# Activity–Multiple Linear Regression

## Learning outcomes: Multiple Linear Regression

After completing this exercise, you should be able to understand and perform below tasks.

1. Building predictive model - Regression model using multiple linear regression technique.
2. Validating the model results and optimizing the model.
3. Evaluation of error metrics.
4. Applying the model on un-seen data.
    a. Splitting data into train and test data sets.
    b. Comparing the error metrics.
5. Interpretation of the results
6. Handling multicollinearity and dimensionality reduction
7. Stepwise Regression

### Problem Statement:

A large child education toy company which sells edutainment tablets and gaming systems both online and in retail stores wanted to analyze the customer data. They are operating from last few years and maintaining all transactional information data. The given data

'CustomerData.csv' is a sample of customer level data extracted and processed for the analysis from various set of transactional files.

The objectives of today's activity are

> Building a regression model to predict the customer revenue based on other factors and understand the influence of other attributes on revenue

### Steps:

### Regression Model:

1. Read the data 'CustomerData.csv' into R.
2. Understand the structure of the data and pre-process
    a. Drop the attribute 'CustomerID'
    b. Convert 'City' as factor variable
3. Split the data into train and test data sets

```
rows=seq(1,nrow(data),1)
set.seed(123)
trainRows=sample(rows,(70*nrow(data))/100)
```

```
train = data[trainRows,]

test = data[-trainRows,]
```

4. Build linear regression and interpret the results #Input attributes by selection

```
LinReg<-lm(TotalRevenueGenerated~City+NoOfChildren+

Tenure+NoOfUnitsPurchased, data=train) summary(LinReg)


#Input all attributes into model

LinReg1<- lm(TotalRevenueGenerated ~ ., data=train)
summary(LinReg1)
# study the results of LinReg, LinReg1 models
```

5. Error metrics evaluation on train data and test data

```
library(DMwR)

#Error verification on train data
regr.eval(train$TotalRevenueGenerated,
LinReg$fitted.values) #Error verification on test data

Pred<-predict(LinReg,test)
regr.eval(test$TotalRevenueGenerated, Pred)
```

6. Spend the time on understanding the summary of the results and perform experiments with multiple combinations of attributes (by dropping the attributes which are not significant)

## II. Assignment:

Consider the dataset that depicts the wine-quality and its independent variables details.

a. Read the data into R
b. Do the required preprocessing steps
c. Build the model with all the attributes and study the model and then decide the influencing attributes. Build the model with those selected attributes. Check the results.
d. Split the dataset into train and test datasets

e. Build the model with the selected attributes on train dataset
f. Predict on train and test datasets
g. Compute the error metrics on train and test datasets