**Assignment 01: Evaluate the Ad Budget Dataset of XYZ Firm**

**Description:**

**Problem:**

The given dataset contains ad budgets for different media channels and the corresponding ad sales of XYZ firm. Evaluate the dataset to:

* Find the features or media channels used by the firm
* Find the sales figures for each channel
* Create a model to predict the sales outcome
* Split as training and testing datasets for the model
* Calculate the Mean Square Error (MSE)

**Assessment:**

(Note: For code refer attached notebook as pdf.)

**1. Import & Analyze the dataset:**

* Import the pandas library.
* Read and assign the ‘Advertising Budget and Sales.csv’ file using read\_csv() function to the variable ‘df\_ad\_data’.
* View the initial five records of the dataset using df\_ad\_data.head()
* Using the .size method the total number of elements in the dataset is obtained as ‘**800**’.

**2. Find the features or media channels used by the firm**

* Using shape method, the number of observations (rows) and attributes (columns) in the dataset is obtained as 200 Rows and 4 Columns.
* Using columns method, the name of each of the attribute is obtained.
* The media channel used by the firm are,

1. TV Ad
2. Radio Ad &
3. Newspaper Ad

**3. Create a model to predict the sales outcome & Find the sales figures for each channel**

* Create a feature object ‘X\_features’ from the columns such as 'TV Ad Budget ($)', 'Radio Ad Budget ($)', 'Newspaper Ad Budget ($)'.
* Create a target object ‘Y\_target’ using ‘Sales ($)’ column.
* 'Y\_target' holds the values of Sales figures for each channel.

**4. Split as training and testing datasets for the model**

* Import model selection from sklearn.
* Use the X\_feature and Y\_tagert objects in model selection function to split the dataset in to training and testing datasets.

**5. Create a model to predict the sales outcome**

* From sklearn linear\_model import LinearRegression to create a Linear Regression model.
* Fit the training datasets to the model.
* Using ‘intercept\_’ and ‘coef\_’ methods the Intercept is calculated as [2.87696662] and Coefficients is obtained as [[0.04656457 0.17915812 0.00345046]].
* Using .predict() function, the outcome for the testing dataset is predicted.

**6. Calculate the Mean Square Error (MSE)**

* From sklearn import metrics, and use the Mean\_Squared\_Error function to calculate the MSE.
* The calculated the MSE(Mean Squared Error) value is 1.9730456202283364.
* Variance Score obtained is 0.9156213613792233.