**A**

**PROJECT REPORT ON**

**“Analysis of How Worldwide Happiness is Affecting Covid19 Infection”**

**Submitted by:**

**Manasi Dhokare**

**mmdhokare@mitaoe.ac.in**

#### **PROBLEM STATEMENT**

To Analysis of How Worldwide Happiness is Affecting Covid19 Infection.

#### **OBJECTIVES AND SCOPE**

* Extracting features of data which is required in analysis.
* Visualizing how covid19 infection is spreading over the world.
* Finding out the good measures which are responsible for the issue.
* Analyzing and visualizing the effect of world happiness parameters on covid19 infection.

### **DATASET**

**3.1 Dataset Source**

I am working with COVID19 dataset, published by John Hopkins University, consisting of the data related to cumulative number of confirmed cases, per day, in each Country.

Also we have a “Worldwide happiness report” consisting of various life factors, scored by the people living in each country over the world.

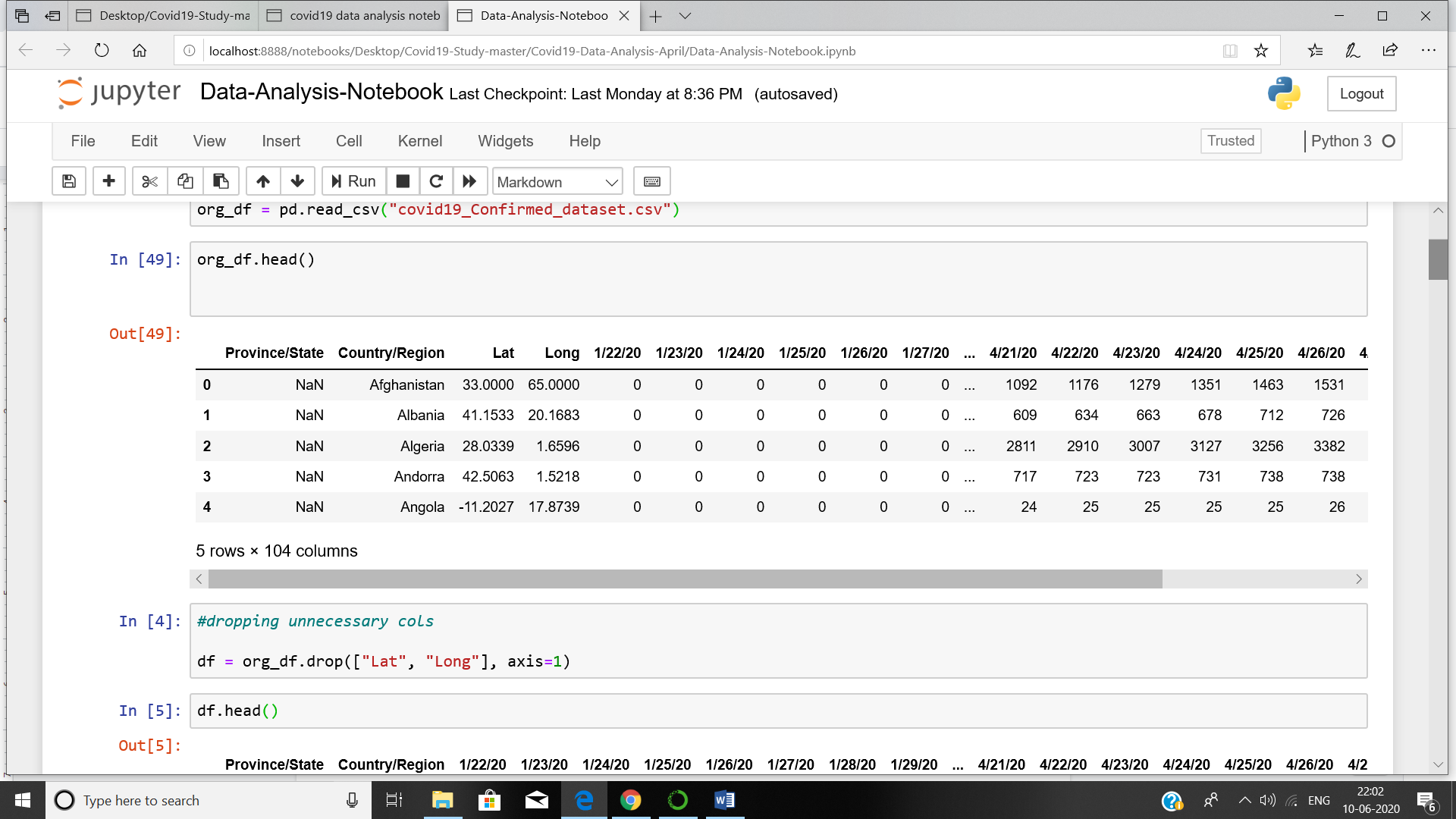
I am going to merge these two datasets to see if there is any relationship between the spread of the virus in a country and how happy people are, living in that country.

**3.2 DATASET DESCRIPTION**

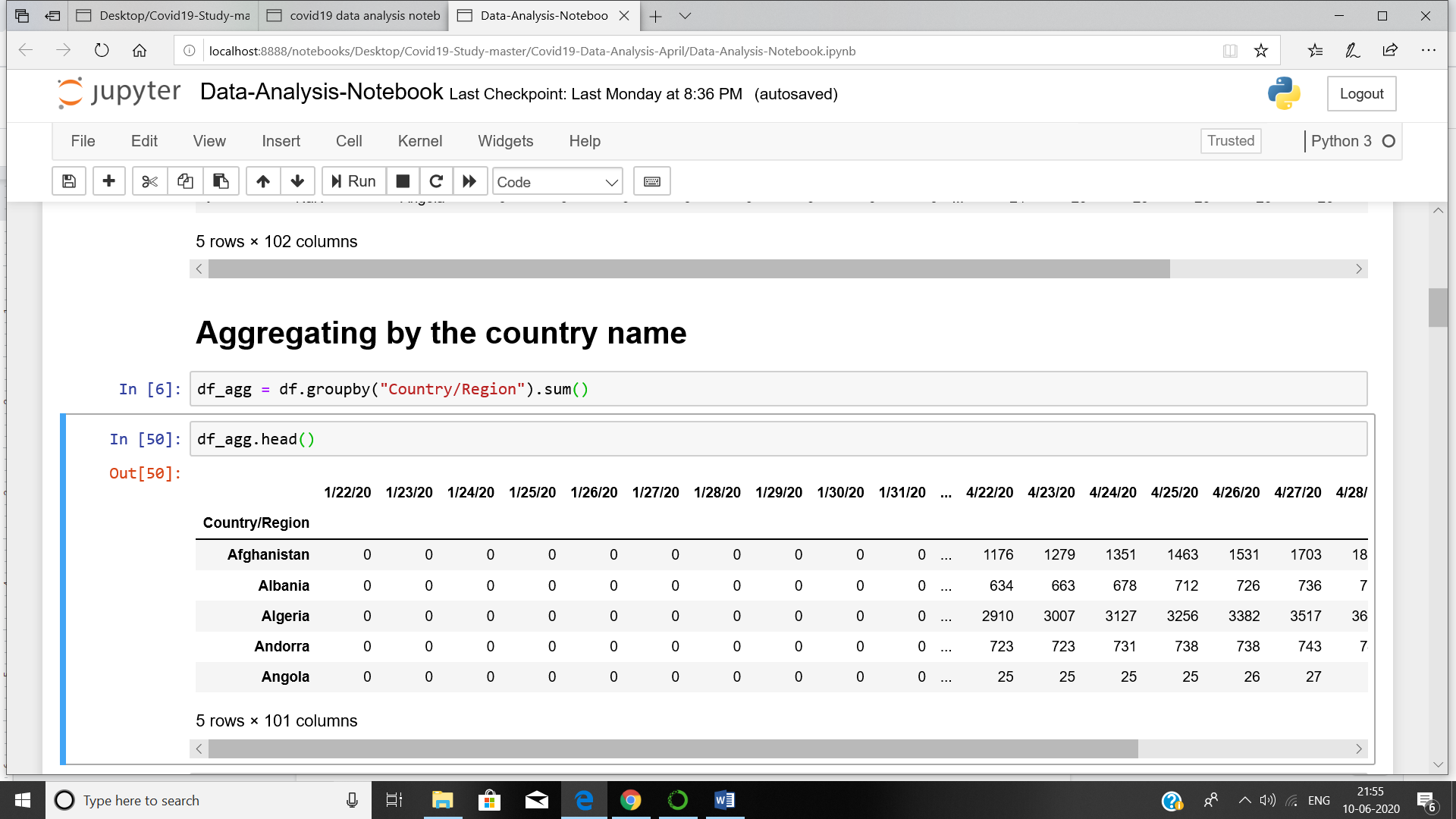
1. **Covid19 infected cases dataset.**

In this dataset there are 104 columns and 266 entries(rows).

In which the first four columns are of Provision/State, Region/Country, Latitude and Longitude and remaining are the dates from 22nd January to 26th April.

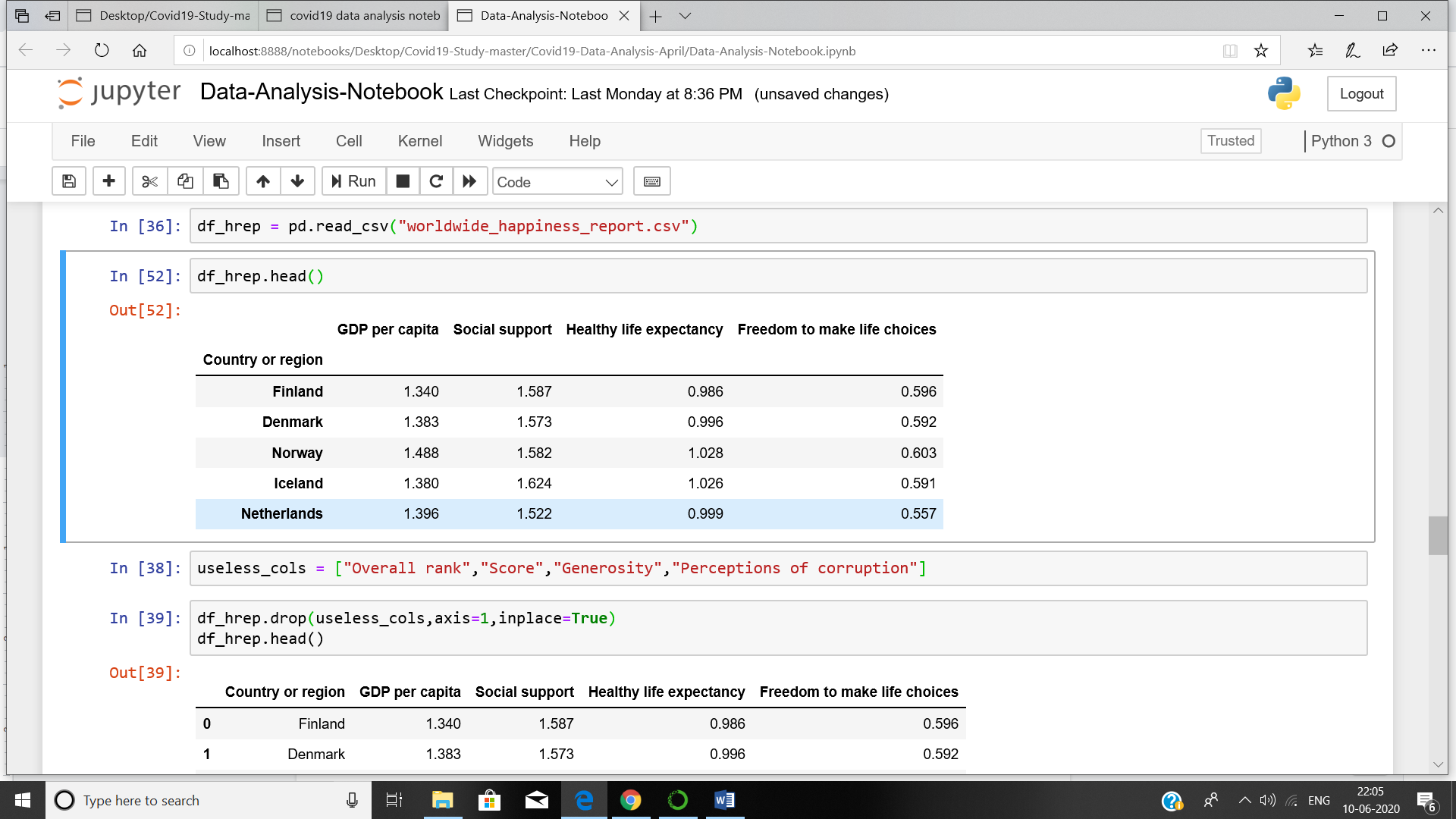
****

After dropping unnecessary columns that is Latitude and Longitude and also aggregating the data by country name we get data as follow:

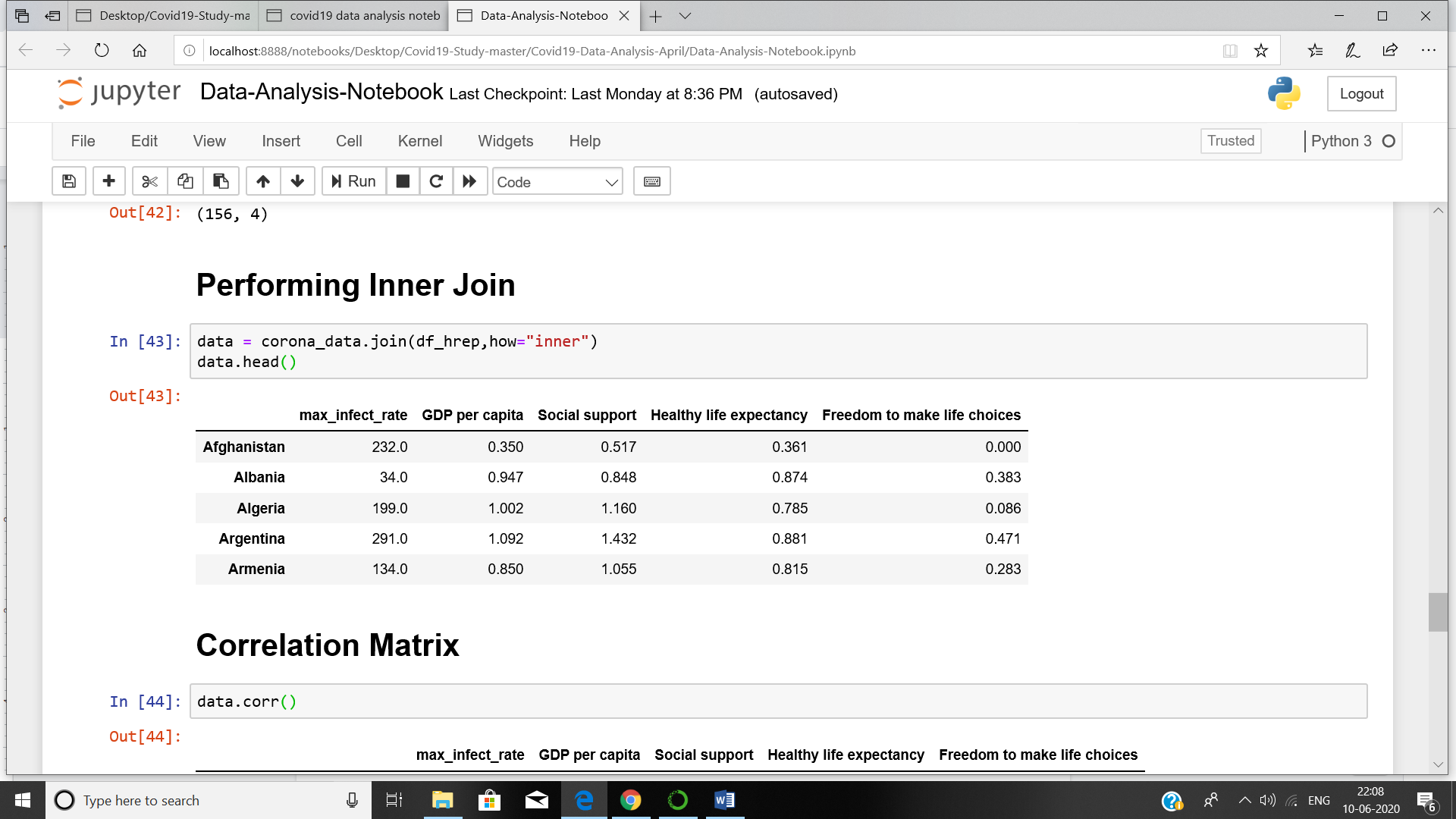


1. **Worldwide happiness report dataset.**

In this dataset there are 4 columns and 156 entries. The columns are GDP per capita, Social support, Healthy life expectancy and freedom to make life choices.



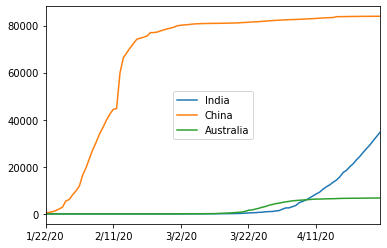
After joining two datasets by country we get our final dataset for visualization and analysis.



**4. EXPLORATORY DATA ANALYSIS**

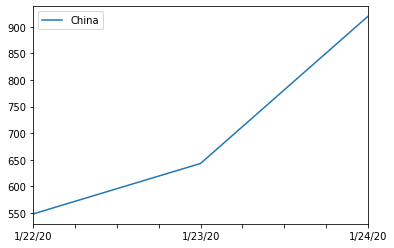
**1. Visualizing data related to a country.**

Visualization always helps for better understanding of our data. By this visualization we can see how the infection trend is going across any country.



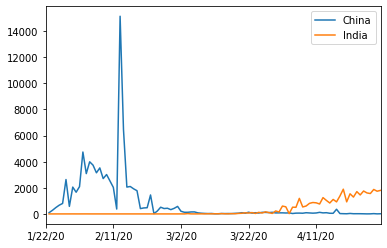
Here on X axis there are dates in 4 months and on Y axis there is a number of infected people.

If we take only one country and see the infection trend over a period with only three days.



From this we can see that how the trend is going day by day.

For finding out the maximum infected value we have to plot differential curve like:



From this we can find out the maximum infection rate for each country.

countries = list(df\_agg.index)

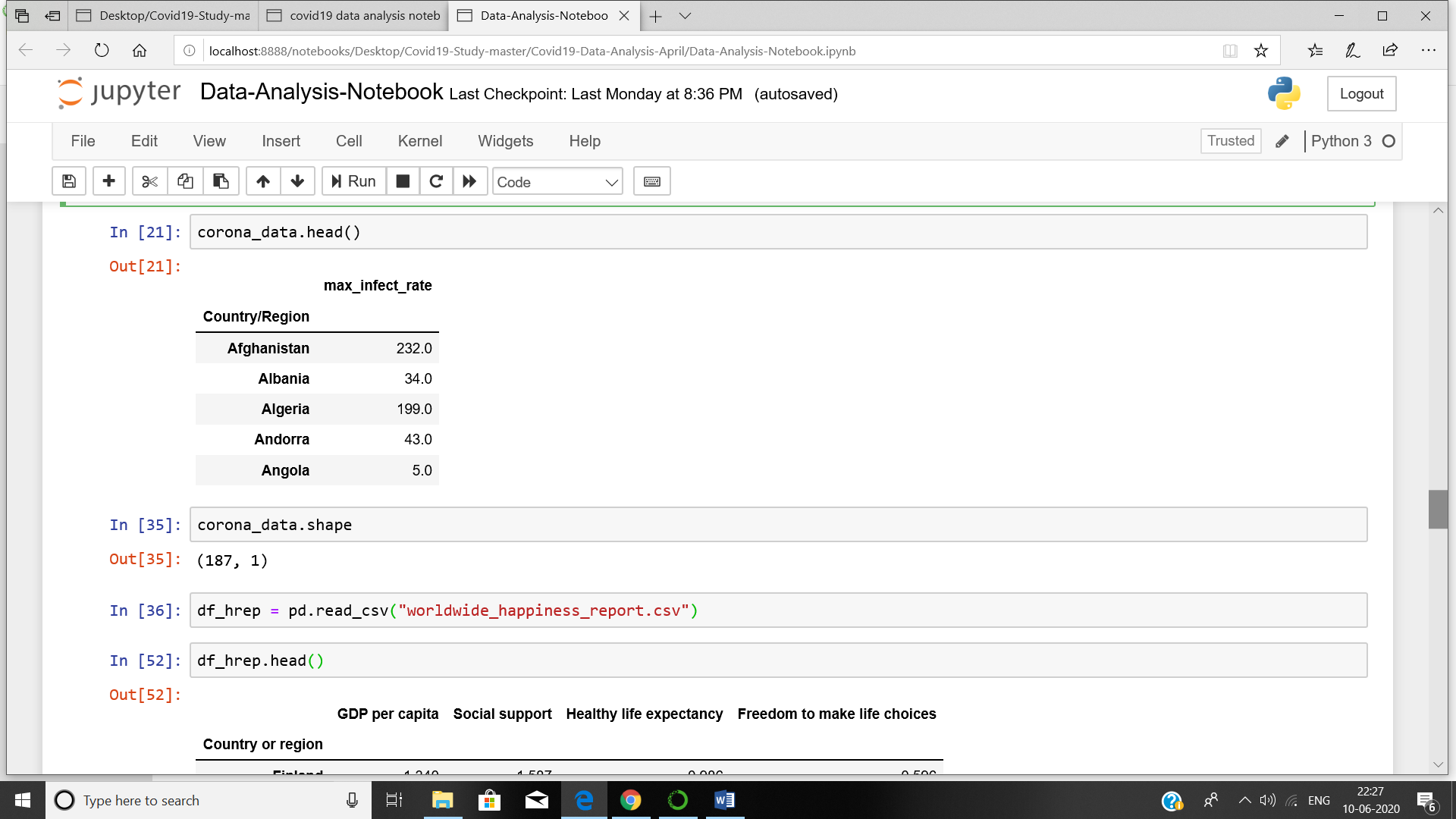
max\_rate = []

for i in countries:

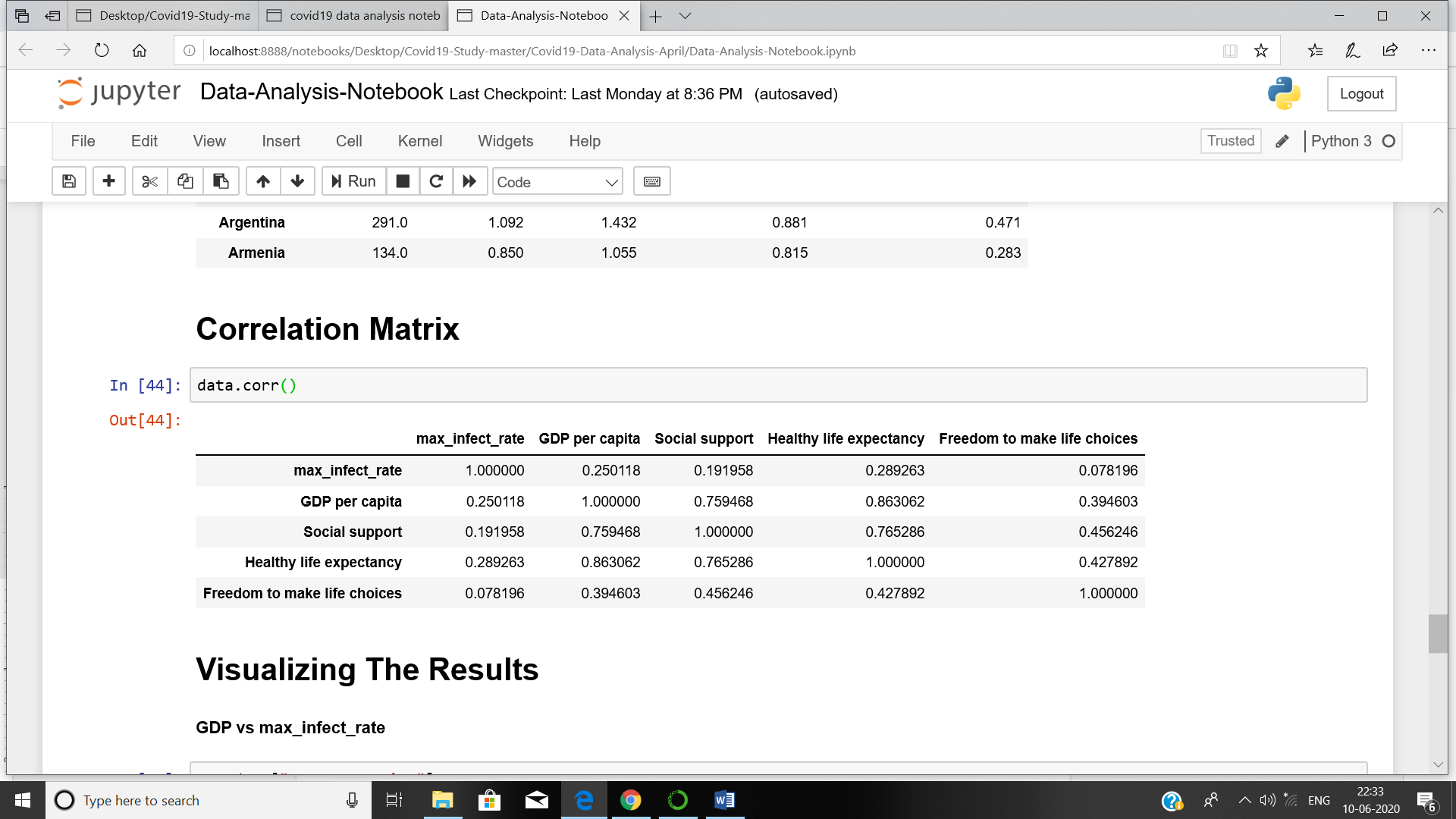
max\_rate.append(df\_agg.loc[i].diff().max())

df\_agg["max\_infect\_rate"] = max\_rate

corona\_data = pd.DataFrame(df\_agg["max\_infect\_rate"])



**2. Finding out the correlation matrix for maximum infection rate and other happiness attributes like GDP per capita and other.**

****

**5. VISUALIZING THE RESULT**

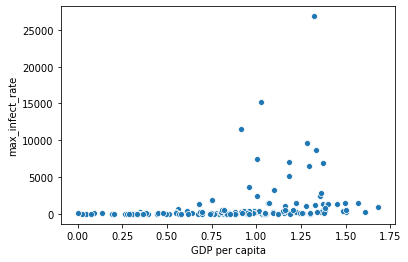
Our Analysis is not finished unless we visualize the results in terms of figures and graphs.

1. **Plotting GDP vs maximum Infection rate**

x = data["GDP per capita"]

y = data["max\_infect\_rate"]

sns.scatterplot(x,y)



Here we have the maximum infection rate in each country and we can see how the GDP per capita affects the infection.

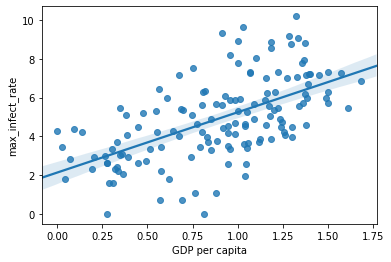
From here we cannot see the result properly as the number of infections are more so by applying log scaling we can see proper results.

x = data["GDP per capita"]

y = data["max\_infect\_rate"]

sns.scatterplot(x,np.log(y))

sns.regplot(x,np.log(y))



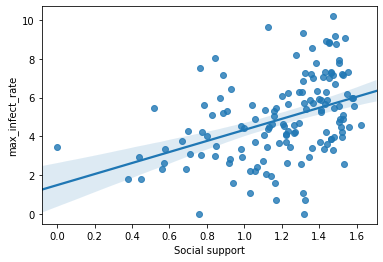
So from here we can properly see our result. We can see that as GDP is increasing infection is also increasing.

1. **Plotting Social support vs maximum Infection rate**

x = data["Social support"]

y = data["max\_infect\_rate"]

sns.regplot(x,np.log(y))

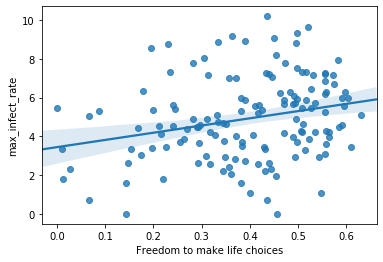


1. **Plotting Healthy life expectancy vs maximum Infection rate**

x = data["Healthy life expectancy"]

y = data["max\_infect\_rate"]

sns.regplot(x,np.log(y))

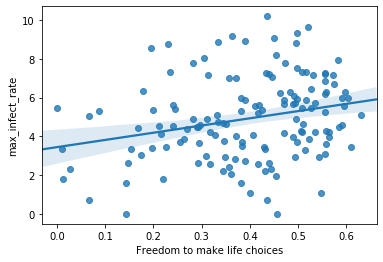


1. **Plotting Freedom to make life choices vs maximum Infection rate**

x = data["Freedom to make life choices"]

y = data["max\_infect\_rate"]

sns.regplot(x,np.log(y))



**6. CONCLUSION**

I have successfully completed a project on Data analysis and visualization of covid19 dataset with Worldwide happiness report. From this we can conclude that the worldwide happiness which is more in developed countries and the people in developed countries are more affected than undeveloped and developing countries.

This may be because the tests conducted in developed countries are more compared to the test conduction in other countries. To overcome this we take covid19 death cases dataset so more analysis can be done which will be future scope.