dataY<-read.csv("E:/VBA/RegressionExample.csv")

names(dataY)

str(dataY)

dataY$City<-NULL

dataY$Rider<-as.numeric(dataY$Rider)

dataY$Population<-as.numeric(dataY$Population)

dataY$Mincome<-as.numeric(dataY$Mincome)

table(dataY$Climate)

dataY$Climate\_Num<-ifelse(dataY$Climate=='Sunny',1,ifelse(dataY$Climate=='Cold',2,3))

dataY$Climate<-NULL

dataY$Climate\_Num<-as.numeric(dataY$Climate\_Num)

str(dataY)

#to check outliers

boxplot(dataY$Rider)

boxplot(dataY$Price)

boxplot(dataY$Population)

boxplot(dataY$Mincome)

boxplot(dataY$Prate)

#treatment of outlier for Price per week

summary(dataY$Price)

upper<-75+1.5\*IQR(dataY$Price);upper

dataY$Price[dataY$Price>upper]<-upper #replace Outlierars

boxplot(dataY$Price)

summary(dataY$Price)

summary(dataY$Prate)

upper<-140+1.5\*IQR(dataY$Prate);upper

dataY$Prate[dataY$Prate>upper]<-upper #replace Outlierars

boxplot(dataY$Prate)

summary(dataY$Prate)

#data subset

abcd<-dataY[,-c(6)] #delete columne from data and in new data

names(abcd)

#data Partition

library(caret)

TrainY<-createDataPartition(abcd$Price,p=0.70,list=FALSE)

trainingY<-abcd[TrainY,] #match row of 70% and make in new data

testingY<-abcd [-TrainY,]

str(trainingY)

#model building and may leads to overfitting

str(trainingY)

cor(trainingY)

summary(trainingY)

model2<-step(lm(Rider~.,data=trainingY),direction = "both")

summary(model2) #showing 100% r thats why its overfitting

#assumption

par(mfrow=c(2,2))

plot(model2)

library(lmtest)

dwtest(model2)

library(car)

ncvTest(model2)