**Malaria**

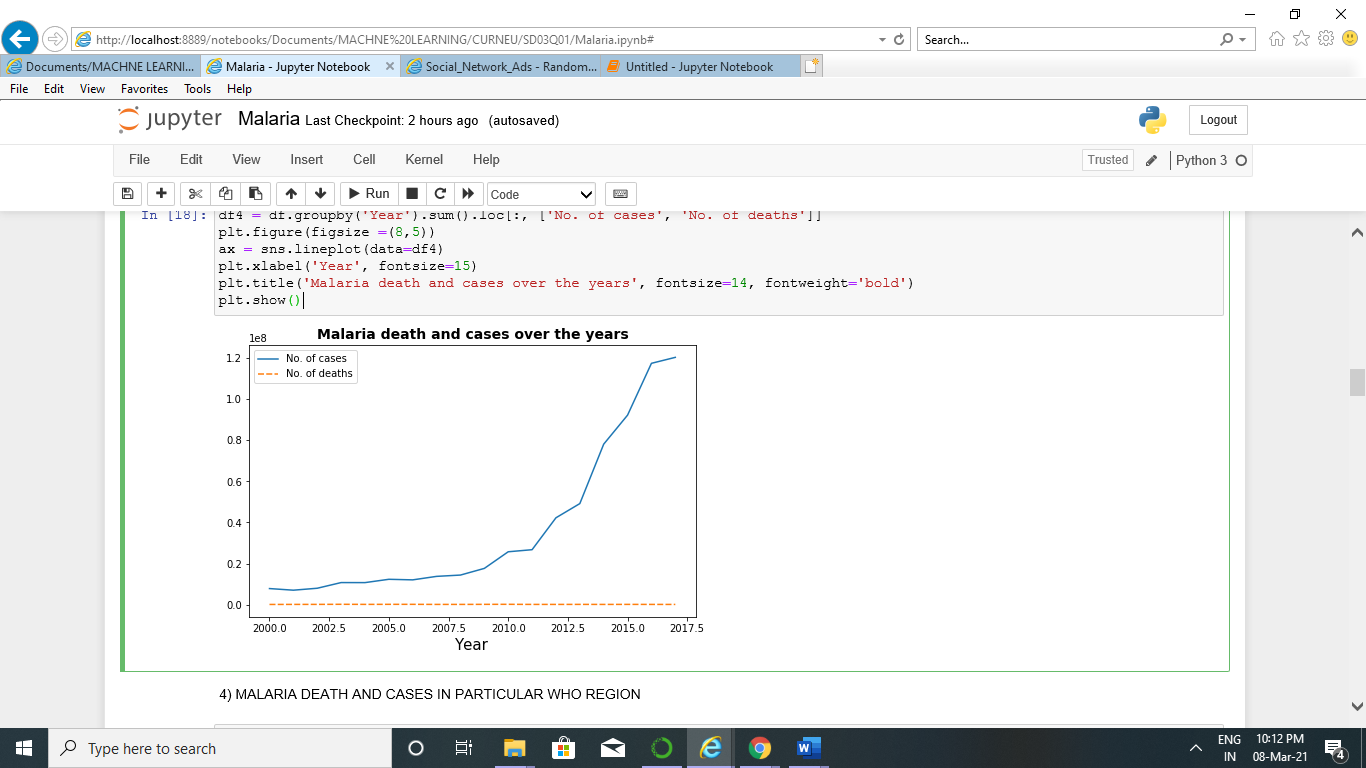
**Dataset:**

There are three datasets given

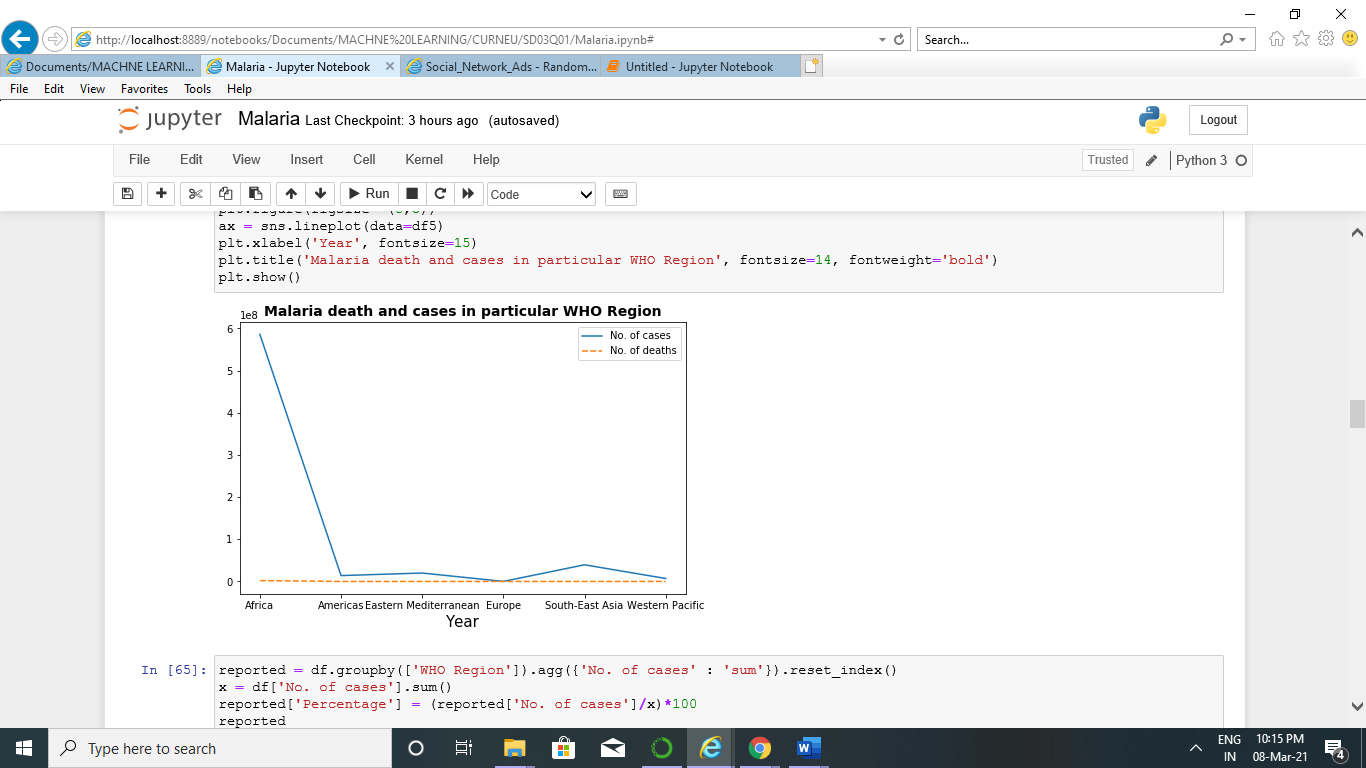
1. Reported numbers
2. Estimated numbers
3. Incidence numbers

**Interpretation**

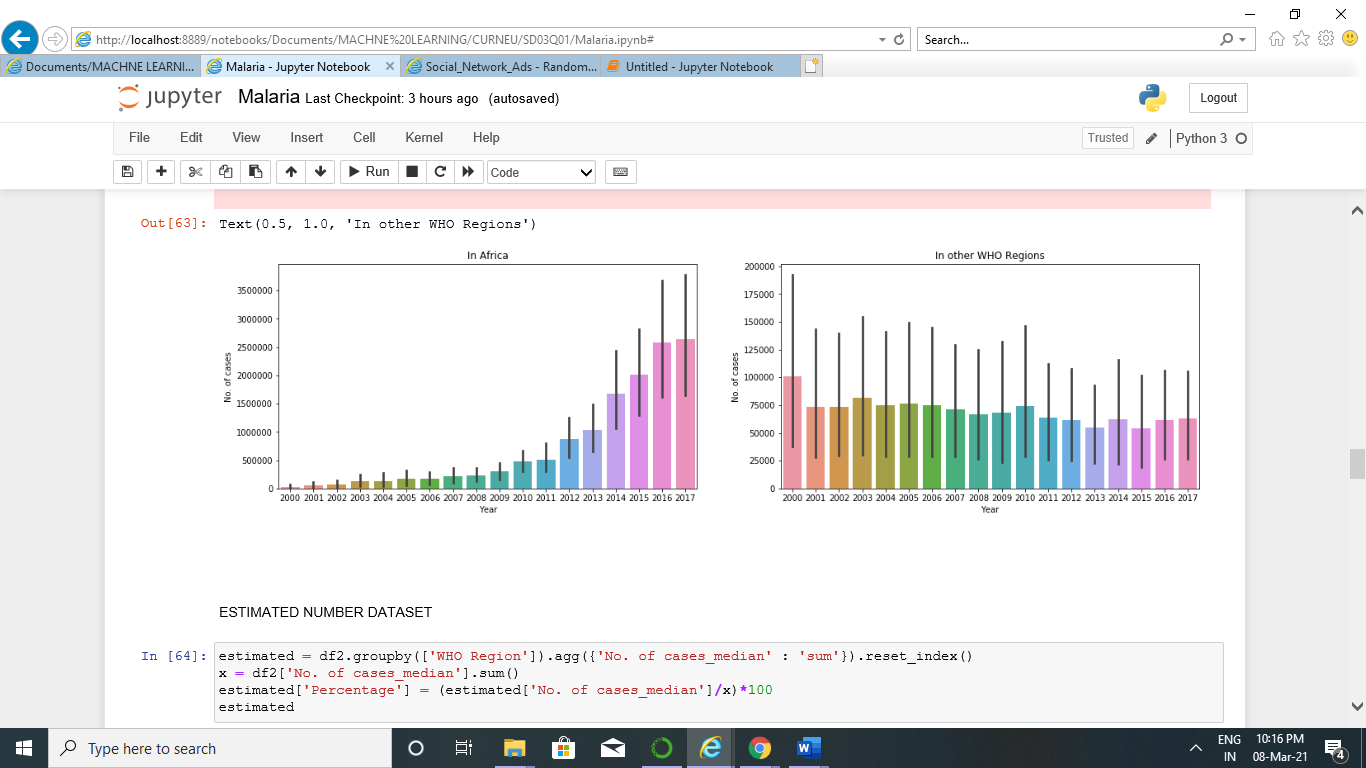
* The libraries are imported, the dataset is loaded, and data preprocessing is done.
* After preprocessing the dataset exploratory data analysis is done for all the three datasets.
* The first plot is bar plot, which is plotted with the attributes no of cases and countries in the year 2016. From this plot we infer that Democritic republic of the congo has the highest no of cases.
* The second plot is also an bar plot which is plotted with the same attributes no. of cases and countries in the year 2017. Even from this plot we infer that Democritic republic of the congo highest no of cases.
* Then we plot the graph for malaria cases and deaths together in the same graph.



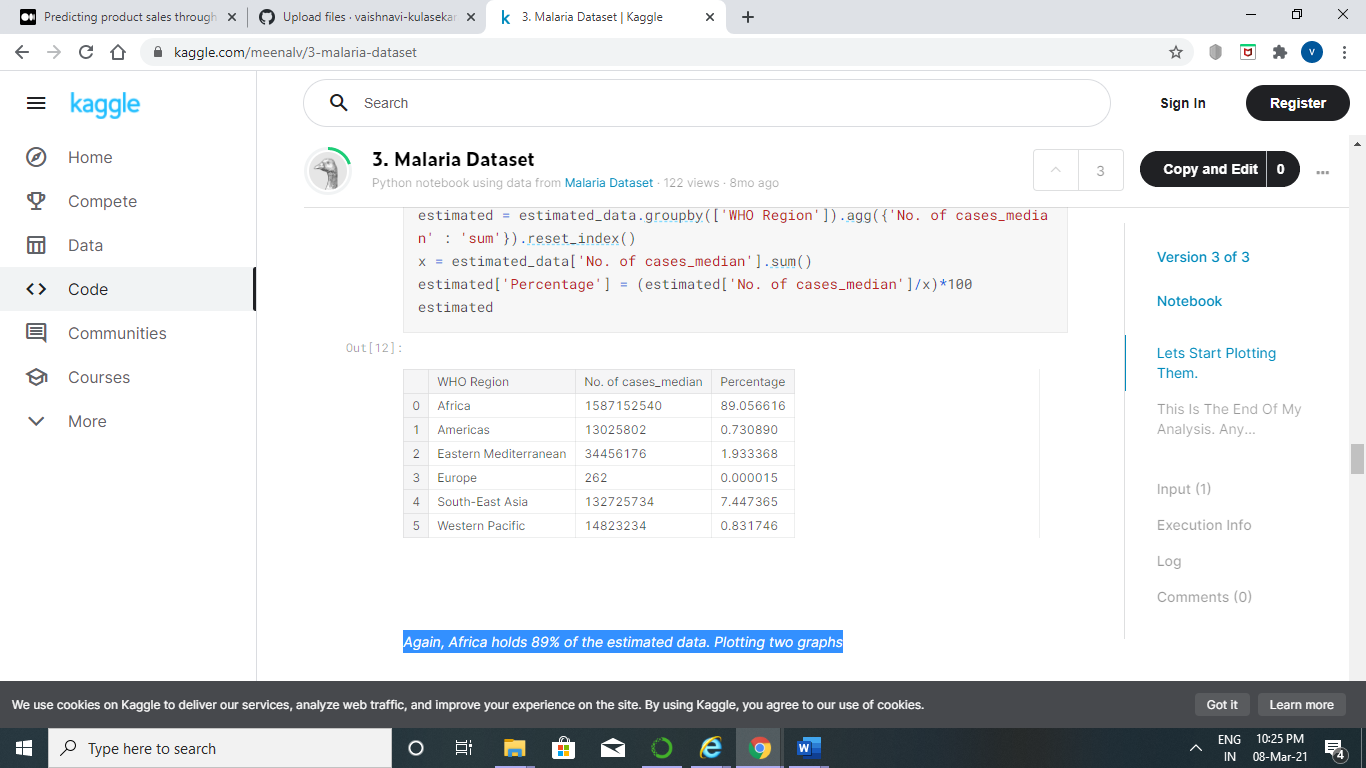
* Next we plot malaria cases and deaths with particular to who region. From this plot we can see that AFRICA constitutes 88% of the reported data. This means, if we include this, most of our data would be biased, so lets do the analysis in two groups, one includes AFRICA, and one excludes AFRICA



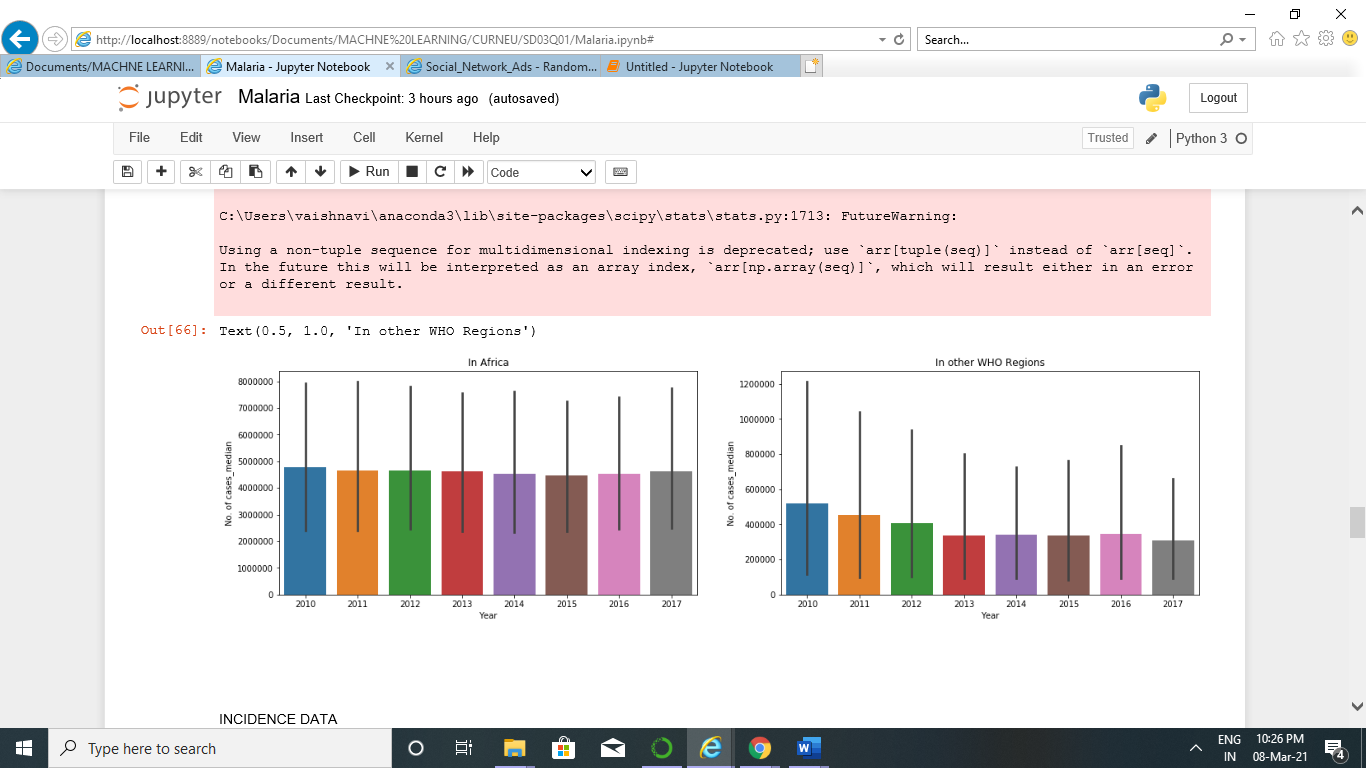
* This gives us the trend that, cases are increasing yearly in Africa, whereas, it is declining for the other Regions



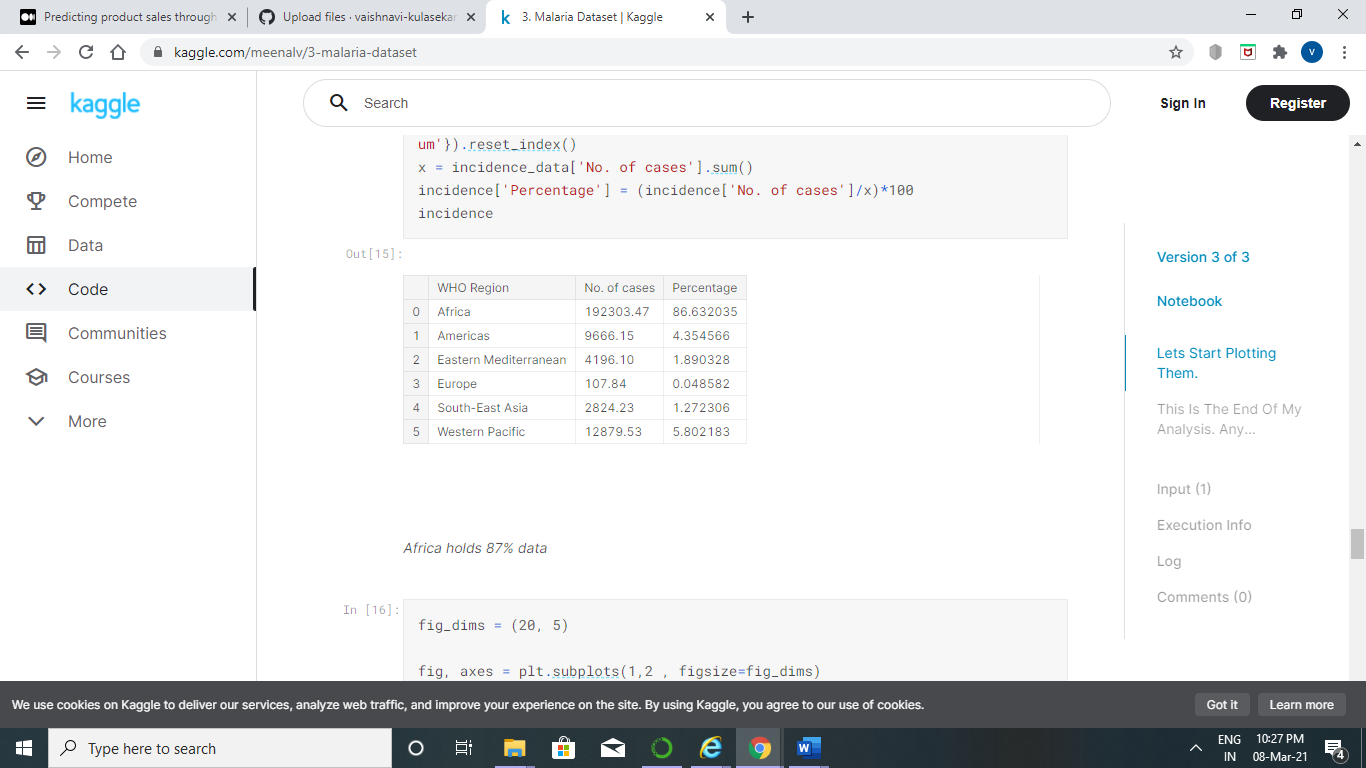
* Again, Africa holds 89% of the estimated data. Plotting two graphs



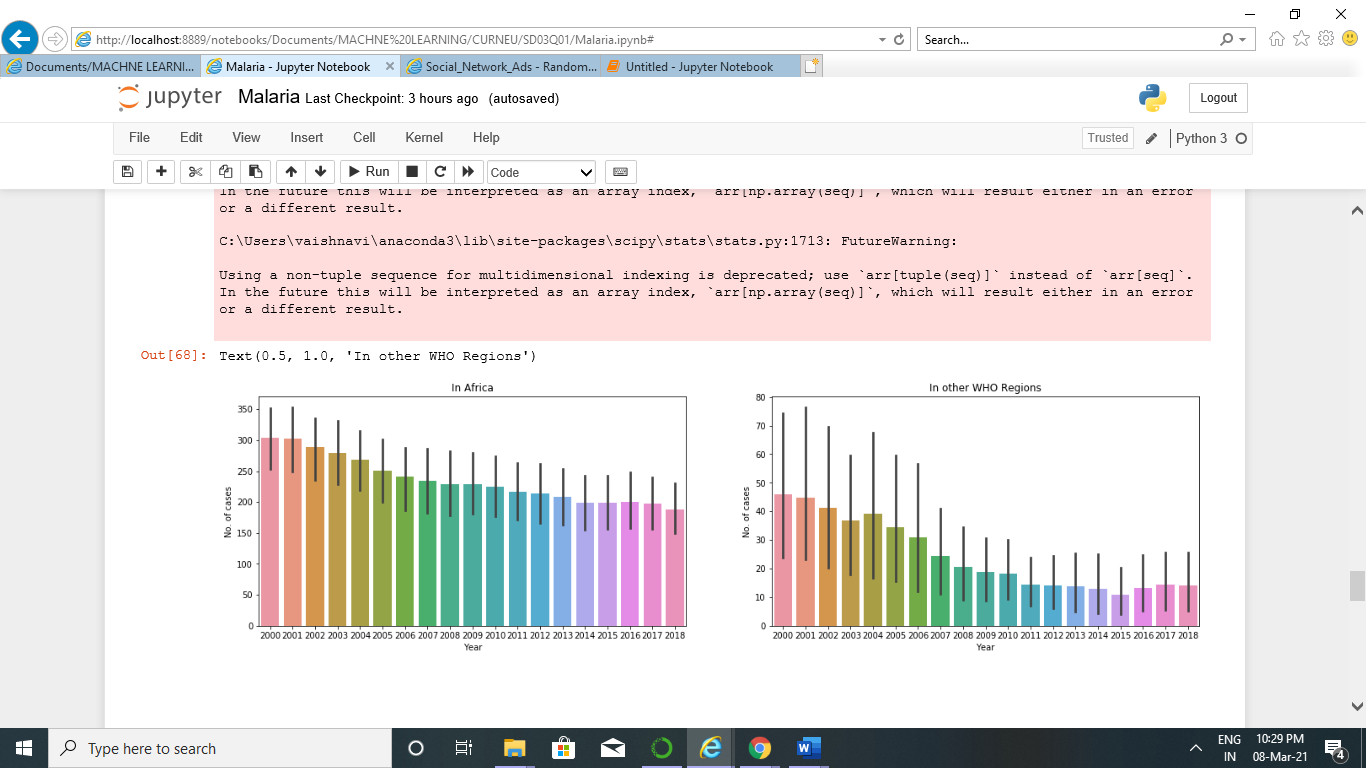
* The estimated data shows a stangnant number of increase in cases, not much. Where as for the other regions, it is declining.



* Africa holds 87% data for incidence number dataset.



* Observing the graph, the trends show that the numbers are declining for both. Its small yearly decrease for Africa region, whereas, for other regions, it's a good yearly decrease



* Observing all the graphs, the following conclusions can be drawn, For all the regions, except Africa, the cases are declining yearly. Africa shows a steady decrease, but not a huge decrease.
* After performing linear exploratory data analysis we perform simple linear regression analysis for no of cases and no of deaths. Taking no of cases as predictor variable we predict no of deaths. Hence the algorithm is coded from scratch and the death rate is predicted.

