**Introduction**

The COVID-19 pandemic has created a global health crisis, with profound social and economic implications. As the number of confirmed cases and deaths continues to rise, it is critical to develop effective strategies for mitigating the spread of the disease and managing its impact on healthcare systems and society as a whole. In this context, data analysis has emerged as a powerful tool for understanding the epidemiology of COVID-19, predicting its spread, and identifying effective interventions.

The purpose of this coursework is to practice modern methods of data analysis to explore the COVID-19 India dataset and gain insights into the patterns and trends of the disease. Specifically, we will use machine learning techniques to predict the number of active, passive, confirmed cases and deaths in different states in India with the help of number of doses given to them.

To accomplish this goal, we will use a publicly available dataset containing information on active, paasive,confirmed cases, deaths, and the different doses present as dose1, dose2, dose3, precautionary dose for different states. We will preprocess and clean the data to prepare it for analysis, and then apply a range of machine learning algorithms to develop predictive models and explore the underlying patterns and relationships.

Overall, this coursework aims to contribute to the ongoing efforts to understand and combat the COVID-19 pandemic through the application of modern data analysis methods and techniques.

**Dataset Description**

This dataset provides a thorough compendium of information about the COVID-19 outbreak in India. It records the quantity of confirmed cases and fatalities in each Indian state and union territory from the time of the initial case report in January 2020 until the present. To understand the scope of the COVID-19 epidemic in India, the dataset was developed. It is significant because it gives researchers, decision-makers, and citizens insight into the numerous elements that might be causing the virus to spread in various Indian states and areas. Additionally, it offers important data for academics attempting to comprehend the dynamics of the pandemic in India.

This dataset is significant because it enables us to comprehend the pandemic's current stage in India and to track the virus's development in each state. It can also be used to gauge how well the efforts taken by the Indian government to stop the virus' spread are working. Anyone interested in learning more about the dynamics of the COVID-19 pandemic in India, including policymakers, researchers, individuals, NGOs, and the media, can use the dataset. It can be used to gather information about the current state of affairs and to monitor the virus's development in each state. It can also be used to assess how well the Indian government's policies are working to stop the virus's spread.

This dataset offers a complete picture of the COVID-19 pandemic in India overall. It offers vital information that is helpful to scholars, decision-makers, and people and is updated daily. It is a priceless tool that can be used to comprehend the dynamics of the infection and track its development in each state.

**Target features:**

Columns:

0. Name of the state (not a feature)

1. Total confirmed cases

2.Total active cases

3.Total passive cases

4. Total deaths

5. Total dose 1 administered

6. Total dose 2 administered

7. Total dose 3 administered

8. Total precaution dose administered

9. Total Doses administered

**Descriptive feature:**

**State** : contains name of states in India

**Confirmed**: consist of total number of confirmed case in the particular state

**Active**: consist of total number of active case in the particular state

**Passive**: consist of total number of passive case in the particular state

**Deaths**: total number of deaths in the particular state

**Dose1**: total number of dose1 give in a particular state

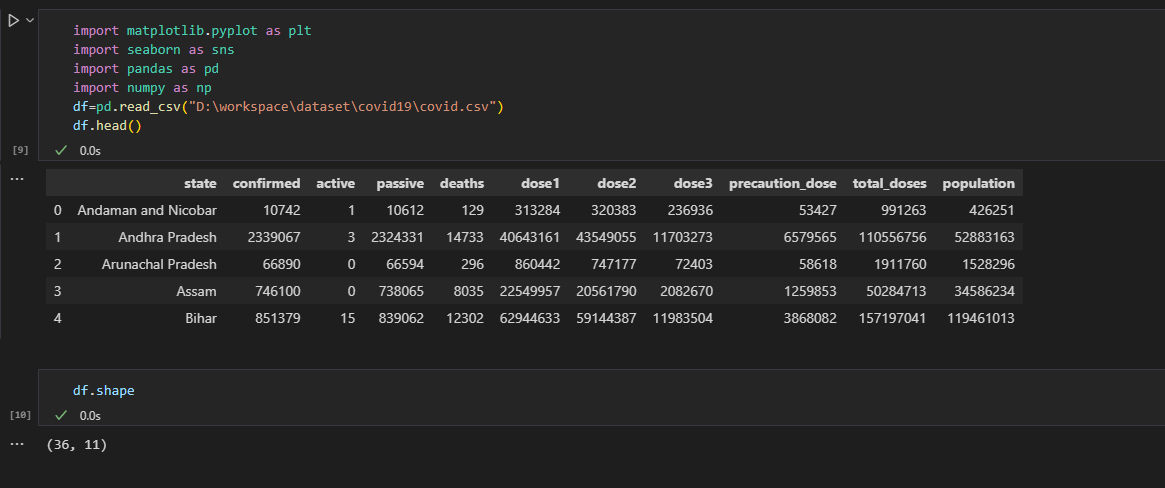
**Dose2**: total number of dose2 give in a particular state

**Dose3**: total number of dose3 give in a particular state

**Precaution\_dose**: total number of precaution dose give in a particular state

**Total\_doses**: total number of doses given in a particular state

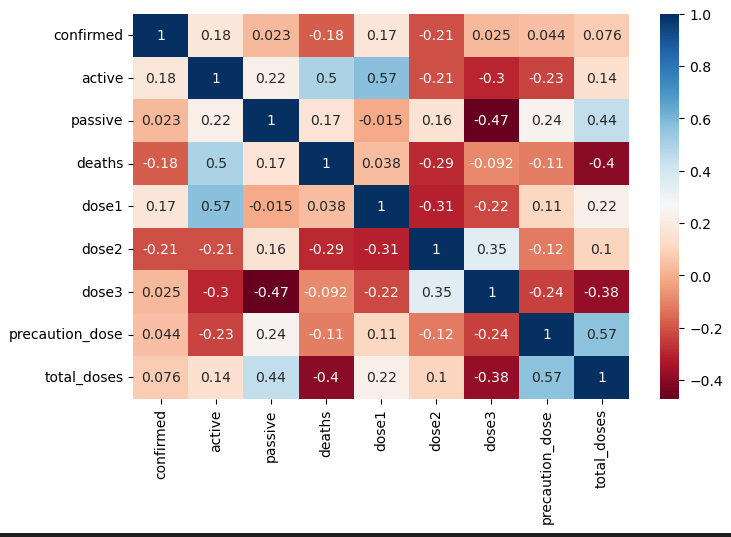
**Population**: total number of population present in a particular state

****

**Correlation**

We have utilized a variety of graphs and plots to discover the relationship between the features because it is crucial to comprehend the correlation between distinct features before performing the linear regression.

**Heatmap of correlation between features**

****

**Analysis through the heatmap is**

1. The number of confirmed cases has positive relation with dose1 , this describes that as the dose 1 is given the confirmed case was present while confirmed case has negative relation with dose2 this means that as the dose increases the confirmed cases decreases.

2. The number of active cases has positive relation with dose1 , this describes that as the dose 1 is given the active case was present while active case has negative relation with dose2 which means after dose 2 the active cases reduces .As we move further to dose 3 and precaution dose the active cases further reduces this means that as the dose increases the active cases decreases which implies the disease reduces.

3. The number of passive cases has negative relation with dose1 , this describes that as the dose 1 is given the passive case was decreases while passive case has positive relation with dose2 this means that as the dose increases the passive cases increases but as the dose 3 is given the passive cases reduces with a larger amount .

4.The number of deaths has positive relation with dose1 , this describes that as the dose 1 is given the number of deaths is also changes while number of death has negative relation with dose3 this means that as the dose increases the number of death reduces as well as the dose 3 is given the number of deaths futher reduces.

With the above analysis we can say that as the number of doses increases the number of deaths reduces.

**Linear Regression**