Background

The 2019 Novel Coronavirus, or 2019-nCoV, is a new respiratory virus first identified in Wuhan, Hubei Province, China. A novel coronavirus (nCoV) is a new coronavirus that has not been previously identified. The 2019 novel coronavirus (2019-nCoV), is not that same as the coronaviruses that commonly circulate among humans and cause mild illness, like the common cold.

A diagnosis with coronavirus 229E, NL63, OC43, or HKU1 is not the same as a 2019-nCoV diagnosis. These are different viruses and patients with 2019-nCoV will be evaluated and cared for differently than patients with common coronavirus diagnosis.

Public health officials and partners are working hard to identify the source of the 2019-nCoV. Coronaviruses are a large family of viruses, some causing illness in people and others that circulate among animals, including camels, cats and bats. Analysis of the genetic tree of this virus is ongoing to know the specific source of the virus. SARS, another coronavirus that emerged to infect people, came from civet cats, while MERS, another coronavirus that emerged to infect people, came from camels. More information about the source and spread of 2019-nCoV is available on the 2019-nCoV Situation Summary: Source and Spread of the Virus.

This virus probably originally emerged from an animal source but now seems to be spreading from person-to-person. It's important to note that person-to-person spread can happen on a continuum. Some viruses are highly contagious (like measles), while other viruses are less so. At this time, it's unclear how easily or sustainably this virus is spreading between people.

Business Problem

As the coronavirus outbreak spreads, the world's biggest companies have begun painting a bleak picture of broken supply chains, disrupted manufacturing, empty stores and flagging demand for their wares.

The announcements by businesses like Mastercard, Microsoft, Apple and United Airlines offer a reading on how the virus is affecting consumer behavior and business sentiment. These corporate bulletins — and what executives do in response — could determine how much economic damage the outbreak inflicts and whether a recession looms.

So, at this point, it is very important for businesses to understand the trend of novel coronavirus, predict its damage and take effective measures against it.

Target Audience

- Government Executives
- Business Professionals
- Any individual currently in lockdown condition

Data

1. Novel Corona Virus 2019 Dataset

Main file in this dataset is `covid_19_data.csv` and the detailed descriptions are below.

- Sno - Serial number
- ObservationDate - Date of the observation in MM/DD/YYYY
- Province/State - Province or state of the observation (Could be empty when missing)
- Country/Region - Country of observation
- Last Update - Time in UTC at which the row is updated for the given province or country. (Not standardised and so please clean before using it)
- Confirmed - Cumulative number of confirmed cases till that date
- Deaths - Cumulative number of of deaths till that date
- Recovered - Cumulative number of recovered cases till that date

    Features and Examples: Recovery Trend vs Death trend can be compared and analysed
    Country wise mortality rate analysis can be done

2. Foursquare Location Dataset

    It provides data about different places all around the world.

    Features and Examples:
    Analysis of different geographic locations including exploring places, finding trending locations near a venue, exploring users and their reviews about places can be done through Foursquare location dataset.

3. Indian Sensex Data from 22/1/30 to 15/3/2020

    The description of the data is as follows:
    - Date
    - Opening Value
    - Highest Value
    - Lowest Value
    - Closing Value
    -
4. Time series dataset of confirmed worldwide cases

Data Acquisition

Novel Corona Virus dataset is available on Kaggle. It can be found on the link
https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset.

Foursquare Location Database can be accesses through API.

The sensex data for targeted dates (in csv form) can be downloaded from
https://www.bseindia.com/Indices/IndexArchiveData.html

Time series data of confirmed worldwide cases can be found at
https://raw.githubusercontent.com/CSSEGISandData/COVID-
19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Confirmed.csv

Data Pre-Processing

Finding Null values in the dataset


```
data.isna().sum()

SNo                0
ObservationDate    0
Province/State     2315
Country/Region     0
Last Update        0
Confirmed          0
Deaths             0
Recovered          0
dtype: int64
```

We see that Province/State column has 2315 null values.
The steps we perform after this are:
- Create a column "Date" by converting "ObservationDate" to datetime format
- Drop the original "ObservationDate" column
- Drop the "SNo" column as this purpose is already fulfilled by the index of dataframe
- Drop the "Province/State" column as it has a large no. of null values

Methodology

In this project I will direct my efforts on analyzing and predicting the outcomes of Covid19

In first step I have collected the required datasets which are available on Kaggle. It has various important features such as country wise confirmed cases, deaths, recovered cases and observation date as time forms an important parameter in analyzing virus properties.

In the second step I will analyse the trends of Novel CoronaVirus which include:
• Rise in cases with respect to time
• Recovery and Motality rate over time
• Density of cases per country
• Fall in Indian Economy with arrival of Covid19

In the third and final step I will predict the future number of cases of coronavirus using Support Vector Machines and Linear Regression and compare the results. I will explore the most affected places with Covid19 and find what are the trending venues during the pandemic using Foursquare location database.

Exploratory Data Analysis

Earliest Recorded Cases

The five earliest recorded cases are as follows:

|	Country/Region	Last Update	Confirmed	Deaths	Recovered	Date
0	Mainland China	1/22/2020 17:00	1.0	0.0	0.0	2020-01-22
1	Mainland China	1/22/2020 17:00	14.0	0.0	0.0	2020-01-22
2	Mainland China	1/22/2020 17:00	6.0	0.0	0.0	2020-01-22
3	Mainland China	1/22/2020 17:00	1.0	0.0	0.0	2020-01-22
4	Mainland China	1/22/2020 17:00	0.0	0.0	0.0	2020-01-22

Latest Recorded Cases (As on 16th March)

The five latest recorded cases are as follows:

```
[ ] data.tail()
```

	Country/Region	Last Update	Confirmed	Deaths	Recovered	Date
5885	UK	2020-03-14T16:33:03	1.0	0.0	1.0	2020-03-15
5886	Uzbekistan	2020-03-15T18:20:19	1.0	0.0	0.0	2020-03-15
5887	Australia	2020-03-14T02:33:04	0.0	0.0	0.0	2020-03-15
5888	US	2020-03-10T02:33:04	0.0	0.0	0.0	2020-03-15
5889	occupied Palestinian territory	2020-03-11T20:53:02	0.0	0.0	0.0	2020-03-15

Corona Virus Trend Analysis

The analysis shows that the number of Corona Virus cases in the world are increasing as a function that can be represented by a polynomial with degree greater than unity.

Corona Virus Trend Overview
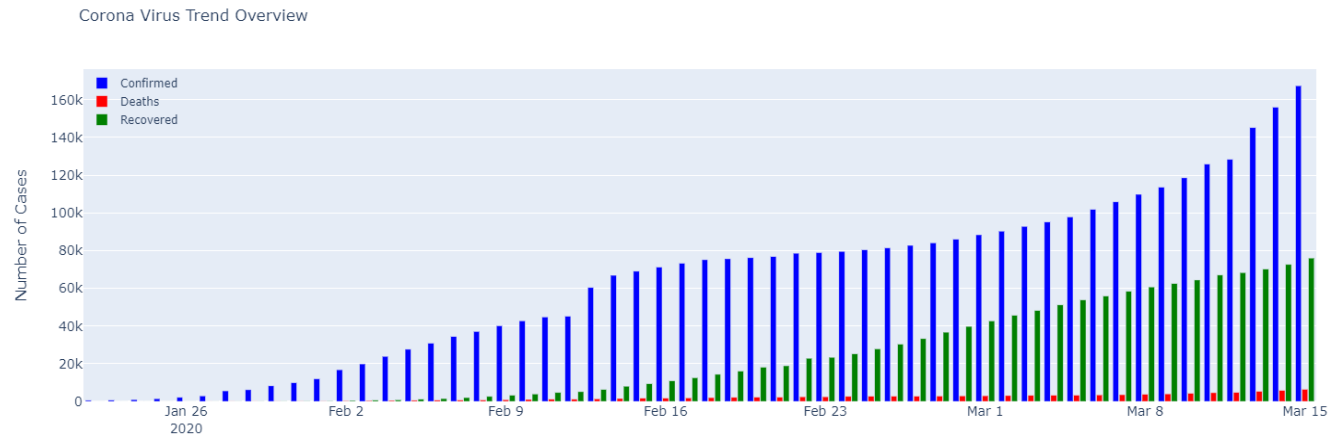
*Folium Map showing countries affected with Corona Virus*

There are no visible clusters of Corona Virus hit countries on map as nearly every country seems to be affected by it.

Chloropleth Map showing density of Corona Virus cases in different countries

The chloropleth map suggests that the worst Covid19 hit countries in the world are China, Italy, Iran, Spain and a part of United States of America

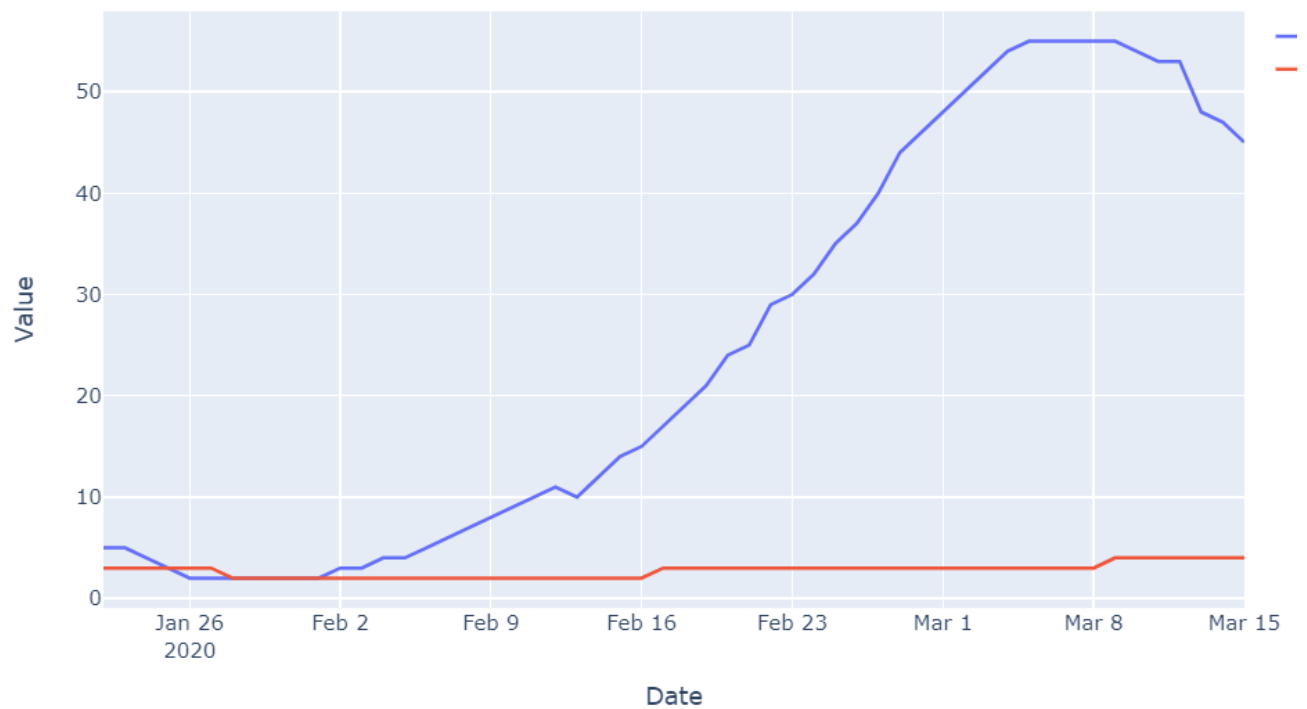## Countries with Confirmed Cases
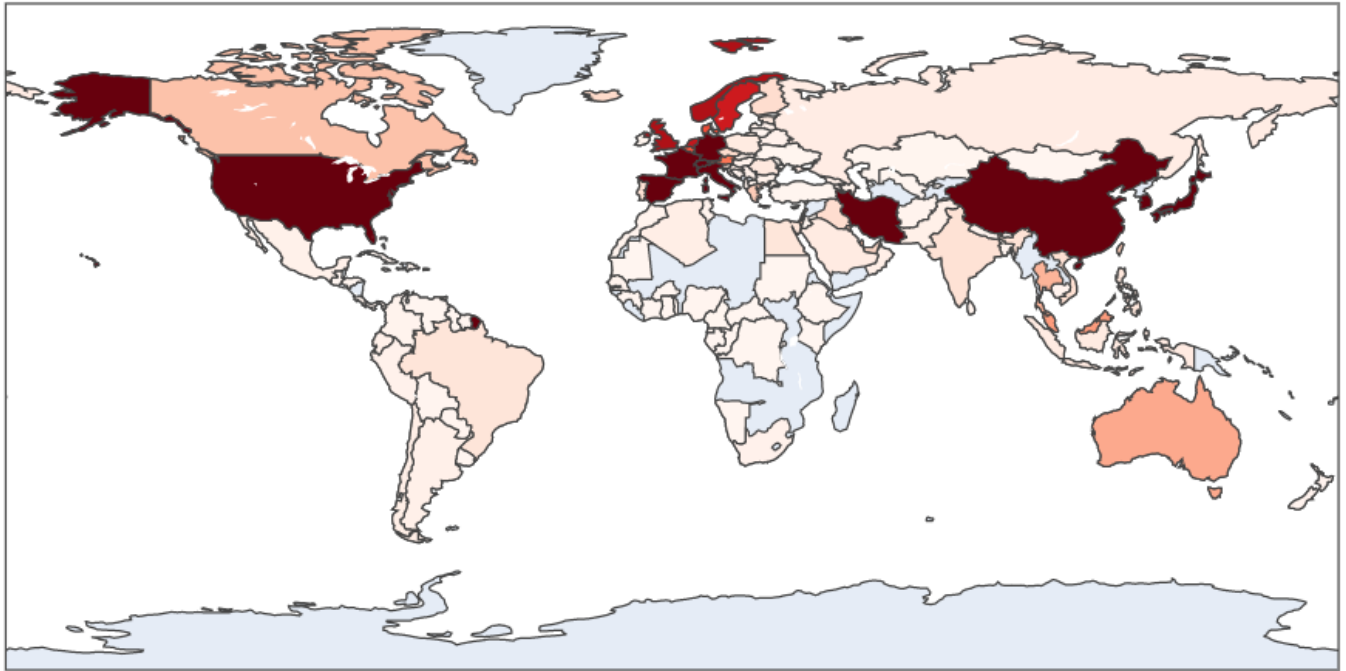
Recovery Rate VS Mortality Rate as a function of time

The mortality rate has remained constant since beginning but the good news is that recovery rate has increased significantly over time.

——— Ratio=No. of Recoveries per 100 Confirmed Cases
——— Ratio=No. of Deaths per 100 Confirmed Cases

Recovery and Mortality Rate Over The Time

Analyzing Indian Economy after Corona Virus outbreak

Indian economy is in a state of decline (worst hit after 2008) as the sensex (benchmark index of Bombay Stock Exchange which is an accurate gauge of India's economy) is continuously decreasing after arrival of Covid19.

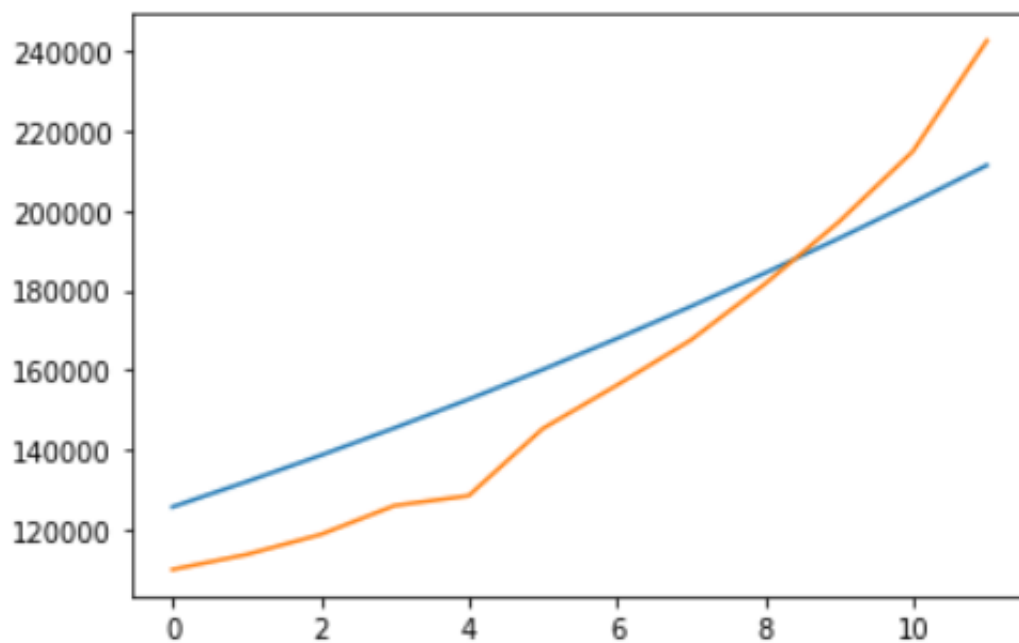Indian Sensex trend after CoronaVirus spread
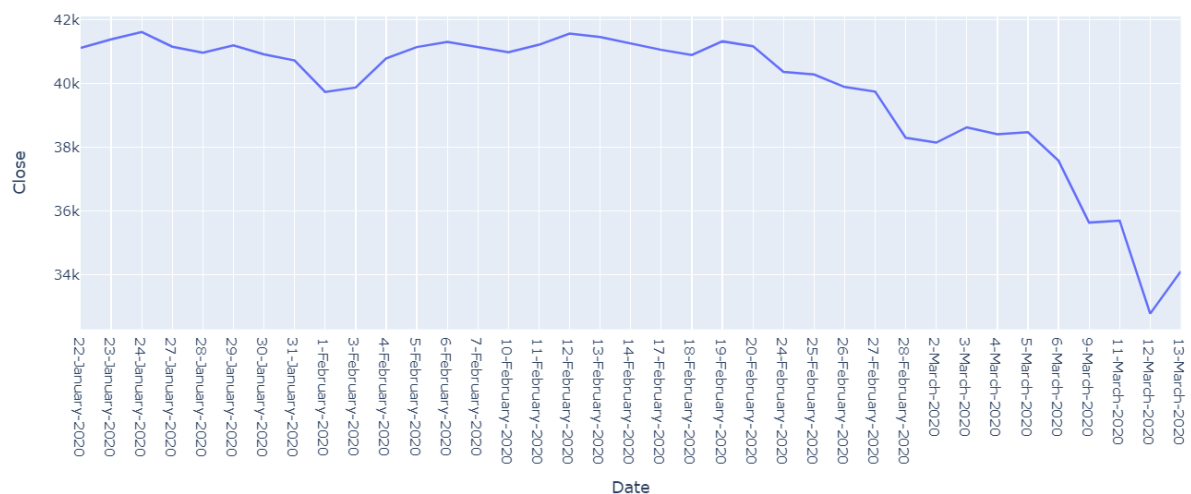
Predicting future No. of Corona Virus Cases (Machine Learning)

Support Vector Machine

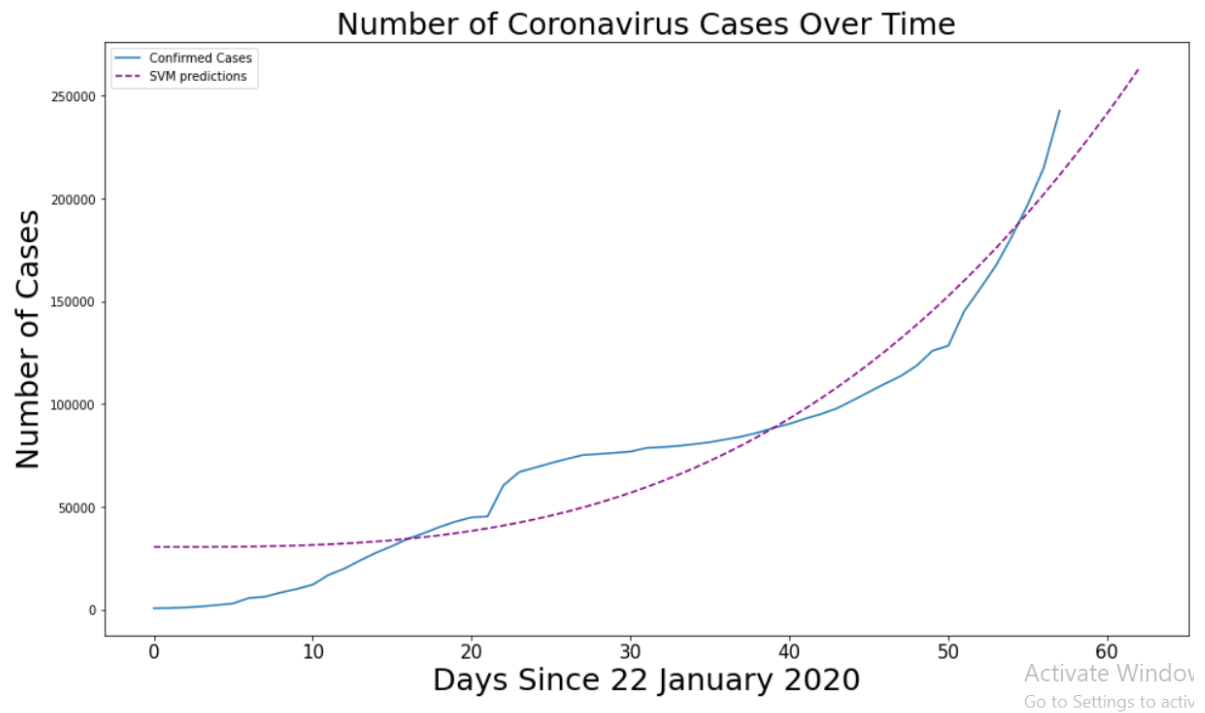Its parameters are as follows:
- `shrinking=True`
- `kernel='poly'`
- `gamma=0.1`
- `epsilon=1`
- `C=0.01`

Testing SVM predictions on test set

SVM Prediction including next 5 days

SVM model predicts that the cases will increase quadratically.
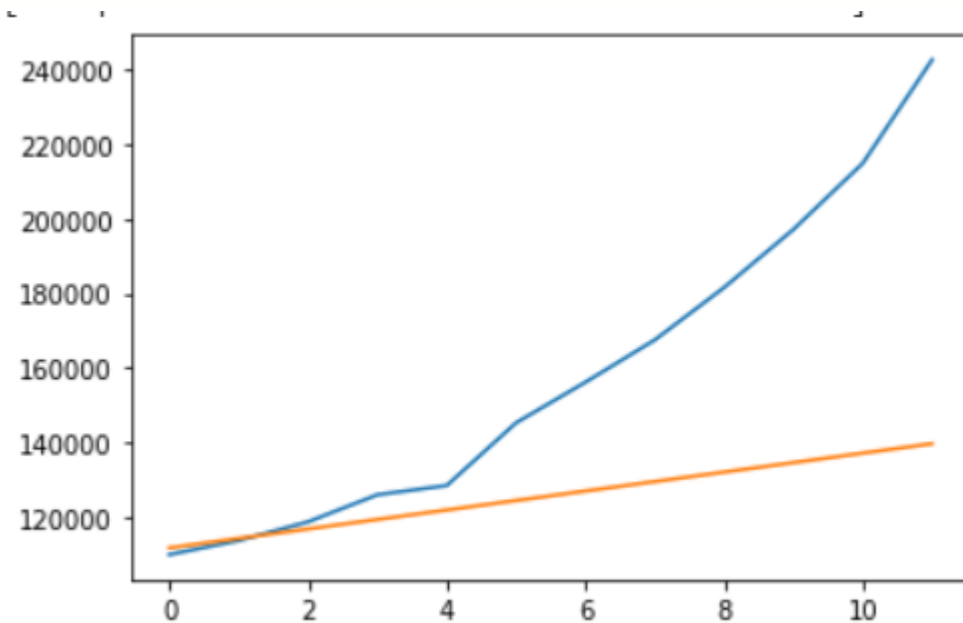
Number of Coronavirus Cases Over Time

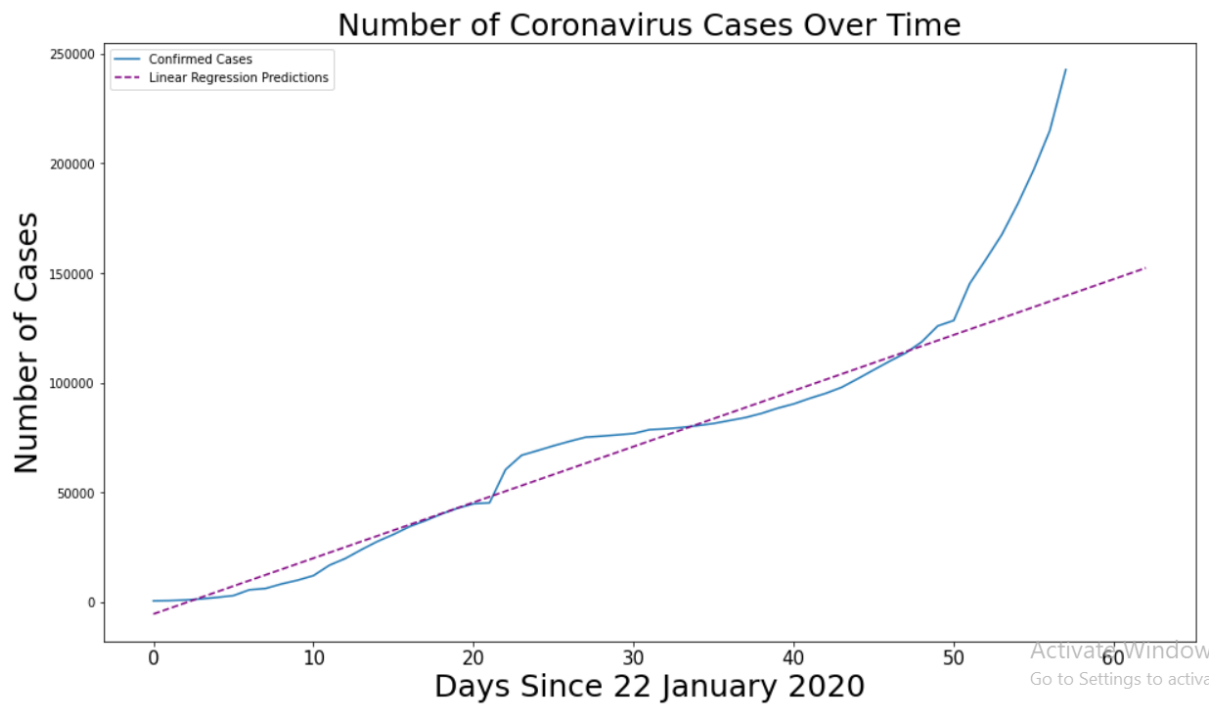*Linear Regression*

Its parameters are as follows:
- normalize=True
- fit_intercept=True

Testing Linear Regression predictions on test set

Linear Regression Prediction including next 5 days

```
It tries to fit a line on the following dataset.
```


It is evident that Support Vector Machine is a better fit to the Corona Virus dataset due to non-linear nature of the number of cases. The nearly quadratic slope of the no. of cases is worrisome as it induces risk of bringing the entire world into lockdown.

Exploring Corona Virus hit places and finding trending venues at the time of pandemic with Foursquare Location API

Mumbai (India)

Let's take an example of Mumbai. It is India's economic capital and one of the worst affected cities in the country.

Some famous places in Mumbai are as follows:

```
<bound method NDFrame.head of                                        name  ...
0                        Royal China  ...  4caa0096d971b1f7ccca23e1
1                     Town House Cafe  ...  5263e1ba11d265711e8024bf
2                         Sher-E-Punjab  ...  4b0587d9f964a52023a422e3
3                      Cafe Excelsior  ...  4c714f4fd7fab1f715d760c9
4      Chhatrapati Shivaji Maharaj Terminus  ...  4babe4fff964a520f8d23ae3
5                    Mahesh Lunch Home  ...  4b0587d6f964a520a5a322e3
6                      Britannia & Co.  ...  4d15a8e1816af04db4d444c2
7                       Yazdani Bakery  ...  4d8ea9e0788c54812b363ffd
8                           Starbucks  ...  507fcb7a498e5d07b1604ea2
9                         Ideal Corner  ...  4bdaa0332a3a0f47ad8eabb6
10                   Sterling Cineplex  ...  4bb483c6737d76b0e86f3b7c
11                    Food for Thought  ...  4cfe22077f2db1f745873bd4
12                   Pratap Lunch Home  ...  4dc777edcc3ff3b304696c5a
13                      Residency Hotel  ...  4bcf4de40ffdce72215fb2c0
14                     Bombay Gymkhana  ...  4b0587e6f964a52068a622e3
15              Dakshin Bar And Kitchen  ...  5215d00311d235bf4fe158e9
16                      Café Universal  ...  4b585e02f964a520a35428e3
17                     Horniman Circle  ...  4b0587d1f964a520d4a222e3
18                      Flora Fountain  ...  4b0587d1f964a520d0a222e3
19                    Wankhede Stadium  ...  4b0587dbf964a5207da422e3
20                     Taste of Kerala  ...  4c190902d4d9c928e755f029
21                        Kyani & Co.  ...  4b0587cef964a52081a222e3
22                                Zara  ...  590ef458f00a703cb8bbf81b
23                         Kitab Khana  ...  4d6e1cf829586dcb8d4badf1
24                      Fashion Street  ...  4b0587ddf964a520cba422e3
25                   Kayani Restaurant  ...  4f1ed1b7e4b0ab7784c2db4c
26           H Parmesha Fast Food Corner  ...  4dde5caed16415421840ebf8
27                     Dark Temptations  ...  4b0587e3f964a520aca522e3
28                    Pancham Puriwala  ...  4e33a472227111ae76912f24
29                      Crawford Market  ...  4b0587def964a520d5a422e3
```

But when I tried finding currently trending places in Mumbai, the result was:

```
'No trending venues are available at the moment!'
```

It reinforces the fact that the entire city has gone into lockdown and there are no trending places as footprint density has reduced in each and every place. The results repeated itself when I tried analysing more Corona Virus hit cities such as Wuhan (China) and Lombardy (Italy) as shown below:

Wuhan (China)

Famous Places:

```
<bound method NDFrame.head of                                    name  ...
0  Howard Johnson Pearl Plaza Wuhan (江城明珠豪生大酒店)  ...  4cc9336c5c7e8cfa1af94266
1                          River Park (汉口江滩公园)  ...  4bfa1443b182c9b627ee795a
2                        Marco Polo Hotel (马可孛罗酒店)  ...  4dc175a08877c00d6ad8e225
3                Dazhi Rd Metro Station (地铁大智路站)  ...  4e4538096365d05debfc0124
4              Sanyang Road Metro Station (地铁三阳路站)  ...  4ca42d489cbf9c748960d54b
5                           Royal Suites and Towers  ...  4f73a38be4b0ae979e5aca3a
```

Trending Places:

```
'No trending venues are available at the moment!'
```

Lombardy (Italy)

Famous Places:

```
<bound method NDFrame.head of                            name  ...
0                       Castello Malpaga  ...  4de12fb5fa76425c542bd22b
1  Locanda Dei Nobili Viaggiatori  ...  535b6c89498ebb3fe28b796a
2                    Bowling Il Triangolo  ...  4eda8fd0b8f7fcd0408184df
3              Ristorante Pizzeria Jolly  ...  4efa2517722e340611a37e85
4                              Pollicino  ...  511157bfe4b00ee78e34318a
5              Locanda del Santo Bevitore  ...  4e3e58eecc3f6517531d195b
```

Trending Places:

```
'No trending venues are available at the moment!'
```

Results and Discussion

=

- The analysis shows that the number of Corona Virus cases in the world are increasing as a function that can be represented by a polynomial with degree greater than unity.

- The chloropleth map suggests that the worst Covid19 hit countries in the world are China, Italy, Iran, Spain and a part of United States of America.

- The mortality rate has remained constant since beginning but the good news is that recovery rate has increased significantly over time.

- There are no visible clusters of Corona Virus hit countries on map as nearly every country seems to be affected by it.

- Indian economy is in a state of decline (worst hit after 2008) as the sensex (benchmark index of Bombay Stock Exchange which is an accurate gauge of India's economy) is continuously decreasing after arrival of Covid19

- Support Vector Machines (SVM's) are better at predicting the number of Corona virus cases than Linear Regression. The nearly quadratic slope of the no. of cases is worrisome as it can bring the entire world into lockdown.

- Analysis done through Foursquare Location Database confirms that the worst hit places are into lockdown as there are no trending places currently.

-

Conclusion

The purpose of this project was to identify the trends and predict the future of Corona Virus. The trends show the danger that is about to come in the near future. However, several measures can be taken by an individual to protect oneself and human race as a whole. They are as follows:

1. Wash your hands often with soap and water for at least 20 seconds especially after you have been in a public place, or after blowing your nose, coughing, or sneezing.
2. If soap and water are not readily available, use a hand sanitizer that contains at least 60% alcohol. Cover all surfaces of your hands and rub them together until they feel dry.
3. Avoid touching your eyes, nose, and mouth with unwashed hands.
4. If surfaces are dirty, clean them: Use detergent or soap and water prior to disinfection.

The stakeholders all over the world are doing their best to flatten the curve of this pandemic and various relief packages are released for all affected people. We, on an individual level, can help fight this pandemic by keeping ourselves healthy and following the directives laid by respective governments.