COVID-19 is a pandemic of respiratory illness caused by SARS-CoV-2 virus, which was discovered in 2019. COVID-19 is severe and has caused millions of deaths around the world. The spread of COVID-19 occurs via air borne particles and droplets. When people with COVID exhale, they can discharge particles and droplets of respiratory fluids into the air. It is spread from person to person and is diagnosed with a laboratory test. The most common symptoms of COVID-19 are fever, dry cough, fatigue. To prevent the spread of COVID-19, we should use simple precautions such as wearing a mask, physical distancing, avoiding crowds, getting vaccinated, monitoring our health daily. In this project, we represent the pattern of COVID-19 pandemic which can be utilized to develop preventive measures in order to prevent or handle future pandemics. We develop a model to predict the COVID-19 pandemic's death rate and impact of vaccination on the COVID-19 death rate.

As of 31 December 2020, COVID-19 had infected over 82 million people and killed more than 1.8 million worldwide. But preliminary estimates suggest the total number of global "excess deaths" directly and indirectly attributable to COVID-19 in 2020 amount to at least 3 million, 1.2 million higher than the official figures reported by countries to WHO[5].

The term "excess deaths" describes deaths beyond what would have been expected under "normal" conditions. It captures not only confirmed deaths, but also COVID-19 deaths that were not correctly diagnosed and reported as well as deaths attributable to the overall crisis conditions. This provides a more comprehensive and accurate measure when compared with confirmed COVID-19 deaths alone.

For example, some countries only report COVID-19 deaths occurring in hospitals or the deaths of people who have tested positive for COVID-19. In addition, many countries cannot accurately measure or report cause of death due to inadequate or under-resourced health information systems.

The spread of SARS-CoV-2, the causative agent of COVID-19, has resulted in an unprecedented global public health and economic crisis. The outbreak was declared a pandemic by the World Health Organization on 11 March 2020, and development of COVID-19 vaccines has been a major undertaking in fighting the disease. As of December 2020, many candidate vaccines have been shown to be safe and effective at generating an immune response, with interim analysis of phase III trials suggesting efficacies as high as 95%7. At least two vaccine candidates have been authorized for emergency use in the USA, the UK, the European Union and elsewhere, with more candidates expected to follow soon. For these COVID-19 vaccines to be successful, they need to be not only be proven safe and efficacious, but also widely accepted.

It is estimated that a novel COVID-19 vaccine will need to be accepted by at least 55% of the population to provide herd immunity, with estimates reaching as high as 85% depending on country and infection rate. Reaching these required vaccination levels should not be assumed given well-documented evidence of vaccine hesitancy across the world, which is often fueled by online and offline misinformation surrounding the importance, safety or effectiveness of vaccines. There has been widely circulating false information about the pandemic on social media platforms, such as that 5G mobile networks are linked with the virus, that vaccine trial participants have died after taking a

candidate COVID-19 vaccine, and that the pandemic is a conspiracy or a bioweapon. Such information can build on pre-existing fears, seeding doubt and cynicism over new vaccines, and threatens to limit public uptake of COVID-19 vaccines.

A precise prediction considering all the factors leading to COVID and its correlation with vaccine pertinence is the foundation for building an arsenal to fight against any future pandemic.

The COVID data was gathered from CDC (Centers for Disease Control and Prevention), where they have a separate section for COVID-19 death data and resources. Dataset includes information on COVID-19 deaths, state, age, place of death and also includes information on vaccinations.

Initially the dataset is pre-processed by creating a proper pipeline which involves data cleaning and feature engineering.

The main objective is to build a predictive model that'll help analyze future trend in COVID-19 deaths by using a supervised model. The supervised method would be time series forecasting. The suitable algorithm will be chosen to work on the dataset and finetuned to achieve better results.

Data has been collected from Centre of Disease Control and Prevention (CDC). In the initial Exploratory Data Analysis, multiple null values were found, especially in the Month column which will be eliminated (Refer Image 1).

Important predictors like age group, gender, ethnicity with respect to the overall COVID-19 cases were determined using feature selection.

In the Graph 1, we have plotted a graph for the number of deaths based with respect to the age group. Marking with respect to each State in the United States. From this, we can analyze that the most affected states are Vermont, Washington, Texas, Utah. The results are still not very accurate as there are more Data Preprocessing to be done, like removal of outliers, removal of null values etc.,

In order to predict the future trend of the COVID-19 cases, a supervised learning model will be created, which will be a time series forecasting. Analysis based on COVID-19 cases affected gender, race, prior health conditions will also be done.