**Lab 3 summary:**

In this lab, we had to analyse whether smaller data sets generates a better outcome or larger data sets give a better model .

For this lab we had to add three data sets to our portfolio.

* The first data set contains top 2 features
* The second data set contains top 5 features
* The third data set contains top 10 features

We used standard correlation efficient method to choose the best correlated features.

All the datasets went through the same process i.e.,

* Dividing data set into training and testing models.
* Create a logistic regression function modelled on the training data.
* Print the precision score

After evaluating all the precision scores these are the following results:

Precision score [Data set 1]: 0.7342

Precision score [Data set 2]: 0.7352

Precision score [Data set 3]: 0.8013

We can clearly observe that as the number of features increases the accuracy and quality of the model increases.

We even observed with all the features of the dataset to achieve a precision score of 0.8322.

Conclusions:

More top features of the data set clearly increases performance level. Any of our trained models will yield better predictions on reliable data as see from the results above.

Although too many features may lead to overfitting and therefore may be dealt with a lot of regularization. Hence, a set of well reliable attributes of a dataset can be made useful if there is a correlation between the additional features and the target variable.