Group 8 Final Project

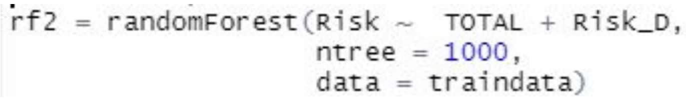
Task 1

Create a random forest model with Risk as the target and all other variables as inputs.

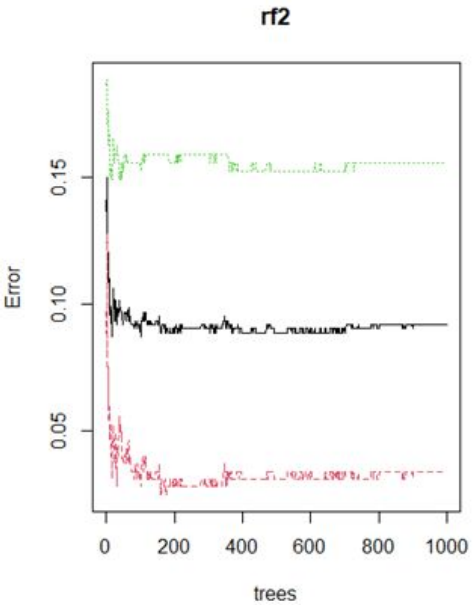
Create random forest models by adding the ten most significant variables in forward stepwise selection.

Choose model with highest accuracy.

Task 1 model



```
rf2 = randomForest(Risk ~  TOTAL + Risk_D,
                   ntree = 1000,
                   data = traindata)
```



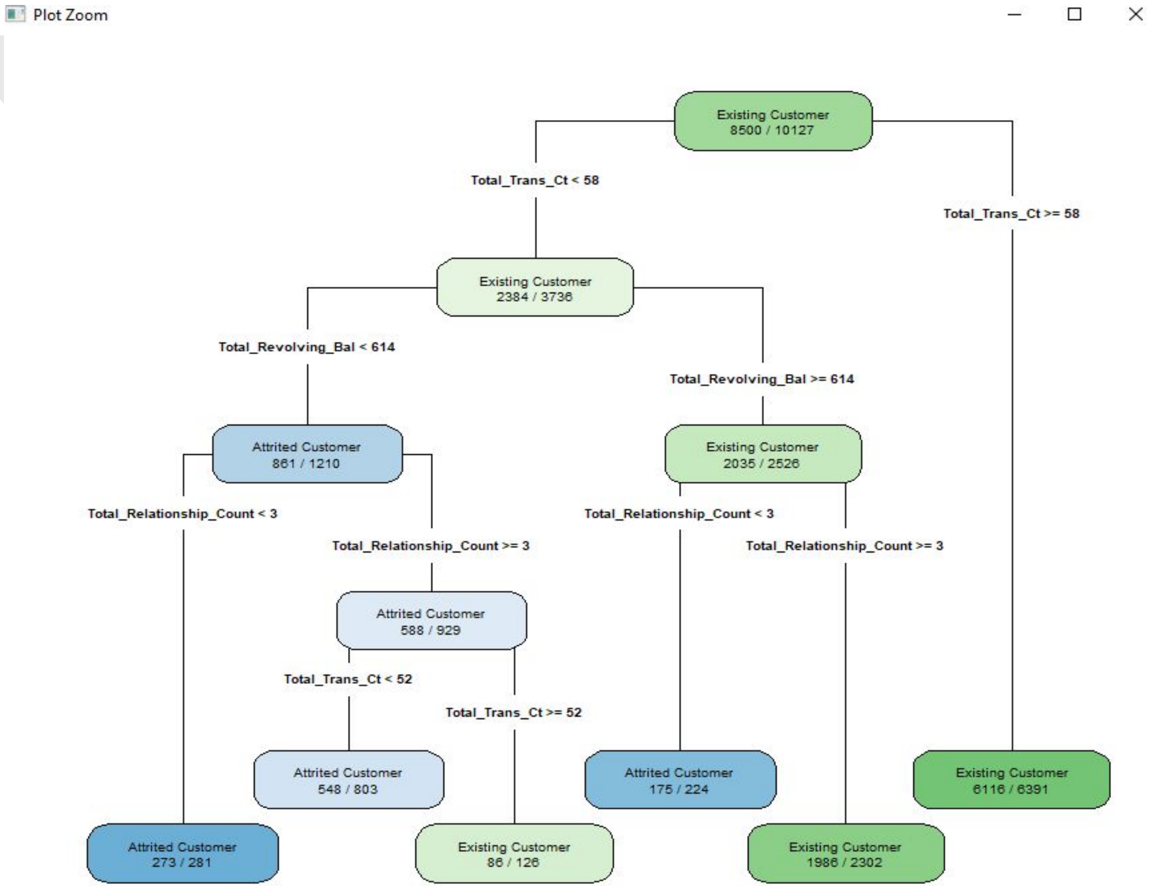
Task 2 (Natalie Windisch)

Predictive model for customer churn using Decision Tree model.

Decision Tree contains Root, Branch, Internal Nodes and Leaf Nodes.

If customers make more than 1 transaction, have been with bank for 3 or more years and have revolving balances are considered existing customers and anything less are attrited customers or they have had bad experiences with the bank.

Decision Tree



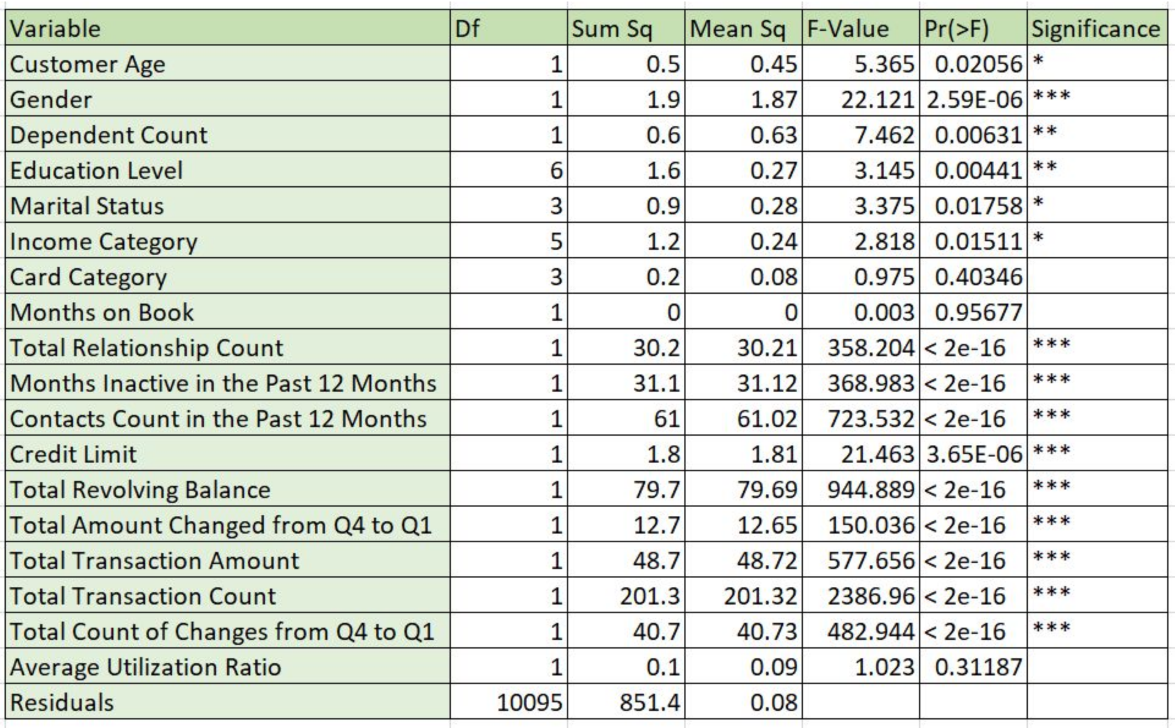
Task 3

Predictive Model for Customer Churn using ANOVA Analysis

Null Hypothesis: Variable have no effect on customer churn

Alternate Hypothesis: Variable has effect on customer churn

Raw Data



Variable	Df	Sum Sq	Mean Sq	F-Value	Pr(>F)	Significance
Customer Age	1	0.5	0.45	5.365	0.02056	*
Gender	1	1.9	1.87	22.121	2.59E-06	***
Dependent Count	1	0.6	0.63	7.462	0.00631	**
Education Level	6	1.6	0.27	3.145	0.00441	**
Marital Status	3	0.9	0.28	3.375	0.01758	*
Income Category	5	1.2	0.24	2.818	0.01511	*
Card Category	3	0.2	0.08	0.975	0.40346	
Months on Book	1	0	0	0.003	0.95677	
Total Relationship Count	1	30.2	30.21	358.204	< 2e-16	***
Months Inactive in the Past 12 Months	1	31.1	31.12	368.983	< 2e-16	***
Contacts Count in the Past 12 Months	1	61	61.02	723.532	< 2e-16	***
Credit Limit	1	1.8	1.81	21.463	3.65E-06	***
Total Revolving Balance	1	79.7	79.69	944.889	< 2e-16	***
Total Amount Changed from Q4 to Q1	1	12.7	12.65	150.036	< 2e-16	***
Total Transaction Amount	1	48.7	48.72	577.656	< 2e-16	***
Total Transaction Count	1	201.3	201.32	2386.96	< 2e-16	***
Total Count of Changes from Q4 to Q1	1	40.7	40.73	482.944	< 2e-16	***
Average Utilization Ratio	1	0.1	0.09	1.023	0.31187	
Residuals	10095	851.4	0.08			

Strength of Evidence

Strong Evidence Against Null Hypothesis (Factors that strongly affect customer churn): Gender, Total Relationship Count, Months Inactive in the Past 12 Months,, Contacts Count in the Past 12 Months, Credit Limit, Total Revolving Balance, Total Amount Changed from Q4 to Q1, Total Transaction Amount, Total Transition Count, Total Count of Changes from Q4 to Q1

Moderate Evidence Against Null Hypothesis (Factors that Moderately affect customer churn): Dependent Count, Education Level

Weak Evidence Against Null Hypothesis (Factors that weakly affect customer churn): Marital Status, Income Category

No Evidence Against Null Hypothesis (Factors that do not noticeably affect customer churn): Card Category, Months on Book, Credit Utilization Ratio

Task 4

Predictive model using multiple regression with Price as outcome variable

Variable selection performed on 22 variables

Linear model to create simple regression model

Determine which variables influence the price of cars

Car Price Variable Selection

Stepwise Model

Variables kept (14):

- Aspiration
- Carbody
- Carlength
- Carwidth
- Carheight
- Curbweight
- Enginetype
- Cylindernumber
- Enginesize
- Stroke
- Compressionratio
- Horsepower
- Peakrpm
- Highwaympg

Variables removed (8):

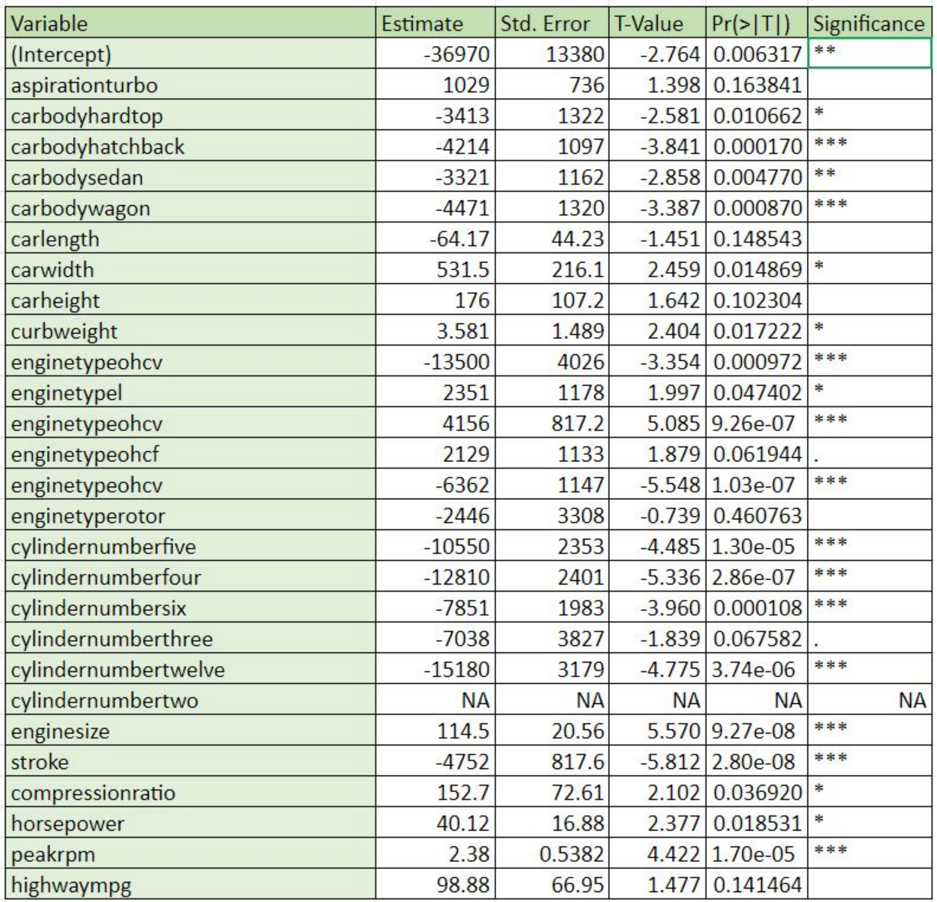
- Fueltype
- Doornumber
- Drivewheel
- Enginelocation
- Wheelbase
- Fuelsystem
- Boreratio
- Citympg

Multiple Regression Analysis

Residual standard error: 2241 on 178 degrees of freedom

Multiple R-squared: 0.9313, Adjusted R-squared: 0.9213

F-statistic: 92.83 on 26 and 178 DF, p-value: < 2.2e-16



Variable	Estimate	Std. Error	T-Value	Pr(>|T|)	Significance
(Intercept)	-36970	13380	-2.764	0.006317	**
aspirationturbo	1029	736	1.398	0.163841	
carbodyhardtop	-3413	1322	-2.581	0.010662	*
carbodyhatchback	-4214	1097	-3.841	0.000170	***
carbodysedan	-3321	1162	-2.858	0.004770	**
carbodywagon	-4471	1320	-3.387	0.000870	***
carlength	-64.17	44.23	-1.451	0.148543	
carwidth	531.5	216.1	2.459	0.014869	*
carheight	176	107.2	1.642	0.102304	
curbweight	3.581	1.489	2.404	0.017222	*
enginetypeohcv	-13500	4026	-3.354	0.000972	***
enginetypel	2351	1178	1.997	0.047402	*
enginetypeohcv	4156	817.2	5.085	9.26e-07	***
enginetypeohcf	2129	1133	1.879	0.061944	.
enginetypeohcv	-6362	1147	-5.548	1.03e-07	***
enginetyperotor	-2446	3308	-0.739	0.460763	
cylindernumberfive	-10550	2353	-4.485	1.30e-05	***
cylindernumberfour	-12810	2401	-5.336	2.86e-07	***
cylindernumbersix	-7851	1983	-3.960	0.000108	***
cylindernumberthree	-7038	3827	-1.839	0.067582	.
cylindernumbertwelve	-15180	3179	-4.775	3.74e-06	***
cylindernumbertwo	NA	NA	NA	NA	NA
enginesize	114.5	20.56	5.570	9.27e-08	***
stroke	-4752	817.6	-5.812	2.80e-08	***
compressionratio	152.7	72.61	2.102	0.036920	*
horsepower	40.12	16.88	2.377	0.018531	*
peakrpm	2.38	0.5382	4.422	1.70e-05	***
highwaympg	98.88	66.95	1.477	0.141464	

Significant Variables that affect Price

- Carbody
- Enginetype
- Cylindernumber
- Enginesize
- Stroke
- Peakrpm

The End