**Introduction**:

* Dimensionality reduction is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data.
* In some datasets the dependent variable is not affected by the features present in the dataset or they have very less correlation. So, these features should be recognized and removed from the dataset while developing the machine learning models.

e.g. While predicting the customer going to buy the car or not, is not much correlated with the weight of the customer. So, impact of weight over customer should be find in the dataset. If no impact found, then it should be removed from the dataset.

* In high dimensional dataset having much amount of feature variables while analysing the data we face the problem of multicollinearity (One feature variable is correlated with other feature variables) which causes the machine learning model unreliable. To overcome the problem of multicollinearity we perform the dimensionality reduction methods.

**Dimensionality Reduction Techniques:**

There are two commonly used dimensionality reduction techniques are:

1. Feature Selection: We need to be very careful while selecting the features within the dataset for developing the reliable machine learning models. In feature selection techniques we can perform regression techniques. For this we perform the hypothesis testing.

Null Hypothesis H0: Variable have no effect on target variable and can be removed from the model.

Alternative Hypothesis H1: Variable have some effect on target variable and can be retained into the model.

We perform the statistical methods and find the p values of statistics (probability of obtaining the result).

If p-value > 0.05: Accept null hypothesis: variable have no significant impact over target variable.

If p-value < 0.05: Reject null hypothesis: variable have significant impact over target variable.

Please go through the Feature\_Selection\_Advertising\_Data.ipynb file.

1. Factor Analysis Process: To deal with the problem of multicollinearity within the dataset we perform this technique. There are two commonly main factor analysis processes are:

* Principal Component Analysis (PCA). (Please go through the PCA\_Analysis\_Iris\_Dataset.ipynb file).
* Linear Discriminant Analysis (LDA). (Please go through LDA\_IRIS\_XTEND\_Dimensionality\_Reduction.ipynb file).