**Project Name:** Covid-19 Death Rate Analysis

**Github Link:** https://github.com/projectsforstudents2022/Covid19\_Death\_Rate\_Analysis.git

**Why was this project created?**

According to numerous reports, Wuhan (in the Chinese province of Hebei) was the site of the first COVID-19 detection in December 2019. The virus was in South China for about a month before spreading to the rest of the world via Europe. Every day, a huge number of people around the world contract diseases and pass away. Additionally, information about cases of infection, death, and recovery is made available on numerous websites, including Google and web pages. But only national and international levels are covered by this data.

**What problem is it solving?**

In light of the global COVID -19 outbreak, the WHO is regularly disseminating information via websites, newspapers, Facebook, and other media. But the degree of analysis for these data is merely national. These websites highlight infected, recovered, and fatal cases on a countrywide level. However, it doesn't offer systematically collected global statistics. Therefore, these sites do not provide accurate analysis of global data through pictorial depiction. As a result, it is important to highlight national and local-level data and perform efficient analysis and review.

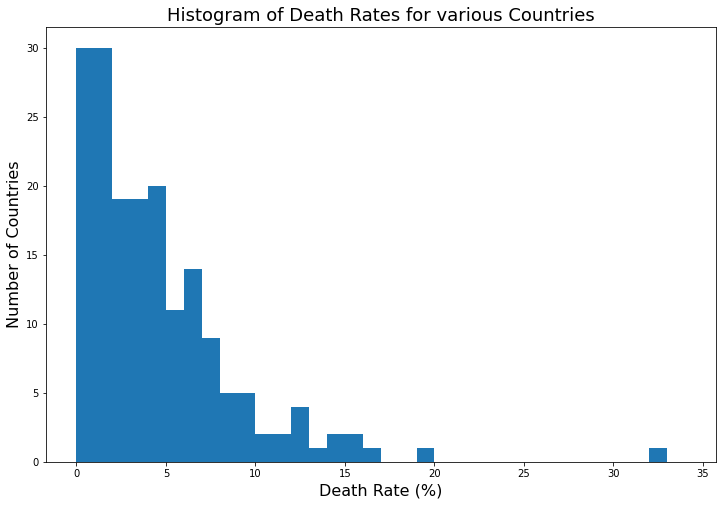
**Entire explanation of project**

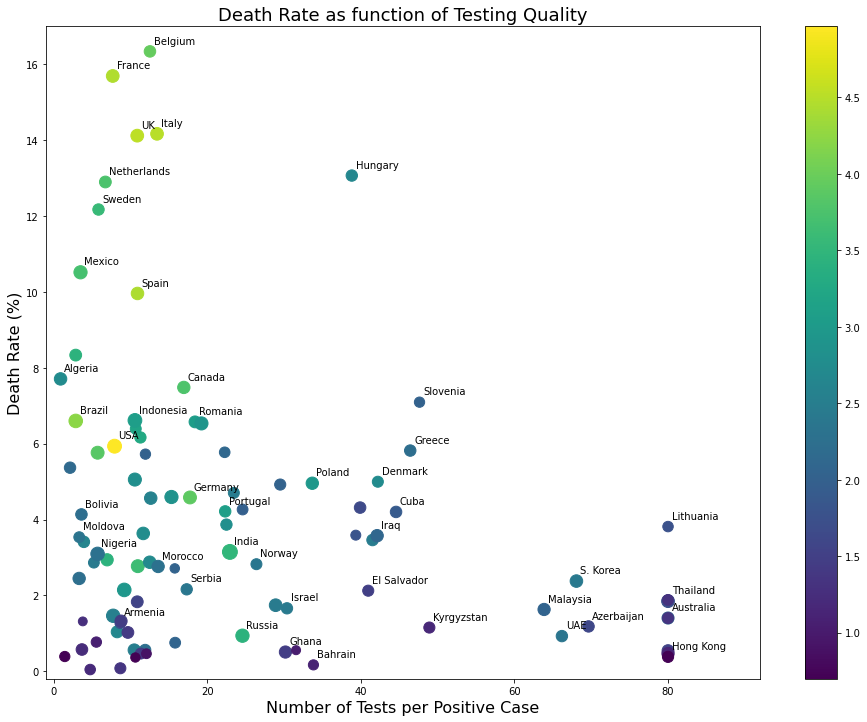
* **PROPOSED APPROACH**

The CDC's case surveillance data were used as the data source for this investigation . This dataset includes 32 features and observations of over 22 million COVID-19 occurrences in the United States. The observations span the time period from 1 January 2020 to 30 March 2021. We mainly rely on Breiman's random forest (RF) classifier in this project's strategy for forecasting COVID-19 patient death. Therefore, a succinct explanation of this strategy is necessary. Numerous decision trees make up the RF classifier itself. By successively dividing our data at decision nodes according to feature values, a decision tree classifier is created. According to a cutoff value for one of the data attributes, our initial decision node divides the data into two groups. Then another decision node divides these groups once more, and so on, creating the "branches" of the decision tree. The decision tree's final groups, or "leaves," serve as a marker for the class to which each member of that group belongs when splitting ceases. Typically, the feature utilized at each decision node to divide the data is selected to reduce error at that stage. To make its classifications, the RF classifier employs an ensemble "forest" of these decision trees.

A bootstrapped random sample of the available data is used to construct each tree in the RF ensemble, and only a random selection of the available features is taken into account when splitting nodes are created. Each decision tree in the ensemble "votes" for the class it predicts, and the class that the RF classifier predicts is determined by the majority vote of the ensemble's decision trees.

* **DATA VISUALIZATION**





* **CONCLUSION**

In this project, we gathered information on COVID-19 infections across the globe, both qualitatively and quantitatively. The qualitative data assess the socioeconomic circumstances, health facilities, infected patients, and individual perceptions toward this pandemic. The quantitative data analyzed the COVID cases in several aspects such as gender, age, etc.