# **Sales Prediction**

# ( Simple Linear Regression)

**Problem Statement**

Build a model which predicts sales based on the money spent on different platforms for marketing.

**Data**

Use the advertising dataset given in ISLR and analyse the relationship between 'TV advertising' and 'sales' using a simple linear regression model.

1. Load the data from ‘ex1data1.txt’. The first column is the population of a city and the second column is the profit of a food truck in that city. A negative value for profit indicates a loss

Perform the following:

a. Visualize the data using the appropriate plot.

b. Print the description of the data.

c. Check if attributes have a linear relationship, and apply the Linear Regression model. (Train/Test split = 80/20)

d. Find and print the regression parameters.

e. Find the goodness of the model using R-squared (Co-efficient of Determination).

f. Compute MSE for the instances in the test set.

2. Load the data from ‘ex1data2.txt’ contains a training set of housing prices in Portland, Oregon. The first column is the size of the house (in square feet), the second column is the number of bedrooms, and the third column is the price of the house.

Perform the following:

a. Print the description of the data.

b. Apply the Linear Regression model. (Train/Test split = 80/20)

c. Find and print the regression parameters.

d. Find the goodness of the model using R2.

e. Compute MSE for the instances in the test set.

3. Load the data from ‘canada\_per\_capita\_income.csv’. Use this to build a regression model and predict the per capita income for Canadian citizens in the year 2020. Predict Canada’s per capita income in the year 2020. (Expected Output - 41288.69409442)

4. Suppose that you are the administrator of a university department and you want to determine each applicant’s chance of admission based on their results on two exams. You have historical data (“ex2data1.txt”) from previous applicants that you can use as a training set for logistic regression. For each training example, you have the applicant’s scores on two exams and the admissions decision. Your task is to build a classification model that estimates an applicant’s probability of admission based the scores from those two exams.

5. Load the data from ‘HR\_comma\_sep.csv’. Use this to build a logistic regression model and compute the accuracy of model.