A cloth manufacturing company is interested to know about the segment or attributes causes high sale.

Approach - A Random Forest can be built with target variable Sales (we will first convert it in categorical variable) & all other variable will be independent in the analysis.

**Ans :**

**R Code :**

## Random Forest

########## Company Data Set #########

install.packages("randomForest")

library(randomForest)

Company\_Data\_r <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Random Forests\\Company\_Data\_r.csv')

model<-randomForest(Company\_Data\_r$Target.Sales~.,data=Company\_Data\_r,ntree=1000)

# View the forest results.

print(model)

#Importance of the variable - Lower Gini

print(importance(model))

#Prediction

pred<- predict(model,Company\_Data\_r[,-12])

table(pred,Company\_Data\_r$Target.Sales)

**Results :**

> model<-randomForest(Company\_Data\_r$Target.Sales~.,data=Company\_Data\_r,ntree=1000)

> # View the forest results.

> print(model)

Call:

randomForest(formula = Company\_Data\_r$Target.Sales ~ ., data = Company\_Data\_r, ntree = 1000)

Type of random forest: classification

Number of trees: 1000

No. of variables tried at each split: 3

OOB estimate of error rate: 0.5%

Confusion matrix:

High Low Medium class.error

High 79 0 0 0.000000000

Low 0 77 0 0.000000000

Medium 1 1 242 0.008196721

> #Importance of the variable - Lower Gini

> print(importance(model))

MeanDecreaseGini

Sales 150.4084833

CompPrice 6.6028814

Income 6.5253149

Advertising 6.0149788

Population 6.3749124

Price 16.2377282

ShelveLoc 14.2818014

Age 8.2079814

Education 3.5404514

Urban 0.5547058

US 1.0634071

> #Prediction

> pred<- predict(model,Company\_Data\_r[,-12])

> table(pred,Company\_Data\_r$Target.Sales)

pred High Low Medium

High 79 0 0

Low 0 77 0

Medium 0 0 244

**Inference :**

Sales was found as the important feature as it was having the highest Index.