COVID OUTBREAK ANALYSIS

A Report Submitted in Partial Fulfilment for the Course Entitled Case Study for the Degree in Bachelors of Computer Science and Engineering.

By

Kallamadi Rutva- 160118733187

BE 3/4 Computer Science and Engineering

Bhima Maanvitha - 160118733309

BE 3/4 Computer Science and Engineering

SUBMISSION DATE-28 MAY 2021

Computer Science and Engineering Department,

Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad
(Affiliated to Osmania University, Hyderabad)

DECLARATION

We hereby declare that the case study entitled "**Covid Outbreak Analysis**" submitted for the B. E (CSE- VI Semester) Case Study has not formed the basis for the award of any other degree, diploma, fellowship or any other similar titles.

Name(s) and Signature(s) of the Student

Kallamadi Rutva

Bhima Maanvitha

Place : Hyderabad

Date:28-05-2021

ACKNOWLEDGEMENTS

We would like to express our heartfelt gratitude to **Sri.R.Srikanth sir(Assistant Professor)** and **Ms.Madhurima Rana Mam(Assistant professor)** mentors, for their invaluable guidance and constant support, along with his capable instruction and persistent encouragement.

We are grateful to our Head of Department, **Dr. Y.Ramadevi(Professor),** for her steady support and the provision of every resource required for the completion of this case study.
.

We would like to take this opportunity to thank our Incharge, **Sri. Kolla Morarjee**, as well as the management of the institute, for having designed an excellent learning atmosphere.

Our thanks are due to all members of the staff and our friends for providing us with the help required to carry out the groundwork of this project.

Abstract

Corona Virus disease (COVID-19) is an infectious disease caused by a newly discovered virus, which emerged in Wuhan, China in December of 2019.

Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness.

Global coronavirus epidemic (COVID19) has brought new challenges to the scientific community

Analysing the Outbreak of COVID 19 using Machine Learning

Problem Statement

We need a strong model that predicts how the virus could spread across different countries and regions. The goal of this task is to build a model that predicts the spread of the virus in the coming days.

1. Introduction

Coronavirus Disease (COVID-19) is a viral infection that was proclaimed a global epidemic by the World Health Organization (WHO) in March 2020 reflecting the extent of its worldwide transmission. The declaration of the pandemic of the virus also highlighted the growing fear of the alarming spread and severity of COVID-19. It is characterised by its existence as a public health concern that has spread throughout the world

The outbreak of the 2019 novel coronavirus disease (COVID-19) has adversely affected many countries in the world. The unexpected large number of COVID-19 cases has disrupted the healthcare system in many countries and resulted in a shortage of bed spaces in the hospitals. Consequently, predicting the number of COVID-19 cases is imperative for governments to take appropriate actions. The number of COVID-19 cases can be accurately predicted by considering historical data of reported cases alongside some external factors that affect the spread of the virus.

1.1 Purpose

By the rapid increase in the cases of covid-19 around the world its important to build a model that predict the number of cases based on the daily data available.

Machine Learning (ML) can be deployed very effectively to track the disease, predict growth of the epidemic and design strategies and policies to manage its spread. This study applies an improved mathematical model to analyse and predict the growth of the epidemic.

2. Background information

Machine learning is an innovative approach that has extensive applications in prediction. This technique needs to be applied for the COVID-19 pandemic to identify patients at high risk, their death rate, and other abnormalities. It can be used to understand the nature of this virus and further predict the upcoming issues.

This paper shows how machine learning algorithms and methods can be employed to fight the COVID-19 virus and the pandemic. It further discusses the primary machine learning methods that are helpful during the COVID-19 pandemic. We further identified and discussed algorithms used in machine learning and their significant applications.

The framework analyses data sets containing real data from the past day to day and uses machine learning algorithms to make forecasts about future days. The study findings show that in view of the existence and dimension of the dataset.

The outbreak of Coronavirus 2019 (COVID-19) has impacted everyday lives globally. The number of positive cases is growing and India is now one of the most affected countries. This paper builds predictive models that can predict the number of positive cases with higher accuracy. Regression-based, Decision tree-based, and Random forest-based models have been built on the data from China and are validated on India's sample. The model is found to be effective and will be able to predict the positive number of cases in the future with minimal error. The developed machine learning model can work in real-time and can effectively predict the number of positive cases.

3.Scope of the Case study

<u>FINDINGS AND SUGGESTIONS:</u>

This project can be done and spread on a broader scale and will be successful if we create a webpage so that everyone can have access to it and it can be used on a larger scale. It can be accessible to everyone by just entering few data inputs and gives the predicted value

OUTBREAK ANALYSIS: Impact of COVID-19 on Operations of Different Sectors

Unorganised Sector

This pandemic affected workers of unorganised sector mostly who are daily wager or those working in Micro, Small and Medium Enterprises (MSMEs) rand left them jobless, and rapidly increased the unemployment rate, left no alternate income source. Everyone is witnessing their painful migration on foot and cycles to their homes but now some sort of help has been extended by states by way of running some special trains. After lockdown, giving them employment is a very necessary step, lack of which forced them to leave their home. They may not die from corona but will die definitely from starvation.

Agriculture and Food Processing

Agriculture is considered the backbone of the Indian economy. As Inter-state transportation services have shut down, farmers are unable to sell their crops in the market. They are incurring huge losses and forced to throw out their crops. They don't have any other source of income. The poultry sector which is the fastest-growing subsector of the Indian economy has also incurred huge losses due to social media where misinformation has been spread by correlating the infection of COVID-19 with the consumption of meat and poultry products. Though the government is providing a helping hand still their conditions are miserable

E-Commerce

The government has issued a special advisory for maintaining social distancing to prevent the community transfer of COVID-19 and asked the corporates to allow their employees to work from home. The nationwide lockdown will tremendously affect the operations of the E-commerce industry especially at a time when there is a huge demand for home delivery of goods. Their losses can be recovered if the government brings some policies like lossmaking E-commerce companies can get a GST refund and can grant the permission of some operation with restrictions.

Education

Due to the outbreak of the pandemic, most schools and educational institutions have closed down to prevent the transfer of disease among children. Though, we are safeguarding them, this will also negatively impact their academic progress. Now, we need to shift our focus from traditional to the virtual classroom. There are many technology-enabled educational institutions that are providing live classes like byjus, extramarks etc. This pandemic forced the Government to boost edtech sector. COVID-19 has changed the way of learning in the long term. Higher education has also got affected as universities and colleges are shut down, most higher education institutions are not equipped with digital technology. There will be a delay in the admission process, as most of the entrance exams are scheduled around April and May. Despite online education, platforms helped students in learning but if this could continue then there can be seen drastic unemployment in the education sector.

Tourism and Hospitality Sector

The revenue of the tourism sector got down due to a strict ban on both domestic and international flights. Even many tourists got themselves cancelled. Meetings, conferences and major international events got cancelled like mobile world congress, Olympic, Wimbledon, Cannes international film festival and Facebook F8 which lead to huge losses. Earlier there were a huge number of Indian travellers to both domestic and international destinations but now nobody is willing to go anywhere. According to the Indian Association of tour operators, the hotel, aviation and travel sector together may incur a loss of around 8,500 crores due to the restriction imposed by the Indian government on the movement of flights.

Healthcare Industry

COVID-19 has exposed the vulnerabilities of healthcare systems. As we know that access to healthcare is a fundamental right but the fear of COVID-19 everywhere has in turn affected may people’s primary healthcare provisions. This pandemic has made impossible for the pregnant women to visit obstetrician for prenatal checkups and instead of this, opting for telemedicine. Many hospitals are mainly focusing only on COVID-19 patients and due to this, they are ignoring other people who are suffering from some other major problems like cancer and found it difficult to get proper treatment. If this will be continued the death rate from corona will be lower than the death rate from other diseases. This pandemic has taught a lesson that temples, statues and museums are not a necessary requirement but the hospital with world-class infrastructure is. Even there can be seen an adverse impact on the profitability of medical device manufacturer who imports consumables, disposables and capital equipment from china.

Defense and Security

The COVID-19 impacted the supply chains and production/manufacturing facilities of defense companies. As they have to depend on different components on different sources located in affected countries. This will lead to a decrease in demand for defense equipment. The current scenario is not even good for business development as we know that many high-value procurement programmes were finalised during defense shows which are now cancelled. Military exercises, which expose foreign equipment and their capabilities to the prospective buyer also affected business development as many countries like the USA, UK have cancelled travel plans, deployments and exercise for troops. Even the assembled equipment which are ready for dispatch are also held up due to the lockdown of airspace. Due to non-dispatchment on time their sales value will substantially reduce which in turn will affect the balance sheet of the manufacturing companies. COVID-19 has taught a lesson to defense industry that they need to explore the different aspects of risk planning. They need to shift themselves toward technological platforms or start using an unmanned system.

In the end, it can be said that almost every sector got affected from the pandemic including aviation where all domestic and international flights got cancelled sports where cancellation of events lead to huge loss to organisers, textile sector, as India depend for its raw material on Chinese textile factories which are shut down, even the revenue of transport sector got affected due to ban on public transport including metros and railways. The government has announced a package of twenty lakh crores to stimulate the economy. Post COVID-19, the focus of the government should be on increasing the employment level, provide financial help to industrial units and streamlining the GST regime to ensure ease of doing business. Each and every country need to find the way of living with corona as the vaccine is not likely to be available for mass use for at least a year or two.

Multiple sectors affected in six months of lockdown in India:

The lockdown restriction halted the most economic activities and led to job loss of millions of people and revenue streams. The government took precautionary measures to curb the virus spread by restricting the movement but the virus wreaked havoc in the country.

The lockdown restriction halted the most economic activities and led to job loss of millions of people and revenue streams. The government took precautionary measures to curb the virus spread by restricting the movement but the virus wreaked havoc in the country. The virus killed many, including those who were fighting the battle against it to safeguard others' life. Let's take a look at the impact of COVID-19 during the six months since the lockdown.

Healthcare Sector | Healthcare is the epicenter during this unprecedented global pandemic. Lack of medical investment and healthcare infrastructure are the biggest challenge for an effective response in India in a battle against the novel pandemic. Amid the rising cases lack of healthcare facilities like shortage of beds, lack of protection equipment. According to a survey conducted by LocalCircles, which included responses from over 17,000 individuals located in over 211 districts of the country, only 4 percent patients who needed an ICU bed were able to find one by going through the routine process while 78 percent are forced to use connections, clout to secure an ICU bed.

GDP Contracted | In the first quarter of the fiscal year 2021 India's Gross Domestic Product collapsed by 23.9 percent amid the coronavirus lockdown. The pandemic has led to an unprecedented shutdown of business, industries and services. From manufacturing to real estate, hospitality to mining, has been impacted as the economy records its sharpest drop in 41 years.

Job Loss | The pandemic has wreaked havoc on the job landscape in India. According to the Centre for Monitoring Indian Economy (CMIE) about 21 million salaries employees lost their jobs during April-August. There were 86 million salaried jobs in India during 2019-20. In August 2020, the count was down to 65 million after 3.3 million jobs were gone in the particular month, mostly among the industrial workers and white collar workers, as per the CMIE

Income Loss for Vulnerable Section | The lockdown has impacted the disadvantaged groups with a greater magnitude. According to an analysis by Scroll, during the first two months of the lockdown, India's vulnerable section lost incomes amounting to as much as Rs 4 lakh crores, or nearly 2 percent of the country's annual GDP

Travel & Tourism Industry | Indian travel and tourism industry is one of the worst-impacted sectors by the coronavirus pandemic. According to the Confederation of Indian Industry (CII) and hospitality consulting firm Hotelivate. Tour operators, including both online and offline as well as inbound and outbound will lose $4.77 billion. The entire value chain linked to Travel & Tourism is likely to lose around 5 lakh crore or US $65.57 billion, with the organized sector alone likely to lose US $25 billion.

Indian Railways | Amid the coronavirus crisis Railways had suspended all passenger services from March till May 3. During a web conference on July 28, Railway Board chairman VK Yadav said that the Indian Railways expects to incur revenue loss of up to Rs 35,000 from passenger train segment during the FY21 due to a fallout of train travel following the restriction to prevent the spread of COVID-19.

Education Sector | The pandemic has disrupted the education sector as people are forced to stay at home. The schools and colleges remain closed ever since the nationwide lockdown was imposed. Many children are struggling to keep up with the challenges of online classes. Students and teachers in cities, towns and villages scramble to cope with the demands of the times.

SOFTWARE SPECIFICATIONS:

Data Science
Jupyter Notebook
Python for programming
CSV file for Dataset

DATA ANALYSIS:
Source Of The Dataset : Kaggle

From the original data examples with missing values were removed .
Total number of observations in dataset: 33
Total number of attributes removed in the dataset : 8 (As they are least dependent columns they are not required . After removing the unrequired data)

Total Number of Rows : 79477

Total Number of Rows After Removing : 194 (Restricting the dataset only to India)

Total Number of Columns After Removing : 25

4. Design and Implementation

4.1 Data Pre-processing

The dataset consists of columns with the data being the Date, String, and Numeric type. We also have categorical variables in the dataset. Since the ML model requires all the data that is passed as input to be in the numeric form, we performed label-encoding of the categorical variables. This assigns a number to every unique categorical value in the column.

The dataset consists of multiple missing values which cause an error when passed directly as an input. Thus, we fill the missing values with “NA.” Certain patient data records contain missing values for both the “death” and “recov” columns, such patient records have been separated from the main dataset and compiled into the test dataset, while the remaining records have been compiled into the train dataset.

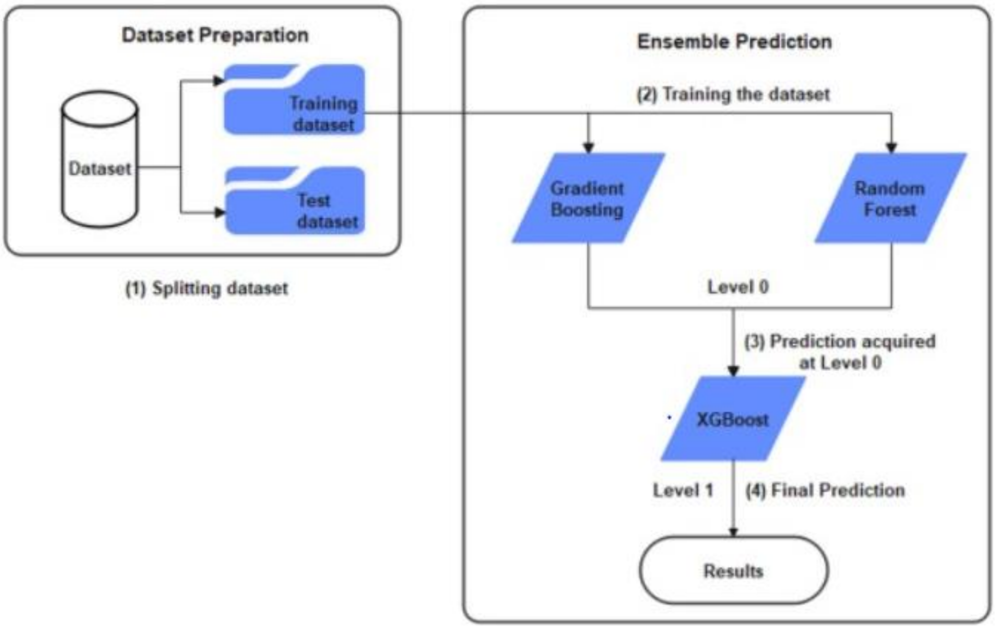


Figure 4.1.1

Data set: https://covid.ourworldindata.org/data/owid-covid-data.csv

Logistic regression and random forest (RF) algorithms were utilized to build classification models.

Random forest

A Random Forest is a technique used to perform both regression and classification models with the support of multiple decision trees and is called Bootstrap and Aggregation, traditionally call as bagging. The basic idea of this technique is to combine multiple decision trees in the process of getting final output rather than relying on individual decision trees .Random Forest has multiple decision trees as base learning models. In Decision Tree and Random Forest models MSE is used to estimate the error. It is a set of Decision Trees. Each Decision Tree is a set of internal nodes and leaves. In the internal node, the selected feature is used to make decision how to divide the data set into two separate sets with similars responses within.

Method Used : Root Mean Squared Error , Random forest regressor

Linear Regression

Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. It’s used to predict values within a continuous range, (e.g. sales, price) rather than trying to classify them into categories (e.g. cat, dog). There are two main types:

Simple regression

Simple linear regression uses traditional slope-intercept form, where mm and bb are the variables our algorithm will try to "learn" to produce the most accurate predictions. xx represents our input data and yy represents our prediction.

y=mx+by=mx+b

Multivariable regression

A more complex, multi-variable linear equation might look like this, where ww represents the coefficients, or weights, our model will try to learn.

f(x,y,z)=w1x+w2y+w3zf(x,y,z)=w1x+w2y+w3z

The variables x,y,zx,y,z represent the attributes, or distinct pieces of information, we have about each observation. For sales predictions, these attributes might include a company's advertising spend on radio, TV, and newspapers.

Sales=w1Radio+w2TV+w3News

DFD OF LINEAR REGRESSION:

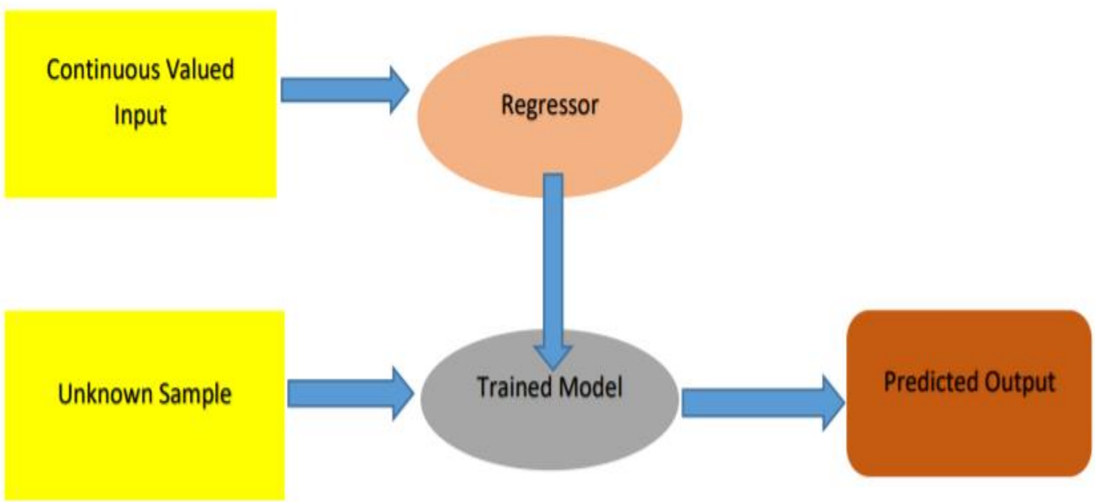


Figure 4.1.2

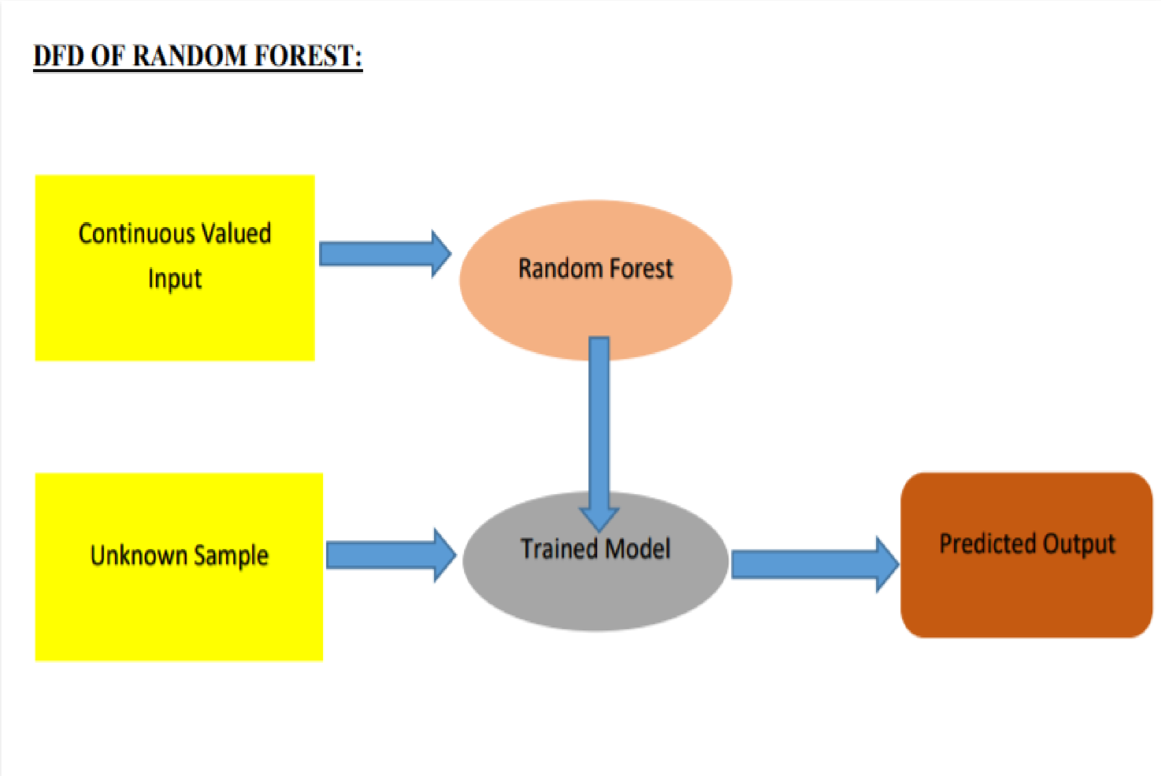


Figure 4.1.3

E-R DIAGRAM

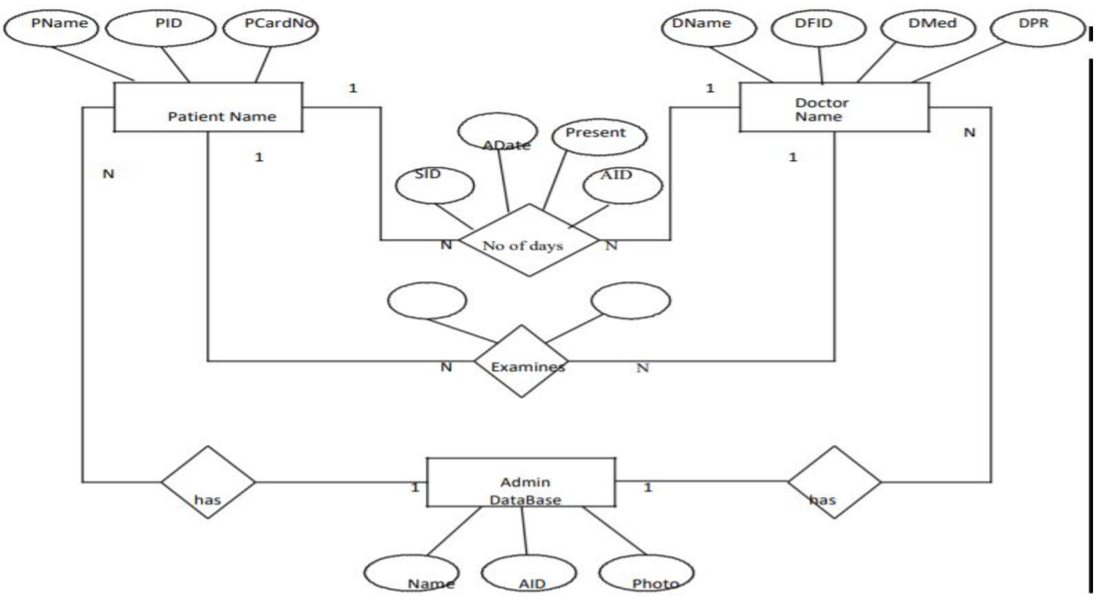


Figure 4.1.4

SEQUENCE AND COLLABORATION DIAGRAMS

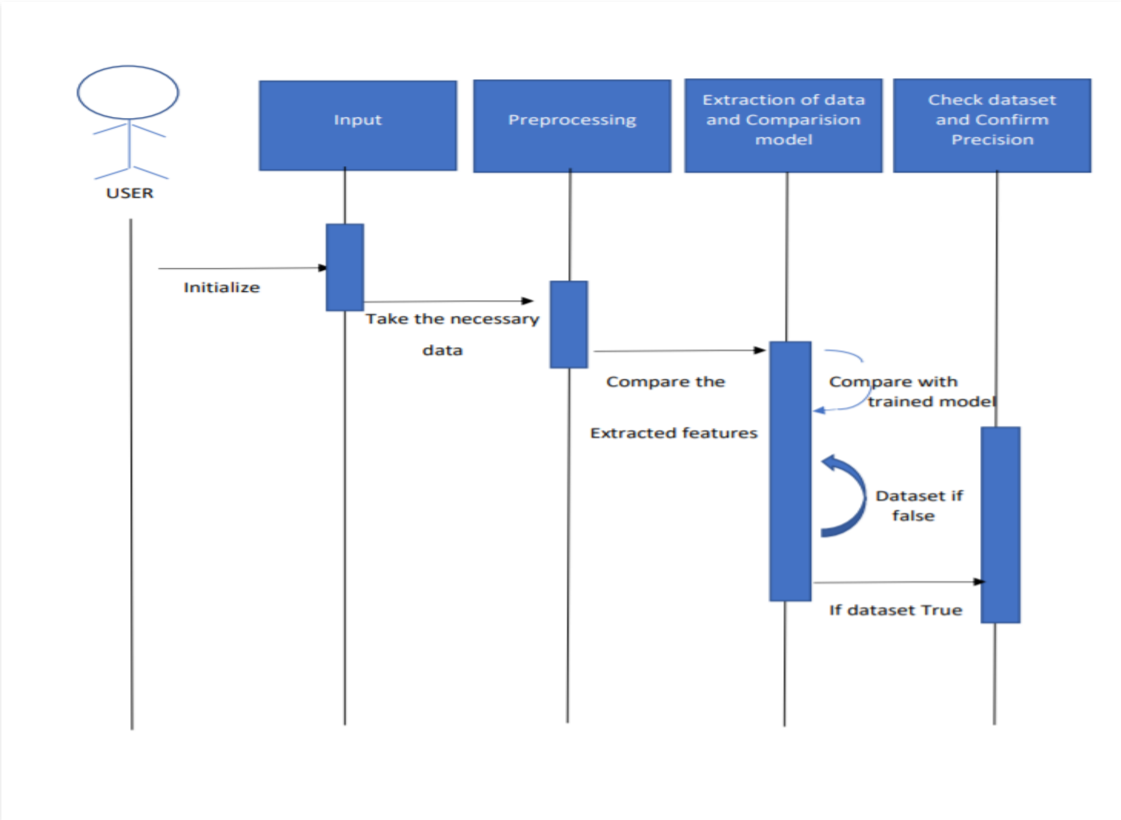


Figure 4.1.5

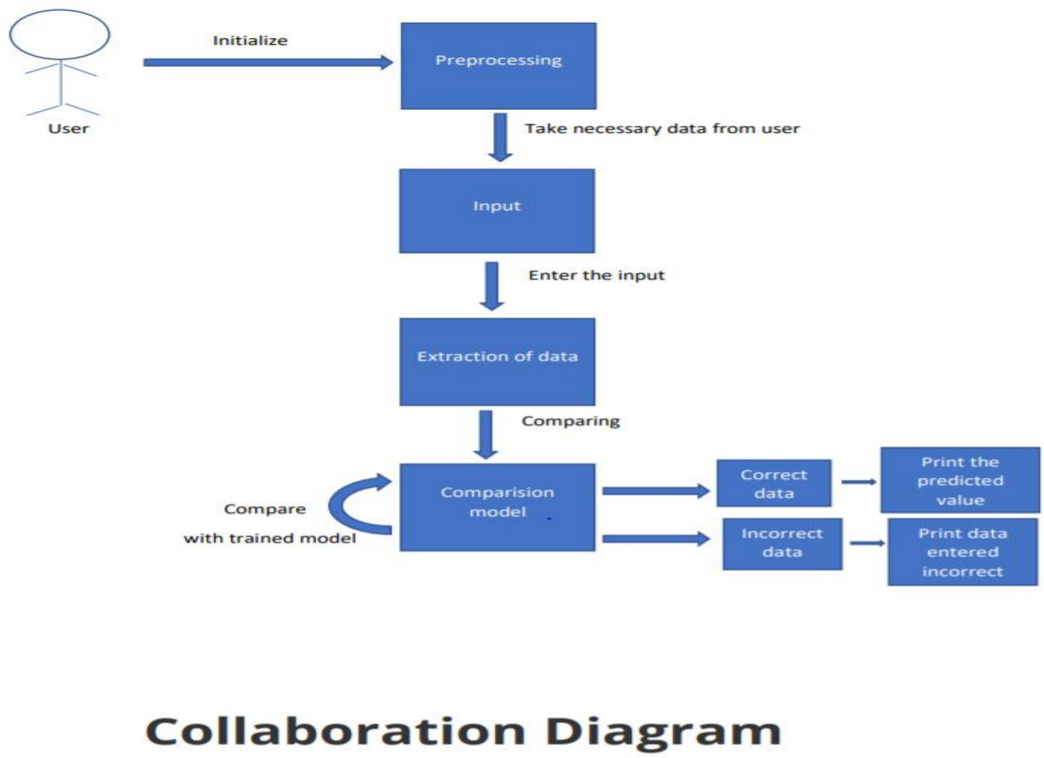


Figure 4.1.6

5. Result Analysis and alternative Solutions

Thus built a model of random forest , linear regression models with **99%** accuracy. We built a successful model.

5.1 PREDICTED GRAPHS

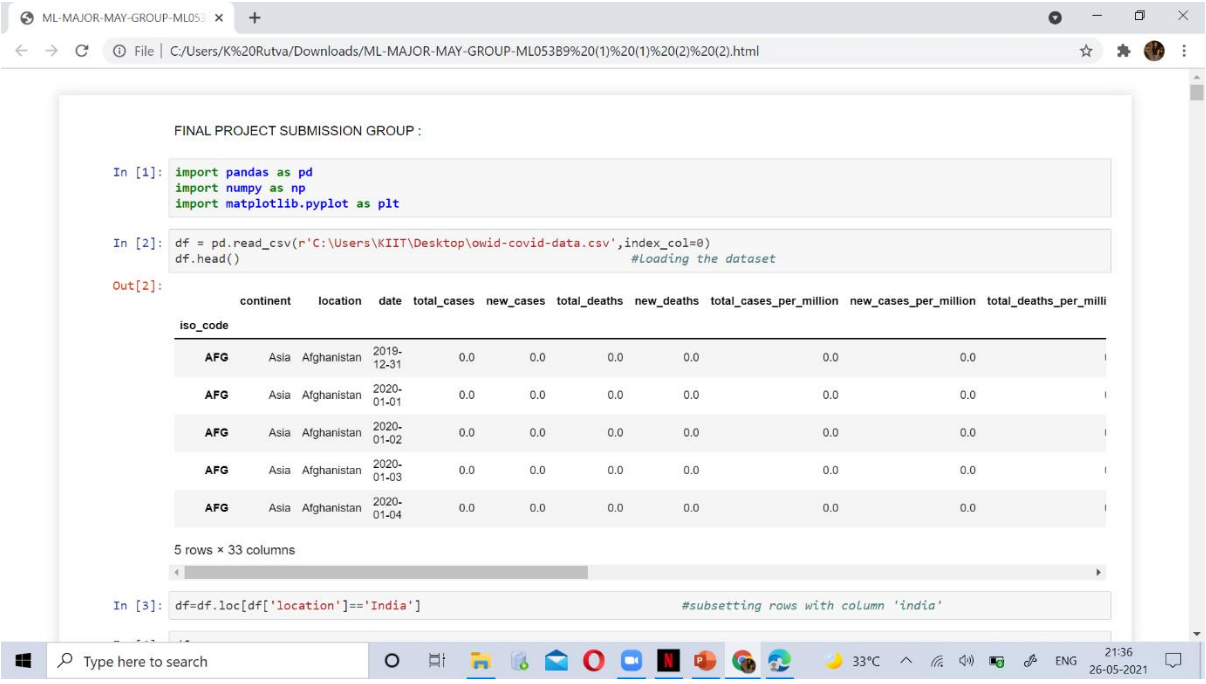


Figure 5.1.1

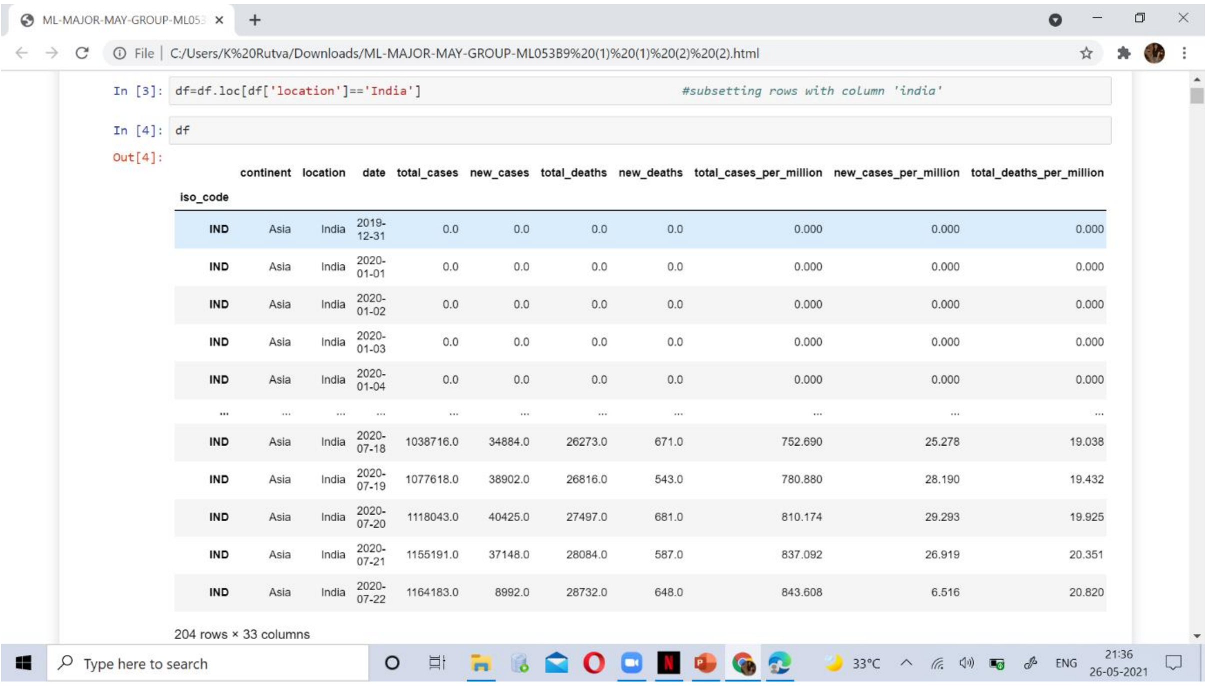


Figure 5.1.2

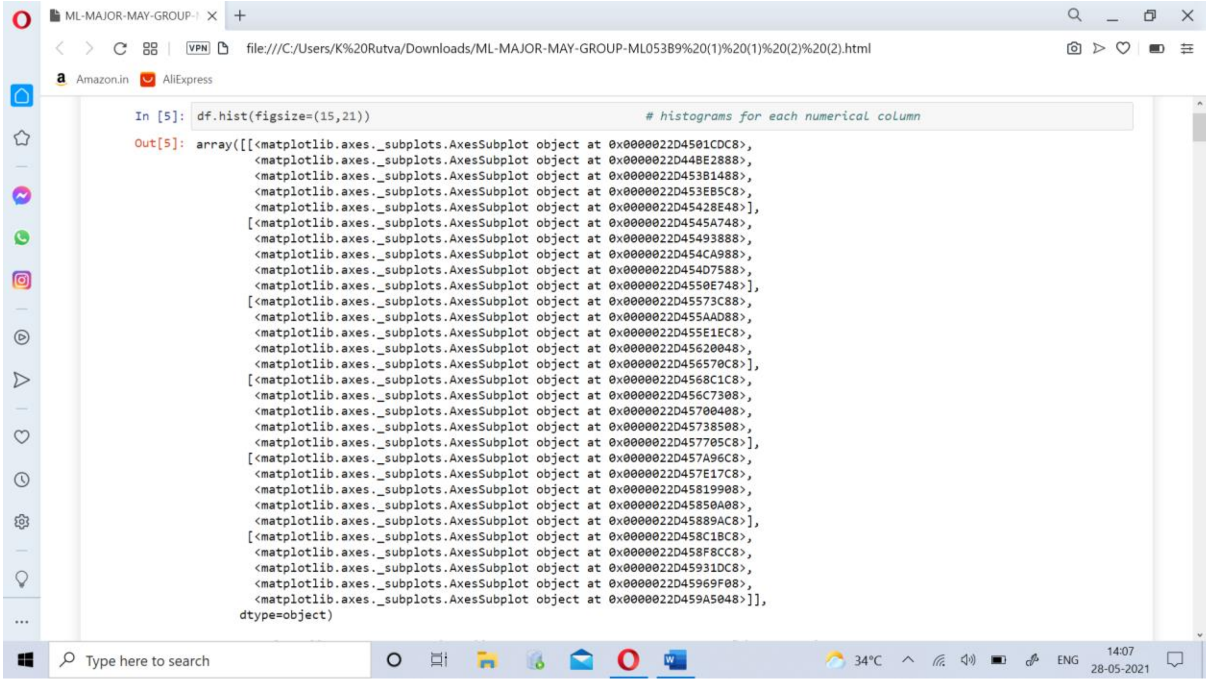


Figure 5.1.3

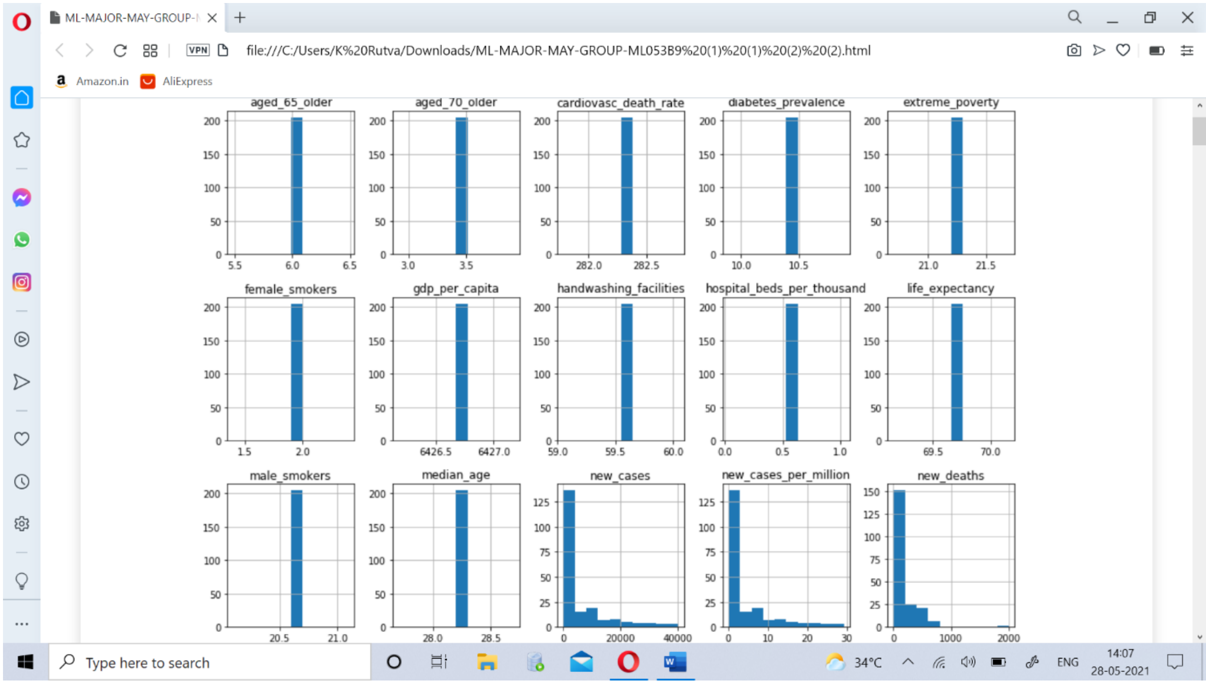


Figure 5.1.4

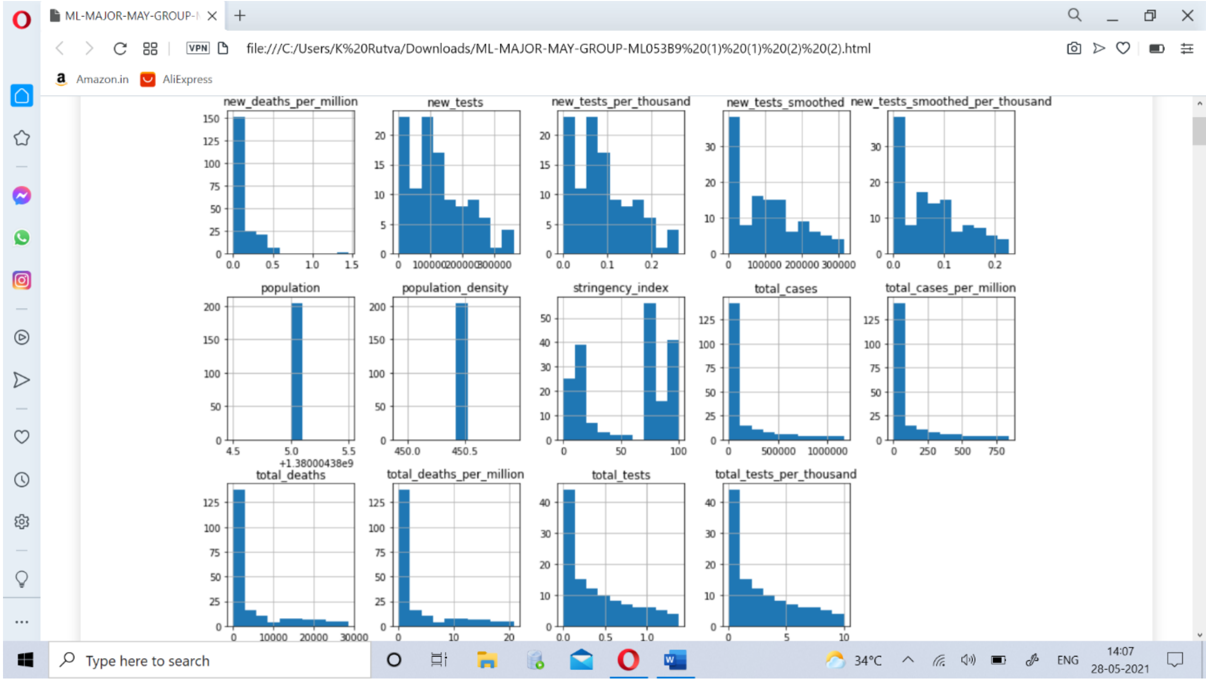


Figure 5.1.5

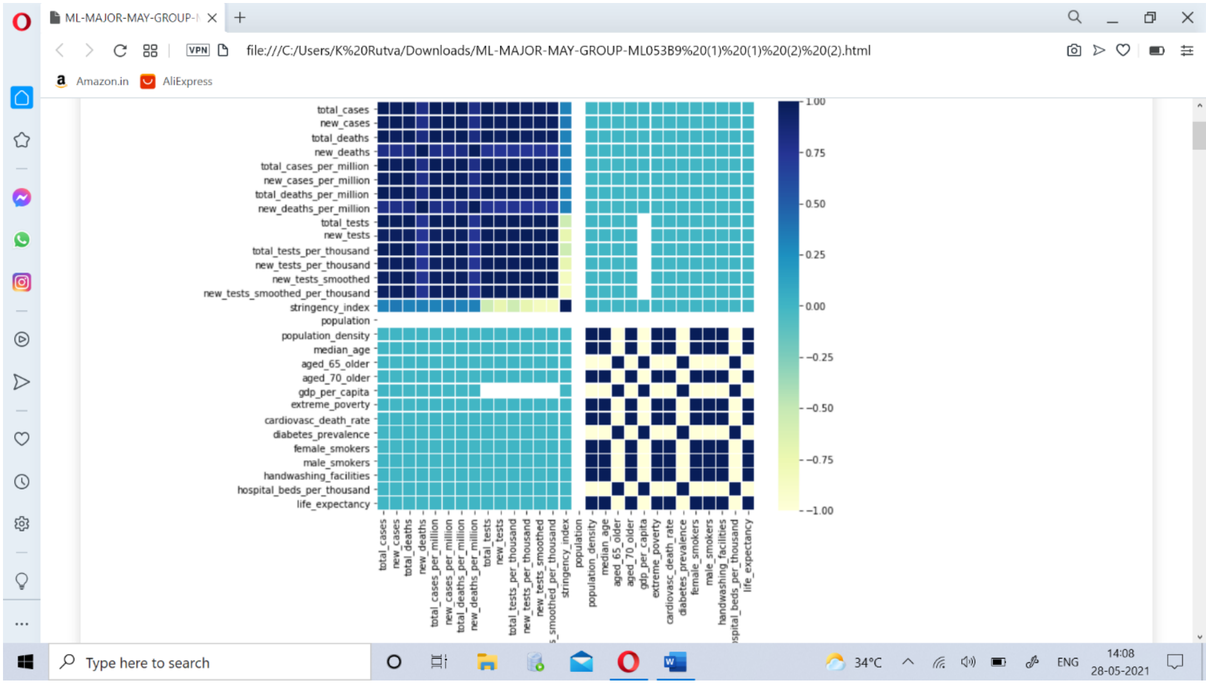
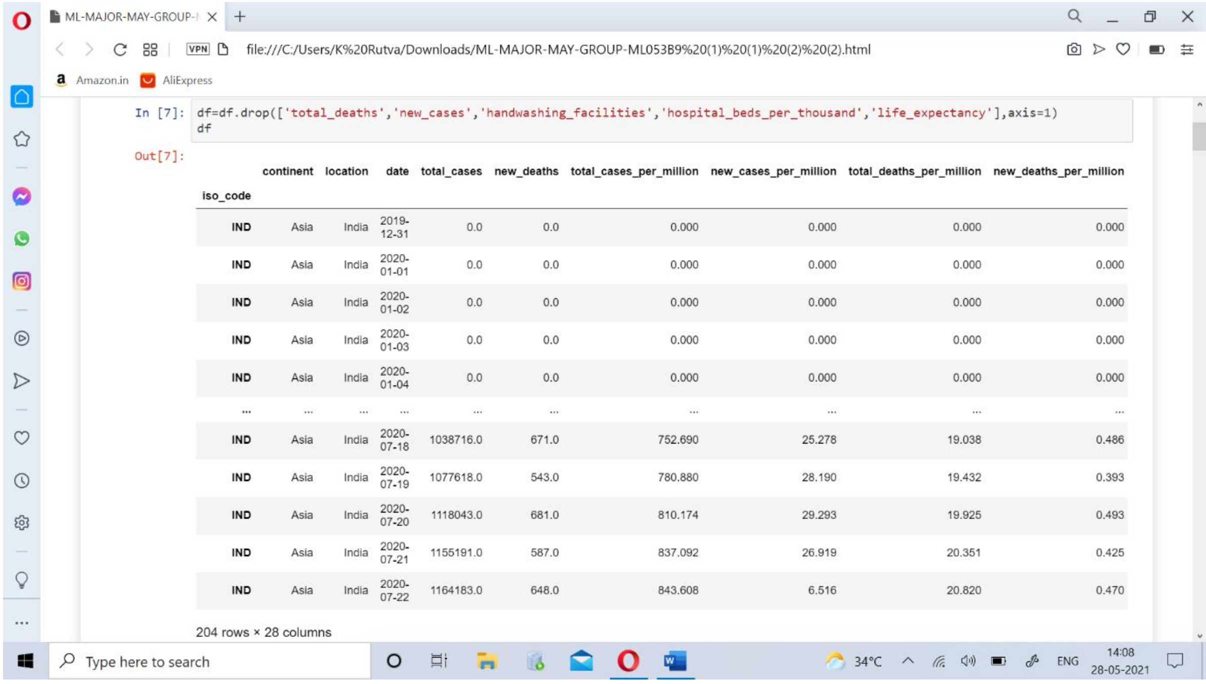


Figure 5.1.6

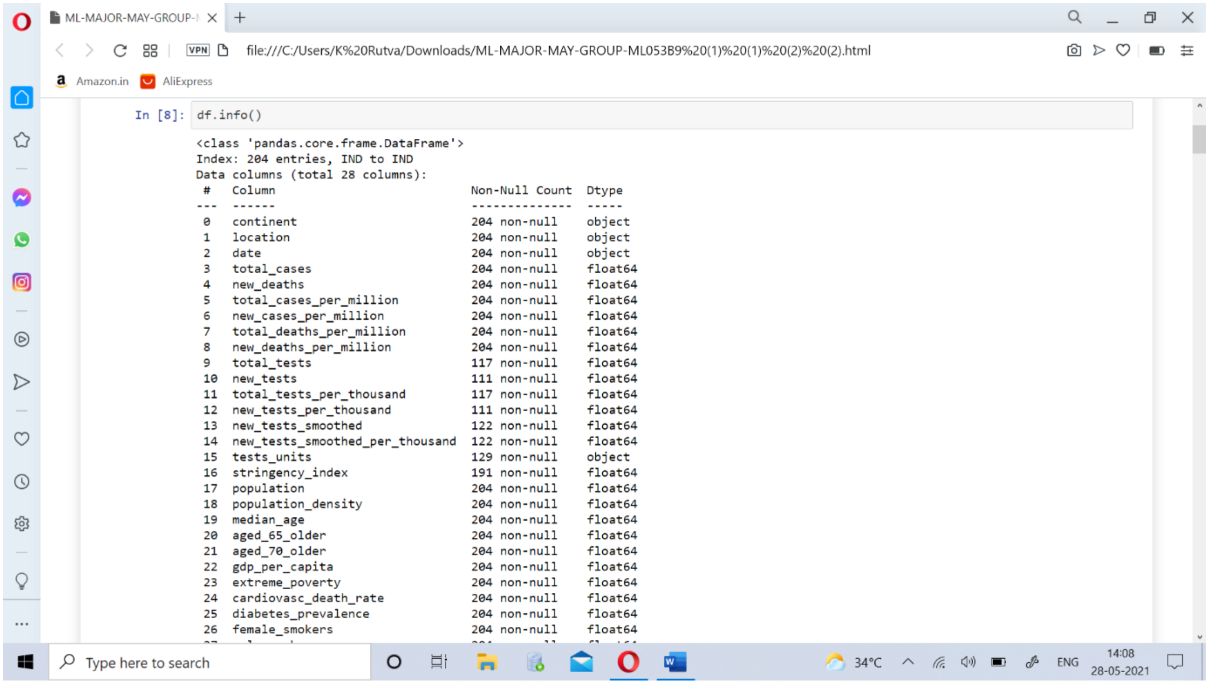


```
In [7]: df=df.drop(['total_deaths','new_cases','handwashing_facilities','hospital_beds_per_thousand','life_expectancy'],axis=1)
df
Out[7]:
```

iso_code	continent	location	date	total_cases	new_deaths	total_cases_per_million	new_cases_per_million	total_deaths_per_million	new_deaths_per_million
IND	Asia	India	2019-12-31	0.0	0.0	0.000	0.000	0.000	0.000
IND	Asia	India	2020-01-01	0.0	0.0	0.000	0.000	0.000	0.000
IND	Asia	India	2020-01-02	0.0	0.0	0.000	0.000	0.000	0.000
IND	Asia	India	2020-01-03	0.0	0.0	0.000	0.000	0.000	0.000
IND	Asia	India	2020-01-04	0.0	0.0	0.000	0.000	0.000	0.000
...	...	...	...	...	...	...	...	...	...
IND	Asia	India	2020-07-18	1038716.0	671.0	752.690	25.278	19.038	0.486
IND	Asia	India	2020-07-19	1077618.0	543.0	780.880	28.190	19.432	0.393
IND	Asia	India	2020-07-20	1118043.0	681.0	810.174	29.293	19.925	0.493
IND	Asia	India	2020-07-21	1155191.0	587.0	837.092	26.919	20.351	0.425
IND	Asia	India	2020-07-22	1164183.0	648.0	843.608	6.516	20.820	0.470

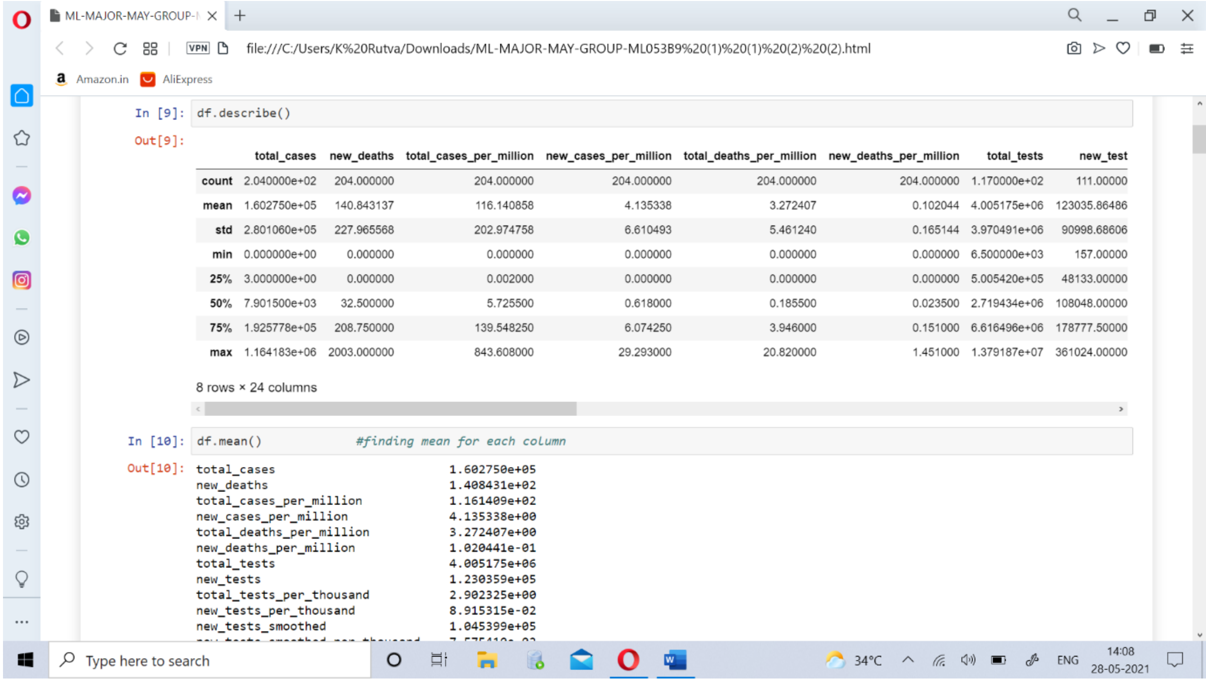
204 rows × 28 columns

Figure 5.1.7



```
In [8]: df.info()

<class 'pandas.core.frame.DataFrame'>
Index: 204 entries, IND to IND
Data columns (total 28 columns):
 #   Column                           Non-Null Count  Dtype
---  ------                           --------------  -----
 0   continent                        204 non-null    object
 1   location                         204 non-null    object
 2   date                             204 non-null    object
 3   total_cases                      204 non-null    float64
 4   new_deaths                       204 non-null    float64
 5   total_cases_per_million          204 non-null    float64
 6   new_cases_per_million            204 non-null    float64
 7   total_deaths_per_million         204 non-null    float64
 8   new_deaths_per_million           204 non-null    float64
 9   total_tests                      117 non-null    float64
 10  new_tests                        111 non-null    float64
 11  total_tests_per_thousand         117 non-null    float64
 12  new_tests_per_thousand           111 non-null    float64
 13  new_tests_smoothed               122 non-null    float64
 14  new_tests_smoothed_per_thousand  122 non-null    float64
 15  tests_units                      129 non-null    object
 16  stringency_index                 191 non-null    float64
 17  population                       204 non-null    float64
 18  population_density               204 non-null    float64
 19  median_age                       204 non-null    float64
 20  aged_65_older                    204 non-null    float64
 21  aged_70_older                    204 non-null    float64
 22  gdp_per_capita                   204 non-null    float64
 23  extreme_poverty                  204 non-null    float64
 24  cardiovasc_death_rate            204 non-null    float64
 25  diabetes_prevalence              204 non-null    float64
 26  female_smokers                   204 non-null    float64
```

Figure 5.1.8



In [9]:
```
df.describe()
```

Out[9]:

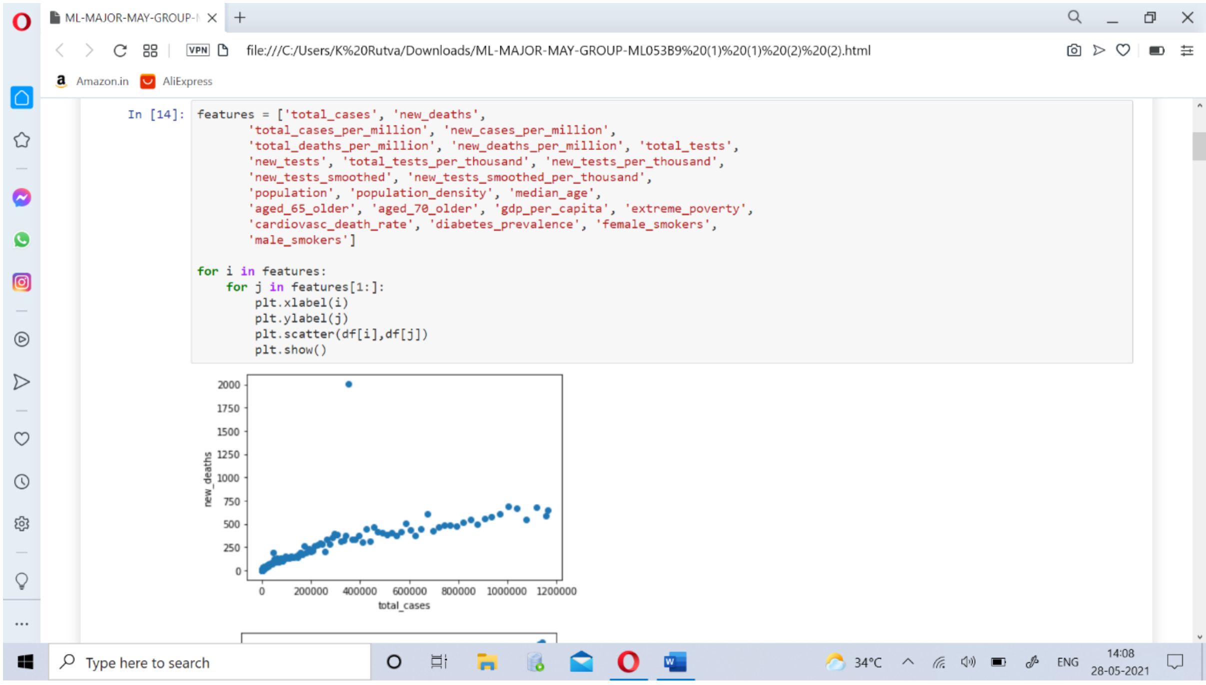
	total_cases	new_deaths	total_cases_per_million	new_cases_per_million	total_deaths_per_million	new_deaths_per_million	total_tests	new_test
count	2.040000e+02	204.000000	204.000000	204.000000	204.000000	204.000000	1.170000e+02	111.00000
mean	1.602750e+05	140.843137	116.140858	4.135338	3.272407	0.102044	4.005175e+06	123035.86486
std	2.801060e+05	227.965568	202.974758	6.610493	5.461240	0.165144	3.970491e+06	90998.68606
min	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	6.500000e+03	157.00000
25%	3.000000e+00	0.000000	0.002000	0.000000	0.000000	0.000000	5.005420e+05	48133.00000
50%	7.901500e+03	32.500000	5.725500	0.618000	0.185500	0.023500	2.719434e+06	108048.00000
75%	1.925778e+05	208.750000	139.548250	6.074250	3.946000	0.151000	6.616496e+06	178777.50000
max	1.164183e+06	2003.000000	843.608000	29.293000	20.820000	1.451000	1.379187e+07	361024.00000

8 rows × 24 columns

In [10]:
```
df.mean()              #finding mean for each column
```

Out[10]:
```
total_cases                        1.602750e+05
new_deaths                         1.408431e+02
total_cases_per_million            1.161409e+02
new_cases_per_million              4.135338e+00
total_deaths_per_million           3.272407e+00
new_deaths_per_million             1.020441e-01
total_tests                        4.005175e+06
new_tests                          1.230359e+05
total_tests_per_thousand           2.902325e+00
new_tests_per_thousand             8.915315e-02
new_tests_smoothed                 1.045399e+05
```

Figure 5.1.9



In [14]:
```
features = ['total_cases', 'new_deaths',
       'total_cases_per_million', 'new_cases_per_million',
       'total_deaths_per_million', 'new_deaths_per_million', 'total_tests',
       'new_tests', 'total_tests_per_thousand', 'new_tests_per_thousand',
       'new_tests_smoothed', 'new_tests_smoothed_per_thousand',
       'population', 'population_density', 'median_age',
       'aged_65_older', 'aged_70_older', 'gdp_per_capita', 'extreme_poverty',
       'cardiovasc_death_rate', 'diabetes_prevalence', 'female_smokers',
       'male_smokers']

for i in features:
    for j in features[1:]:
        plt.xlabel(i)
        plt.ylabel(j)
        plt.scatter(df[i],df[j])
        plt.show()
```


Figure 5.1.10

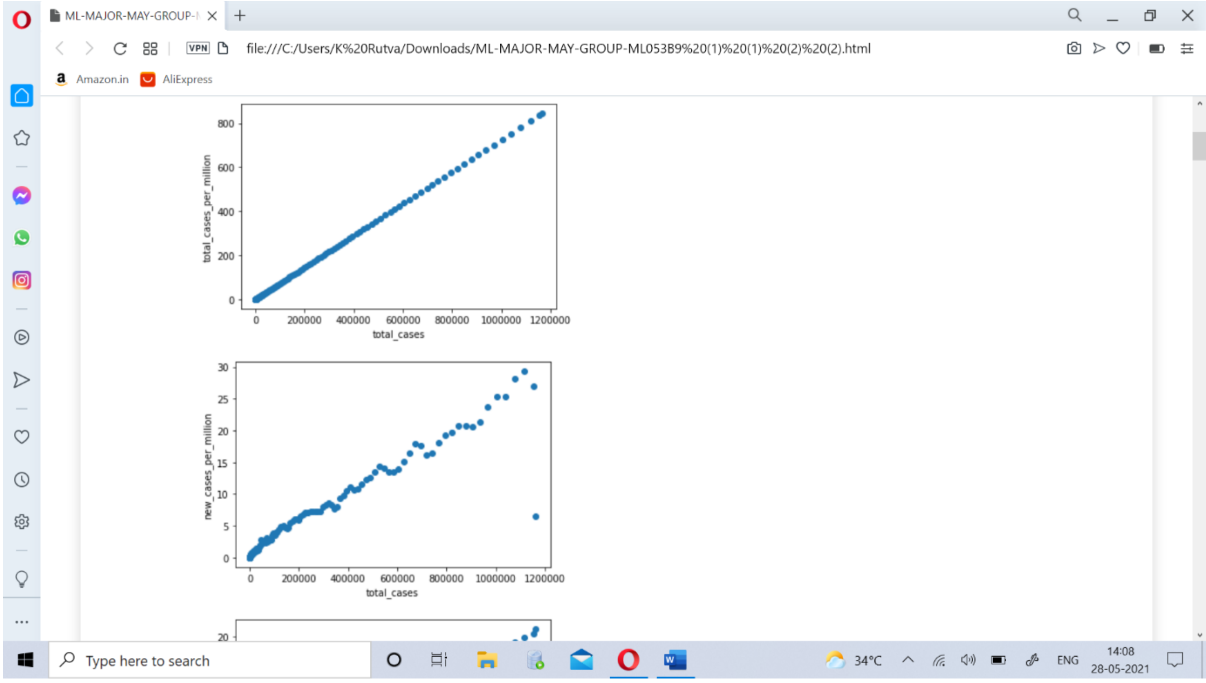


Figure 5.1.11

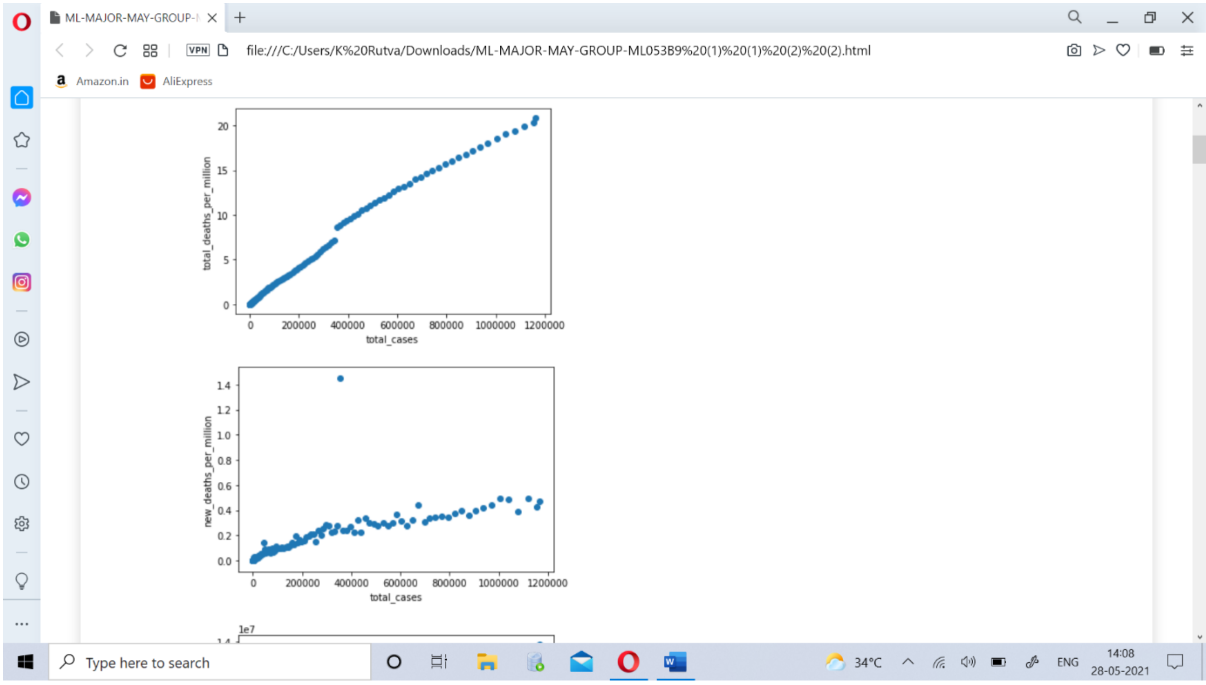


Figure 5.1.12

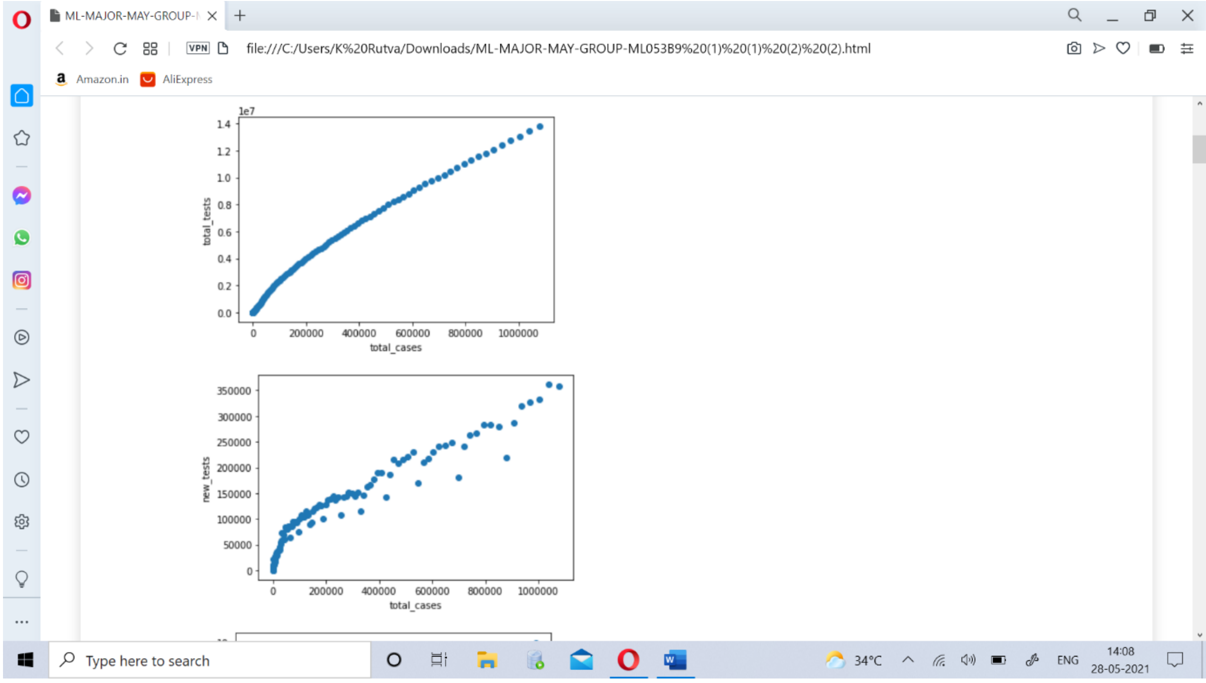


Figure 5.1.13

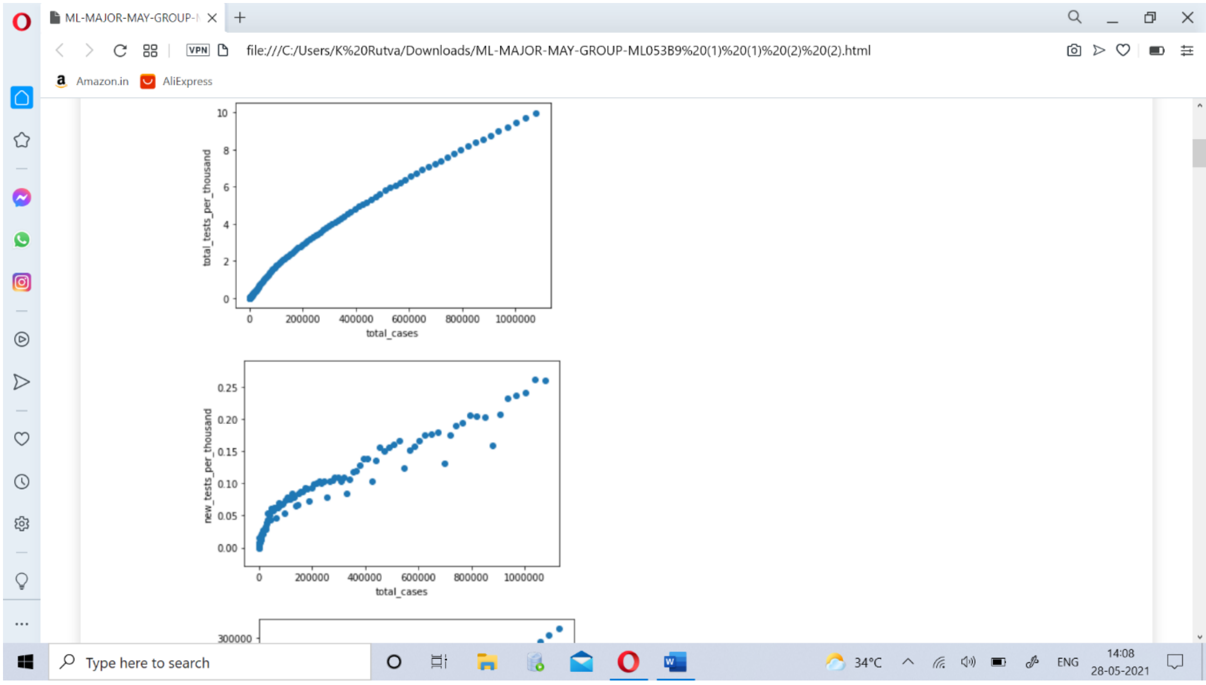


Figure 5.1.14

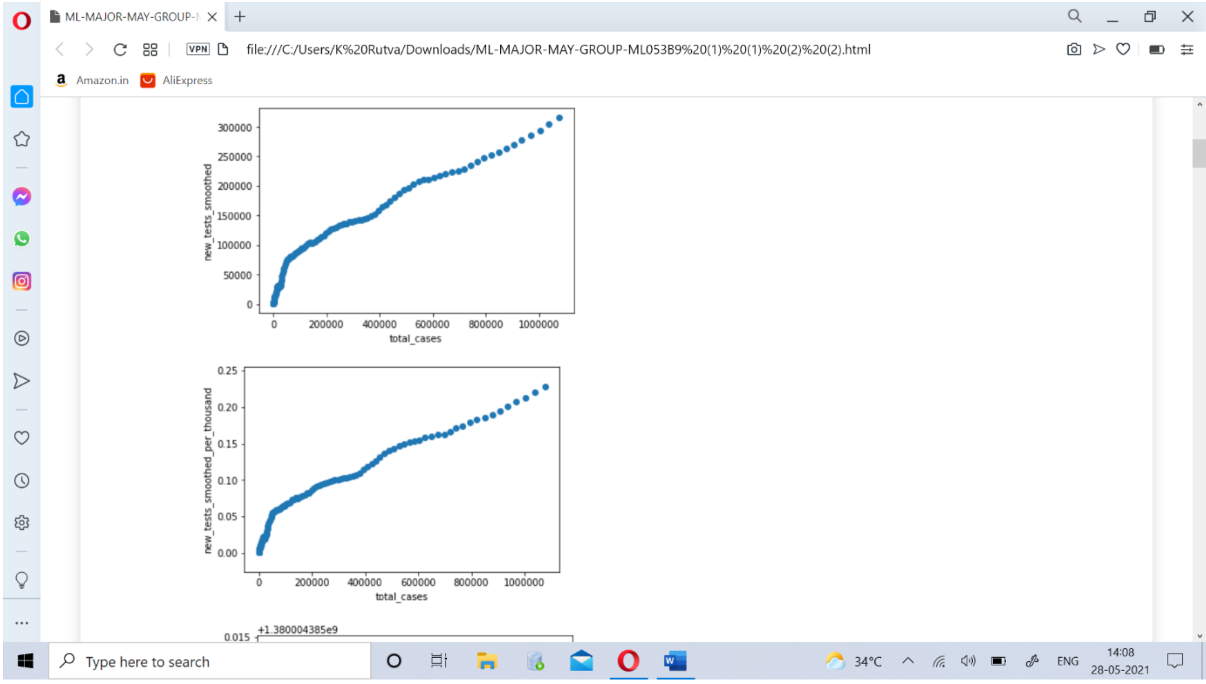


Figure 5.1.15

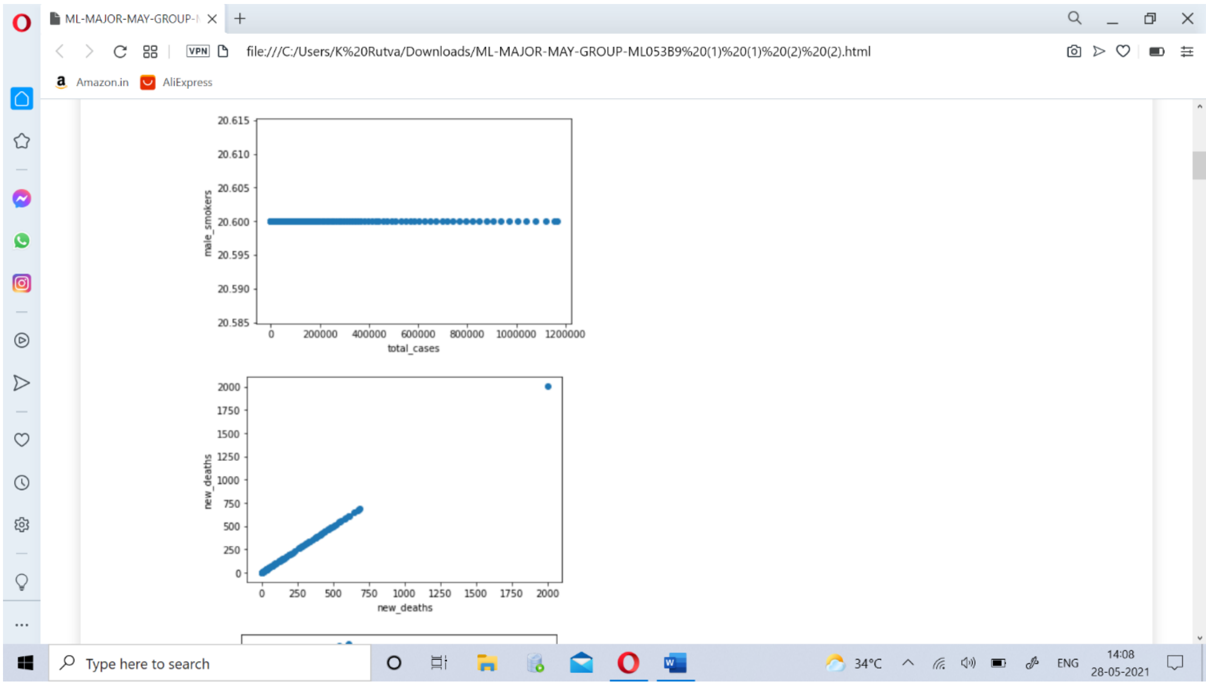


Figure 5.1.16

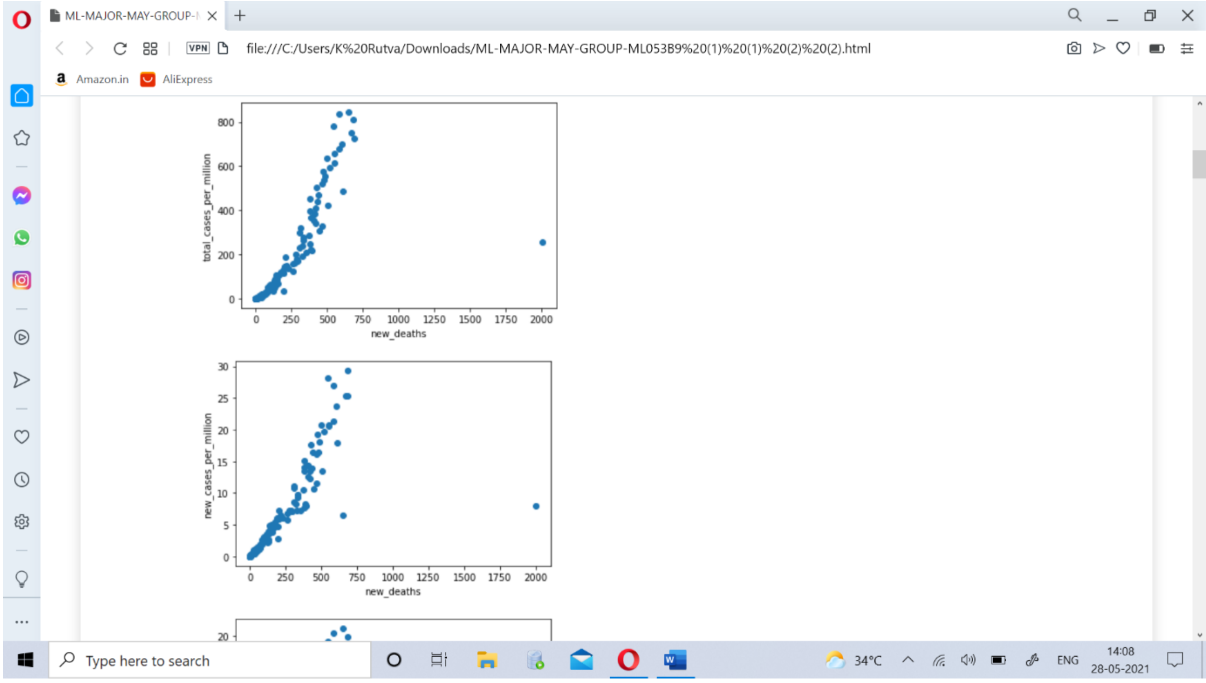


Figure 5.1.17

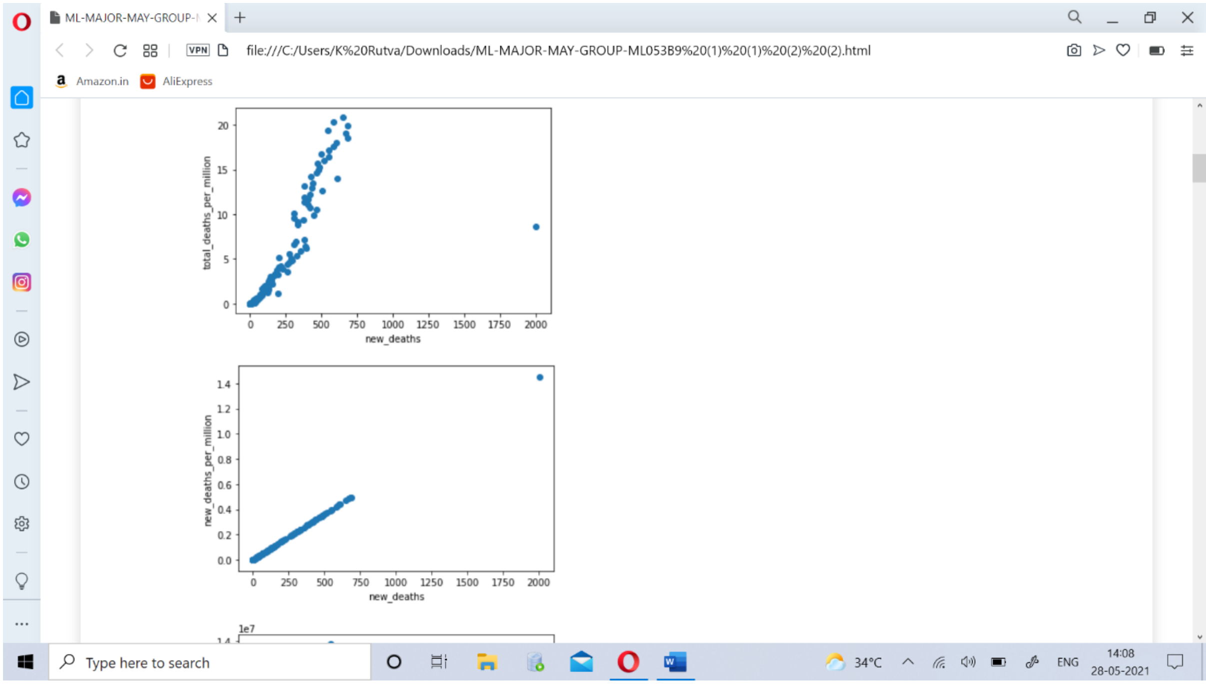


Figure 5.1.18

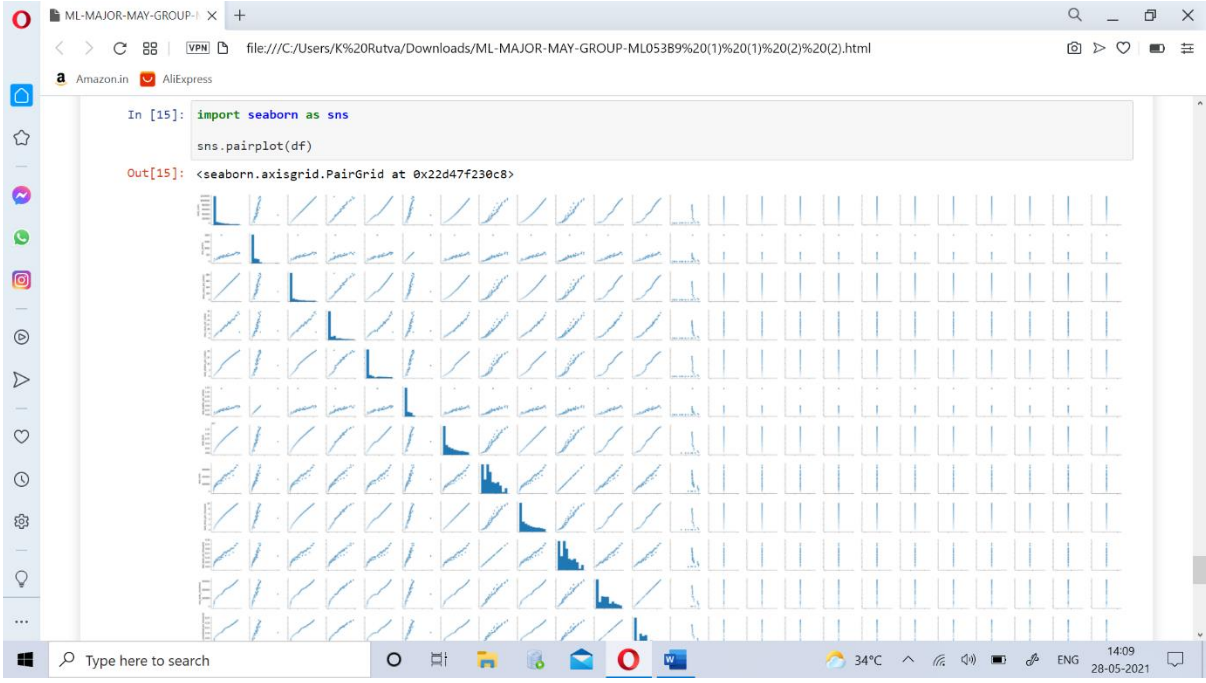


Figure 5.1.19

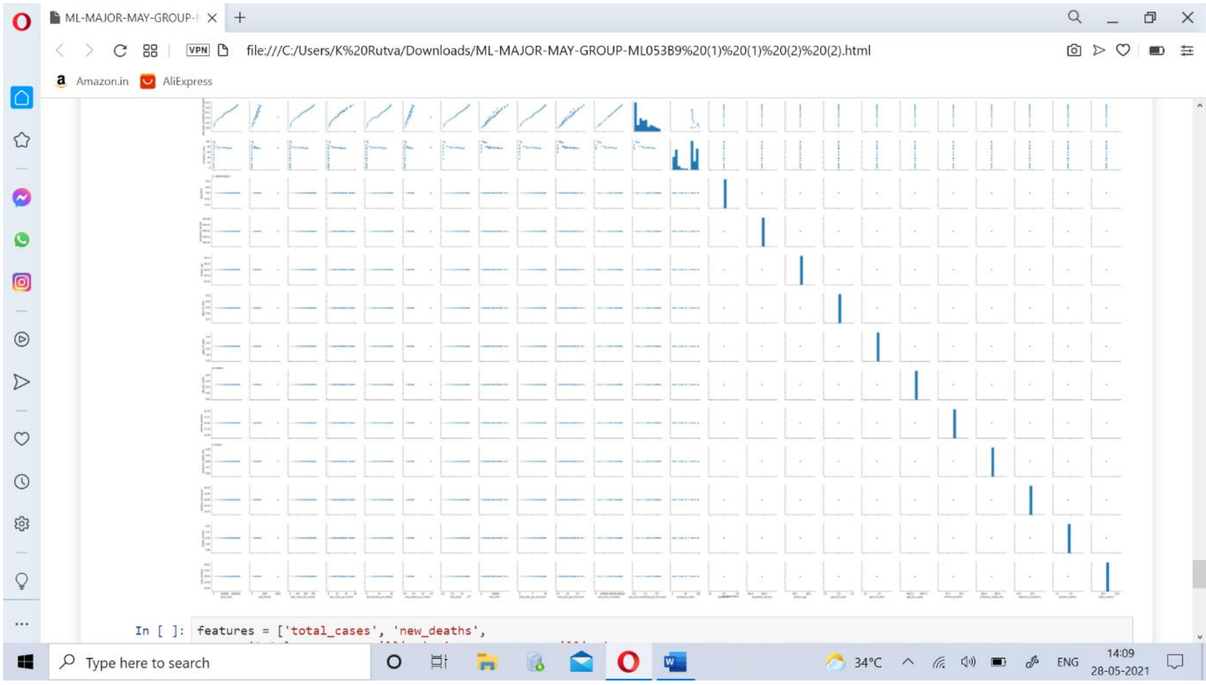


Figure 5.1.20

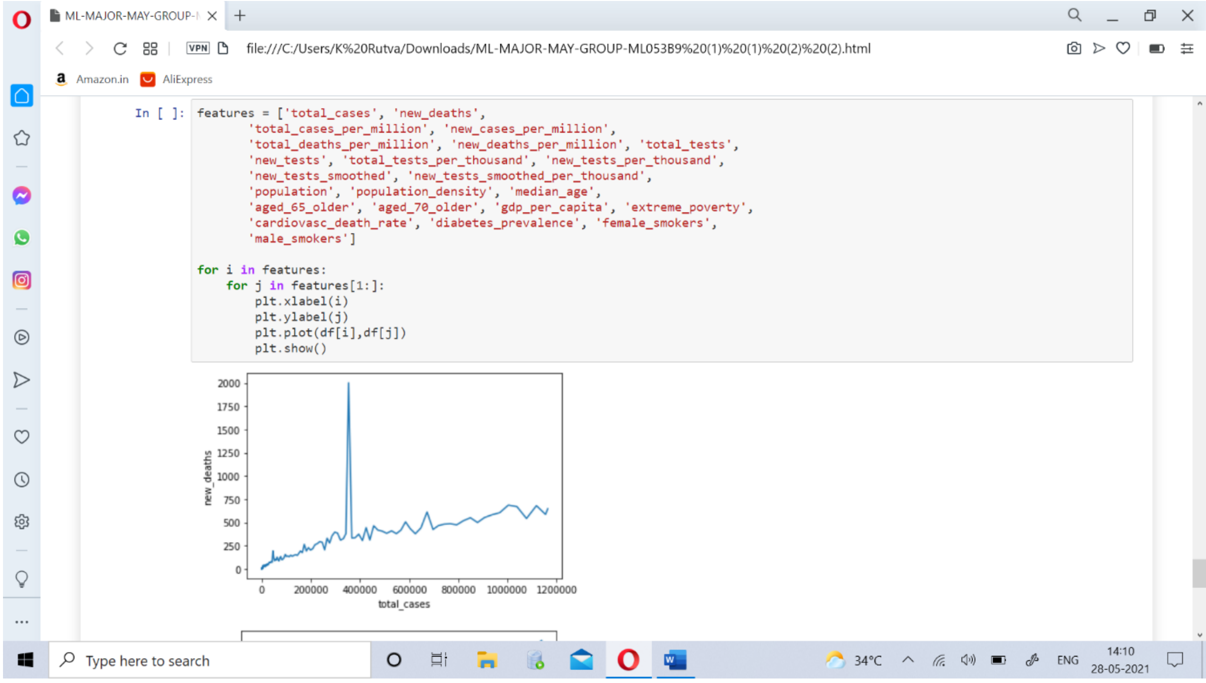


Figure 5.1.21

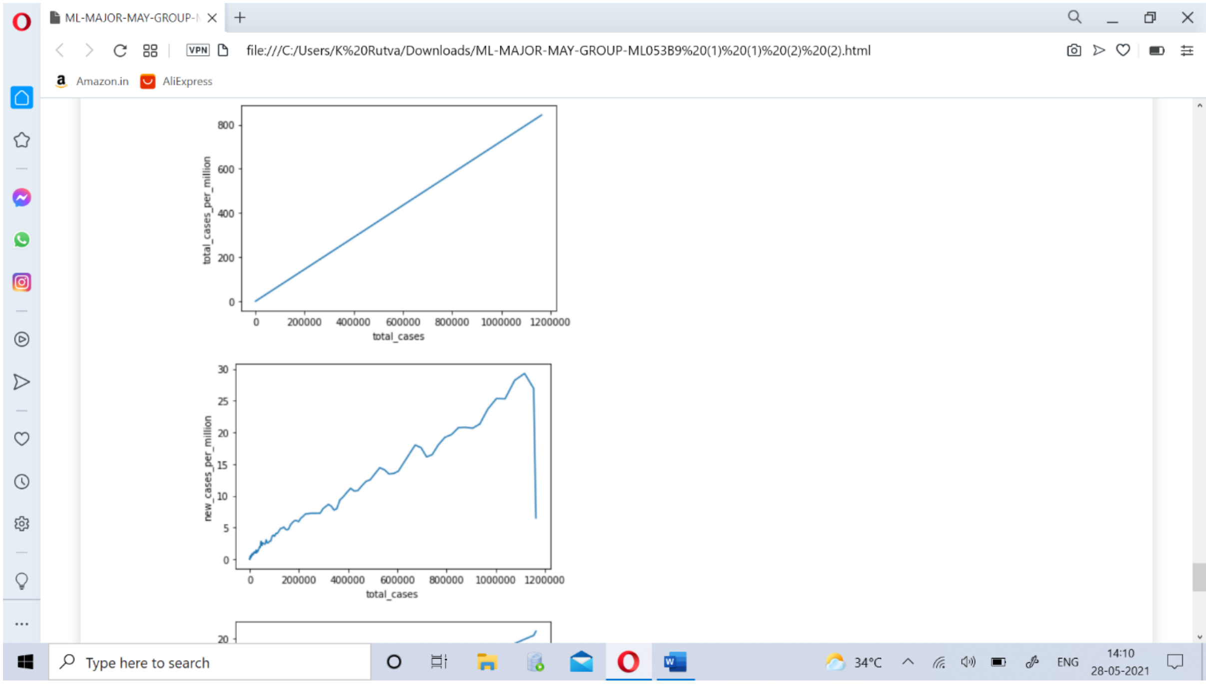


Figure 5.1.22

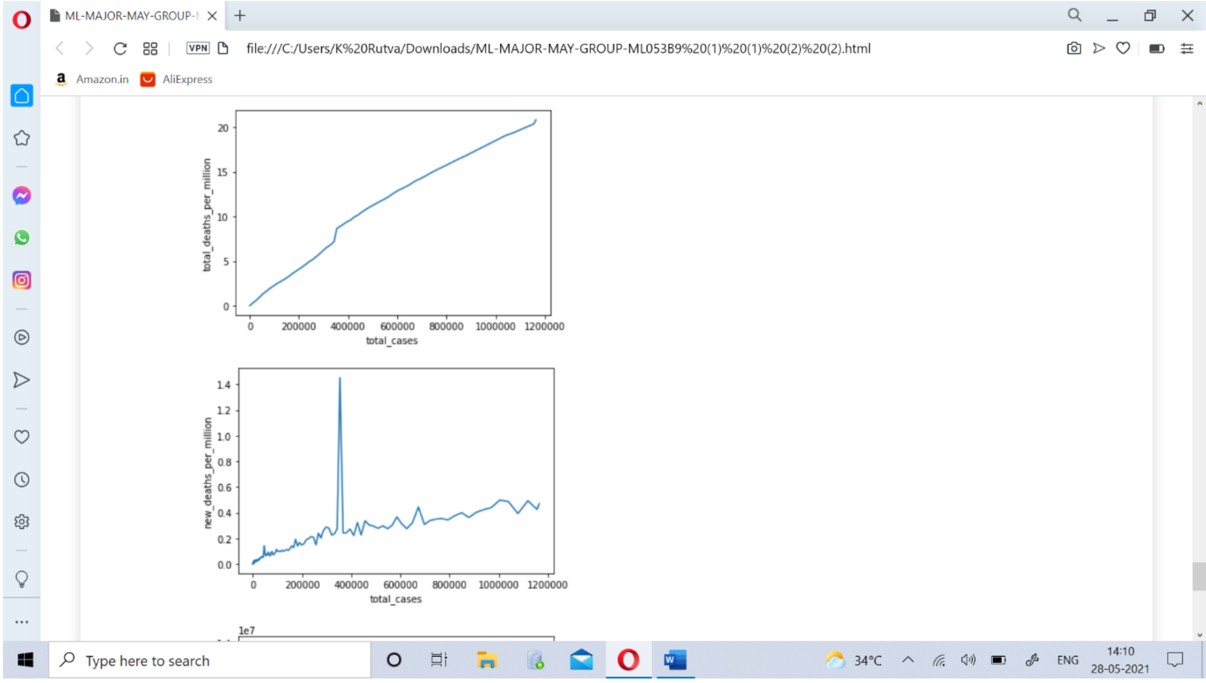


Figure 5.1.23

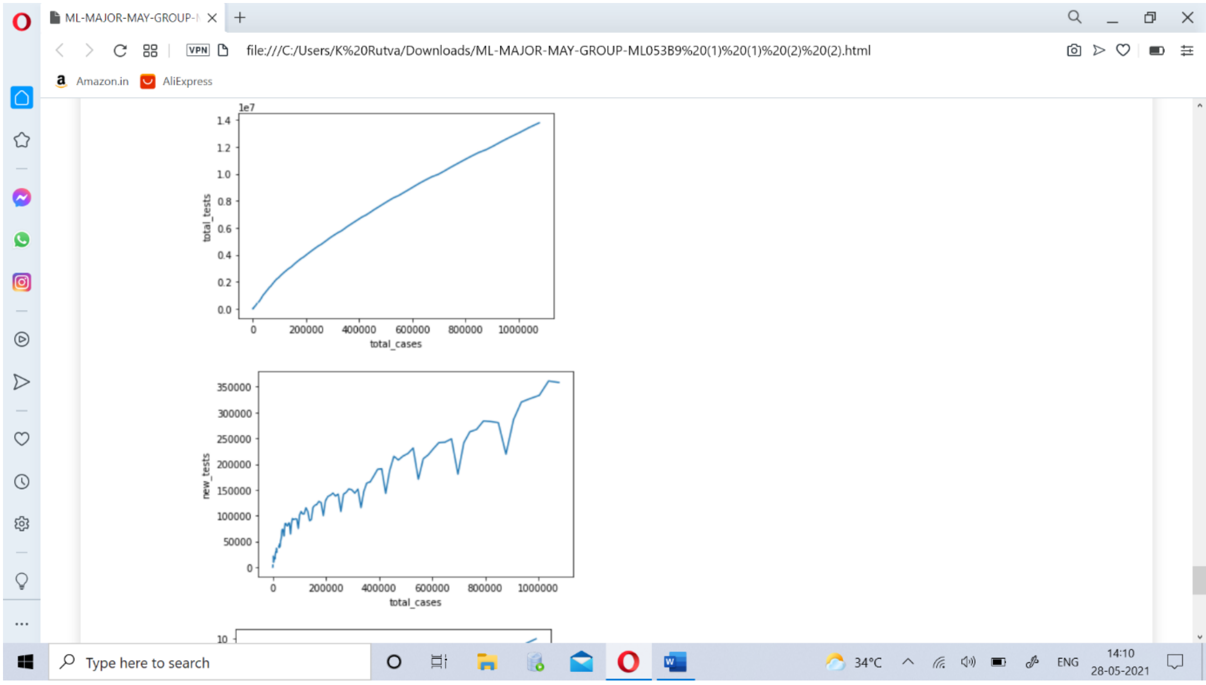


Figure 5.1.24

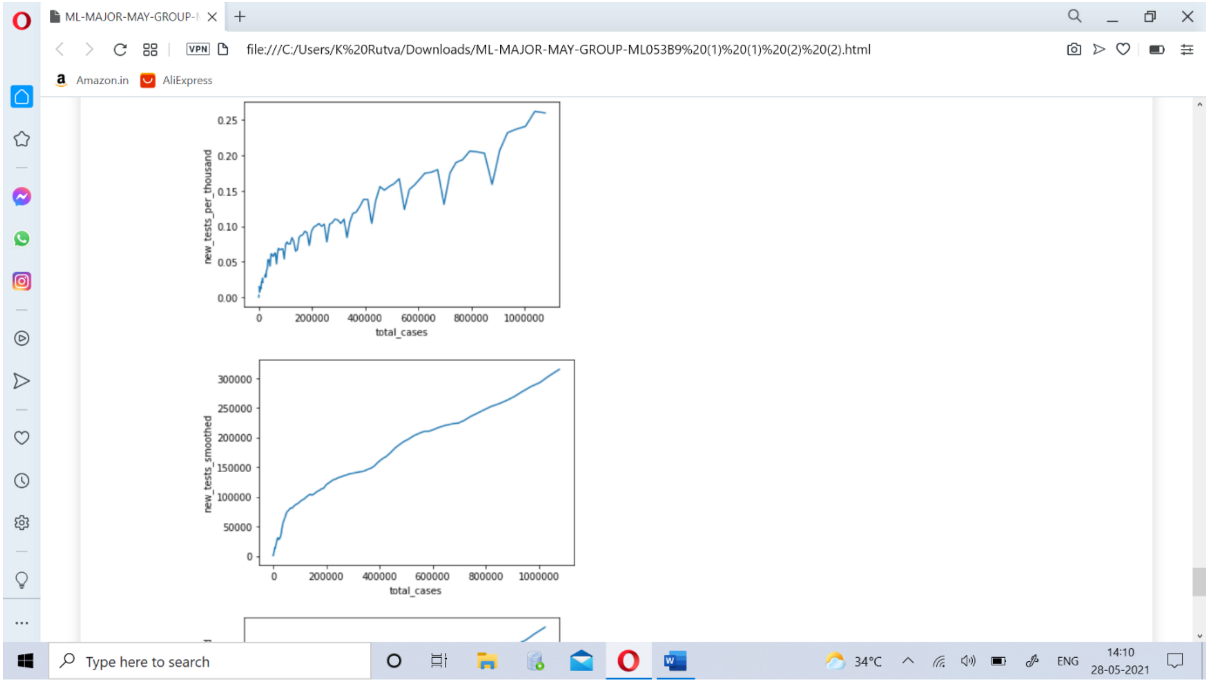


Figure 5.1.25

6. Conclusion and Discussion

Thus built a model of random forest , linear regression and multi-linear regression thus built a model with **99%** accuracy built a successful model

6.1 CODE AND IMPLEMENTATION


```
from sklearn.linear_model import LinearRegression
from sklearn.metrics      import  mean_squared_error
from sklearn.model_selection  import   train_test_split


# Create training and test sets
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size =.3, random_state=42)

# Create the regressor: reg_all
reg_all =LinearRegression()

# Fit the regressor to the training data
reg_all.fit(X_train, y_train)

# Predict on the test data: y_pred
y_pred = reg_all.predict( X_test)
```

```
# Compute and print R^2 and RMSE
print("R^2: {}".format(reg_all.score(X_test, y_test)))
rmse = np.sqrt(y_test,y_pred)
print("Root Mean Squared Error: {}".format(rmse))
```

```
R^2: 0.9999999999936375
```

Figure 6.1.1

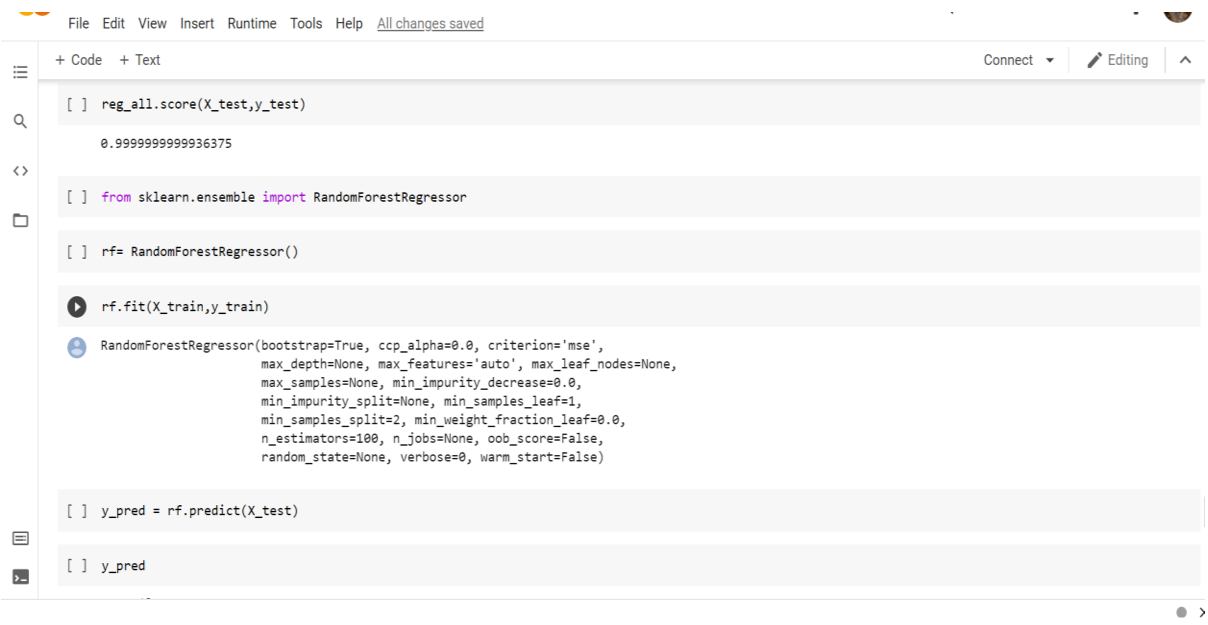


Figure 6.1.2

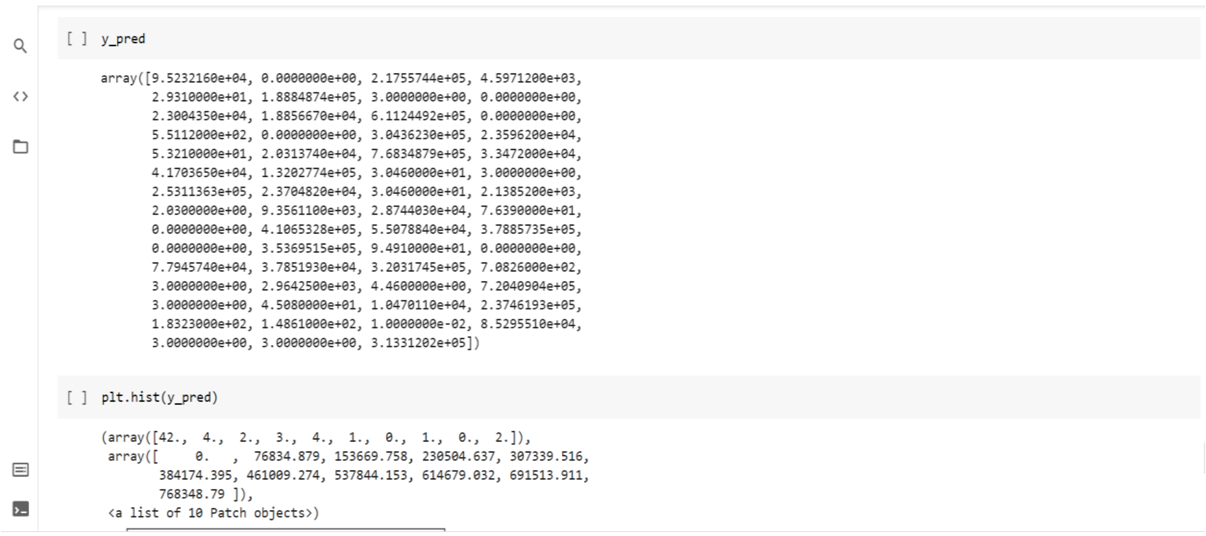
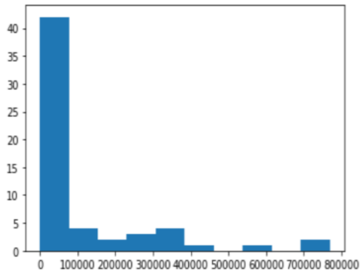


Figure 6.1.3

```
[ ] plt.hist(y_pred)

(array([42.,  4.,  2.,  3.,  4.,  1.,  0.,  1.,  0.,  2.]),
 array([     0.   ,  76834.879, 153669.758, 230504.637, 307339.516,
        384174.395, 461009.274, 537844.153, 614679.032, 691513.911,
        768348.79 ]),
 <a list of 10 Patch objects>)
```




```
[ ] rf.score(X_test,y_test)

0.9986691266667814
```

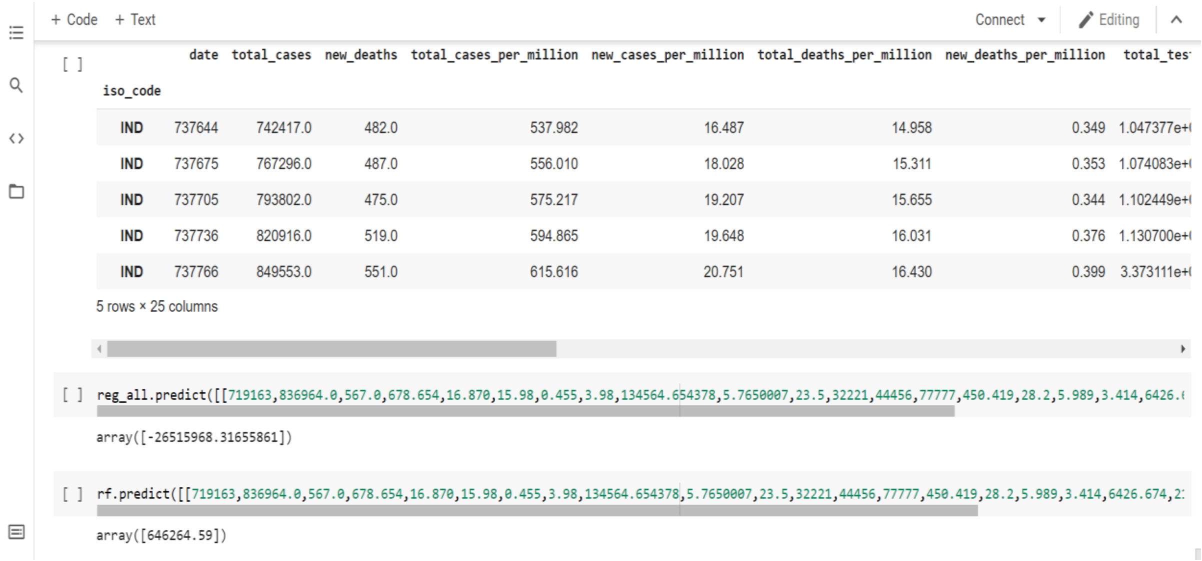
Figure 6.1.4

+ Code + Text Connect Editing

```
[ ] new_date = pd.date_range('15/08/2020',periods = 1,freq = 'H')
    new_date

DatetimeIndex(['2020-08-15'], dtype='datetime64[ns]', freq='H')

[ ] import datetime as dt
    new_date=pd.to_datetime(new_date)
    new_date = new_date.map(dt.datetime.toordinal)
    new_date

Int64Index([737652], dtype='int64')

[ ] df.shape

(194, 25)

[ ] df.tail()
```

	date	total_cases	new_deaths	total_cases_per_million	new_cases_per_million	total_deaths_per_million	new_deaths_per_million	total_tes
iso_code								
IND	737644	742417.0	482.0	537.982	16.487	14.958	0.349	1.047377e+

Figure 6.1.5



+ Code + Text   Connect   Editing

iso_code	date	total_cases	new_deaths	total_cases_per_million	new_cases_per_million	total_deaths_per_million	new_deaths_per_million	total_tes
IND	737644	742417.0	482.0	537.982	16.487	14.958	0.349	1.047377e+
IND	737675	767296.0	487.0	556.010	18.028	15.311	0.353	1.074083e+
IND	737705	793802.0	475.0	575.217	19.207	15.655	0.344	1.102449e+
IND	737736	820916.0	519.0	594.865	19.648	16.031	0.376	1.130700e+
IND	737766	849553.0	551.0	615.616	20.751	16.430	0.399	3.373111e+

5 rows × 25 columns

```
reg_all.predict([[719163,836964.0,567.0,678.654,16.870,15.98,0.455,3.98,134564.654378,5.7650007,23.5,32221,44456,77777,450.419,28.2,5.989,3.414,6426.
array([-26515968.31655861])
```

```
rf.predict([[719163,836964.0,567.0,678.654,16.870,15.98,0.455,3.98,134564.654378,5.7650007,23.5,32221,44456,77777,450.419,28.2,5.989,3.414,6426.674,2
array([646264.59])
```

Figure 6.1.6

7. FUTURE WORK/ SCOPE:

It can be used only on jupyter notebook now. But, in future we will develop more like creating a web page and App. We can also explore/include more different type of models for better performance and understanding of the models.