#RCODE

# First get the training and testing data set from folder

train<-read.csv("investmentsData.csv")

test <- read.csv("samplesubmission.csv")

test$Date <- 7

library(rpart)

#Normalize the variables that you think are varying too much and also separate the outliers.

train$Adv <- (train$Advisor\_Id - mean(train$Advisor\_Id)) / sd(train$Advisor\_Id)

test$Adv <- (test$Advisor\_Id - mean(train$Advisor\_Id)) / sd(train$Advisor\_Id)

test$Morningstar.Category <- test$Transaction\_Type

train$Morningstar.Category <- as.integer(as.factor(train$Morningstar.Category))

test$Morningstar.Category <- as.integer(as.factor(test$Morningstar.Category))

train1 <- subset(train,train$Morningstar.Category == 1)

train2 <- subset(train,train$Morningstar.Category != 1)

test2 <- subset(test,test$Morningstar.Category != 1)

test1 <- subset(test,test$Morningstar.Category == 1)

test1$zVar1 <- (test1$Investment\_Id - mean(train1$Investment\_Id)) / sd(train1$Investment\_Id)

train1$zVar1 <- (train1$Investment\_Id - mean(train1$Investment\_Id)) / sd(train1$Investment\_Id)

test2$zVar2 <- (test2$Investment\_Id - mean(train2$Investment\_Id)) / sd(train2$Investment\_Id)

train2$zVar2 <- (train2$Investment\_Id - mean(train2$Investment\_Id)) / sd(train2$Investment\_Id)

#Start predicting the things which are not given in test dataset on basis of data of training dataset

model <- glm(Rating ~ Adv+Morningstar.Category + zVar2 + Date,data = train2)

pred <- predict(model,newdata = test2)

test2$Rating <- round(pred)

test1$Rating <- " "

model1 <- glm(X1.Yr...Rank ~ Adv+zVar2+Rating + Morningstar.Category + Date, data = train2)

pred <- predict(model1,newdata = test2)

test2$X1.Yr...Rank<-round(pred)

model1 <- glm(X3.Yr...Rank ~Adv+zVar2+X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train2)

pred <- predict(model1,newdata = test2)

test2$X3.Yr...Rank<-round(pred)

model1 <- glm(X5.Yr...Rank ~ Adv+zVar2+X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train2)

pred <- predict(model1,newdata = test2)

test2$X5.Yr...Rank<-round(pred)

############################################################

train3 <- subset(train2,is.na(train2$X10.Yr...Rank))

train4 <- subset(train2,!is.na(train2$X10.Yr...Rank))

a<-c(intersect(test2$Investment\_Id,train3$Investment\_Id))

a<-as.vector(a)

test3<-subset(test2,test2$Investment\_Id == 106023 | test2$Investment\_Id ==106364 | test2$Investment\_Id ==106366 |test2$Investment\_Id == 111277 | test2$Investment\_Id ==120816 | test2$Investment\_Id ==120819 | test2$Investment\_Id ==120821 |test2$Investment\_Id == 120823 | test2$Investment\_Id ==122057 )

b<-c(intersect(test2$Investment\_Id,train4$Investment\_Id))

b<-as.vector(b)

test4<-subset(test2,test2$Investment\_Id == 105561 |test2$Investment\_Id == 105564 |test2$Investment\_Id == 106020 )

test3$X10.Yr...Rank<- " "

model4 <- glm(X10.Yr...Rank ~ Adv+zVar2+X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model4,newdata = test4)

test4$X10.Yr...Rank<-pred

model5 <- glm( X1.Yr.Return ~ Adv+zVar2+ X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$X1.Yr.Return<-pred

model5 <- glm( X3.Yr.Return ~ Adv+zVar2 + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$X3.Yr.Return <-pred

model5 <- glm( X5.Yr.Return ~ Adv+zVar2+ X3.Yr.Return + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$X5.Yr.Return <-pred

model5 <- glm( X10.Yr.Return ~Adv+ zVar2 + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$X10.Yr.Return <-pred

model5 <- glm( X1.Yr.Return ~ Adv+zVar2 + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3)

pred <- predict(model5,newdata = test3)

test3$X1.Yr.Return<-pred

model5 <- glm( X3.Yr.Return ~ Adv+zVar2 + X1.Yr.Return + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3)

pred <- predict(model5,newdata = test3)

test3$X3.Yr.Return <-pred

model5 <- glm( X5.Yr.Return ~ Adv+zVar2 + X3.Yr.Return + X1.Yr.Return + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3)

pred <- predict(model5,newdata = test3)

test3$X5.Yr.Return <-pred

test3$X10.Yr.Return <- " "

######################################

train4$zVar3 <- (train4$Net.Flows - mean(train4$Net.Flows)) / sd(train4$Net.Flows)

train3$zVar4 <- (train3$Net.Flows - mean(train3$Net.Flows)) / sd(train3$Net.Flows)

model5 <- glm( zVar3 ~ X10.Yr.Return +Adv+ zVar2 + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$zVar3<-pred

model5 <- glm( zVar4 ~ X5.Yr.Return +Adv+ zVar2 + X3.Yr.Return + X1.Yr.Return + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3)

pred <- predict(model5,newdata = test3)

test3$zVar4 <-pred

####################################

model5 <- lm( Code\_1 ~ zVar3 +Adv+ zVar2+ X10.Yr.Return + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$Code\_1 <-round(pred)

model5 <- lm( Code\_1 ~ zVar4+ zVar2 +Adv+ X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3)

pred <- predict(model5,newdata = test3)

test3$Code\_1 <-round(pred)

####################################3

model5 <- lm( Code\_2 ~ Code\_1 + zVar3 +Adv+ zVar2 + X10.Yr.Return + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$Code\_2 <-round(pred)

model5 <- lm(Code\_2 ~ Code\_1 + zVar4 +Adv+ zVar2+ X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3)

pred <- predict(model5,newdata = test3)

test3$Code\_2 <-round(pred)

###############################

model5 <- lm(Code\_3 ~ Code\_2 + Code\_1 +Adv+ zVar3 + zVar2 + X10.Yr.Return + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$Code\_3 <-round(pred)

model5 <- lm(Code\_3 ~ Code\_2 + Code\_1 +Adv+ zVar4 + zVar2 + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3)

pred <- predict(model5,newdata = test3)

test3$Code\_3 <-round(pred)

#####################################

model5 <- lm(Code\_4 ~ Code\_3 + Code\_2 +Adv+ Code\_1 + zVar3 + zVar2 + X10.Yr.Return + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$Code\_4 <-round(pred)

model5 <- lm(Code\_4 ~ Code\_3 + Code\_2 +Adv+ Code\_1 + zVar4 + zVar2 + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3)

pred <- predict(model5,newdata = test3)

test3$Code\_4 <-round(pred)

#####################################

model5 <- lm(Code\_5 ~ Code\_4 + Code\_3 +Adv+ Code\_2 + Code\_1 + zVar3 + zVar2 + X10.Yr.Return + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$Code\_5 <-round(pred)

model5 <- lm(Code\_5 ~ Code\_4 + Code\_3 + Code\_2 +Adv+ Code\_1 + zVar4 + zVar2 + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3)

pred <- predict(model5,newdata = test3)

test3$Code\_5 <-round(pred)

model5 <- lm(Amount ~ Code\_5 + Code\_4 + Adv+Code\_3 + Code\_2 + Code\_1 + zVar3 + zVar2 + X10.Yr.Return + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4)

pred <- predict(model5,newdata = test4)

test4$Amount <-pred

model5 <- lm(Amount ~ Code\_5 + Code\_4 +Adv+ Code\_3 + Code\_2 + Code\_1 + zVar4 + zVar2 + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3)

pred <- predict(model5,newdata = test3)

test3$Amount<-pred

train4$Amt <- (train4$Amount - mean(train4$Amount)) / sd(train4$Amount)

test4$Amt <- (test4$Amount - mean(train4$Amount)) / sd(train4$Amount)

train3$Amt <- (train3$Amount - mean(train3$Amount)) / sd(train3$Amount)

test3$Amt <- (test3$Amount - mean(train3$Amount)) / sd(train3$Amount)

model5 <- glm(Transaction\_Type ~ Amt+ Code\_5 + Code\_4 + Adv+Code\_3 + Code\_2 + Code\_1 + zVar3 + zVar2 + X10.Yr.Return + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X10.Yr...Rank + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train4,family = binomial)

pred <- predict(model5,newdata = test4,type="response")

test4$Transaction\_Type1 <-pred

model5 <- glm(Transaction\_Type ~ Amt+Code\_5 + Code\_4 +Adv+ Code\_3 + Code\_2 + Code\_1 + zVar4 + zVar2 + X5.Yr.Return + X3.Yr.Return + X1.Yr.Return + X5.Yr...Rank + X3.Yr...Rank + X1.Yr...Rank+ Rating + Morningstar.Category + Date, data = train3,family = binomial)

pred <- predict(model5,newdata = test3,type="response")

test3$Transaction\_Type1 <-pred

test3$Transaction\_Type1[which(test3$Transaction\_Type1<0.5)]<-0

test4$Transaction\_Type1[which(test4$Transaction\_Type1<0.5)]<-0

test3$Transaction\_Type1[which(test3$Transaction\_Type1>=0.5)]<-1

test4$Transaction\_Type1[which(test4$Transaction\_Type1>=0.5)]<-1

######################################################

model5 <- lm( Code\_1 ~ Adv+ zVar1+Morningstar.Category + Date, data = train1)

pred <- predict(model5,newdata = test1)

test1$Code\_1 <-round(pred)

########################################################

model5 <- lm( Code\_2 ~ Code\_1 + Adv+ zVar1+Morningstar.Category + Date, data = train1)

pred <- predict(model5,newdata = test1)

test1$Code\_2 <-round(pred)

#################################################

model5 <- lm( Code\_3 ~ Code\_2 + Code\_1 + Adv+ zVar1+Morningstar.Category + Date, data = train1)

pred <- predict(model5,newdata = test1)

test1$Code\_3 <-round(pred)

##################################################

model5 <- lm(Code\_4 ~ Code\_3 + Code\_2 + Code\_1 + Adv+ zVar1+Morningstar.Category + Date, data = train1)

pred <- predict(model5,newdata = test1)

test1$Code\_4 <-round(pred)

#############################################

model5 <- lm(Code\_5 ~ Code\_4 + Code\_3 + Code\_2 + Code\_1 + Adv+ zVar1+Morningstar.Category + Date, data = train1)

pred <- predict(model5,newdata = test1)

test1$Code\_5 <-round(pred)

library(randomForest)

train1$Amount <- (train1$Amount - mean(train1$Amount)) / sd(train1$Amount)

model5 <- randomForest(Amount ~ Code\_5 + Code\_4 + Code\_3 + Code\_2 + Code\_1 + Adv+ zVar1+Morningstar.Category + Date, data = train1)

pred <- predict(model5,newdata = test1)

test1$Amount <-round(pred)

train1$Transaction\_Type<-as.factor(train1$Transaction\_Type)

model5 <- randomForest(Transaction\_Type ~ Amount + Code\_5 + Code\_4 + Code\_3 + Code\_2 + Code\_1 + Adv+ zVar1+Morningstar.Category + Date, data = train1)

pred <- predict(model5,newdata = test1)

test1$Transaction\_Type <-pred

#############################################

tst1<-test1[,c(1,2,3)]

tst3<-test3[,c(1,2,23)]

tst4<-test4[,c(1,2,23)]

tst1$Transaction\_Type<-as.integer(tst1$Transaction\_Type)

tst1$Transaction\_Type[which(as.integer(tst1$Transaction\_Type)==1)]<-0

tst1$Transaction\_Type[which(as.integer(tst1$Transaction\_Type)==2)]<-1

tst3$Amount[which(tst3$Amount>0)]<-0

tst3$Amount[which(tst3$Amount<0)]<-1

tst4$Amount[which(tst4$Amount>0)]<-0

tst4$Amount[which(tst4$Amount<0)]<-1

tst4$Transaction\_Type<-tst4$Amount

tst3$Transaction\_Type<-tst3$Amount

#Finally when all transaction\_type is predicted combeine all final dataset and get the final submission “.csv” file.

final<-rbind(tst1,subset(tst3[,c(1,2,4)]),subset(tst4[,c(1,2,4)]))