In order to visualize the effect of the lockdown we considered the data from 31st January 2020 to 24th March 2020 i.e when the lockdown was enforced.

We fit a regression model to this data. Once we obtained the fit, we extrapolated the graph to see the number of cases that would have occurred had the lockdown not been enforced.

We used **polynomial regression with a degree of 7** to determine the spread if there was no lockdown.

Using this we can see that the lockdown prevented roughly **167,000** cases.

The methodology for choosing polynomial regression and the specific degree we chose is outlined in the notebook which contains the code.

**What regression to use**

On plotting the scatter plot of the data till 24th March we can clearly see that the data does not follow a linear relationship.

We inferred that the data could follow the following distributions;

1. Exponential log(y)=ax + b
2. Polynomial y=a1x + a2x2 and so on
3. Polynomial and exponential log(y)=a1x + a2x2 and so on

**Exponential**

On plotting the graph for exponential relationship on the data obtained till 24th March 2020 we noticed that the graph does not coincide as well as in the polynomial case. We therefore prefer extrapolating the data using polynomial regression.

The RMSE=71.75430801258224 is much higher than what we get with simple polynomial regression

This is also why the predicted values till 30th April show such a high deviation from the actual curve

**Polynomial Exponential**

On plotting the graph for polynomial exponential relationship using a degree 2 on the data obtained till 24th March 2020 we noticed that the graph does not coincide as well as in the polynomial case. We therefore prefer extrapolating the data using polynomial regression.

The RMSE=30.8852725706685 is much higher than what we get with simple polynomial regression

This is also why the predicted values till 30th April show such a high deviation from the actual curve