INTRODUCTION

Coronaviruses are a group of viruses that cause disease in mammals and birds. In human coronaviruses cause respiratory tract infections that are typically mild, such as common cold .Coronaviruses are transmitted between animals and humans. . The most common symptoms of COVID-19 are fever, tiredness, and dry cough. Some patients may have aches and pains, nasal congestion, runny nose, sore throat or diarrhea. These symptoms are usually mild and begin gradually (WHO, 2019).

In this project, I will use Johns World Health Organisation dataset to achieve the following:

1. Data Preprocessing
2. Preforming exploratory data analysis on the world data using Plotly, Matplotlib ,and Seaborn
3. Visualizing the geospatial data concerning china using Folium and cufflinks
4. Exploring the Venues of the locations that has high number of cases (confirmed, recovered, deaths).

A description of the data and how it will be used to solve the problem.

DATA

As mentioned above, the data used is World Health Organisation dataset which contain the following fields:

1. Country/Region
2. Province/State
3. Latitude
4. Longitude
5. Latitude
6. Confirmed: Number of Confirmed Cases
7. Recovered: Number of Recovered Cases
8. Date:Date of the report

This is the data repository for the 2019 Novel Coronavirus Visual Dashboard operated by the World Health Orgnaisation.

# References

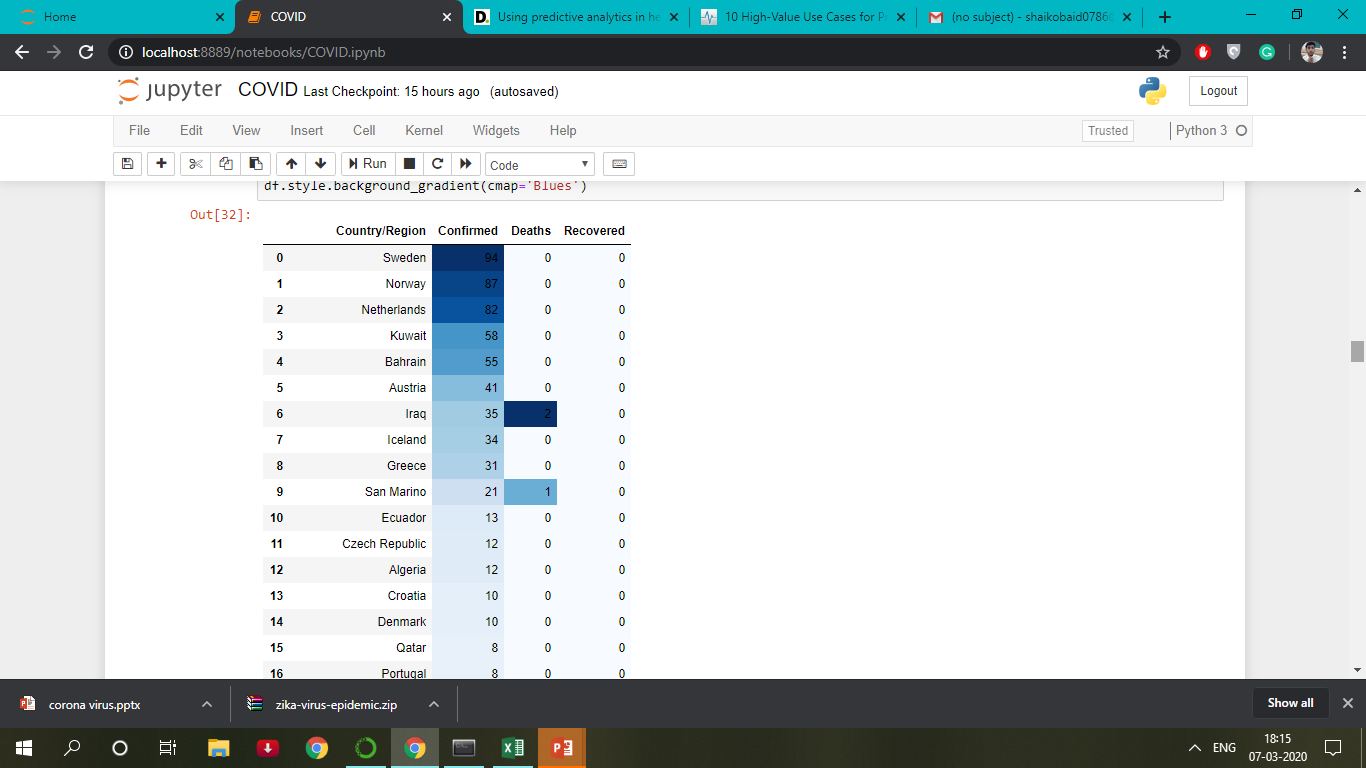
**WHO. 2019.** Coronavirus. *World Health Organization.* [Online] 2019. https://www.who.int/health-topics/coronavirus.

METHODOLOGY

This section is divided into four parts which are :Data processing,Exploratory data analysis(EDA),geospatial data visualization and Linear Regression model to forecast the recovered cases.

I.Data Processing

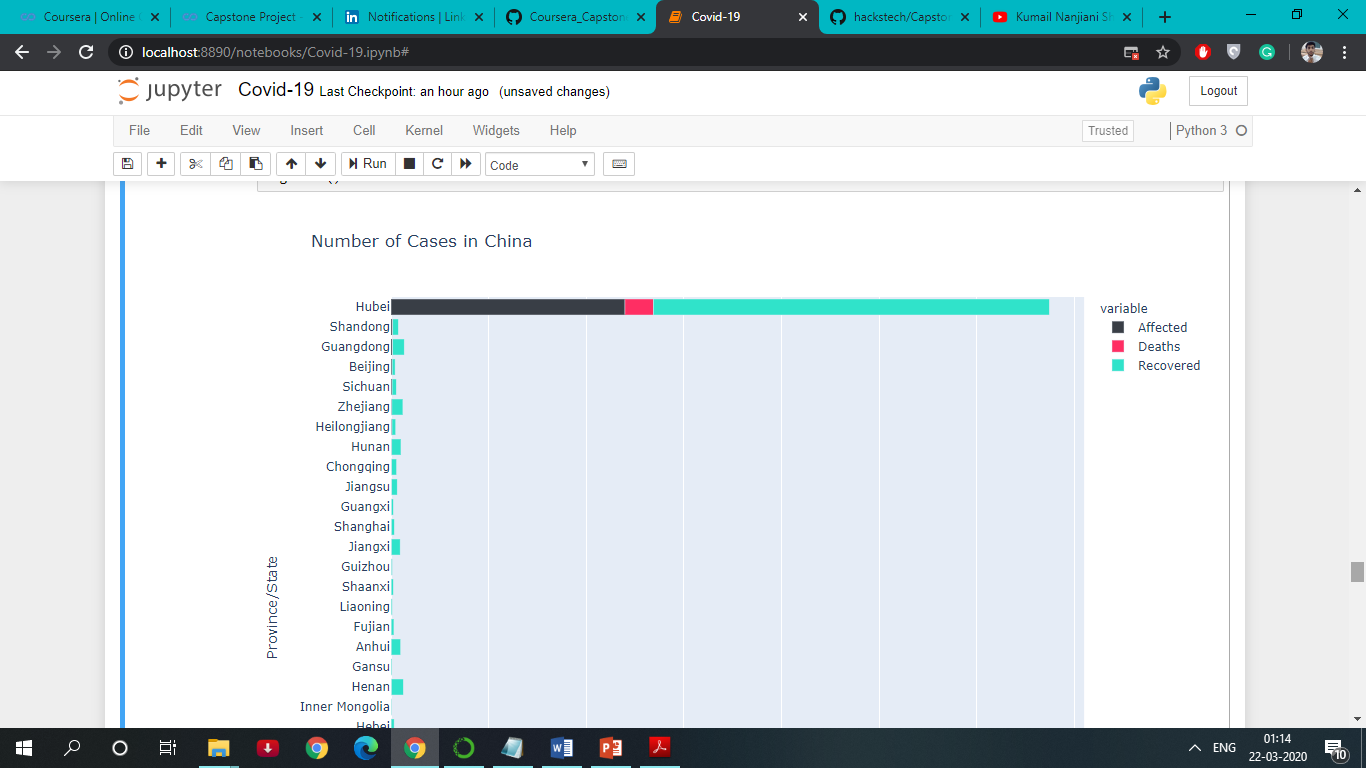
At this stage, I check the missing data.



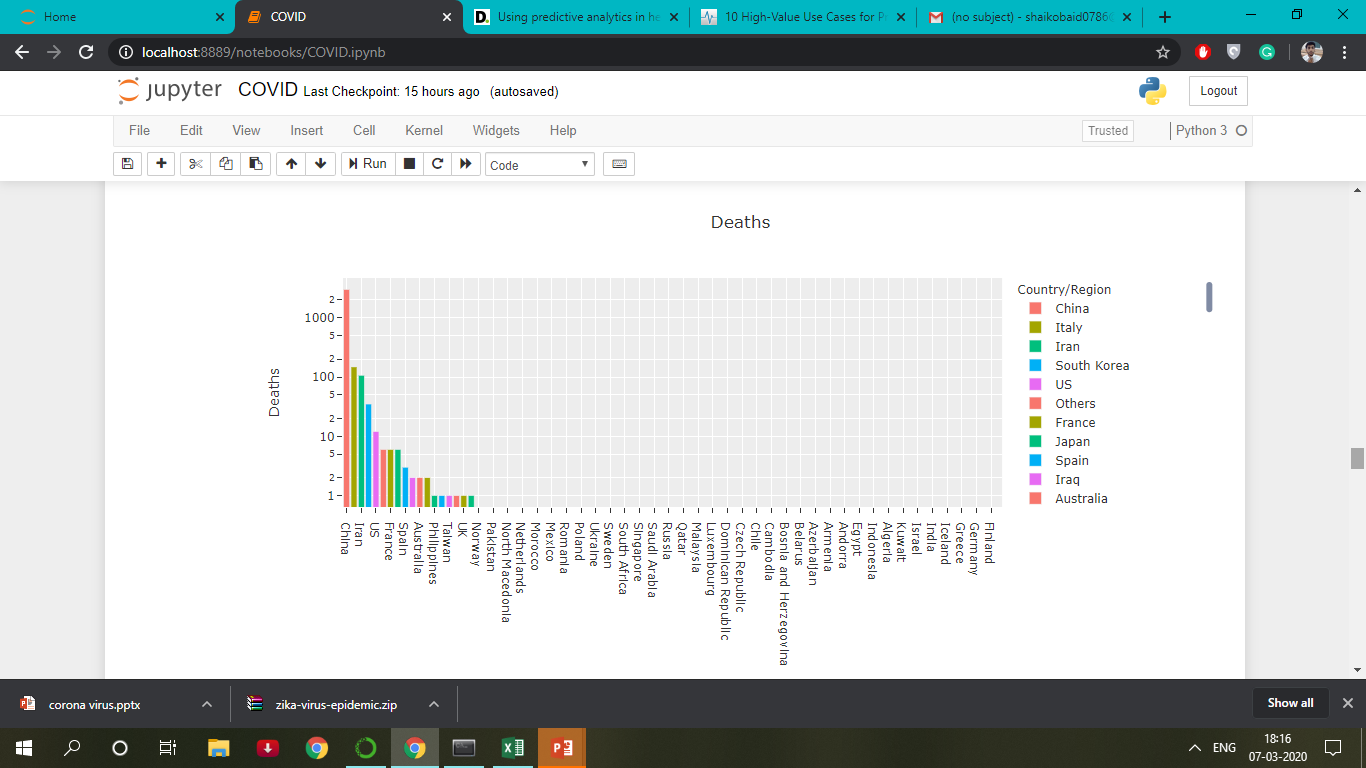
If the Provinces is null ,change the name of the corresponding country/region.Drop the row with missing latitude and longitude.The missing in columns (Confirmed,Recovered,Deaths) were replaced by 0.

II Exploratory Data Analysis(EDA)

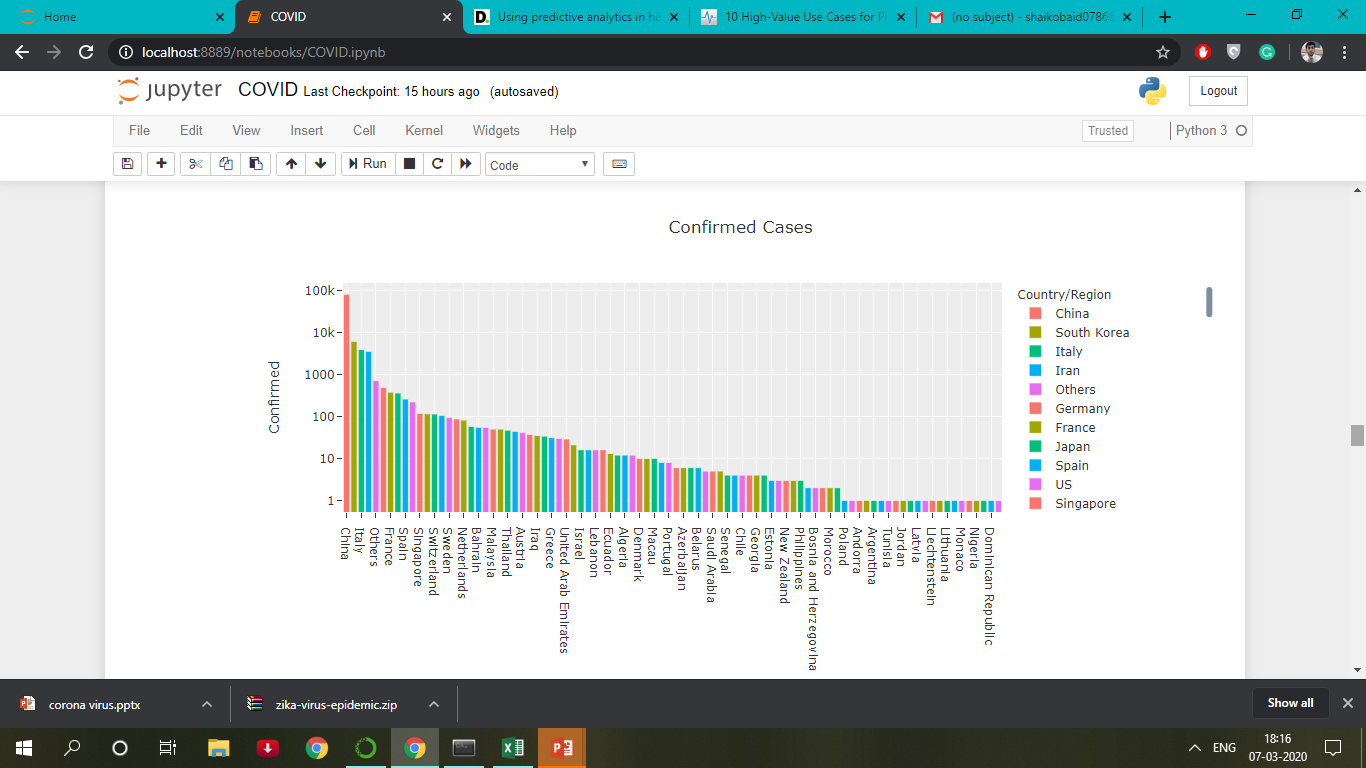
Visualizing the Confirmed, Recovered and Death cases in the provinces of China.



Visualizing the Death Cases Worldwide.

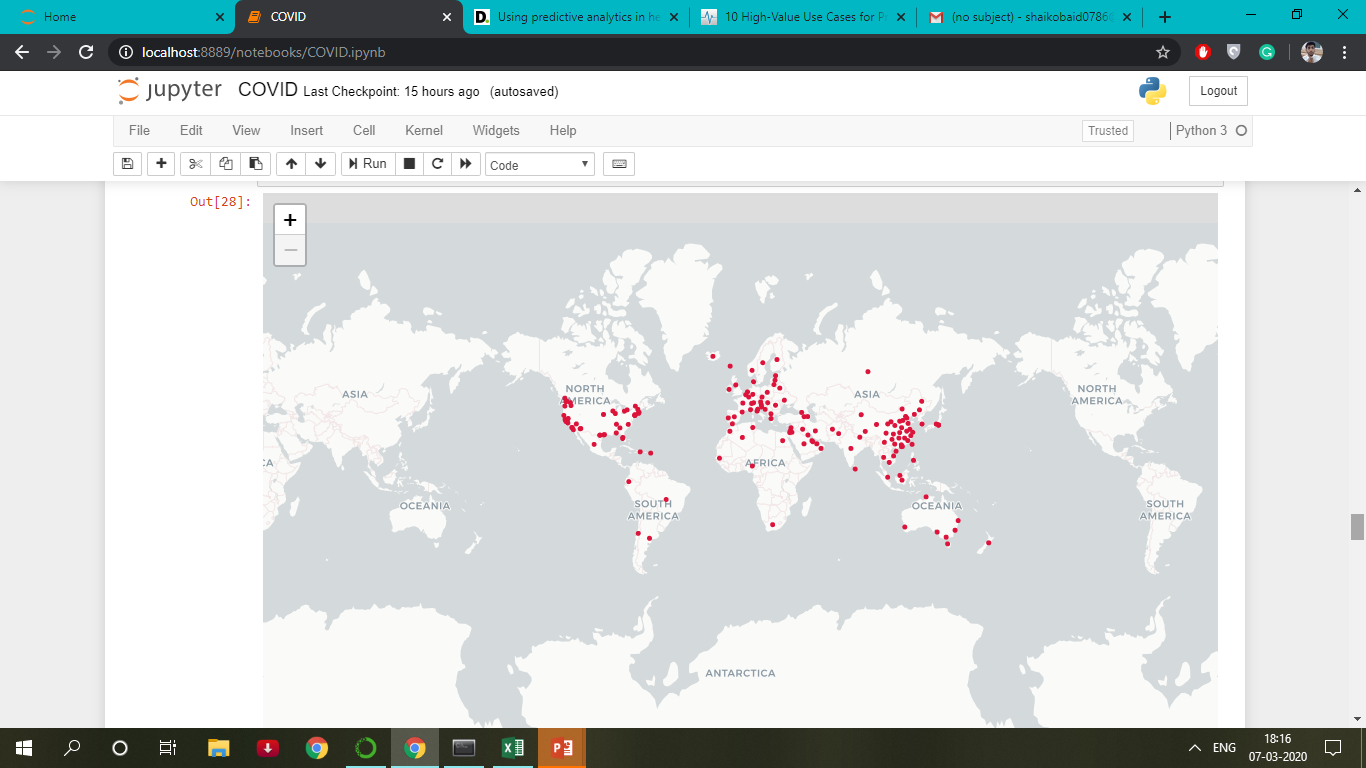


Visualizing the Confirmed Cases Worldwide.

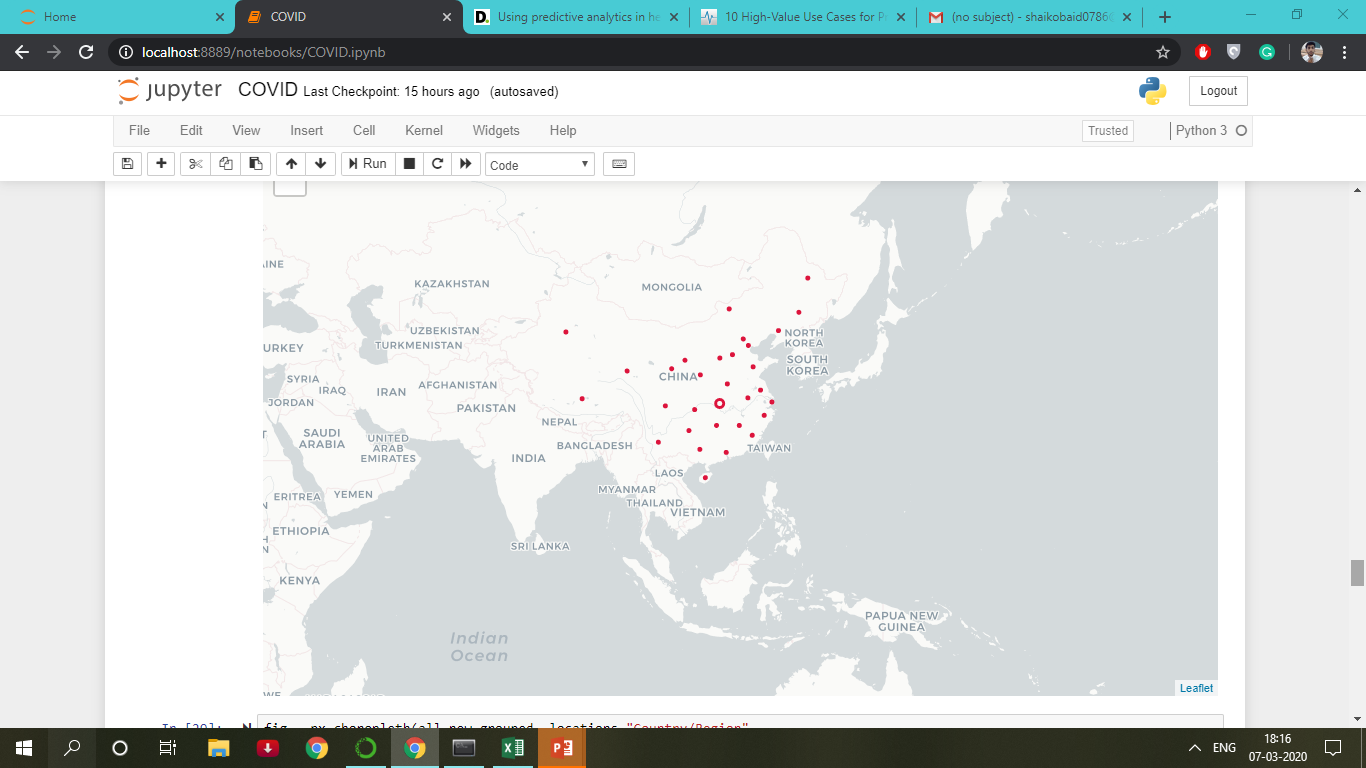


III Geospatial Data Visualization

-Visualization of Virus Outbreak using Folium libarary for World.



-Visualization of Virus Outbreak using Folium libarary for China.



IV Linear Regression Model

To Buils a Linear regression Model to predict the number of recovered cases by dividing the data into two groups: training set which consists of 5100 and test set which consists of 2112.

The following figure shows MSE and MAE.

