

## IC252: Lab Assignment - 8

- 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 midspread, middle 50%, or Hspread. It is defined as the spread difference between the 75<sup>th</sup> and 25<sup>th</sup> percentiles of the data. The lower quartile corresponds with the 25<sup>th</sup> percentile and the upper quartile corresponds with the 75th percentile. So,  $IQR = Q3 - Q1$ . **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.