**DR B.R. AMBEDKAR NATIONAL INSTITUTE**

**OF TECHNOLOGY, JALANDHAR**

****

**SOFTWARE ENGINEERING**

**CONCEPTS**

**ITPC-323**

**Submitted to: Submitted by:**

Dr. Nonita Sharma Bindheshwari Singh (18124009)

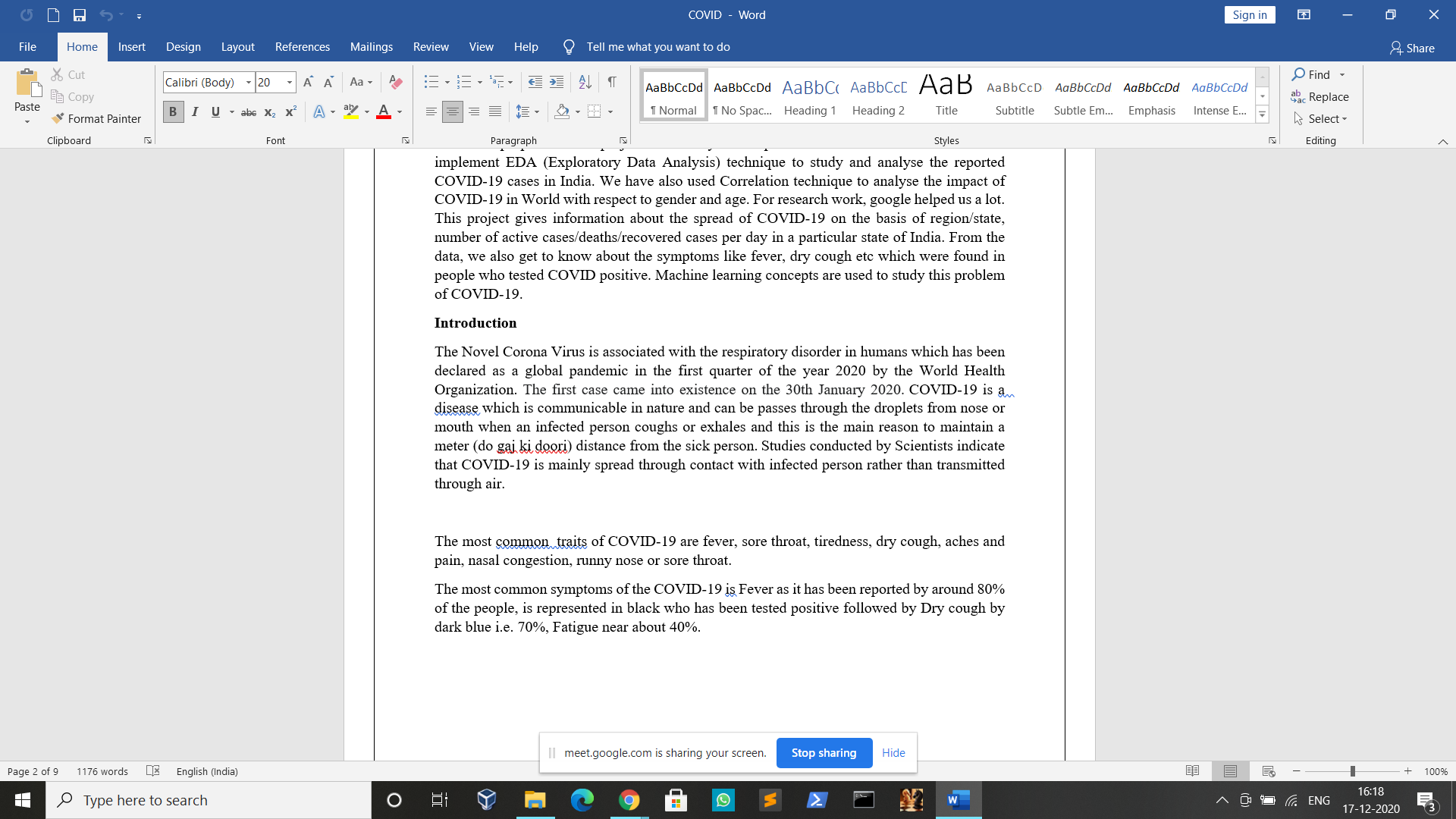
Assistant Professor Jasleen Kaur (18124014)

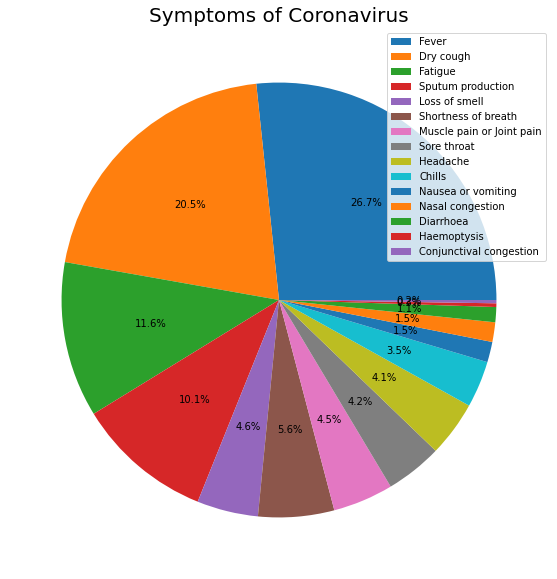
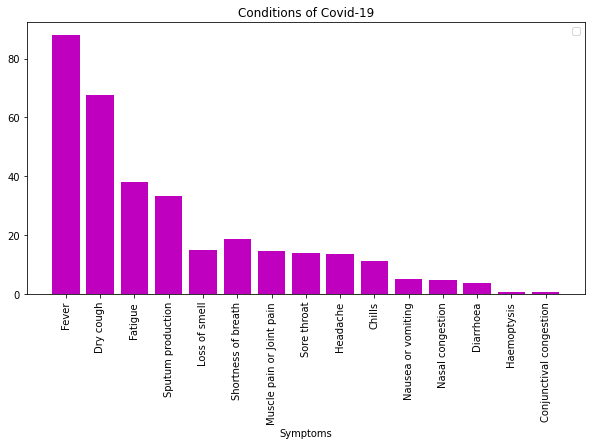
Nandini Verma (18124019)

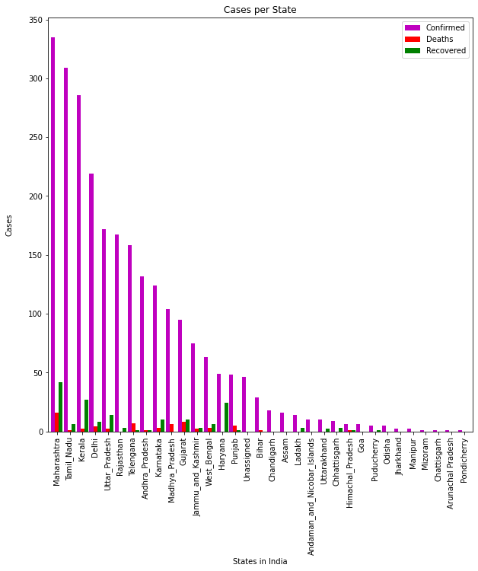
**COVID-19 Spread Analysis (India)**

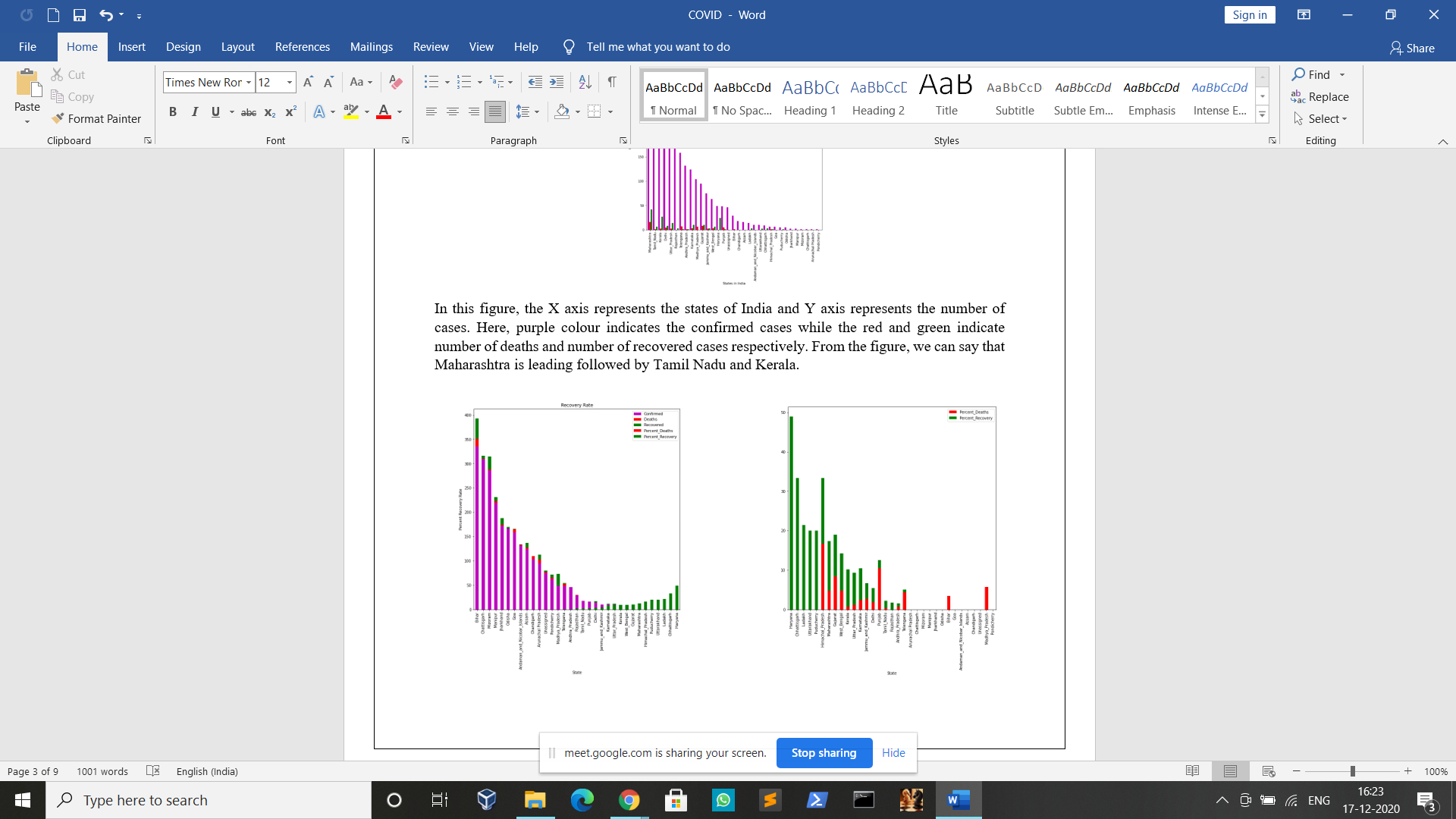
**Abstract**

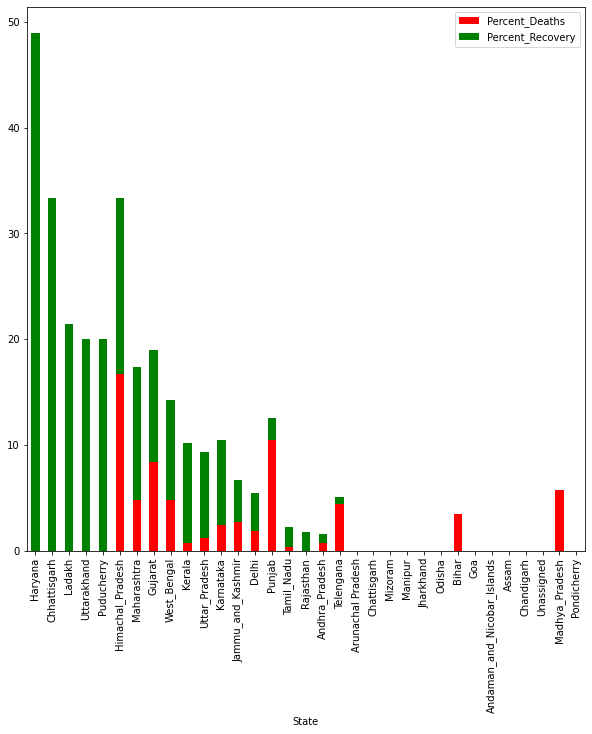
The main purpose of this project is to analyse the spread of covid-19 in India. In this we implement EDA (Exploratory Data Analysis) technique to study and analyse the reported COVID-19 cases in India. We have also used Correlation technique to analyse the impact of COVID-19 in World with respect to gender and age. For research work, google helped us a lot. This project gives information about the spread of COVID-19 on the basis of region/state, number of active cases/deaths/recovered cases per day in a particular state of India. From the data, we also get to know about the symptoms like fever, dry cough etc which were found in people who tested COVID positive. Machine learning concepts are used to study this problem of COVID-19.

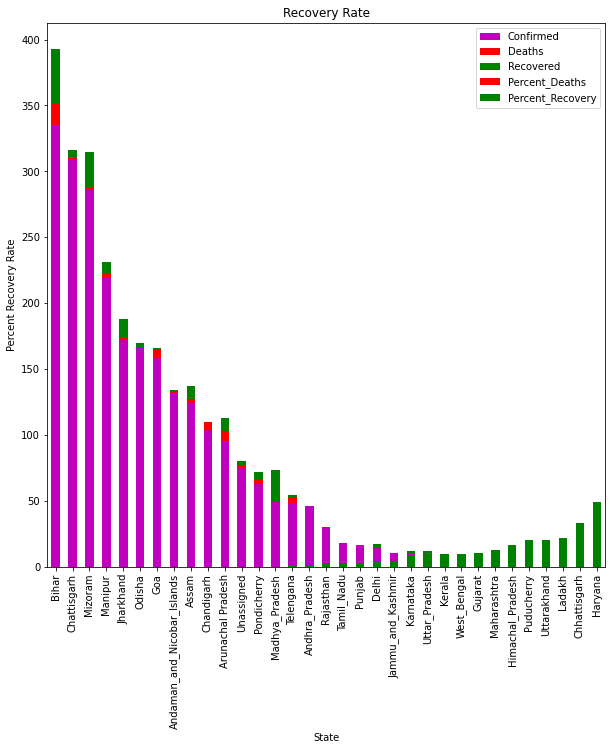
**Introduction**

****



****

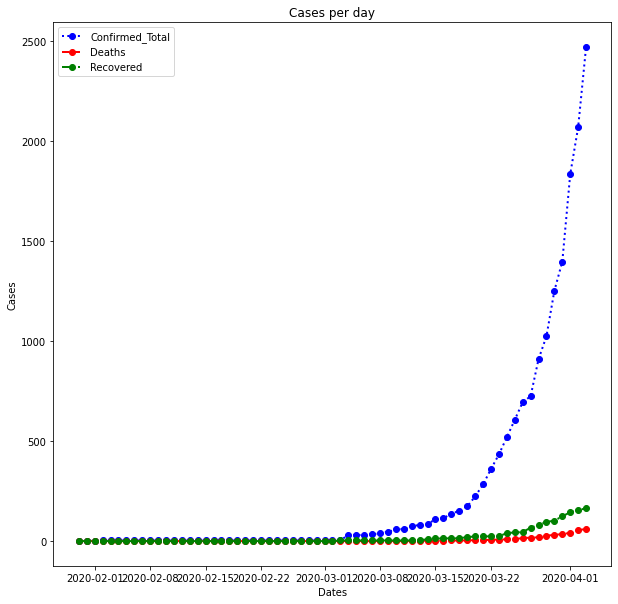
****



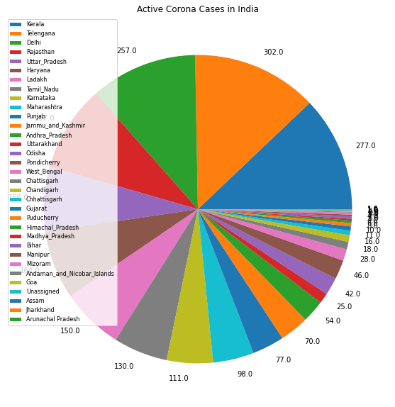
1. (b)

In almost every state the recovery rate is more than rate of increment of cases per day. From (a), in many states such as Karnataka, West Bengal there is no increment in cases so they have successfully recovered from covid-19.

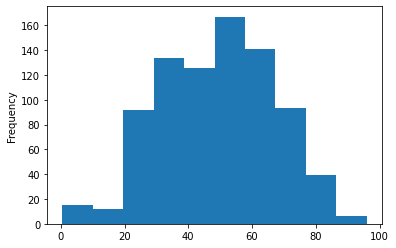
In (b), the overall recovery rate is high in Haryana. Here percentage recovering and percentage deaths in some states like Mizoram, Manipur, Jharkhand etc are nil.



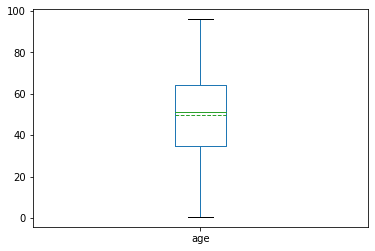
This figure shows the confirmed cases , number of deaths and number of recoveries of COVID-19 per day.

****

This graph shows the active cases of COVID-19 in India on the basis of state.

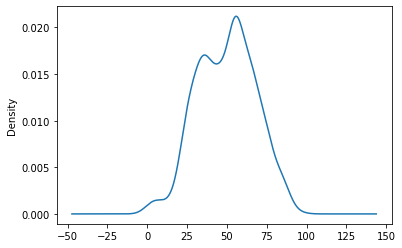


This histogram depicts the age wise impact of COVID-19 in the world. From this, we can conclude that most of the people under the range of 20 to 80 are tested positive for this Novel Coronavirus.

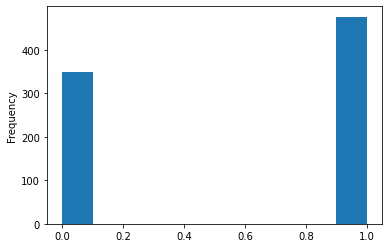


This figure represents the boxplot. Boxplot is a simple method to depict data which is statistical in nature on a plot in which rectangle is drawn to show second and third quartiles. The lower and upper quartiles are horizontal lines on either side of rectangle.

Here the green lines represents mean and median are coinciding from which we can conclude that our distribution of data is normal distribution, that is , there are no or least number of outliers and least or null skewness of the data.

****

From this Kernel distribution graph, we can conclude that old people are impacted more by this Novel Coronavirus.



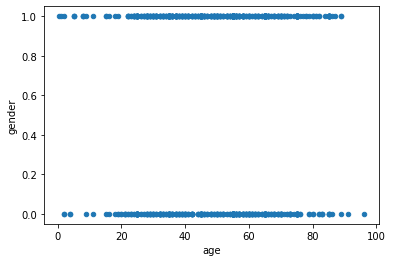
This histogram shows the number of cases of COVID-19 in World with respect to gender.

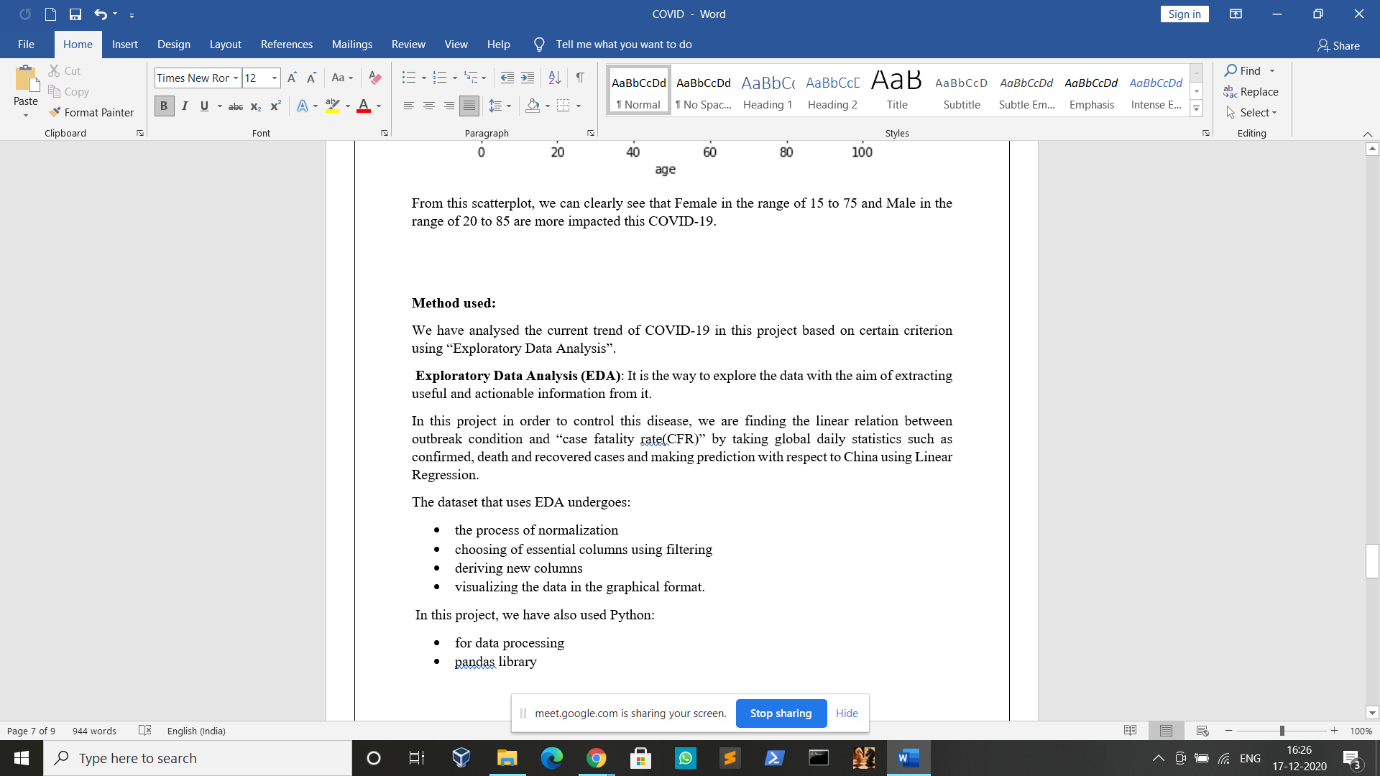
Here 0 represents Female and 1 represents Male.

From this, We can conclude that the Males are more prone to this Novel Coronavirus (COVID-19) than Females.

Using this data, we have found the correlation between age and gender on the dataset of World.

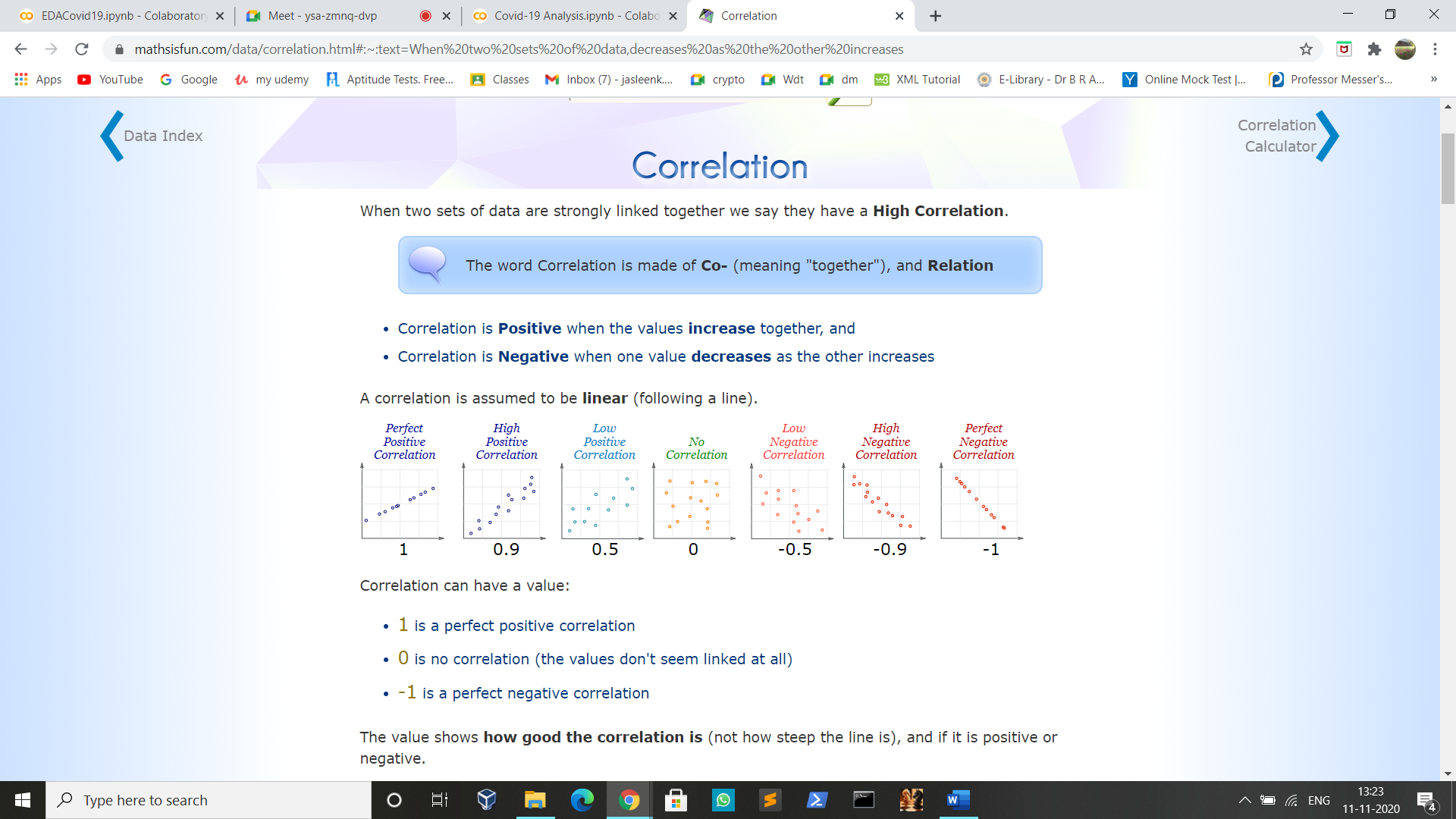
The scatterplot for the same is shown below: (0 is for Female and 1 is for Male)

****

From this scatterplot, we can clearly see that Female in the range of 15 to 75 and Male in the range of 20 to 85 are more impacted this COVID-19.

* Graphs that we created for better visualization are the results of Matplotlib and sklearn library of the Python

**Correlation:**



For correlation, the libraries that we used are:

* Scipy.stats
* Seaborn

**Result:**

From this project, we get to know that if a person belongs to the region where COVID-19 cases are more and he/she is showing the symptoms then we can say that he/she is likely to be prone to COVID-19.

**Motivation:**

In order to track the spread of this disease on the basis of state so that safe travelling can be ensured and to check what are the various symptoms associated with this pandemic and how this virus is leaving its impact on a particular gender we have decided to make this project.

**Tools:**

* Jupyter notebook
* Collaboratory
* Jira Software

**Dataset:**

We have used the csv file:

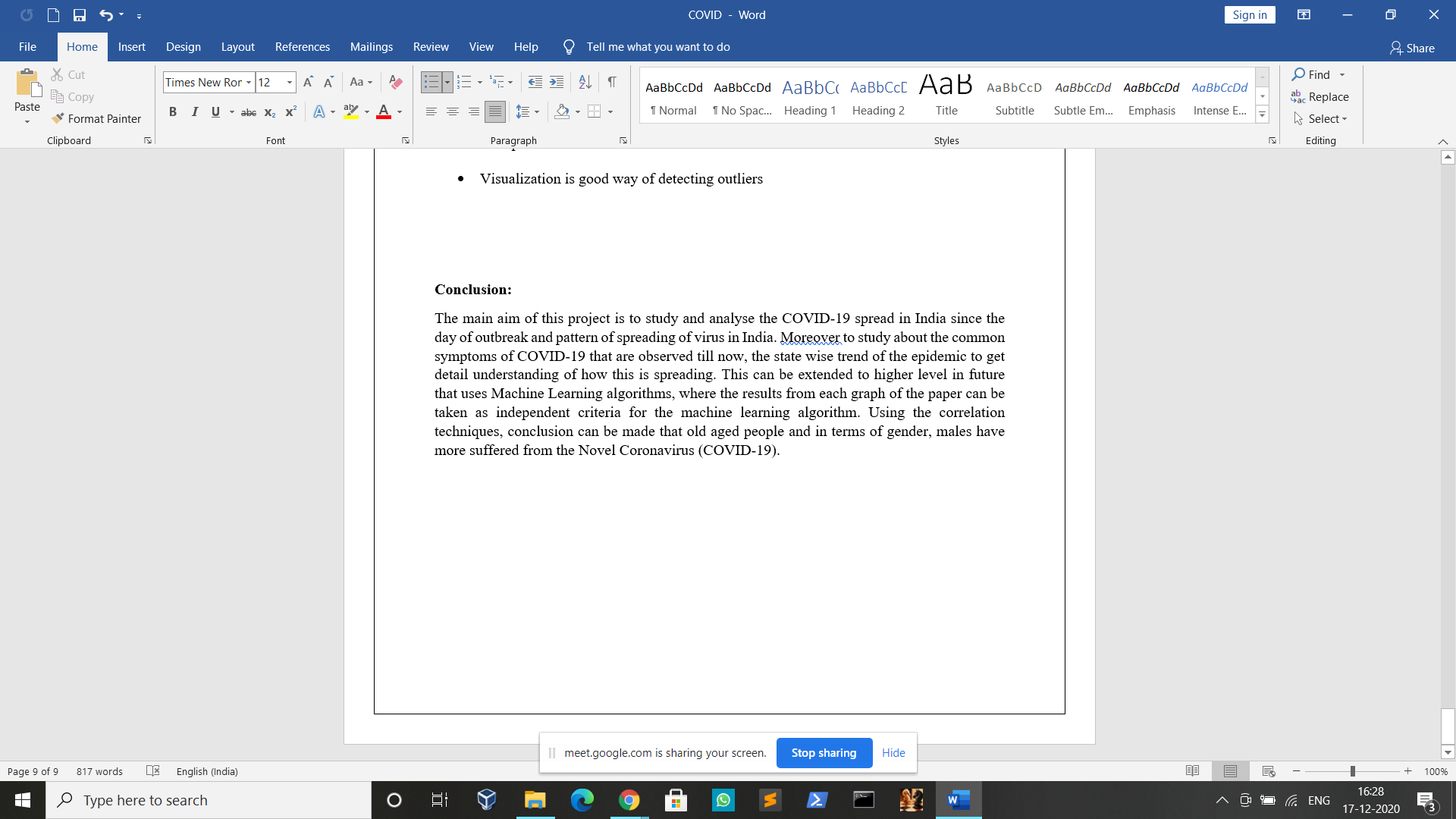
<https://drive.google.com/file/d/11CIKZkzu65Euhb0JYXoOMfvfD6rh8vtG/view?usp=sharing>

<https://drive.google.com/file/d/1r4XMiJZaDIB_cDYdSxbaoDFuw81G_leS/view?usp=sharing>

**Other methods:**

We can also use rmse (root mean square error) in place of linear regression.

But EDA technique has some advantages over other methods. Hence we decided to use this technique.

* It gives us insights about the data
* It helps us for feature selection
* ****Visualization is good way of detecting outliers