**Q7: Linear Regression model**

library(pROC)

library(corrplot)

library(caTools)

library(caret)

library(e1071)

library(rpart)

library(ROCR)

library(reshape2)

library(dplyr)

library(tidyr)

library(magrittr)

library(BSDA)

#Setting working directory

workdir<-"F:\Arpit\Arpit\_Stuff1"

setwd(workdir)

# Import DM\_PRJ\_Q7\_DATA.csv

AvgFare=read.csv("DM\_PRJ\_Q7\_DATA.csv", header=TRUE,sep=",")

# Summary for AvgFare

summary(AvgFare)

#Defining variables.

Duration=AvgFare$Ride.Duration

Days=AvgFare$Days

Days=AvgFare$Days

Category=AvgFare$CATEGORY

Start=AvgFare$START

Stop=AvgFare$STOP

Miles=AvgFare$MILES

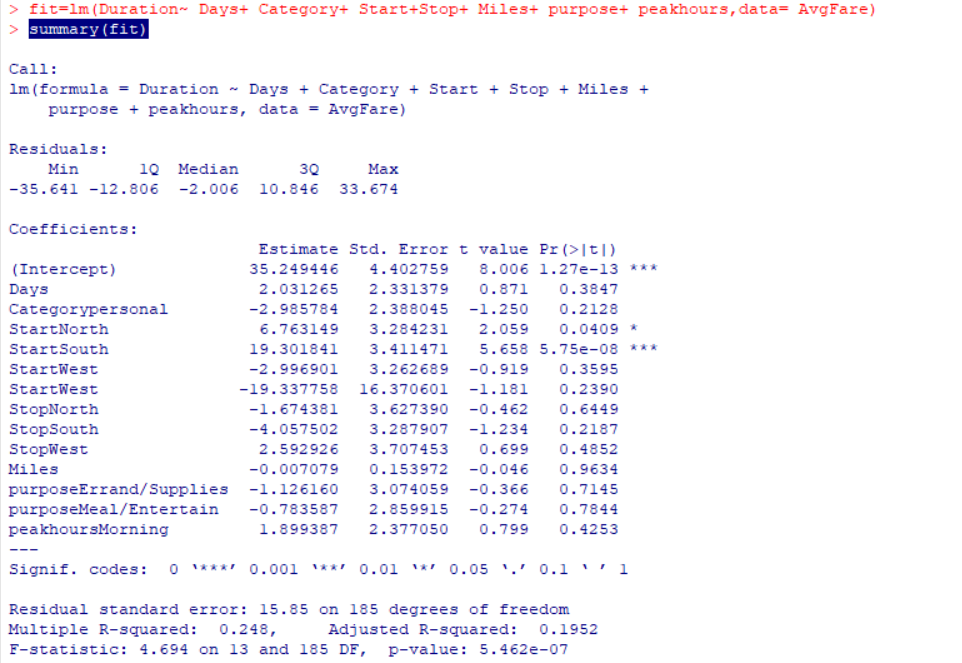
purpose=AvgFare$PURPOSE

peakhours=AvgFare$Peak.hours

# To build a model(fit)

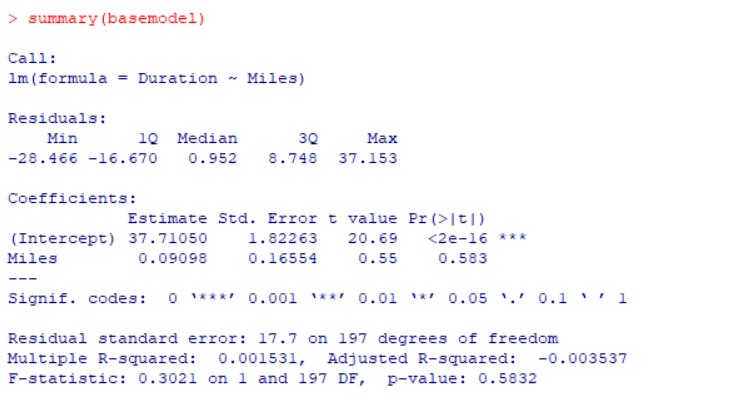
fit=lm(Duration~ Days+ Category+ Start+Stop+ Miles+ purpose+ peakhours,data= AvgFare)

summary(fit)



# Build a base model

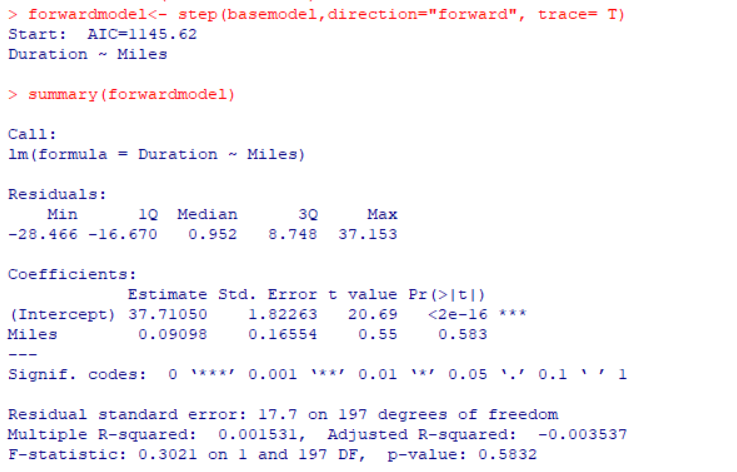
basemodel= lm(Duration~Miles)



# Build a forward model using step function

forwardmodel<- step(basemodel,direction="forward", trace= T)

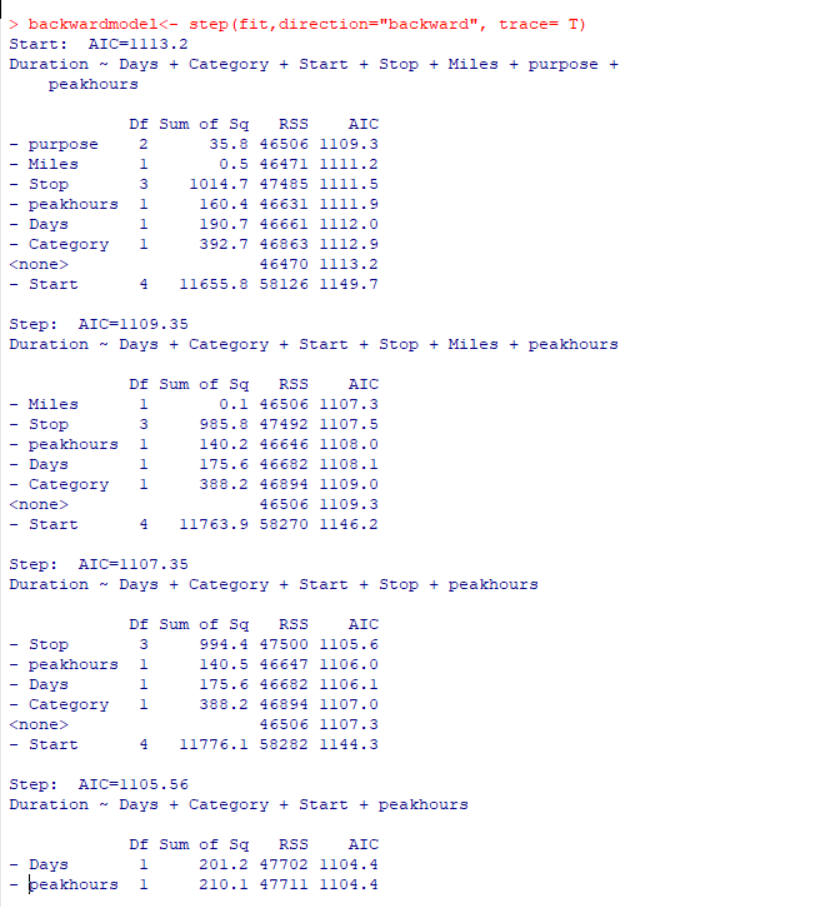
summary(forwardmodel)

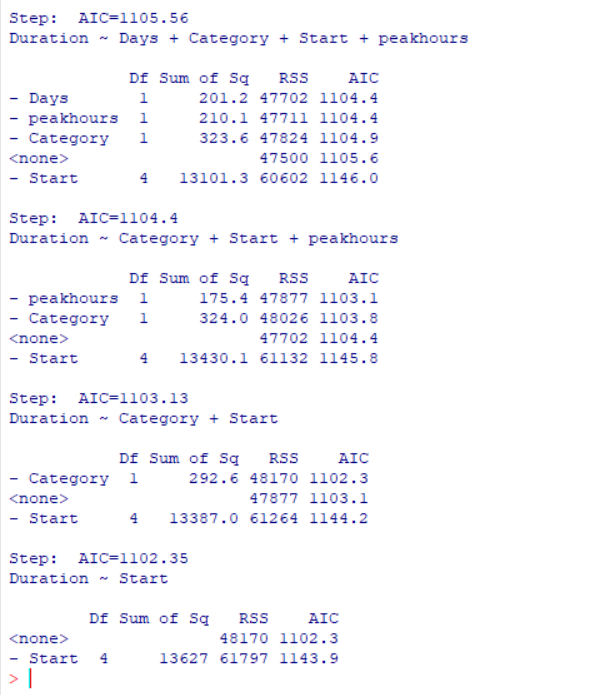


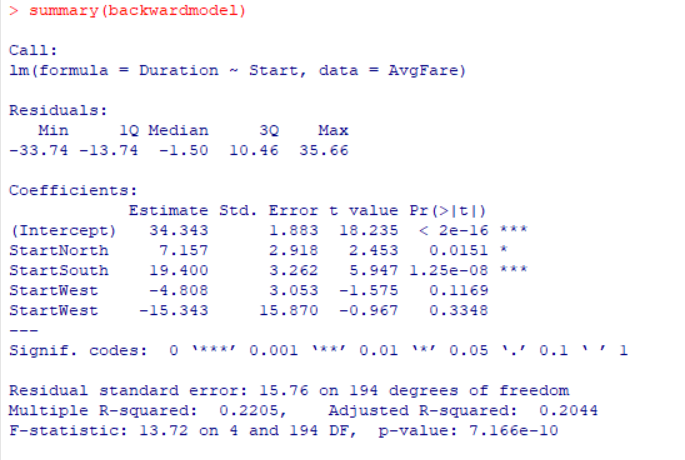
# Build a backward model using step function

backwardmodel<- step(fit,direction="backward", trace= T)

summary(backwardmodel)





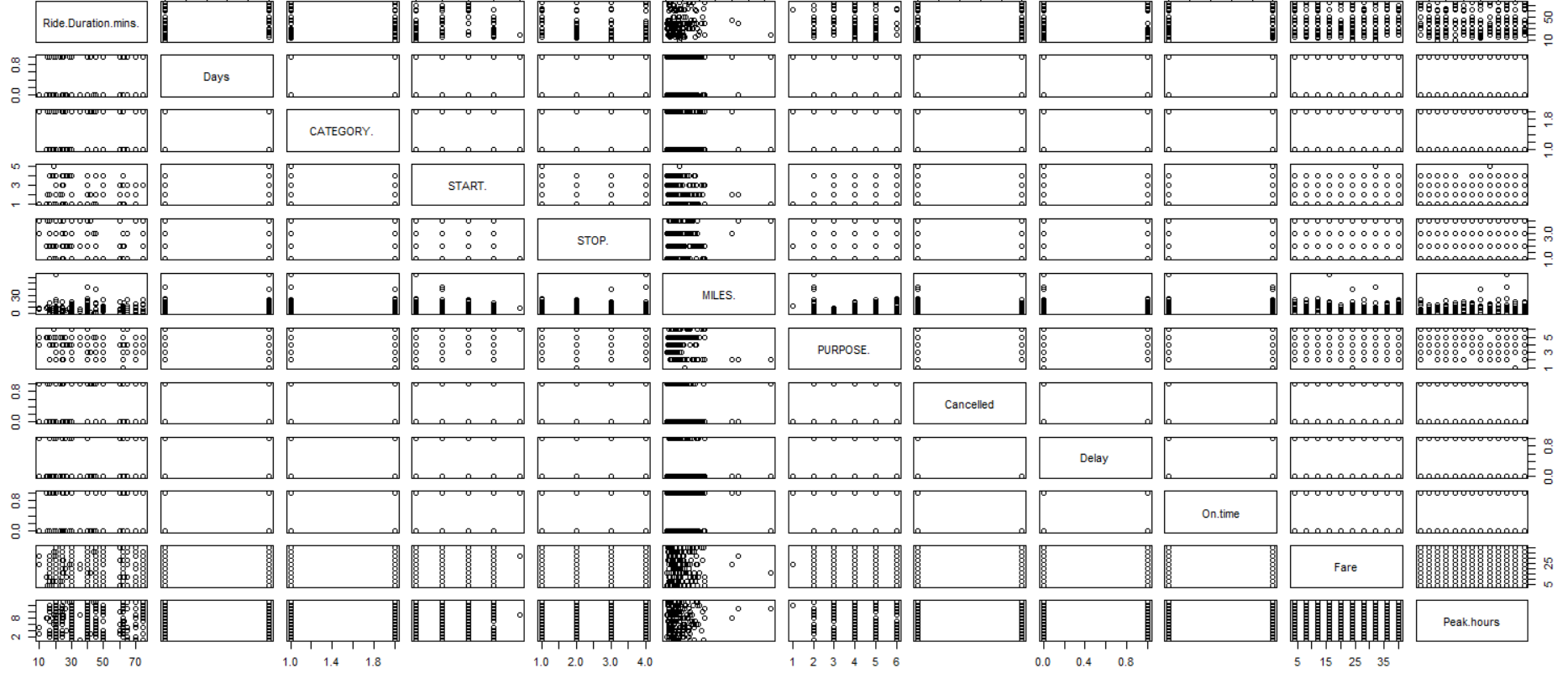


Comparison of different models will be done with the help of AIC values:

We found that AIC value is less for backward model compared to forward model, Hence model having lowest AIC is better model, in our case backward model is better than forward model.

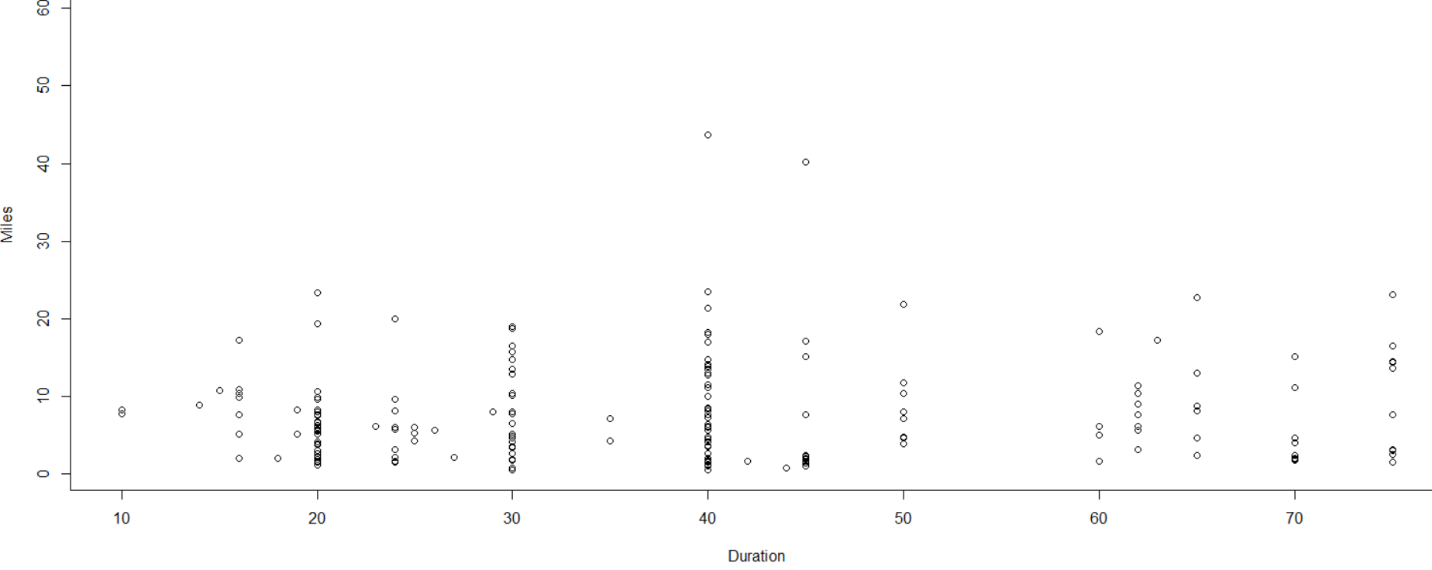
# Scatter Plot of Model

Plot(AvgFare)

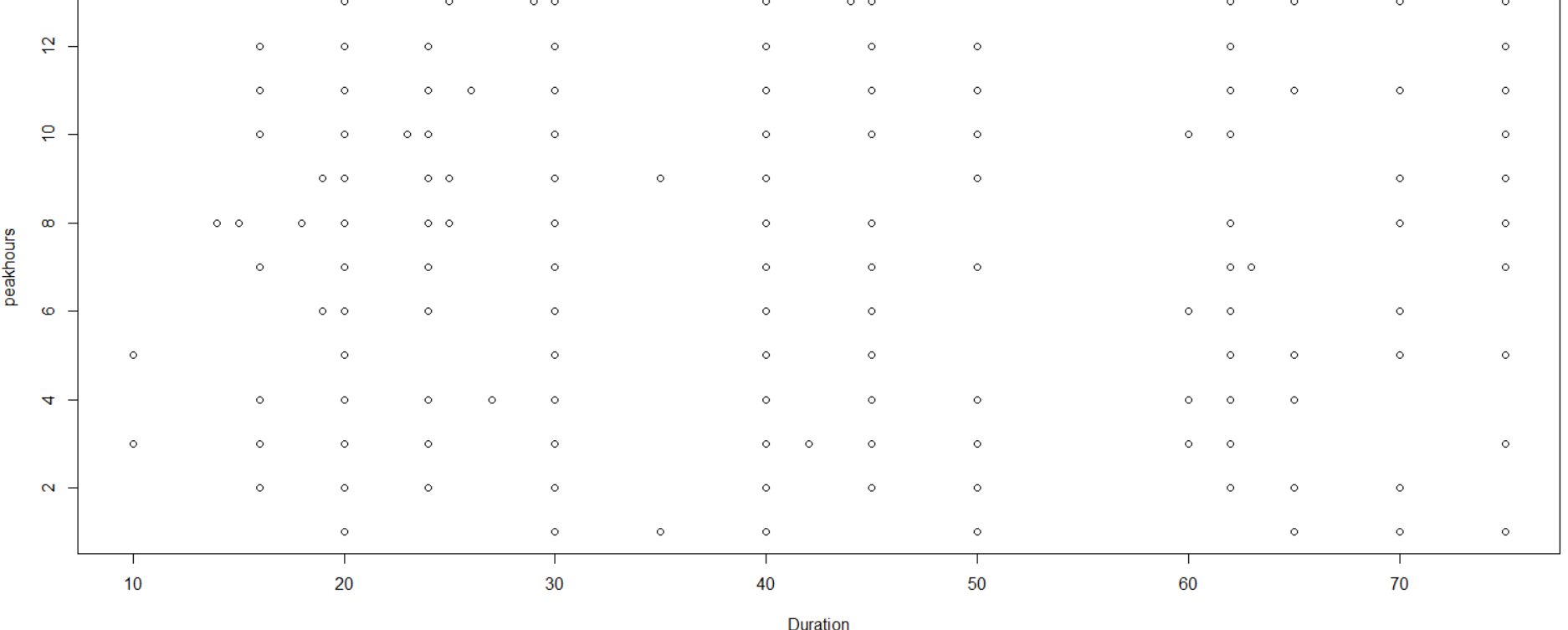


**We will examine the relationship between duration and miles,**

**Plot(Duration,Miles)**

****

**Plot(Duration,Peakhours)**

****