library(glmnet)

dat = read.csv("C:/Users/percy/Dropbox/Datatest/data1.csv")

dat = dat[1:11444,]

train = dat$Made.Year<2016

test = dat$Made.Year >=2016

summary(dat)

dat$TypeProduct\_Desc[dat$TypeProduct\_Desc=='Woven']='WOVEN'

dat$TypeProduct\_Desc[dat$TypeProduct\_Desc=='Knit']='KNIT'

dat$TypeProduct\_Desc[dat$TypeProduct\_Desc=='Shoes']='SHOES'

dat$TypeProduct\_Desc[dat$TypeProduct\_Desc=='Sweater']='SWEATER'

dat$STYLE[dat$STYLE=='Onepieces']='Onepiece'

linearModel = lm(eaCMCost~TypeProduct\_Desc+STYLE+Country\_Cd+Gender\_Desc+Selling.Season,data = dat,subset = train)

linearModel$rank

summary(linearModel)

PRED = predict(linearModel,dat[test,])

results = cbind(PRED,dat[test,])

sum((PRED-dat[test,]$eaCMCost)^2)/length(PRED)

head(dat)

head(linearModel$fitted.values,100)

head(dat$eaCMCost,100)

write.csv(results,file="C:/Users/percy/Desktop/results1.csv")

X = model.matrix(eaCMCost~TypeProduct\_Desc+STYLE+Country\_Cd+Gender\_Desc+Selling.Season+Retail\_SubClass\_Desc-1,data = dat)

Y = dat$eaCMCost

fit.lasso = glmnet(X,Y)

plot(fit.lasso, xvar = "lambda", label = TRUE)

cv.lasso = cv.glmnet(X, Y)

plot(cv.lasso)

coef(cv.lasso,s=1)

predict(cv.lasso, X, s = "lambda.1se")

dat$eaCMCost

results = cbind(predict(cv.lasso, X, s = "lambda.1se"),dat$eaCMCost)

write.csv(results,file="C:/Users/percy/Desktop/results1.csv")