1. Objective: Build Linear models along with difference regularizations to predict total revenue generated by a customer.

2. Understanding Cross-Validation

Steps:

1. Load Data into R
2. Data preparation
   a. Check the structure and summary of the dataset
   b. Check for missing values
   c. Convert necessary columns into factors
3. Split the Data into test and train using "createDataPartition" from caret package.
4. Standardize the Data by using the function "PreProcess" from caret package.

```
PreProc=preProcess(train[,setdiff(names(train),"TotalRevenueGenerated")],method = c("center", "scale"))

train<-predict(preProc,train)
test<-predict(preProc,test)
```

5. Create dummy variables for factors using function "dummyVars" from caret package

```
dummies <- dummyVars(TotalRevenueGenerated~., data = train)


x.train=predict(dummies, newdata = train)
y.train=train$TotalRevenueGenerated
x.test = predict(dummies, newdata = test)
y.test = test$TotalRevenueGenerated
```

6. Build linear regression models with different regularizations.

a) Lasso regression

- Build a linear model with Lasso regression (alpha=1)

```
fit.lasso <- glmnet(x.train, y.train, family="gaussian", alpha=1)
```

- Build a cross- validated linear model with Lasso regression

```
fit.lasso.cv <- cv.glmnet(x.train, y.train, type.measure="mse", alpha=1, family="gaussian",nfolds=10,parallel=TRUE)
```

- Plot the above two models and understand the summary

  plot(fit.lasso, xvar="lambda")
  plot(fit.lasso.cv)

- Understand the coefficients of the model

  coef(fit.lasso.cv,s = fit.lasso.cv$lambda.min)

- Predict on test using the best model picked from cross validated model

  pred.lasso.cv.train <- predict(fit.lasso.cv,x.train,s = fit.lasso.cv$lambda.min)
  pred.lasso.cv.test <- predict(fit.lasso.cv,x.test,s = fit.lasso.cv$lambda.min)

- Report the metrics on test and train set.

b) Repeat the above steps for ridge regression (alpha=1)

c) Repeat the above steps for elastic regression (alpha=0.5)