

|  |
| --- |
| Covid-19 Future Prediction |
| Project Report |
| January 22  Members:  Ali Raza (SP19-BSCS-0100)  Anaaf Jawed (FA20-BSCS-2001)  Fouzan Hassan (SP19-BSCS-0125) |

# Abstract

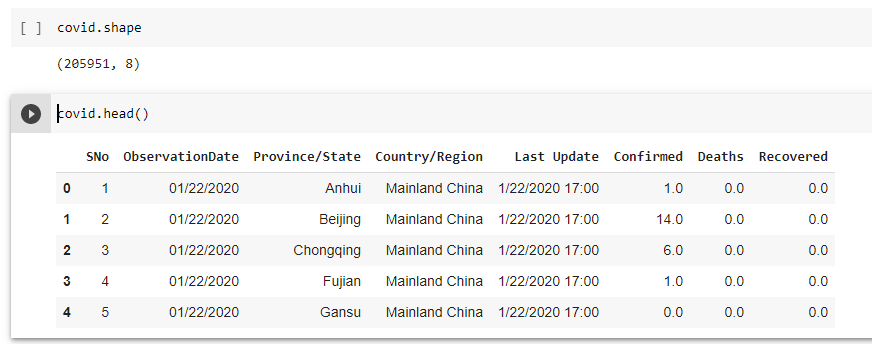
|  |
| --- |
| This Project is about covid-19 deaths prediction in which we will give some features to the model for training and then the model will predict the total No. of deaths.  We used the supervised model for this project. When training a Supervised learning algorithm, the training data will consist of inputs paired with the correct outputs. During training, the algorithm will search for patterns in the data that correlate with the desired outputs. After training, a supervised learning algorithm will take in new unseen inputs and will determine which label the new inputs will be classified as based on prior training data.  We will use the Linear Regression Algorithm for this problem because our data is continuous. Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable. |

Objective

The objective of this prediction is to get the actual no. of deaths of Covid19 infected people. We have train some feature and then test them to get the most optimal no. of fatality rate.

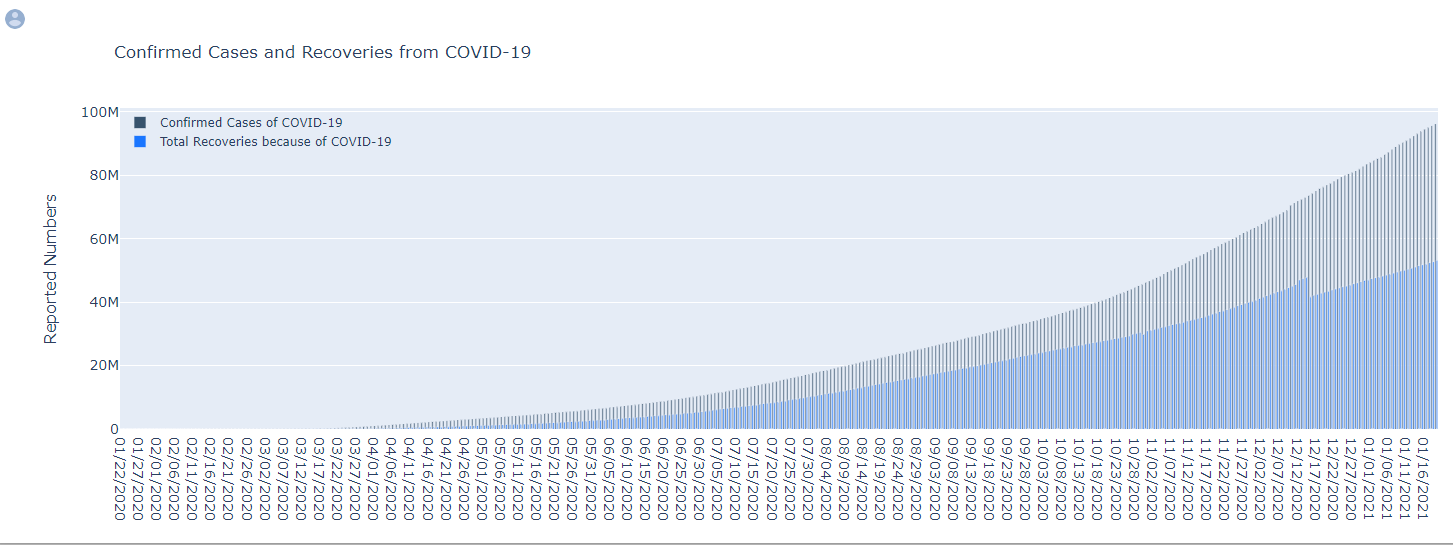
Features

* Observation Date
* Province/State
* Country/Region
* Last Update
* Confirmed
* Deaths
* Recovered

Information regarding Dataset

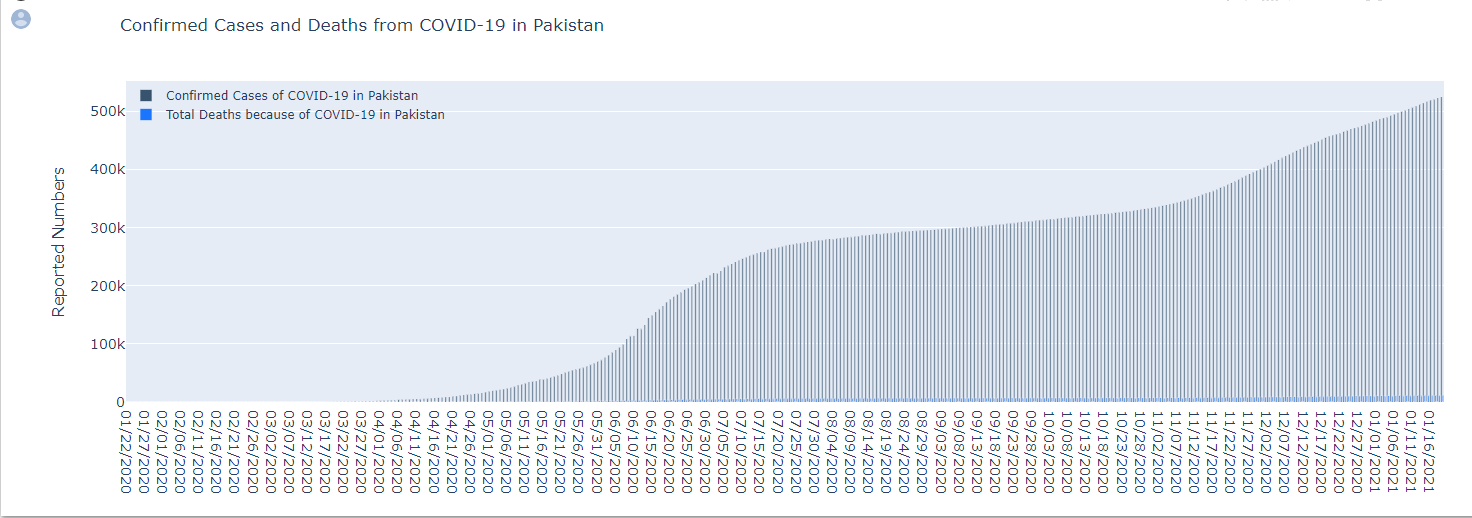
Group & Drop of Columns

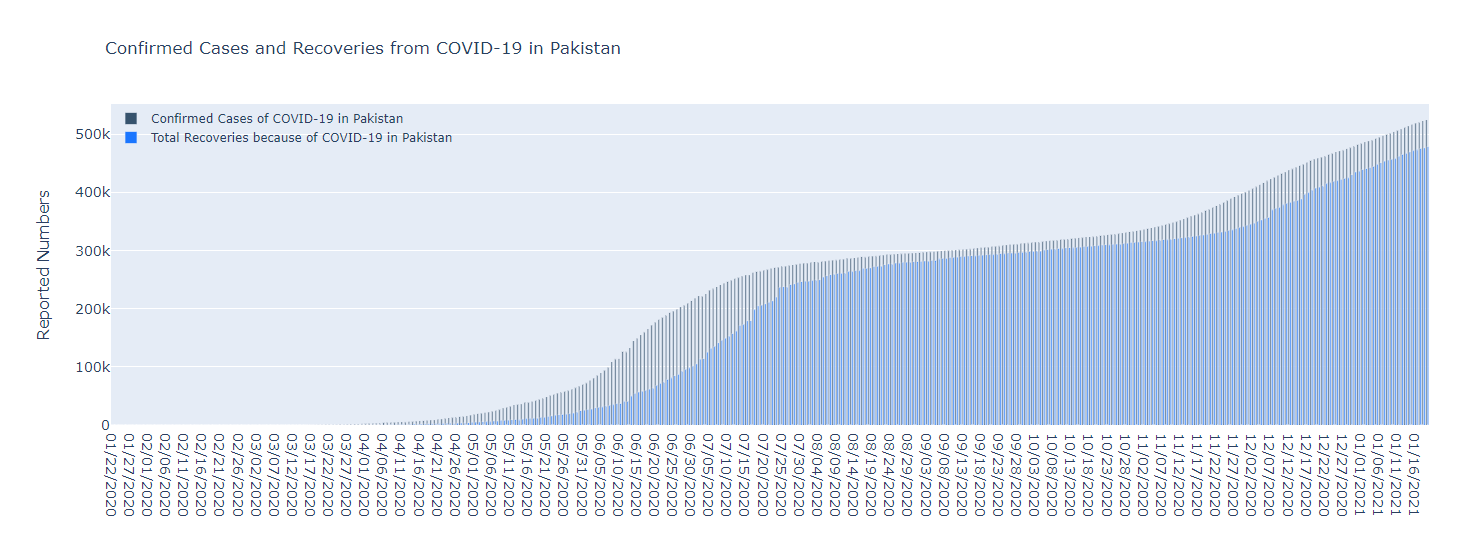
# Bar Graphs of Analysis of Spread and deaths

Confirmed Cases and Recoveries from COVID-19

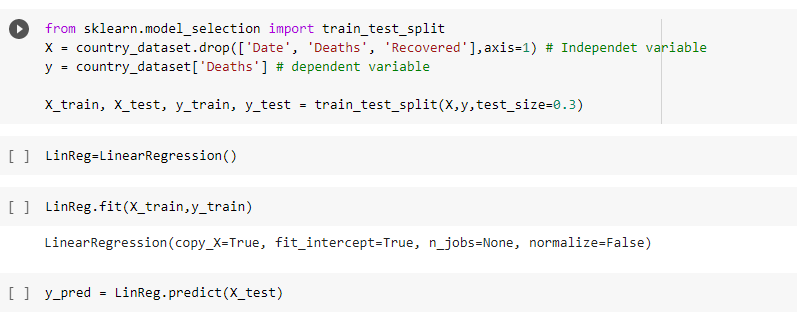
Confirmed Cases and Deaths from COVID-19 in Pakistan

Covid Cases in Pakistan

Confirmed Cases and Deaths from COVID-19 in Pakistan

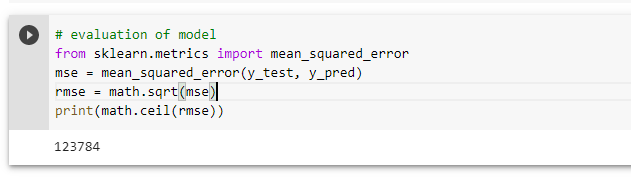
Confirmed Cases and Recoveries from COVID-19 in Pakistan

# Correlation summarize

Applying Linear Regression

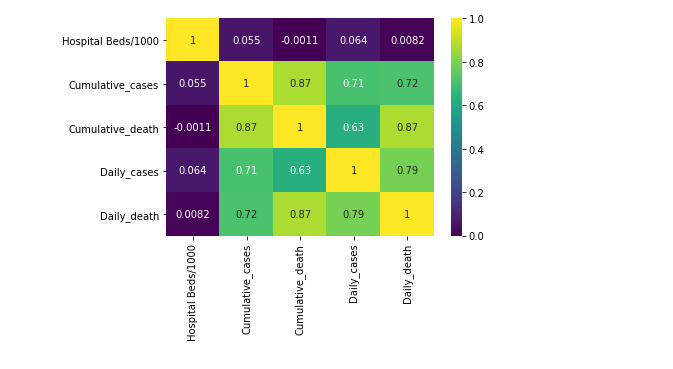
# Prediction on Global Data

Mean Square Error = 123784



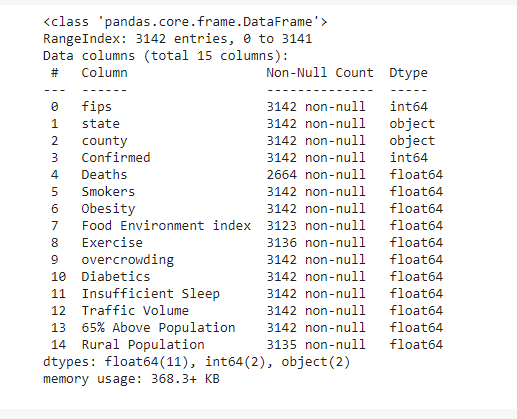
(Because of large no. of deaths occurred due to Covid that’s why it is showing us Mickle no. of mean square error. It might be reduce if the total no. of deaths is decreased)

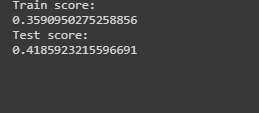
# Working with 2nd Dataset

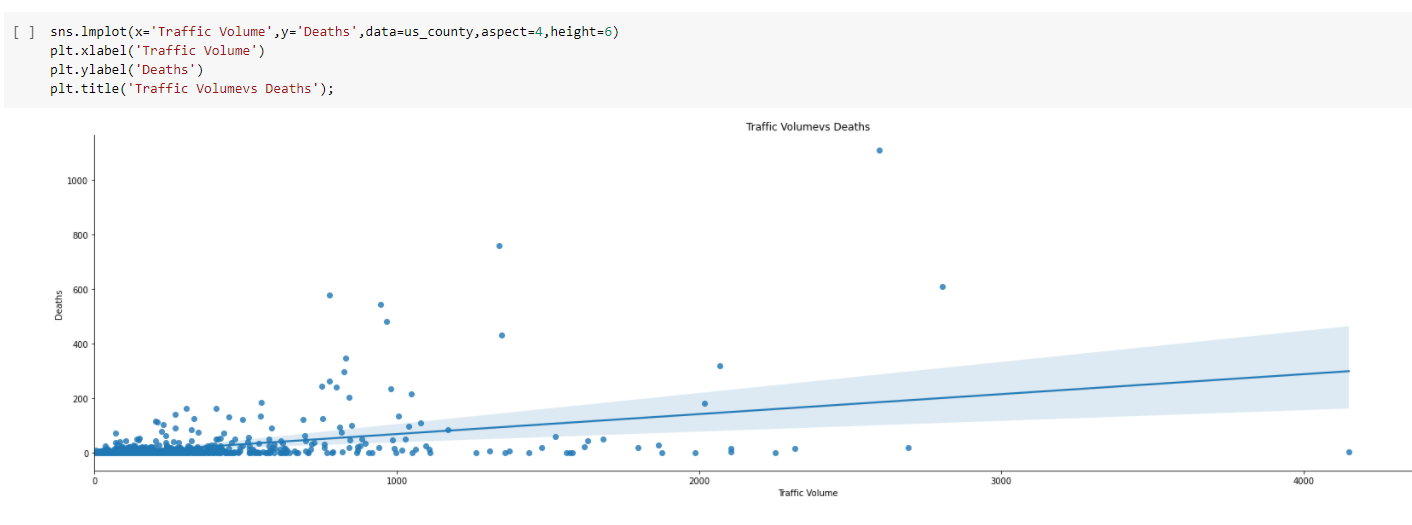
Heatmap

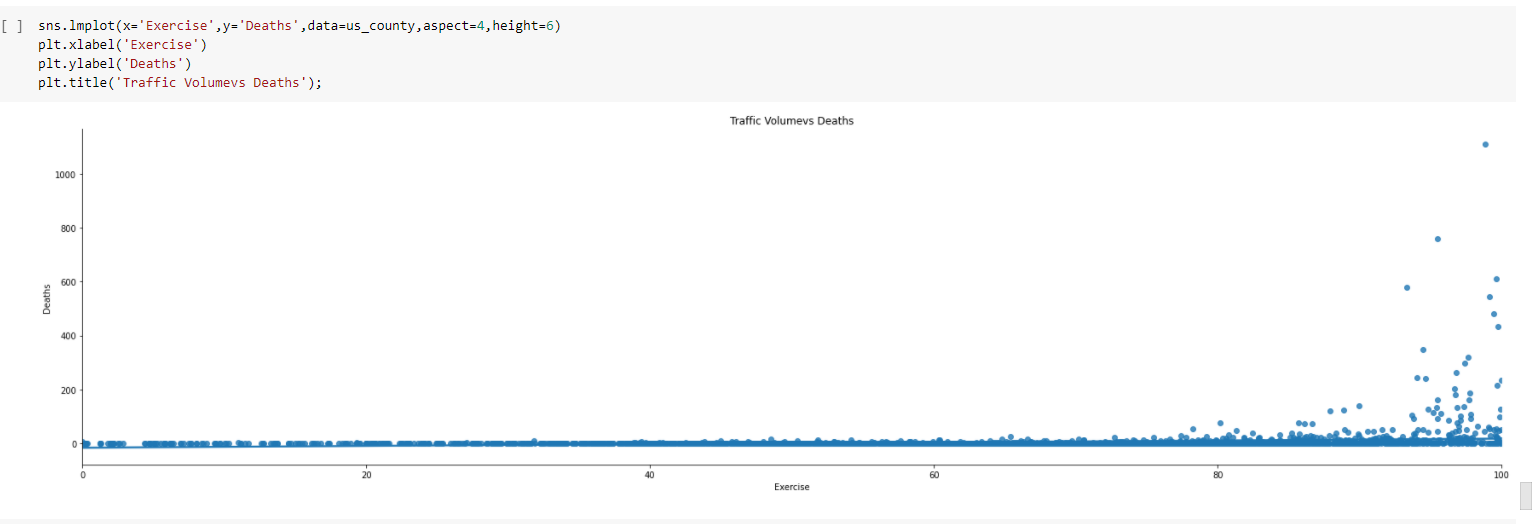
Finding Null values

Working with 3rd Dataset

Dataset Info

Accuracy rate

Traffic Volumes vs Deaths

Exercise Volumes vs Deaths

Train Score : 0.359

Test Score : 0.418

Accuracy Rate : 41 %

Heatmap

Observations of Heatmap

### Observations from the above Heat map over Population Habits

1. None of the figures like Smokers percentage in population, obesity, diabetics tend to affect the spread of COVID-19 Infections in US in general.
2. A certain correlation is observed with the number of confirmed cases in a county and the traffic congestion present for that county (as of 2020). The correlation for the variables are (0.613053). This might be significant as the quarantine and total isolation of people disallowing people movement across US Counties were late in comparison to countries like India/Korea/China/Japan. Hence asymptomatic cases that were carrying the virus might had spread the same, as the moment weren't restricted and the congestion of traffic for the particular counties are high.

CONCLUSION

So far we have used 3 different kind of dataset for this project in which two dataset we used for analysis and one for correlation and for deaths.

We drop those coleus which is not helpful to predict the total no. of deaths. By using some features we are able to predict and know the actual no. of deaths in worldwide and in Pakistan. If someone want to get the least mean square error so for that we must reduce the no. of death in a data.

Our model is having an accuracy of 41% and it is depend on the traffic volume and the no. of deaths. If the correlation between traffic volume and deaths is better so we will get the high accuracy.