Assignment 3: Clustering and Fitting.

# Subject: Applied Data Science 1.

In this assignment we are generating a report on how clustering and fitting works on a dataset.

Clustering: Clustering is statistical method for processing data. It is performed by organizing items into groups or clusters based on how closely they are associated. The objective of clustering is to find similar group of subjects. Clustering is an unsupervised algorithm, meaning you don't know how many clusters exists in the data before running the model.

Fitting: Model Fitting is a measure of how well a model generalizes to similar data to that on which it was trained. A model that is well fitted produces more accurate 0utcomes. A model that is not fitted well doesn't match closely enough.

Clustering:	Fitting:
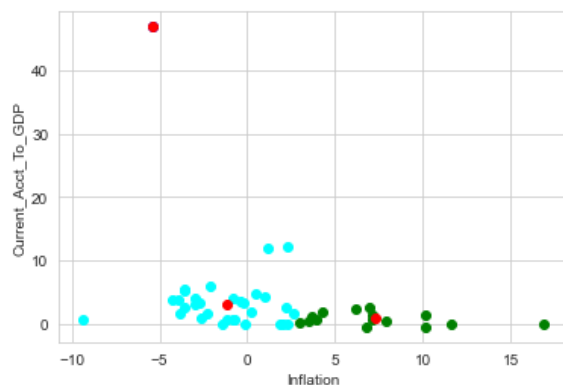
In the above diagram we have used K-means clustering with k=3. The clusters are divided into two groups with green and cyan color based on the 3 centroids in reds. The clusters group in the green one is more scattered from the centroids with the number of clusters less than the cyan color one. As the cyan color-based clusters are more and scattered more than the green one, the data between the two are accurately scattered and grouped based on the centroids.

The two groups' clusters are not similar to one another as they both are not scattered in a similar way. The cyan colour cluster group are closer to each other whereas, the dark green group of clusters are scattered at a distance to each other.
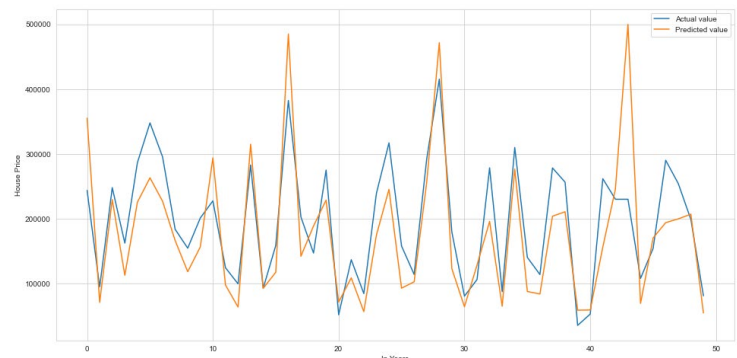
The above figure is of model fitting. The model used is linear regression. As we can see the graph is of actual values vs predicted value. The graph is of median of house value near the ocean. During the first 10 years the actual value was higher than the predicted values. In the next 10 years the predicted values increased until the first half and gradually decreased in the later half. In the next 20 years the, only data that is near to the 30 have predicted values higher than the actual values while the rest of years have actual value more than predicted values.

In the last 10 years the predicted values rose gradually than the previous one while the actual values slightly decreased. In conclusion the graph shows that every 5 years the values of houses raised than the actual values.