LoanLogisticRegModel

Laura

# EDA

#### **Looking into the data**

#### **Statistics**

* ID can be dropped since it is not a useful predictor and just a unique identifier and will affect the results of the model
* Age mean is 45 years old, 23 has is the lowest and 67 is the highest. There seems to be a normal distribution.
* ***There are negative values in Experience which is odd because there can’t be negative years of experience. This will be looked at next***.
* The mean income is 73 while the median is 64 showing skewness which seems normal representation of society. The minimum income is 8,000 and the max is 224,000. These are common Salaries
* ZIP.Code should not be numberic they should be changed to categorical. There are 467 distinct ZIP codes
* There seems to be an equal distribution for families size 1,2,3,4. Each are compose of about 25%
* The CCAVG, Average Spending per 1000 goes from 0 to 10,000, however the median is 1,5000. Here we can see that there are outliers.
* For Education, 41% have have up to highschool, 28% up to under grad studies and interestingly 30% have up to grad school. That seems like a reasonable distribution.
* For mortgage we can see how skewed it is, the mean is 56.5 and the median is 0. Showing the extreme outliers. **This needs to be looked at**
* For the target variable, personal loan, we can see quite a difference, 90% without a loan and only 10% with loan. This makes sense because not many people take loan from banks
* Securities account also has a big difference where 90% does not have a security account while 10% does
* For CD.Account (Ceritificate deposit) once again we see the 94% does not have it while 6% has it.
* As for online (online banking capability), 40% doe not have it while 60% has it. That distribution is a little more balanced and makes sense.
* As for Credit card 70% does not have it while 30% has it.

PersonalLoan   
  
 14 Variables 5000 Observations  
--------------------------------------------------------------------------------  
ID   
 n missing distinct Info Mean Gmd .05 .10   
 5000 0 5000 1 2500 1667 251.0 500.9   
 .25 .50 .75 .90 .95   
 1250.8 2500.5 3750.2 4500.1 4750.1   
  
lowest : 1 2 3 4 5, highest: 4996 4997 4998 4999 5000  
--------------------------------------------------------------------------------  
Age   
 n missing distinct Info Mean Gmd .05 .10   
 5000 0 45 0.999 45.34 13.23 27 30   
 .25 .50 .75 .90 .95   
 35 45 55 61 63   
  
lowest : 23 24 25 26 27, highest: 63 64 65 66 67  
--------------------------------------------------------------------------------  
Experience   
 n missing distinct Info Mean Gmd .05 .10   
 5000 0 47 0.999 20.1 13.23 2 4   
 .25 .50 .75 .90 .95   
 10 20 30 36 38   
  
lowest : -3 -2 -1 0 1, highest: 39 40 41 42 43  
--------------------------------------------------------------------------------  
Income   
 n missing distinct Info Mean Gmd .05 .10   
 5000 0 162 1 73.77 50.91 18 22   
 .25 .50 .75 .90 .95   
 39 64 98 145 170   
  
lowest : 8 9 10 11 12, highest: 203 204 205 218 224  
--------------------------------------------------------------------------------  
ZIP.Code   
 n missing distinct Info Mean Gmd .05 .10   
 5000 0 467 1 93153 2042 90073 90275   
 .25 .50 .75 .90 .95   
 91911 93437 94608 95138 95670   
  
lowest : 9307 90005 90007 90009 90011, highest: 96091 96094 96145 96150 96651  
   
Value 9000 90000 91000 92000 93000 94000 95000 96000 97000  
Frequency 1 573 472 837 626 940 1117 428 6  
Proportion 0.000 0.115 0.094 0.167 0.125 0.188 0.223 0.086 0.001  
  
For the frequency table, variable is rounded to the nearest 1000  
--------------------------------------------------------------------------------  
Family   
 n missing distinct Info Mean Gmd   
 5000 0 4 0.934 2.396 1.279   
   
Value 1 2 3 4  
Frequency 1472 1296 1010 1222  
Proportion 0.294 0.259 0.202 0.244  
--------------------------------------------------------------------------------  
CCAvg   
 n missing distinct Info Mean Gmd .05 .10   
 5000 0 108 0.999 1.938 1.794 0.1 0.3   
 .25 .50 .75 .90 .95   
 0.7 1.5 2.5 4.3 6.0   
  
lowest : 0.0 0.1 0.2 0.3 0.4, highest: 8.8 8.9 9.0 9.3 10.0  
--------------------------------------------------------------------------------  
Education   
 n missing distinct Info Mean Gmd   
 5000 0 3 0.877 1.881 0.9073   
   
Value 1 2 3  
Frequency 2096 1403 1501  
Proportion 0.419 0.281 0.300  
--------------------------------------------------------------------------------  
Mortgage   
 n missing distinct Info Mean Gmd .05 .10   
 5000 0 347 0.668 56.5 88.16 0 0   
 .25 .50 .75 .90 .95   
 0 0 101 200 272   
  
lowest : 0 75 76 77 78, highest: 590 601 612 617 635  
--------------------------------------------------------------------------------  
Personal.Loan   
 n missing distinct Info Sum Mean Gmd   
 5000 0 2 0.26 480 0.096 0.1736   
  
--------------------------------------------------------------------------------  
Securities.Account   
 n missing distinct Info Sum Mean Gmd   
 5000 0 2 0.281 522 0.1044 0.187   
  
--------------------------------------------------------------------------------  
CD.Account   
 n missing distinct Info Sum Mean Gmd   
 5000 0 2 0.17 302 0.0604 0.1135   
  
--------------------------------------------------------------------------------  
Online   
 n missing distinct Info Sum Mean Gmd   
 5000 0 2 0.722 2984 0.5968 0.4814   
  
--------------------------------------------------------------------------------  
CreditCard   
 n missing distinct Info Sum Mean Gmd   
 5000 0 2 0.623 1470 0.294 0.4152   
  
--------------------------------------------------------------------------------

ID Age Experience Income ZIP.Code   
 Min. : 1 Min. :23.00 Min. :-3.0 Min. : 8.00 Min. : 9307   
 1st Qu.:1251 1st Qu.:35.00 1st Qu.:10.0 1st Qu.: 39.00 1st Qu.:91911   
 Median :2500 Median :45.00 Median :20.0 Median : 64.00 Median :93437   
 Mean :2500 Mean :45.34 Mean :20.1 Mean : 73.77 Mean :93152   
 3rd Qu.:3750 3rd Qu.:55.00 3rd Qu.:30.0 3rd Qu.: 98.00 3rd Qu.:94608   
 Max. :5000 Max. :67.00 Max. :43.0 Max. :224.00 Max. :96651   
 Family CCAvg Education Mortgage   
 Min. :1.000 Min. : 0.000 Min. :1.000 Min. : 0.0   
 1st Qu.:1.000 1st Qu.: 0.700 1st Qu.:1.000 1st Qu.: 0.0   
 Median :2.000 Median : 1.500 Median :2.000 Median : 0.0   
 Mean :2.396 Mean : 1.938 Mean :1.881 Mean : 56.5   
 3rd Qu.:3.000 3rd Qu.: 2.500 3rd Qu.:3.000 3rd Qu.:101.0   
 Max. :4.000 Max. :10.000 Max. :3.000 Max. :635.0   
 Personal.Loan Securities.Account CD.Account Online   
 Min. :0.000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
 Median :0.000 Median :0.0000 Median :0.0000 Median :1.0000   
 Mean :0.096 Mean :0.1044 Mean :0.0604 Mean :0.5968   
 3rd Qu.:0.000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:1.0000   
 Max. :1.000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
 CreditCard   
 Min. :0.000   
 1st Qu.:0.000   
 Median :0.000   
 Mean :0.294   
 3rd Qu.:1.000   
 Max. :1.000

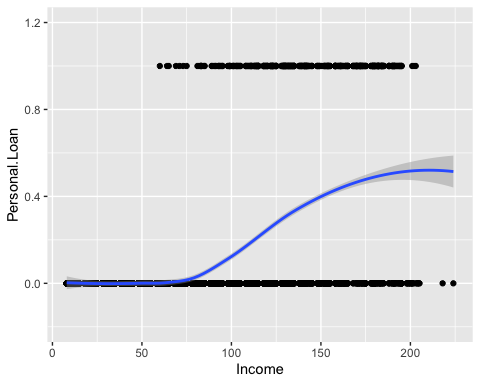
##### Looking into questions obtained in the statistical analysis

* ***Looking at the entries with negative experience***
* The 52 people with negative total experience are between 23 to 29 years old and salary median of 65,000. Sound like these could just young people that just started working. We can change their negative values to 1 since they have an income, thus are are working.

Age Income Personal.Loan  
 Min. :23.00 Min. : 12.00 Min. :0   
 1st Qu.:24.00 1st Qu.: 40.75 1st Qu.:0   
 Median :24.00 Median : 65.50 Median :0   
 Mean :24.52 Mean : 69.94 Mean :0   
 3rd Qu.:25.00 3rd Qu.: 86.75 3rd Qu.:0   
 Max. :29.00 Max. :150.00 Max. :0

# Extra plots

* Not sure what this means….



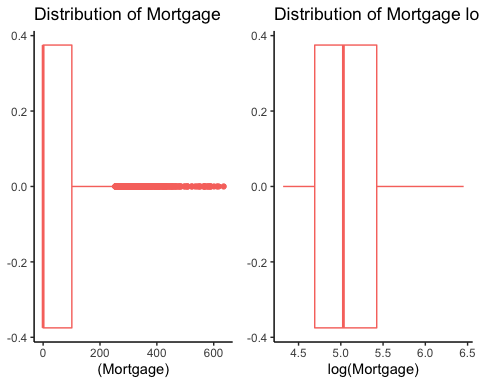
* Setting categories as factors

PersonalLoan$Personal.Loan<-factor(ifelse(PersonalLoan$Personal.Loan==1,1,0),levels=c(0,1)) #last level is the success  
  
PersonalLoan$Securities.Account=as.factor(PersonalLoan$Securities.Account)  
PersonalLoan$CD.Account =as.factor(PersonalLoan$CD.Account)  
PersonalLoan$Online =as.factor(PersonalLoan$Online)  
PersonalLoan$CreditCard =as.factor(PersonalLoan$CreditCard)  
PersonalLoan$Family =as.factor(PersonalLoan$Family)  
str(PersonalLoan)

'data.frame': 5000 obs. of 14 variables:  
 $ ID : int 1 2 3 4 5 6 7 8 9 10 ...  
 $ Age : int 25 45 39 35 35 37 53 50 35 34 ...  
 $ Experience : num 1 19 15 9 8 13 27 24 10 9 ...  
 $ Income : int 49 34 11 100 45 29 72 22 81 180 ...  
 $ ZIP.Code : int 91107 90089 94720 94112 91330 92121 91711 93943 90089 93023 ...  
 $ Family : Factor w/ 4 levels "1","2","3","4": 4 3 1 1 4 4 2 1 3 1 ...  
 $ CCAvg : num 1.6 1.5 1 2.7 1 0.4 1.5 0.3 0.6 8.9 ...  
 $ Education : int 1 1 1 2 2 2 2 3 2 3 ...  
 $ Mortgage : int 0 0 0 0 0 155 0 0 104 0 ...  
 $ Personal.Loan : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 2 ...  
 $ Securities.Account: Factor w/ 2 levels "0","1": 2 2 1 1 1 1 1 1 1 1 ...  
 $ CD.Account : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
 $ Online : Factor w/ 2 levels "0","1": 1 1 1 1 1 2 2 1 2 1 ...  
 $ CreditCard : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 2 1 1 ...

#Droppingzip and ID as there are too many unique values and will affect the model  
PersonalLoan=PersonalLoan %>% select(-ID)  
PersonalLoan=PersonalLoan %>% select(-"ZIP.Code")

* ***Checking mortgage distribution***
* Mortgage may need to be logged as it is very skewed.

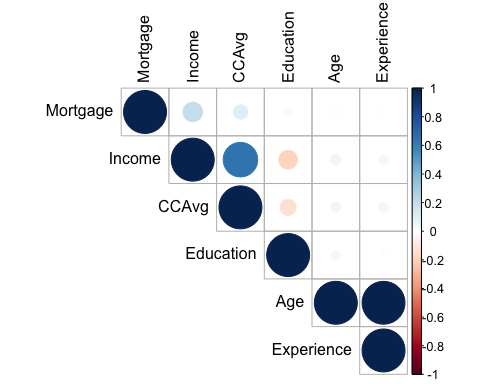
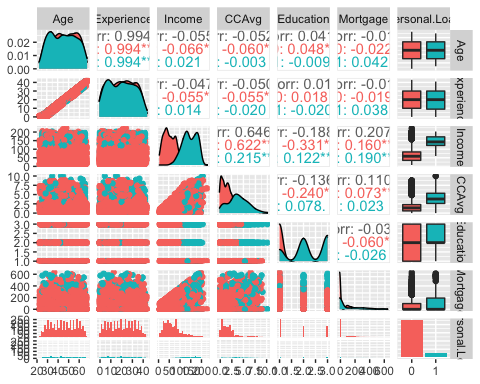


* Creating another data set with Mortgage logged since logging did improve the distribution. The zero’s were replaced to with 1 in order to not get infitiy when logging. Also, removing age since it has a 99% correlation with Experience, therefore only one is needed

#Creating a new data set with the modified attributes  
PersonalLoan2=mutate(PersonalLoan)  
#Updating values  
PersonalLoan2$Mortgage[PersonalLoan2$Mortgage==0]=1  
PersonalLoan2$MortgageLogged=log(PersonalLoan2$Mortgage)  
PersonalLoan2=PersonalLoan2 %>% select(-Mortgage)  
PersonalLoan2=PersonalLoan2 %>% select(-Age)

#### **Relationships and correlations**

* Experience and Age has a Correlation of 99%.Too high. However, they seem to have no relationship with Loan as both, Yes loan and no loan, have correlation of 0.99
* **Make a plot with just loan and experience and experience and loan**
* CCAvg and Income has a 64% correlations and it does seems to have a relationship with Loan since there is 0.62 with no loan and .02 with yes loan
* The rest of the explanatory variables do not seem to have relationship between each other



#### **Checking realtionship between Loan (response variable) and the rest of the predictors**

##### Relationship present

* ***Income***, the more income the more changes to get a loan
* ***CD.Account*** (certificate of deposit) seem to have a relationship with Personal Loan. Those with personal loan tend to have CD.Account more so than those with no CD.Account
* ***Mortgage***, people high mortgages seem to ask for loans more so thatn those with lower mortgages
* ***Personal education***, people with up to highschool tend do not get loans as much as does with an education of higher than highschool
* !!\*\*there seems to be relationship with loan and education but that of education 2 and 3 seem to be very close so maybe making 2 and 3 one single category would be more helpful\*\*!!

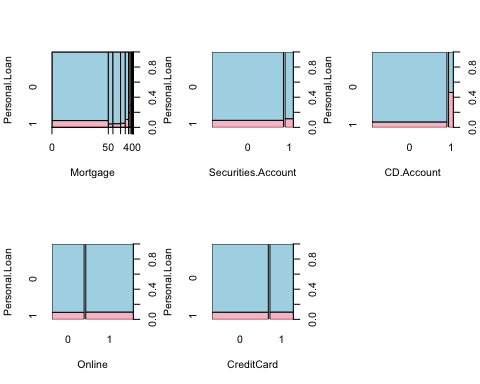
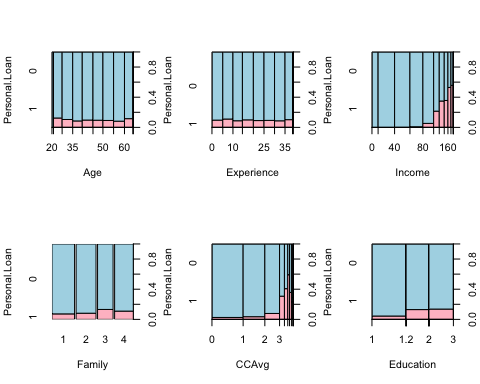
##### Very small relationship

* ***Security account*** and Personal Loan seem to have slight relationship

##### No relationship with response

* ***Online and Credit card*** don’t seem to have a relationship

## Looking closer at each realationship to see if anything was missed



# Model: Objective 1

# Splitting Data  
set.seed(1234)  
index<-sample(1:dim(PersonalLoan)[1],round(.70 \* dim(PersonalLoan)[1]))  
trainPL<-PersonalLoan[index,]  
testPL<-PersonalLoan[-index,]

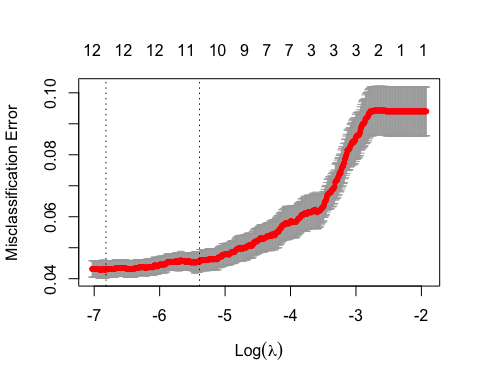
***Performing model with all variables and then another one with Stepwise*** + With the full model with all attributes it showed that the only important were Income, Family,CCavg, Education,Securites.Account, CD.Account, Online, CreditCard

* Once Stepwise was added to the full model it selected all of those that appeared as significant in the full model:Income, Family, CCAvg, Education, Securities.Account, CD.Account, Online, CreditCard. It also selected Experience but that one was not significant as we had seen in the EDA.

Here is the sumary of Stepwise

Call:  
glm(formula = Personal.Loan ~ Experience + Income + Family +   
 CCAvg + Education + Securities.Account + CD.Account + Online +   
 CreditCard, family = "binomial", data = PersonalLoan)  
  
Deviance Residuals:   
 Min 1Q Median 3Q Max   
-3.0597 -0.1964 -0.0746 -0.0261 3.9429   
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -13.157390 0.587811 -22.384 < 2e-16 \*\*\*  
Experience 0.010146 0.006657 1.524 0.127458   
Income 0.057979 0.002768 20.949 < 2e-16 \*\*\*  
Family2 -0.156734 0.216517 -0.724 0.469133   
Family3 2.155621 0.241657 8.920 < 2e-16 \*\*\*  
Family4 1.822492 0.231614 7.869 3.58e-15 \*\*\*  
CCAvg 0.147525 0.041664 3.541 0.000399 \*\*\*  
Education 1.737801 0.116699 14.891 < 2e-16 \*\*\*  
Securities.Account1 -0.948963 0.293090 -3.238 0.001205 \*\*   
CD.Account1 3.762456 0.331928 11.335 < 2e-16 \*\*\*  
Online1 -0.645505 0.160274 -4.028 5.64e-05 \*\*\*  
CreditCard1 -1.083893 0.208429 -5.200 1.99e-07 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 3162.0 on 4999 degrees of freedom  
Residual deviance: 1235.6 on 4988 degrees of freedom  
AIC: 1259.6  
  
Number of Fisher Scoring iterations: 8

***Performing lasso using Cross validation to obtain the optimal penalty*** + LASSO ended up choosing all attributes, not getting rid of any

Here is the output 

15 x 1 sparse Matrix of class "dgCMatrix"  
 s1  
(Intercept) -1.279646e+01  
Age .   
Family1 -1.783827e-02  
Family2 .   
Family3 2.117425e+00  
Family4 2.006844e+00  
Mortgage 8.498061e-04  
CD.Account1 3.305321e+00  
Experience 6.611570e-03  
CCAvg 1.240485e-01  
Online1 -4.534574e-01  
Income 5.583473e-02  
Education 1.603952e+00  
Securities.Account1 -6.084691e-01  
CreditCard1 -9.760266e-01

15 x 1 sparse Matrix of class "dgCMatrix"  
 s0  
(Intercept) -1.280228e+01  
Age .   
Family1 -1.694336e-02  
Family2 .   
Family3 2.119186e+00  
Family4 2.008302e+00  
Mortgage 8.500045e-04  
CD.Account1 3.308019e+00  
Experience 6.621079e-03  
CCAvg 1.241083e-01  
Online1 -4.538357e-01  
Income 5.586015e-02  
Education 1.604694e+00  
Securities.Account1 -6.093215e-01  
CreditCard1 -9.769566e-01

15 x 1 sparse Matrix of class "dgCMatrix"  
 s0  
(Intercept) -4.96022729  
Age .   
Family1 .   
Family2 .   
Family3 .   
Family4 .   
Mortgage .   
CD.Account1 0.87124155  
Experience .   
CCAvg .   
Online1 .   
Income 0.02300902  
Education 0.23011207  
Securities.Account1 .   
CreditCard1 .

***Predicting on models created***

* Using different threholds

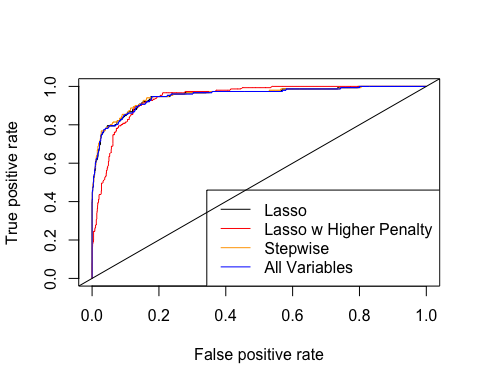
[1] "When setting the threashold to 0.5"  
[1] ""  
[1] "Confusion matrix for Model with all variables"  
   
class.full 0 1  
 0 1327 55  
 1 22 96  
[1] ""  
[1] "Accuracy"  
[1] 0.9486667  
[1] "Specificity"  
[1] 0.6357616  
[1] "Confusion matrix for LASSO"  
   
class.lasso 0 1  
 0 1327 57  
 1 22 94  
[1] "Accuracy"  
[1] 0.9473333  
[1] "Specificity"  
[1] 0.6225166  
[1] "Confusion matrix for Stepwise"  
   
class.step 0 1  
 0 1328 54  
 1 21 97  
[1] "Accuracy"  
[1] 0.95  
[1] "Specificity"  
[1] 0.6423841  
[1] "-----------------------------------"

[1] "When setting the threashold to 0.7"  
[1] ""  
[1] "Confusion matrix for Model with all variables"  
   
class.full 0 1  
 0 1344 75  
 1 5 76  
[1] ""  
[1] "Accuracy"  
[1] 0.9466667  
[1] "Specificity"  
[1] 0.5033113  
[1] "Confusion matrix for LASSO"  
   
class.lasso 0 1  
 0 1344 79  
 1 5 72  
[1] "Accuracy"  
[1] 0.944  
[1] "Specificity"  
[1] 0.4768212  
[1] "Confusion matrix for Stepwise"  
   
class.step 0 1  
 0 1343 75  
 1 6 76  
[1] "Accuracy"  
[1] 0.946  
[1] "Specificity"  
[1] 0.5033113  
[1] "-----------------------------------"

[1] "When setting the threashold to 0.3"  
[1] ""  
[1] "Confusion matrix for Model with all variables"  
   
class.full 0 1  
 0 1309 38  
 1 40 113  
[1] ""  
[1] "Accuracy"  
[1] 0.948  
[1] "Specificity"  
[1] 0.7483444  
[1] "Confusion matrix for LASSO"  
   
class.lasso 0 1  
 0 1311 38  
 1 38 113  
[1] "Accuracy"  
[1] 0.9493333  
[1] "Specificity"  
[1] 0.7483444  
[1] "Confusion matrix for Stepwise"  
   
class.step 0 1  
 0 1313 37  
 1 36 114  
[1] "Accuracy"  
[1] 0.9513333  
[1] "Specificity"  
[1] 0.7549669  
[1] "-----------------------------------"

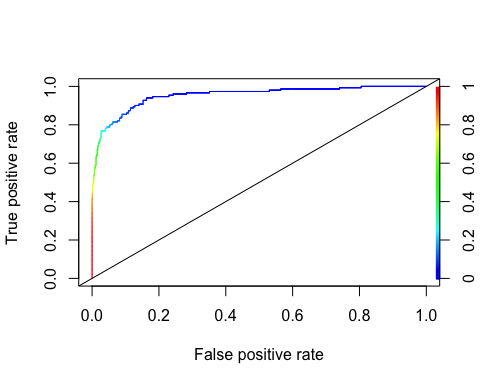
***Running ROC to compare models***

library(ROCR)  
results.lasso<-prediction(fit.pred.lasso, testPL$Personal.Loan,label.ordering=c(0,1))  
roc.lasso = performance(results.lasso, measure = "tpr", x.measure = "fpr")  
  
results.lasso2<-prediction(fit.pred.lasso2, testPL$Personal.Loan,label.ordering=c(0,1))  
roc.lasso2 = performance(results.lasso2, measure = "tpr", x.measure = "fpr")  
  
results.step<-prediction(fit.pred.step, testPL$Personal.Loan,label.ordering=c(0,1))  
roc.step = performance(results.step, measure = "tpr", x.measure = "fpr")  
  
results.origin<-prediction(fit.pred.full,testPL$Personal.Loan,label.ordering=c(0,1))  
roc.origin=performance(results.origin,measure = "tpr", x.measure = "fpr")



* The best model seemed to be the Stepwise

plot(roc.step,colorize = TRUE)  
abline(a=0, b= 1)



# Conclusion from Part 1f

* The best model was step setting the threshold to 0.3 it gave an sensitivity of 94 and specificity of 72
* Due to the imbalance of amount of people with loan and without loan we do see we do see that the model favors no loan due to it but 72 compared to the 55 specificity was a great increase. This model is about trying to predict those who will say yes to Loan therefore Specificity is important.
* The attributes found useful were : Income, Family,CCavg, Education,Securites.Account, CD.Account, Online, CreditCard
* The threshold was set to 0.3 and it lead to a Sensitivity of 0.96 and specificy of 0.71
* These were variables seen in the EDA as related to the loan.
* Coefficients results: For every unit increase in income the odd of getting a loan are e^1.06 times higher For every unit increase in Family the odd of getting a loan are e^0.698209 times higher For every unit increase in CCAvg the odd of getting a loan are e^0.120635 times higher For every unit increase in Education the odd of getting a loan are e^1.713690 times higher For every unit increase in Securities.Account 1 the odd of getting a loan are e^-0.937183 times less likely For every unit increase in CD.Account1 the odd of getting a loan are e^3.840892 times higher For every unit increase in Online1 the odd of getting a loan are e^-0.673230 times less likely For every unit increase in CreditCard1 the odd of getting a loan are e^ -1.122701 times higher

summary(stepWiseAIC)

Call:  
glm(formula = Personal.Loan ~ Experience + Income + Family +   
 CCAvg + Education + Securities.Account + CD.Account + Online +   
 CreditCard, family = "binomial", data = PersonalLoan)  
  
Deviance Residuals:   
 Min 1Q Median 3Q Max   
-3.0597 -0.1964 -0.0746 -0.0261 3.9429   
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -13.157390 0.587811 -22.384 < 2e-16 \*\*\*  
Experience 0.010146 0.006657 1.524 0.127458   
Income 0.057979 0.002768 20.949 < 2e-16 \*\*\*  
Family2 -0.156734 0.216517 -0.724 0.469133   
Family3 2.155621 0.241657 8.920 < 2e-16 \*\*\*  
Family4 1.822492 0.231614 7.869 3.58e-15 \*\*\*  
CCAvg 0.147525 0.041664 3.541 0.000399 \*\*\*  
Education 1.737801 0.116699 14.891 < 2e-16 \*\*\*  
Securities.Account1 -0.948963 0.293090 -3.238 0.001205 \*\*   
CD.Account1 3.762456 0.331928 11.335 < 2e-16 \*\*\*  
Online1 -0.645505 0.160274 -4.028 5.64e-05 \*\*\*  
CreditCard1 -1.083893 0.208429 -5.200 1.99e-07 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 3162.0 on 4999 degrees of freedom  
Residual deviance: 1235.6 on 4988 degrees of freedom  
AIC: 1259.6  
  
Number of Fisher Scoring iterations: 8

(coef(stepWiseAIC))

(Intercept) Experience Income Family2   
 -13.15738998 0.01014613 0.05797945 -0.15673437   
 Family3 Family4 CCAvg Education   
 2.15562141 1.82249194 0.14752545 1.73780123   
Securities.Account1 CD.Account1 Online1 CreditCard1   
 -0.94896276 3.76245576 -0.64550534 -1.08389348

AIC(stepWiseAIC)

[1] 1259.613

# let's predict the same data: use type response to have probability as resulthere you decide the cutoff and put as factor, in one line  
pred\_ <- as.factor(ifelse(predict(stepWiseAIC, testPL, type="response")>0.3,1,0))  
# here we go!  
confusionMatrix(pred\_, as.factor(testPL$Personal.Loan))

Confusion Matrix and Statistics  
  
 Reference  
Prediction 0 1  
 0 1313 37  
 1 36 114  
   
 Accuracy : 0.9513   
 95% CI : (0.9392, 0.9617)  
 No Information Rate : 0.8993   
 P-Value [Acc > NIR] : 1.403e-13   
   
 Kappa : 0.7304   
   
 Mcnemar's Test P-Value : 1   
   
 Sensitivity : 0.9733   
 Specificity : 0.7550   
 Pos Pred Value : 0.9726   
 Neg Pred Value : 0.7600   
 Prevalence : 0.8993   
 Detection Rate : 0.8753   
 Detection Prevalence : 0.9000   
 Balanced Accuracy : 0.8641   
   
 'Positive' Class : 0

AIC(stepWiseAIC)

[1] 1259.613

## EXTRA code that has been commented out

###### Re-run all models but removing age and Logging mortgage