Predicting future Covid-19 cases
Date: 11th April 2020

Introduction

With the current situation in the world, the whole world is seeking answers how to overcome covid-19 pandemic. Therefore it is very important for the whole world to know the effects of the Covid-19 such as demographical and socio-economical effects.

Covid-19 which is also called Corona virus is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease was first identified in December 2019 in Wuhan, the capital of China's Hubei province. Since then the disease spread globally.

Data

For this report I collected data from the following web pages,

- https://en.wikipedia.org/wiki/2019%E2%80%9320_coronavirus_pandemic_by_country_and_territory
- https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset/data

  In the kaggle.com I found the following sets of data,

  - Covid_19_data.csv
  - Covid19_line_list_data.csv
  - Covid19_open_line_list.csv
  - time_series_covid_19_confirmed.csv
  - time_series_covid_19_confirmed_US.csv
  - time_series_covid_19_deaths.csv
  - time_series_covid_19_recovered.csv
- Location data from Foursquare API to get details about the locations

Methodology

Data was imported from above recourses to the notebook initially. Then the data wrangling was done. Thereafter the data formats and the several data was observed in-order to clarify that the data is in the correct format.

After the data wrangling and preprocessing, data was plotted and visualized to observe patterns. Thereafter polynomial regression techniques were used to predict the future cases bases on the related parameters.

Results

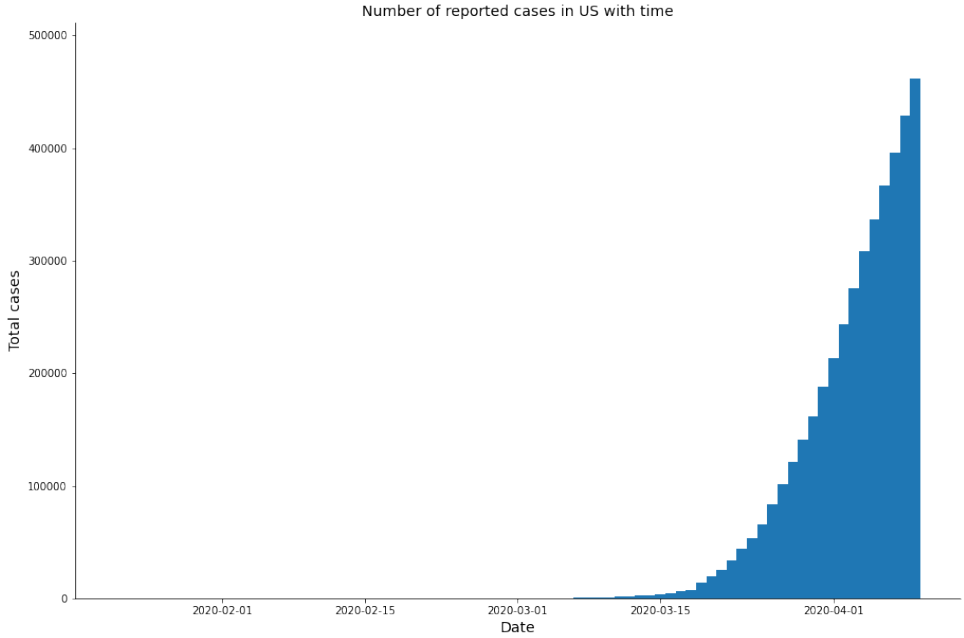

Figure 1: Number of confirmed cases in US

Figure 1 shows the time series variation of the confirmed cases in US, since the start of the pandemic worldwide. We can observe a rapid increase of the number of confirmed cases since 15 March 2020. Thereafter the curve increased exponentially up to more than 450,000 cases up to date.

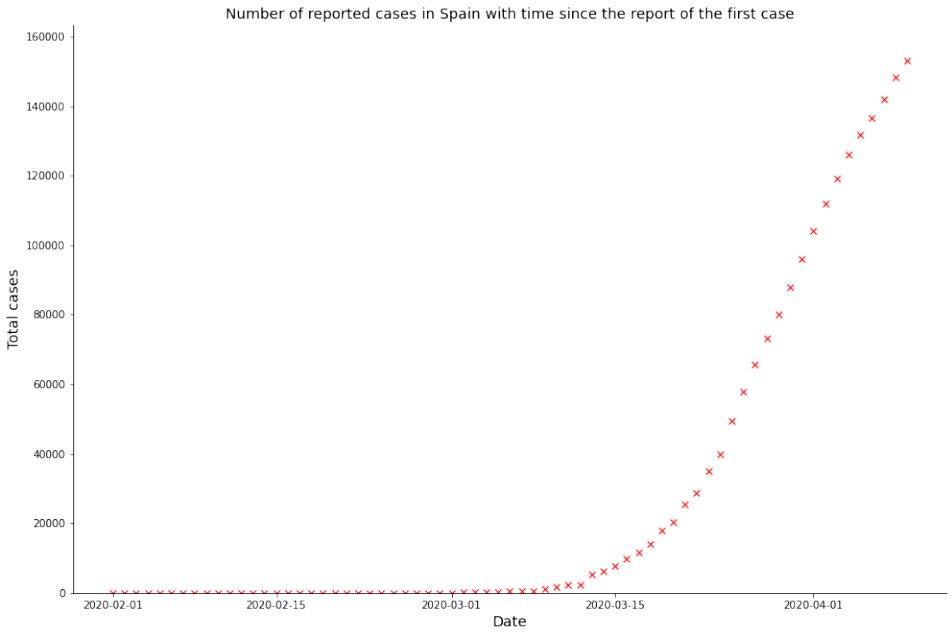

Figure 2: Number of confirmed cases in Spain

As shown in the figure 02, number of confirmed cases in Spain also rapidly increased since 15 March 2020. Since then the curve increased rapidly and now the shape is changing according to the figure 2.

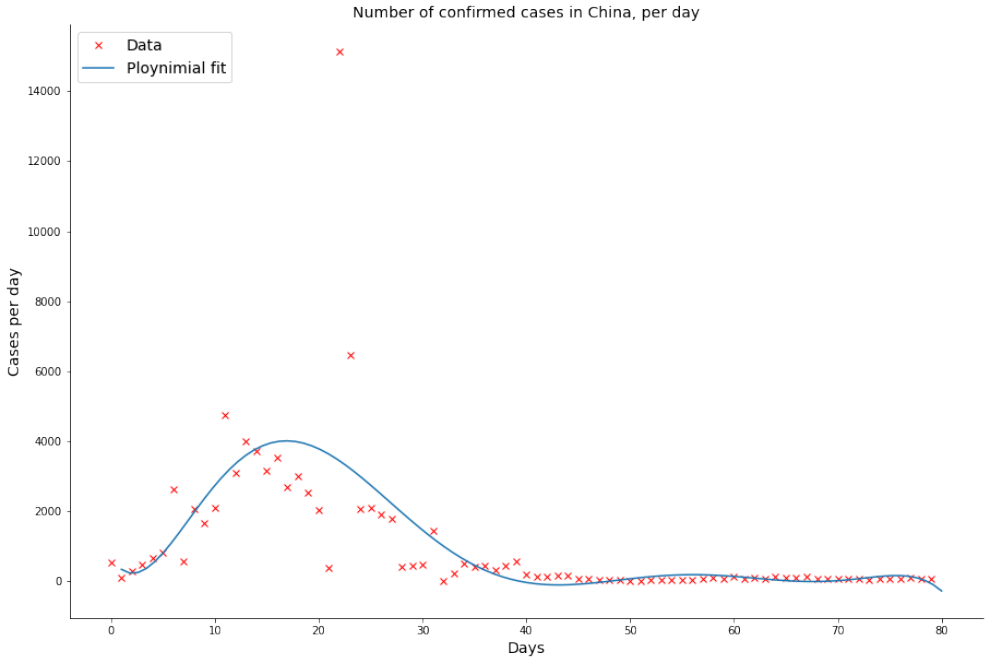

Figure 3: Number of confirmed cases in china, per day

Figure 3 illustrates the variation of the number of cases in china day by day since the pandemic started and the polynomial fitted curve to predict the future cases based on the past data. According to the polynomial fitted curve we can predict that, the number of future day by day cases in china will be reduced in the future. Trend shows that china has overcome the pandemic successfully.

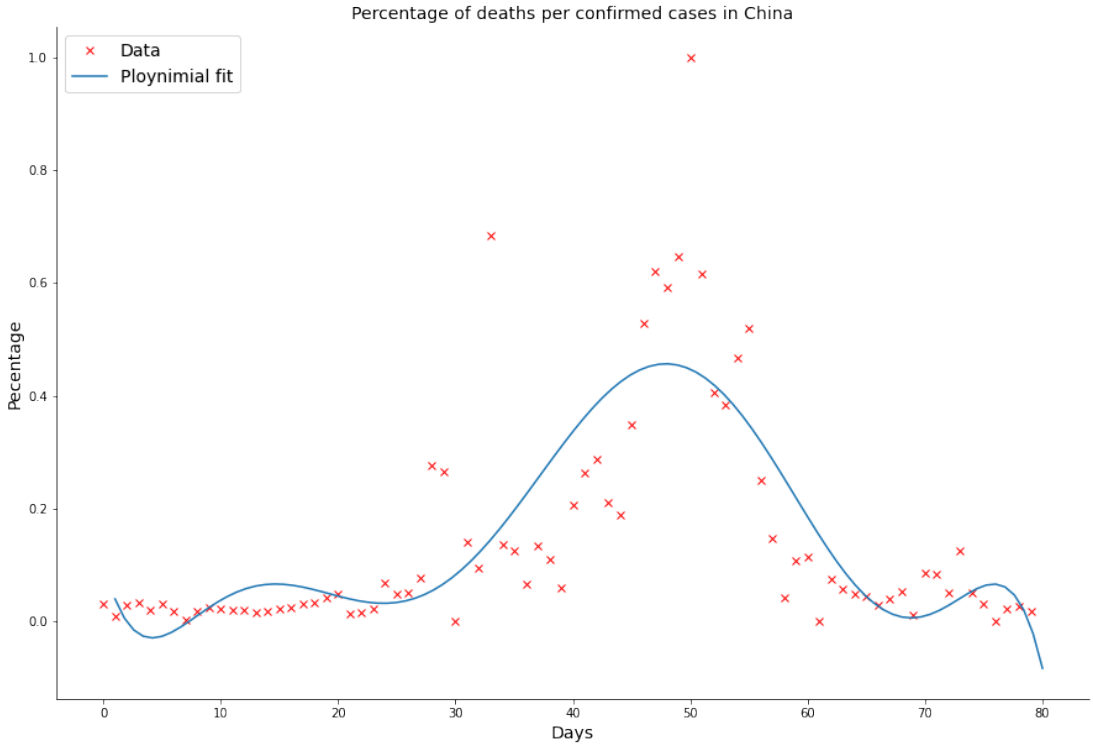

Figure 4: Percentage of deaths per confirmed cases in china

According to the figure 4 also we can observe that the deaths per confirmed cases are reducing in china with time.

Discussion and conclusion

According to the data analysis, the number of cases in US, Spain and Italy will be increased further, while the number of cases and deaths in china will be controlled. With proper health care and governed local attitudes, we can overcome this pandemic globally. Further analysis have to be done in-order to identify parameters and predictions for the number of confirmed cases and deaths due to the covid-19.