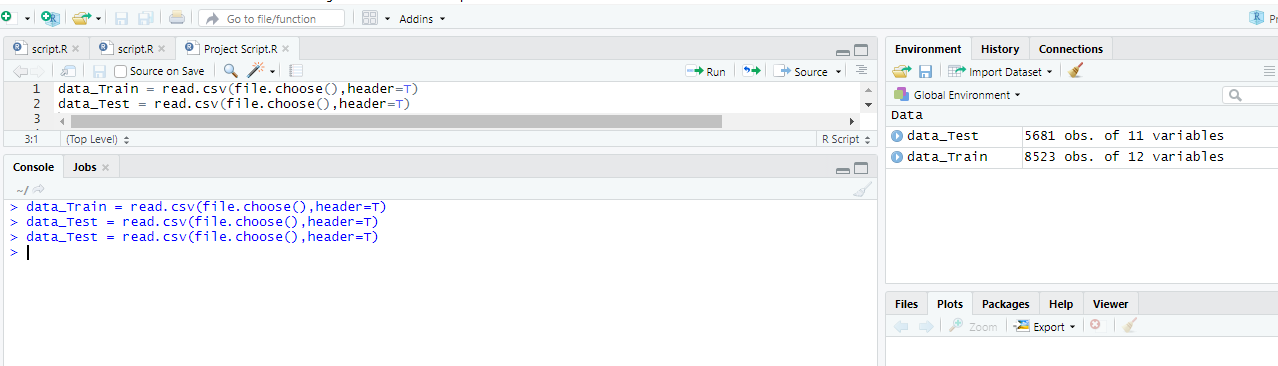
**RESULTS**

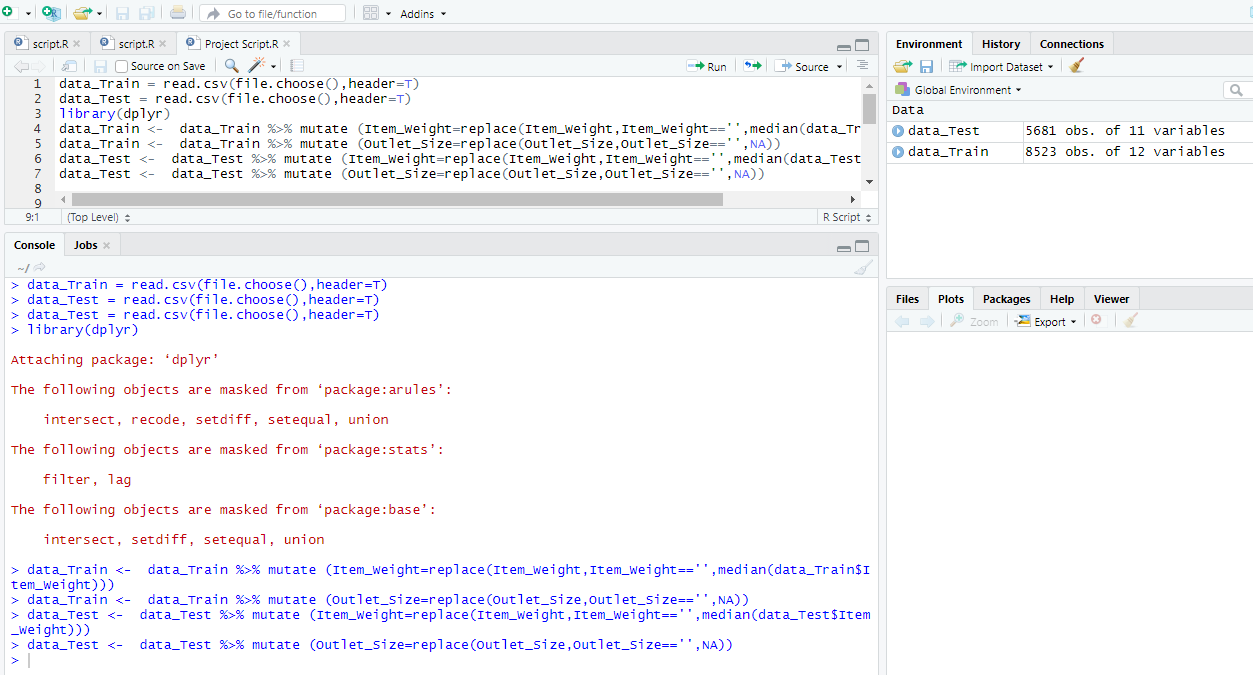
Step 1:

Loaded training and Test data into R studio



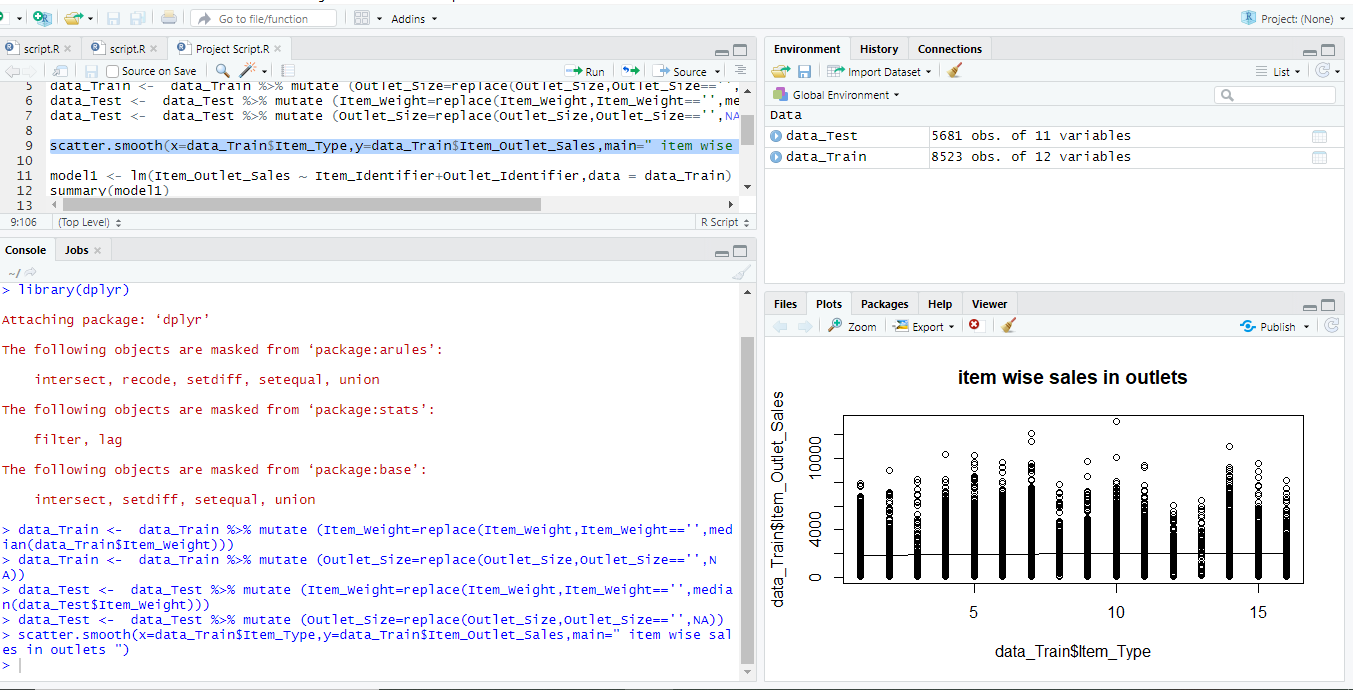
Step 2:

As there are some missing values in both test and train data sets, we used below commands to replace missing values



Step 3: Make a scatter plot on how sales price depends and Item\_Type related

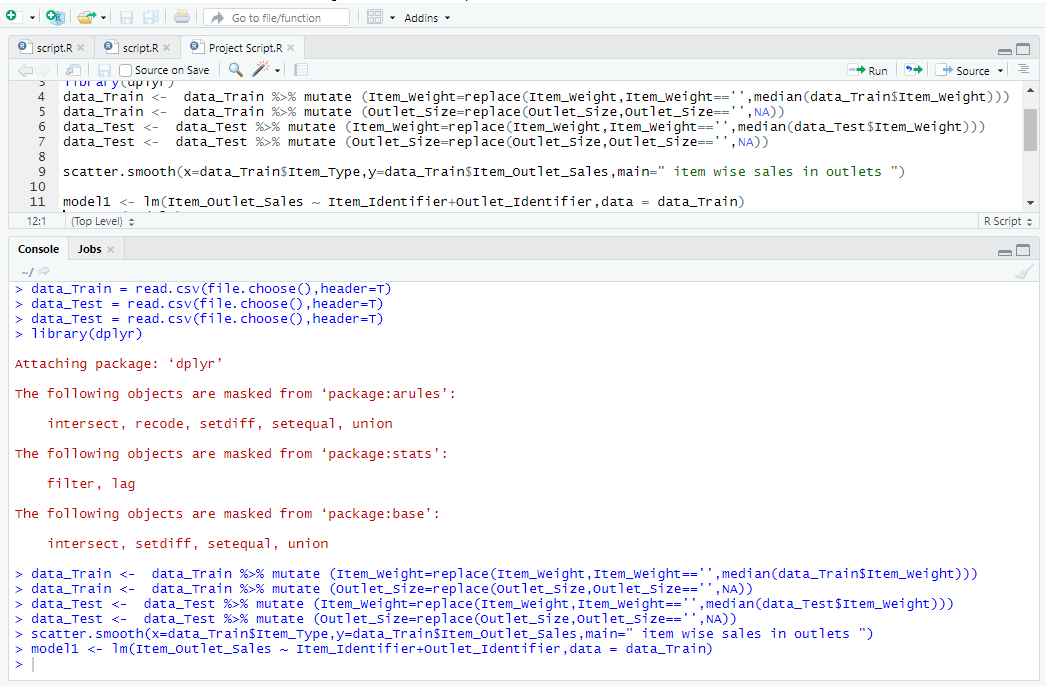
scatter.smooth(x=data\_Train$Item\_Type,y=data\_Train$Item\_Outlet\_Sales,main=" item wise sales in outlets ")



Step 4:

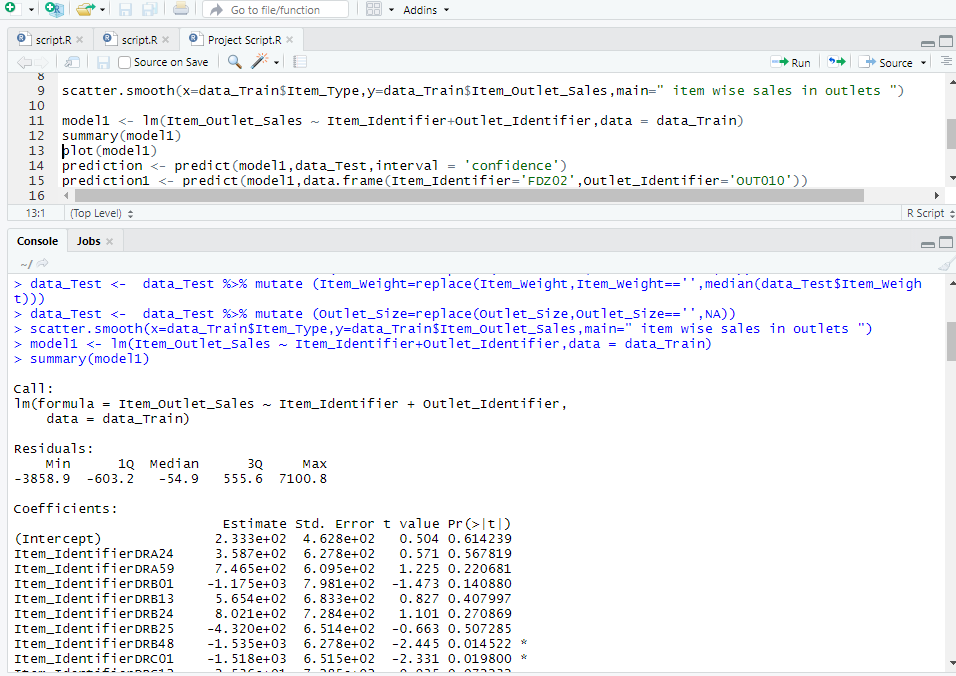
Create a Linear regression model to get Item\_Outlet\_Sales based on Item\_Identifier and Outlet\_Identifier

model1<- lm(Item\_Outlet\_Sales ~ Item\_Identifier+Outlet\_Identifier,data = data\_Train)



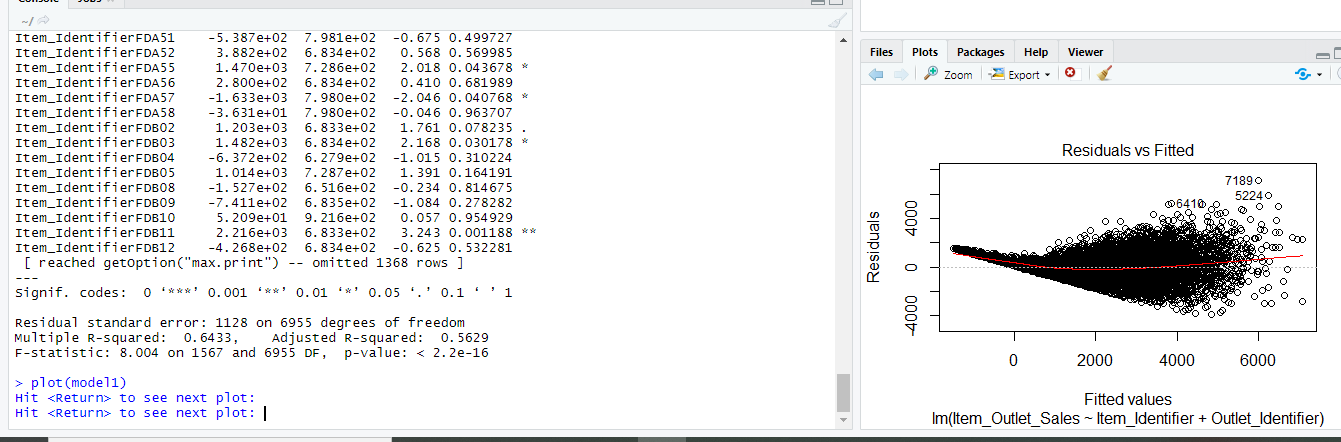
6)Display the summary of model we created

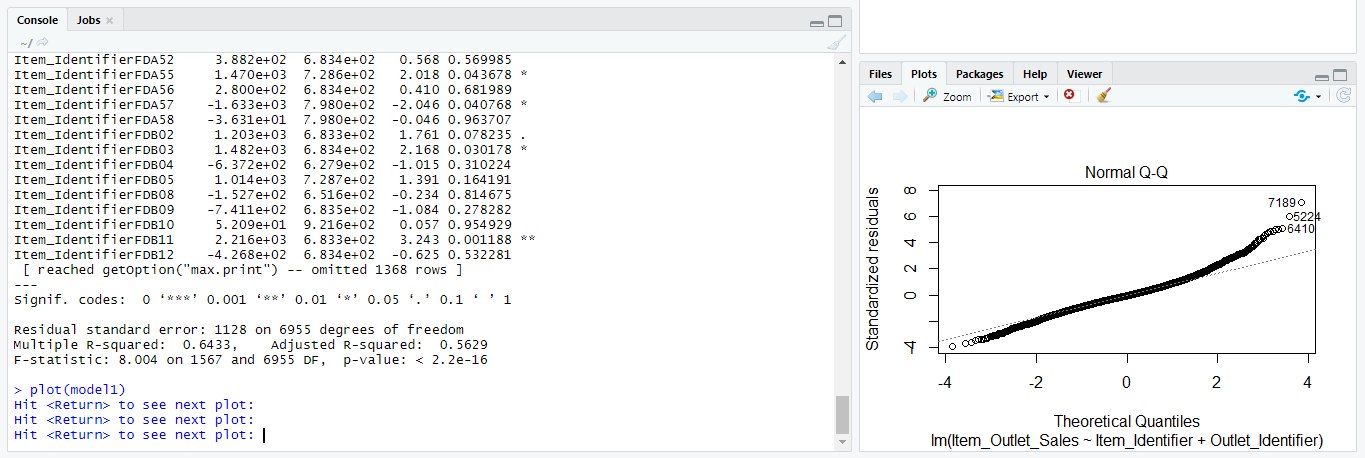
summary(model1)

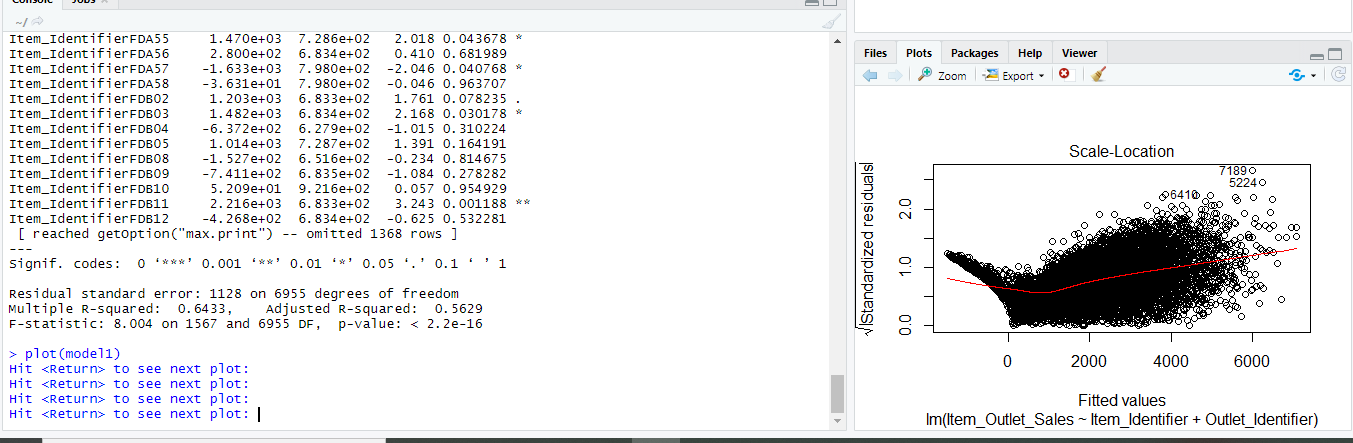


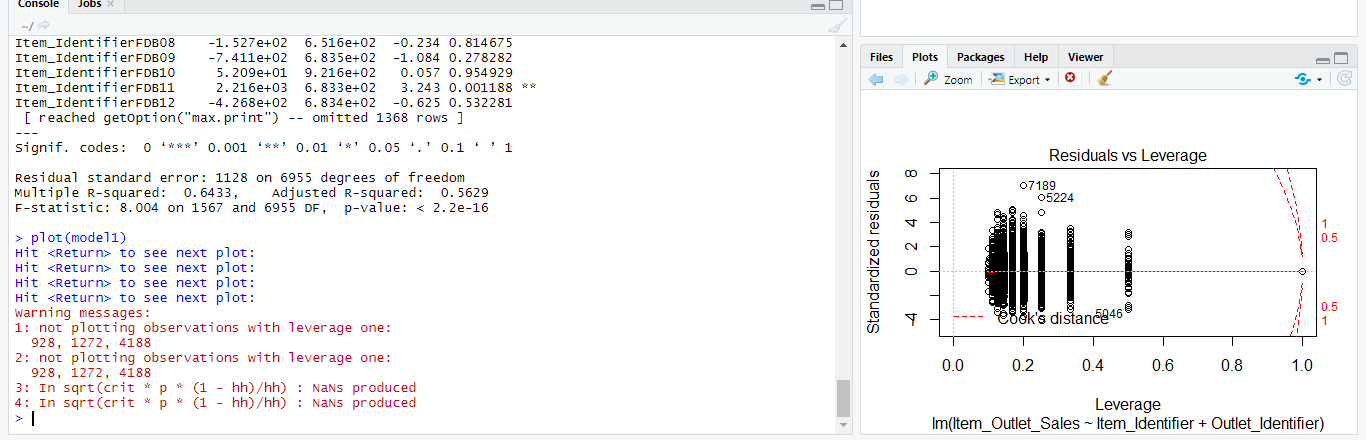
7)plot the model to display the graphical representation

plot(model1)



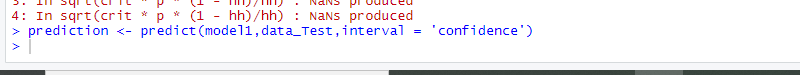






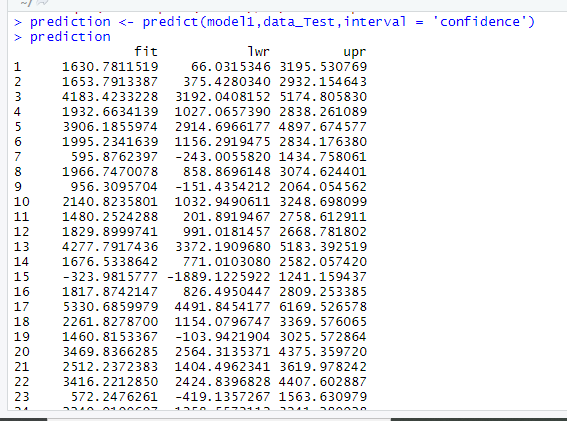
8)Predict the sales from test data set using below predict function. It uses sales values from training data set and it does predict the sales for test data.

prediction <- predict(model1,data\_Test,interval = 'confidence')



9)Display the values what we have prediction .It does give three values. One is approx value, econd one is minimum value and third is maximum value.

> prediction



10) write the values which were predicted into csv format

setwd("~/data mining/project")

write.csv(prediction,"predictionresult.csv")