```
#   Plotting the data after converting cumulative data to instantaneous
data(daily data)
```





Question 2 :

a. Report accuracy (MAPE and MSE) for AR(3), AR(5), EWMA(0.5), EWMA(0.8) for both states

############ Confirmed cases for State 1 ################

AR with p = 3 , MAPE : 27.77652693507591
AR with p = 3 , MSE : 34272.44066158218

AR with p = 5 , MAPE : 28.78680663100255
AR with p = 5 , MSE : 37719.65106856181

EWMA with alpha = 0.5 , MAPE : 35.58806906891773
EWMA with alpha = 0.5 , MSE : 51517.20867454628

EWMA with alpha = 0.8 , MAPE : 33.06123936620476
EWMA with alpha = 0.8 , MSE : 51580.59380203725

############## Deaths for State 1 ################

AR with p = 3 , MAPE : 74.85916207033081
AR with p = 3 , MSE : 39.478994322877135

AR with p = 5 , MAPE : 117.40398967597575
AR with p = 5 , MSE : 40.26156779683938

EWMA with alpha = 0.5 , MAPE : 111.6779025157779
EWMA with alpha = 0.5 , MSE : 50.40351908433928

EWMA with alpha = 0.8 , MAPE : 139.90680918817498
EWMA with alpha = 0.8 , MSE : 71.12060154357297

############## Confirmed cases for State 2 ################

AR with p = 3 , MAPE : 64.33515332948882
AR with p = 3 , MSE : 3166.705499738529

AR with p = 5 , MAPE : 76.79967506396868
AR with p = 5 , MSE : 4306.041969773134

EWMA with alpha = 0.5 , MAPE : 45.5586521405392
EWMA with alpha = 0.5 , MSE : 2510.0911378191518

EWMA with alpha = 0.8 , MAPE : 50.75134943611571
EWMA with alpha = 0.8 , MSE : 2760.2254898752535

############## Deaths for State 2 ################

AR with p = 3 , MAPE : 56.87493280063978
AR with p = 3 , MSE : 12.46122466750738

AR with p = 5 , MAPE : 71.96772356410746
AR with p = 5 , MSE : 14.194830924559804

EWMA with alpha = 0.5 , MAPE : 99.89427659246655
EWMA with alpha = 0.5 , MSE : 14.015098144823298

EWMA with alpha = 0.8 , MAPE : 116.59431482387666
EWMA with alpha = 0.8 , MSE : 18.8856437609379

Question 2.b

Summarized Observation:

One Sample Test for State WI confirmed for March '21

1. One Sample Wald's Test => We **reject** the true hypothesis, i.e Ho not equal to 762.1578947368421
2. One Sample Z Test => We **accept** the true hypothesis, i.e Ho is equal to 762.1578947368421
3. One Sample T Test => We **reject** the true hypothesis, i.e Ho not equal to 762.1578947368421

One Sample Test for State WI Deaths for March '21

1. One Sample Wald's Test => We **reject** the true hypothesis, i.e Ho not equal to 11.105263157894736
2. One Sample Z Test => We **reject** the true hypothesis, i.e Ho not equal to 11.105263157894736
3. One Sample T Test => We **accept** the true hypothesis, i.e Ho is equal to 11.105263157894736

One Sample Test for State WV confirmed for March '21

1. One Sample Wald's Test => We **reject** the true hypothesis, i.e Ho not equal to 340.3157894736842
2. One Sample Z Test => We **reject** the true hypothesis, i.e Ho not equal to 340.3157894736842
3. One Sample T Test => We **reject** the true hypothesis, i.e Ho not equal to 340.3157894736842

One Sample Test for State WV Deaths for March '21

1. One Sample Wald's Test => We **accept** the true hypothesis, i.e Ho is equal to 3.6315789473684212
2. One Sample Z Test => We **accept** the true hypothesis, i.e Ho is equal to 3.6315789473684212
3. One Sample T Test => We **accept** the true hypothesis, i.e Ho is equal to 3.6315789473684212

Two population Test for State WI Confirmed for Feb'21 and March '21

1. Walds=> We **reject** the true hypothesis, i.e mean of confirmed cases for feb'21 and march'21 is not equal for the state WI
2. Unpaired T-test => We **reject** the true hypothesis, i.e mean of confirmed cases for feb'21 and march'21 is not equal for the state WI

Two population Test for State WI Deaths for Feb'21 and March '21

1. Walds=> We **reject** the true hypothesis, i.e mean of deaths cases for feb'21 and march'21 is not equal for the state WI
2. Unpaired T-test => We **accept** the true hypothesis, i.e mean of deaths cases for feb'21 and march'21 is equal for the state WI

Two population Test for State WV Confirmed for Feb'21 and March '21

1. Walds=> We **reject** the true hypothesis, i.e mean of confirmed cases for feb'21 and march'21 is not equal for the state WV
2. Unpaired T-test => We **reject** the true hypothesis, i.e mean of confirmed cases for feb'21 and march'21 is not equal for the state WV

Two population Test for State WV Deaths for Feb'21 and March '21

1. Walds=> We **accept** the true hypothesis, i.e mean of deaths cases for feb'21 and march'21 is equal for the state WV
2. Unpaired T-test => We **accept** the true hypothesis, i.e mean of deaths cases for feb'21 and march'21 is equal for the state WV

Applicability Of Tests

Wald's Test

=> Assumption of Wald's test is to have an asymptotic normal estimator but since the number of data points are very small, we can't even apply CLT. Therefore, the test does not seem to be applicable. Moreover, incase of two population tests, the same rule applies, since we can't apply CLT, thus the test seems to be not applicable.

Z-test

=> Assumption is to know the true standard deviation which is not the case in our problem. Moreover, since N is very small and also data is not normally distributed, we can't ideally apply Z test.

T-test

=>In one sample, the assumption is that the data should be normally distributed, which is not the case in our problem. Thus, we ideally should not apply a T-test. In unpaired T-test, assumption is that the data should be independent and need to be normally distributed, which again is not the case in our problem. Thus, the test is not applicable.

Question 2.c

Summarized Observation:

One Sample K-S Test for WI_confirmed and WV_confirmed

----------Poisson----------

Ho: WI_confirmed distribution is equivalent to WV_confirmed distribution incase when the given distribution is Poisson

H1: WI_confirmed distribution is not equivalent to WV_confirmed distribution incase when the given distribution is Poisson

maximum difference val 1.0

=>**Rejecting** the Hypothesis,i.e WI_confirmed distribution is not equivalent to WV_confirmed distribution incase when the given distribution is Poisson

----------Geometric----------

Ho: WI_confirmed distribution is equivalent to WV_confirmed distribution incase when the given distribution is Geometric

H1: WI_confirmed distribution is not equivalent to WV_confirmed distribution incase when the given distribution is Geometric

maximum difference val 0.7824226825633314

=>**Rejecting** the Hypothesis,i.e WI_confirmed distribution is not equivalent to WV_confirmed distribution incase when the given distribution is Geometric

----------Binomial----------

Ho: WI_confirmed distribution is equivalent to WV_confirmed distribution incase when the given distribution is Binomial

H1: WI_confirmed distribution is not equivalent to WV_confirmed distribution incase when the given distribution is Binomial

maximum difference val 1.0

=>**Rejecting** the Hypothesis,i.e WI_confirmed distribution is not equivalent to WV_confirmed distribution incase when the given distribution is Binomial

One Sample K-S Test for WI_deaths and WV_deaths

----------Poisson----------

Ho: WI_deaths distribution is equivalent to WV_deaths distribution incase when the given distribution is Poisson

H1: WI_deaths distribution is not equivalent to WV_deaths distribution incase when the given distribution is Poisson

maximum difference val 0.957815730505644

=>Rejecting the Hypothesis,i.e WI_deaths distribution is not equivalent to WV_deaths distribution incase when the given distribution is Poisson

----------Geometric----------

Ho: WI_deaths distribution is equivalent to WV_deaths distribution incase when the given distribution is Geometric

H1: WI_deaths distribution is not equivalent to WV_deaths distribution incase when the given distribution is Geometric

maximum difference val 0.6137865128870058

=>**Rejecting** the Hypothesis,i.e WI_deaths distribution is not equivalent to WV_deaths distribution incase when the given distribution is Geometric

----------Binomial----------

Ho: WI_deaths distribution is equivalent to WV_deaths distribution incase when the given distribution is Binomial

H1: WI_deaths distribution is not equivalent to WV_deaths distribution incase when the given distribution is Binomial

maximum difference val 1.0

=>**Rejecting** the Hypothesis,i.e WI_deaths distribution is not equivalent to WV_deaths distribution incase when the given distribution is Binomial

Two Sample K-S Test

----------For WI_deaths and WV_deaths----------

Ho: WI_deaths distribution is equivalent to WV_deaths distribution

H1: WI_deaths distribution is not equivalent to WV_deaths distribution

maximum difference val 0.7037037037037037

=> Rejecting the Hypothesis,i.e WI_deaths distribution is not equivalent to WV_deaths distribution

----------For WI_confirmed and WV_confirmed----------

Ho: WI_confirmed distribution is equivalent to WV_confirmed distribution

H1: WI_confirmed distribution is not equivalent to WV_confirmed distribution

maximum difference val 0.8518518518518519

=>**Rejecting** the Hypothesis,i.e WI_confirmed distribution is not equivalent to WV_confirmed distribution

Permutation Test

----------For WI_deaths and WV_deaths----------

Ho: WI_deaths distribution is equivalent to WV_deaths distribution

H1: WI_deaths distribution is not equivalent to WV_deaths distribution

Tobs: 9.777777777777779

Pval is : 0.0

=> **Rejecting** the Hypothesis,i.e WI_deaths distribution is not equivalent to WV_deaths distribution

----------For WI_confirmed and WV_confirmed----------

Ho: WI_confirmed distribution is equivalent to WV_confirmed distribution

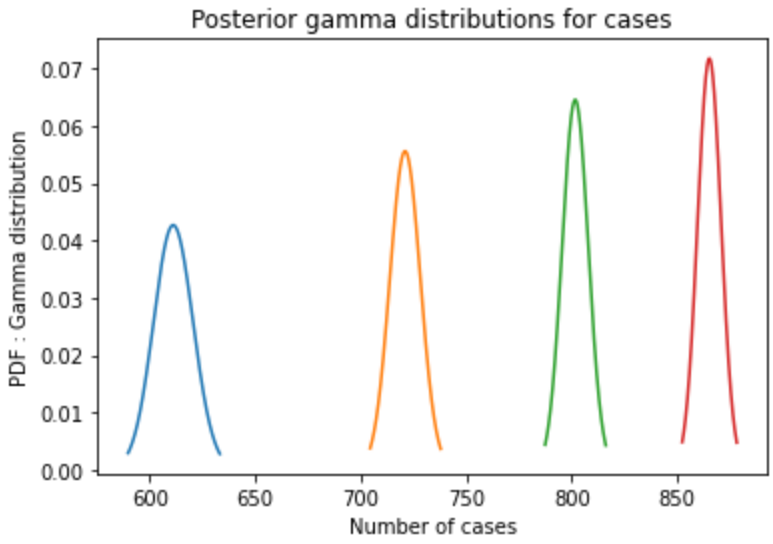
H1: WI_confirmed distribution is not equivalent to WV_confirmed distribution

Tobs: 2262.0

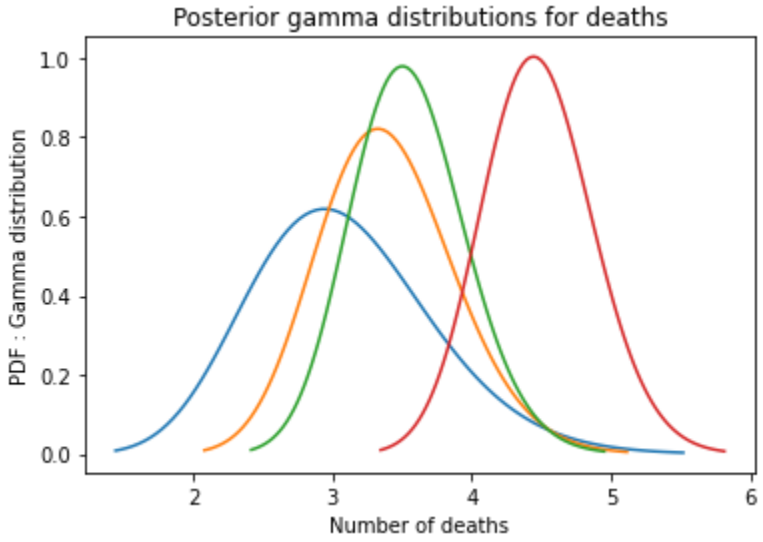
Pval is : 0.0

=>Rejecting the Hypothesis,i.e WI_confirmed distribution is not equivalent to WV_confirmed distribution.

Question 2.d : Plot all the posterior distributions on one graph. Report MAP for all distributions.




```
Confirmed cases week : 1 , MAP : 611.0282338198889
Confirmed cases week : 2 , MAP : 720.9222144870828
Confirmed cases week : 3 , MAP : 801.5496825799343
Confirmed cases week : 4 , MAP : 865.1267993553181
```




```
Deaths week : 1 , MAP : 2.923680827509256
Deaths week : 2 , MAP : 3.335989317260387
Deaths week : 3 , MAP : 3.490901084514456
Deaths week : 4 , MAP : 4.442268958336296
```

Exploratory tasks :

2.A

Hypothesis 1:

Pearson's correlation Confirmed Cases-
H0 : NASDAQ Price and confirmed covid cases is not linearly dependent
H1 : NASDAQ Price and confirmed covid cases is linearly dependent

Sigma hat XY for confirmed Covid cases and NASDAQ price is 0.6563244100547485 which is greater than 0.5. Therefore we accept H0. This implies a positive correlation
Sigma hat XY for Covid deaths and NASDAQ price is 0.5505505322024021 which is also greater than 0.5. So we accept H0. This also implies a positive correlation

Hypothesis 2:

Chi Square Test on confirmed cases

H0 : NASDAQ Price and confirmed covid cases is independent
H1 : NASDAQ Price and confirmed covid cases is dependent
Q Observed value calculated from Chi Square Test is: 3437631.7545028636
Degree of freedom for the data is 302
Calculated p value using q observed and degree of freedom is 0
This means that NASDAQ price and covid cases are dependent by rejecting the independent hypothesis H0

Hypothesis 3:

KS-Test

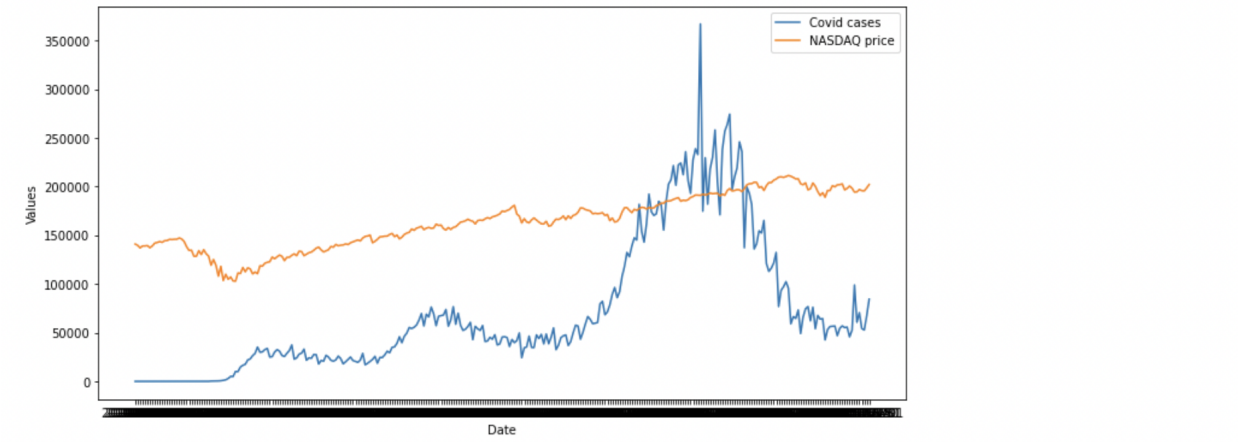
H0: Distribution of NASDAQ Price and confirmed covid cases is same
H1:Distribution of NASDAQ Price and confirmed covid cases is different

Maximum value = 0.6589403973509957 > C: 0.05, So we reject H0.
The distribution of NASDAQ price and covid cases is different.

2.B.

```
plt.figure(figsize=(12, 6))
plt.plot(df1['Date'], df1['sum'])      # blue->covid cases
plt.plot(df1['Date'], df1['Close']*15)   # orange->NASDAQ price
plt.xlabel('Date')
plt.ylabel('Values')

plt.legend(["Covid cases","NASDAQ price"], loc ="upper right")
plt.show()
```



The COVID-19 pandemic caused a severe impact on global financial markets. Initially when Covid cases started to rise in the USA, stock prices started to decrease. We can see a sharp dip in NASDAQ price when Covid started to affect the USA in late February.

By the end of March, the country started to respond. Stricter rules were imposed in terms of social distancing and the manufacturing and industrial sector were kept up and running. This resulted in normalcy in terms of the general rise in stock prices. We can see in the data, the NASDAQ prices started to rise in late March.

From then on, there has been a stable rise in NASDAQ price, even with the rise in covid cases. The country adapted to this situation to ensure a steadily growing economy.

==========