IC252:Data Science II

Lab 8

Submitted by ADARSH SANTORIA(B21176)

1). Given the “cases.csv” dataset, plot the time graph of the Infected Fraction of population (Infected Fraction vs Time in Months). Do this for Delhi, Mumbai and Kolkata.

a) Compare these graphs. What do you infer from these graphs?

b) Calculate the variance of the Infected Fraction.

Hint:-

$$\text{Infected Fraction}=\frac{\text{Confirmed}-\text{Recovered}-\text{Deceased})}{\text{Population}}$$

$$\text{Susceptible}=\frac{(\text{Population}-\text{Confirmed})}{\text{Population}}$$

$$\text{Removed}=\frac{(\text{Recovered}+\text{Deceased})}{\text{Population}}$$

Extra work:- Plot the time graph of Susceptible and Removed population and compare with the Infected Fraction of the population. Plot all these in a single plane (graph). Do this for Delhi and Mumbai.

2). Given the “2021_IN_Region_Mobility_Report.csv” dataset for 2021, plot the following

a) Retail mobility of Delhi and Mumbai. Compare them in the same plane.

b) Transit mobility of Delhi and Mumbai. Compare them in the same plane.

c) What do you infer from these graphs?

d) Calculate the IQR in each case (Interquartile range).

e) What is the expected value of Retail and Transit mobility in Delhi and Mumbai?

Hint:-

In given dataset, “retail_and_recreation_percent_change_from_baseline” column represents Retail mobility and “transit_stations_percent_change_from_baseline” column represents Transit mobility.
Note:-

In descriptive statistics, the interquartile range (IQR) is a measure of statistical dispersion. It is the spread of the data or observations. The IQR may also be called the mid spread, middle 50%, or H spread. It is defined as the spread difference between the 75th and 25th percentiles of the data. The lower quartile corresponds with the 25th percentile and the upper quartile corresponds with the 75th percentile. So, IQR=Q 3−Q 1. First, take the median of the data. Then Q3 (median of the lower half of the data) – Q1 (median of the upper half of the data).

3). Given the “Cowin_Vaccine_Data_Districtwise.csv” dataset, do the following.

a) Plot the vaccination coverage of Delhi and Mumbai. (Basically, for each city you have to plot % of people vaccinated with first dose and % of people vaccinated with second dose in the same plane)

b) Calculate the correlation of first dose coverage with the following:

(i). $\frac{\text{Sites}}{\text{Area of city}}$

(ii). $\frac{\text{Sessions}}{\text{Area of city}}$

What can you infer from these correlations? c) Find the state/ UT with the highest vaccination coverage (first dose).

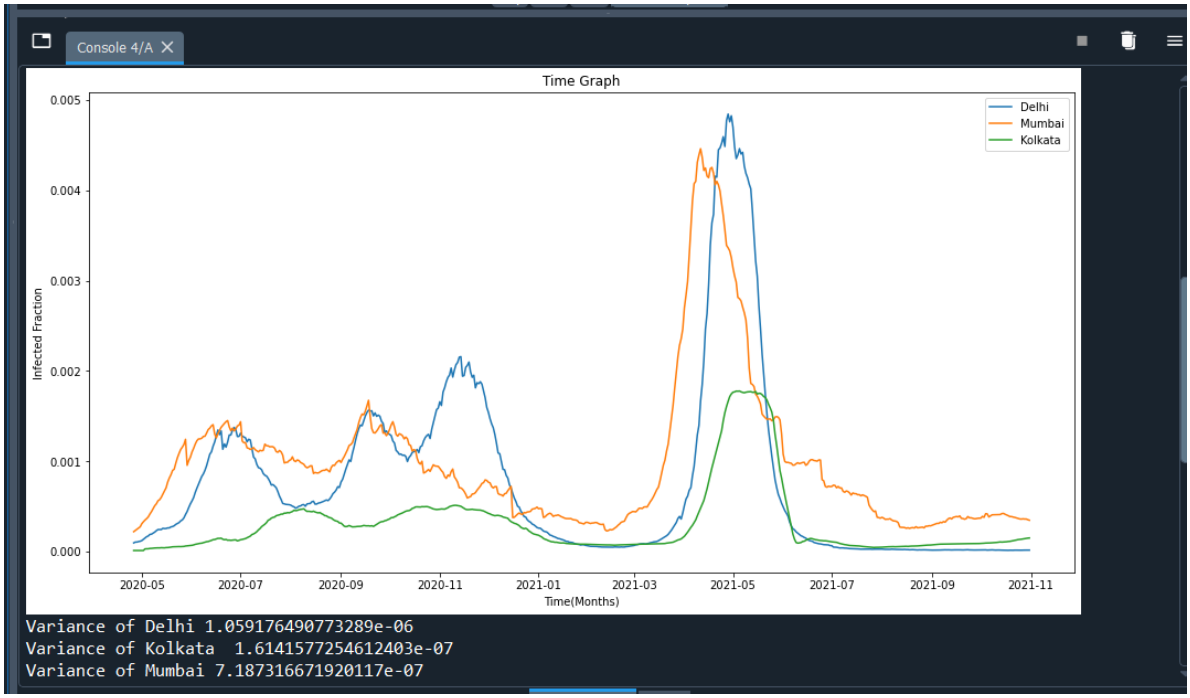
Note:-

• Use the following values of population and area of the cities. (Population of Delhi: 20,591,874, Population of Mumbai: 20,667,656, Population of Kolkata: 14,850,000, Area of Delhi : 1400 sq. km , Area of Mumbai: 670 sq.km, Area of Kolkata: 206 sq.km)

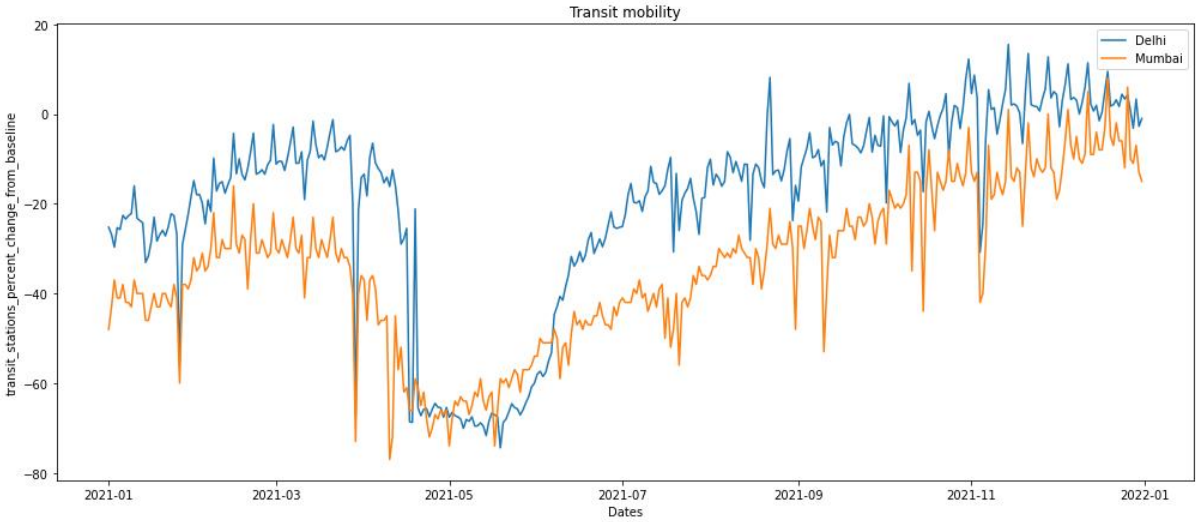
• Empty (None) value should be ignored for any dataset.

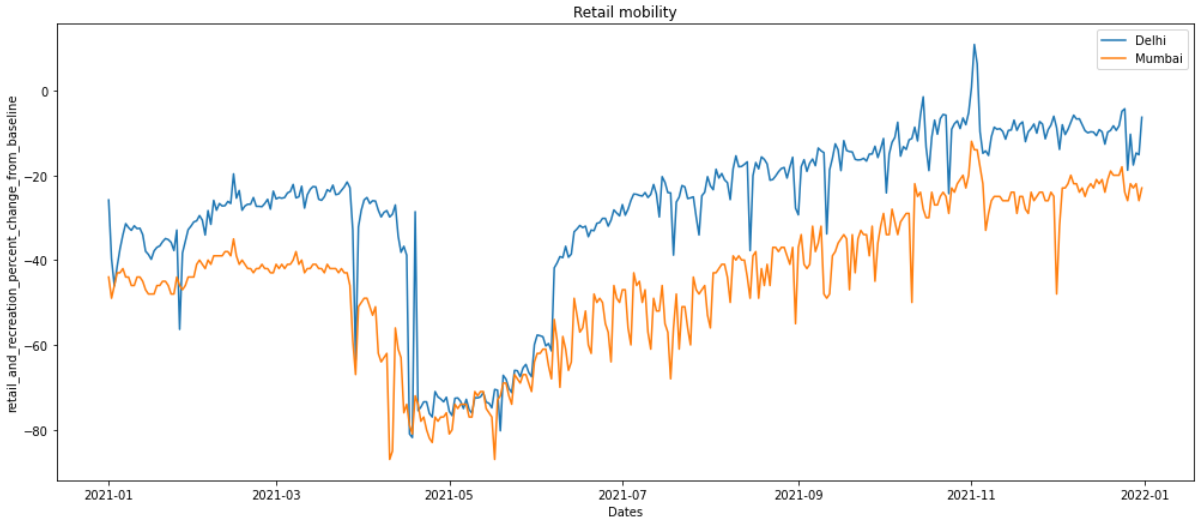
Codes and Graphs are below:-

Q1



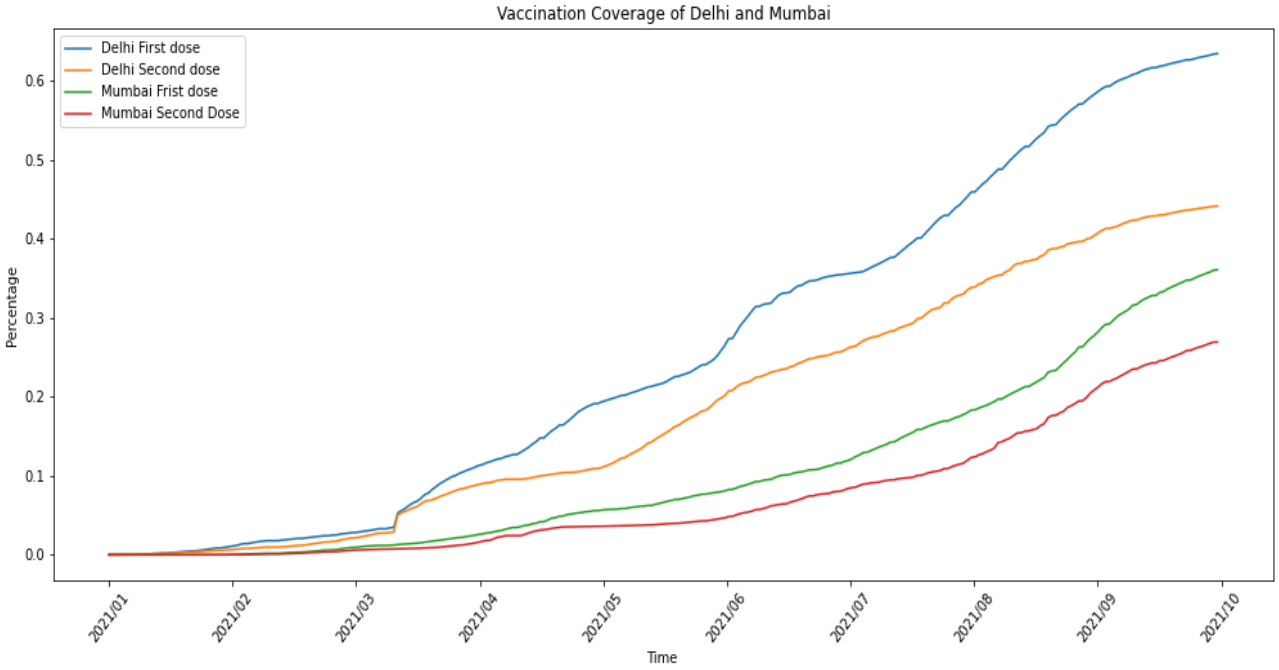
Q2






```
We can infer that the graph of Delhi is more diverse and values are generally greater than that of Mumbai
IQR of Retail mobility of Delhi is 18.374999999999996
IQR of Retail mobility of Mumbai is 20.5
IQR of Transit mobility of Delhi is 18.374999999999996
IQR of Transit mobility of Mumbai is 20.5
Expected Retail mobility of Delhi is -28.310502283105023
Expected Retail mobility of Mumbai is -44.73972602739726
Expected Transit mobility of Delhi is -19.122602739726027
Expected Transit mobility of Mumbai is -34.16986301369863
```

Q3




```
Correlation between First Dose Coverege with sessions in Delhi is  0.7173482549937008
Correlation between First Dose coverage and sites in Delhi is :  0.6949973140036592
Correlation between First Dose Coverege with sessions in Mumbai is  0.8090501065612462
Correlation between First Dose coverage and sites in Mumbai is :  0.5189895731310716
State with highest vaccination Coverege is  Maharashtra
```