

Objective

Find the likelihood of a customer purchasing a RRSP via ad mail. This is done by creating a prediction model and finding the customers that have the highest probability to purchase a RRSP to maximize profits.

Procedures

Variables that were excluded from the prediction model:

gendf: being female would not have impact of opening a RRSP account because RRSP is for retirement and is non-discriminatory for gender,

gendm:being male would not have impact of opening a RRSP account because RRSP is for retirement and is non-discriminatory for gender,

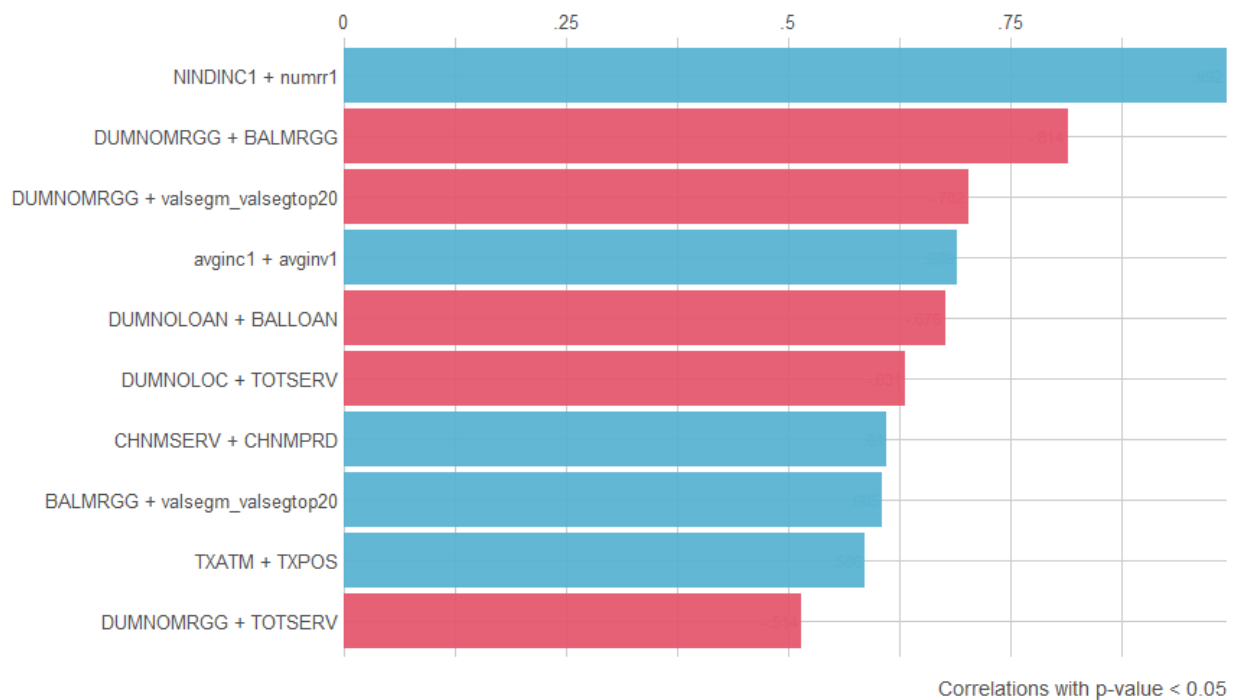
pcode: postal code does not make sense to include this in the model because if someone has postal code that does not increases their odds of getting a RRSP.

unique: unique is an identification number which is useless in our model.

To find variables that highly correlated with each other I used a Ranked Cross-Correlation which shows the variables that have a p-value of 5 or less and are correlated with each other.

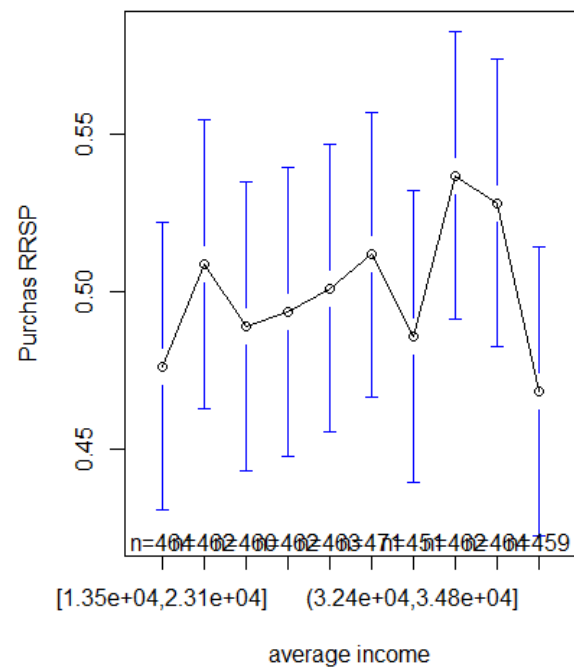
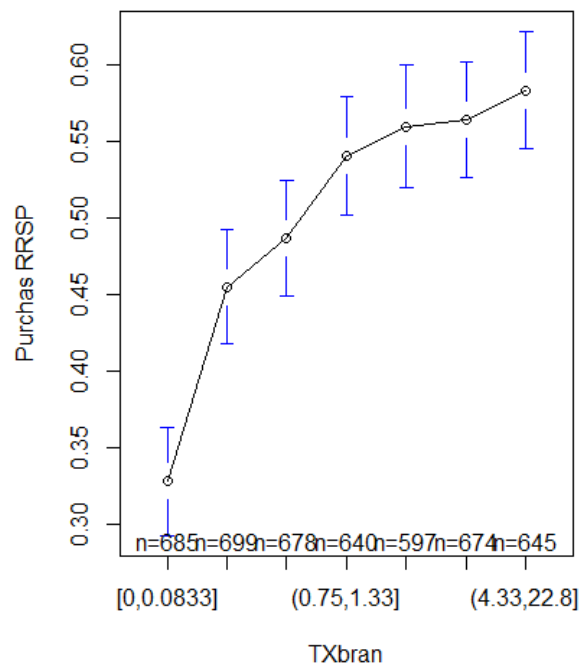
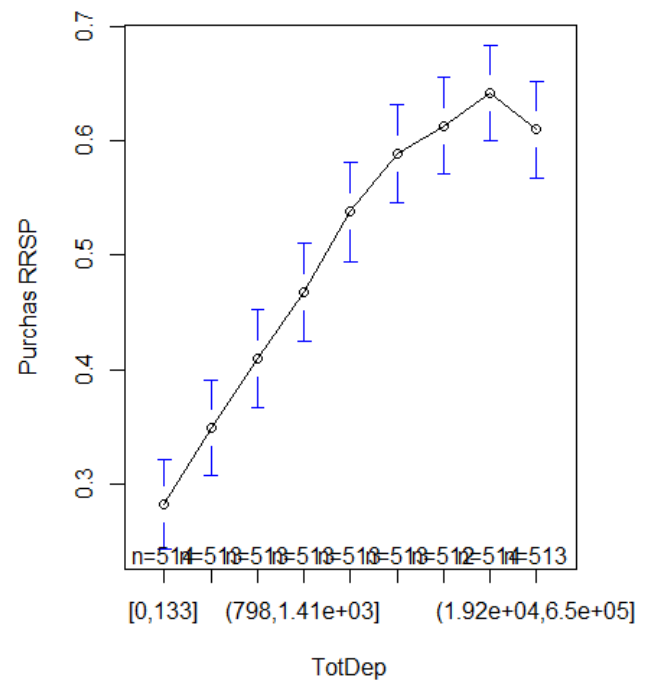
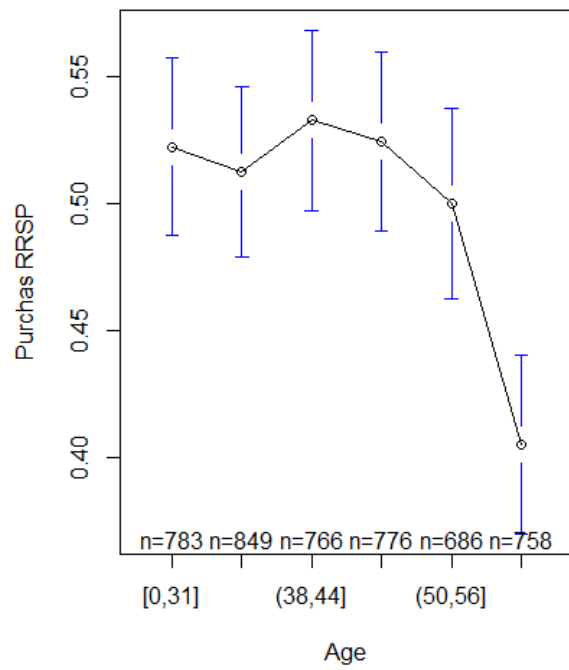
Ranked Cross-Correlations

10 most relevant



NINDINC1 and numrr, and DUMNOMRGG and BALMRGG are highly correlated. The rest have weak correlation.

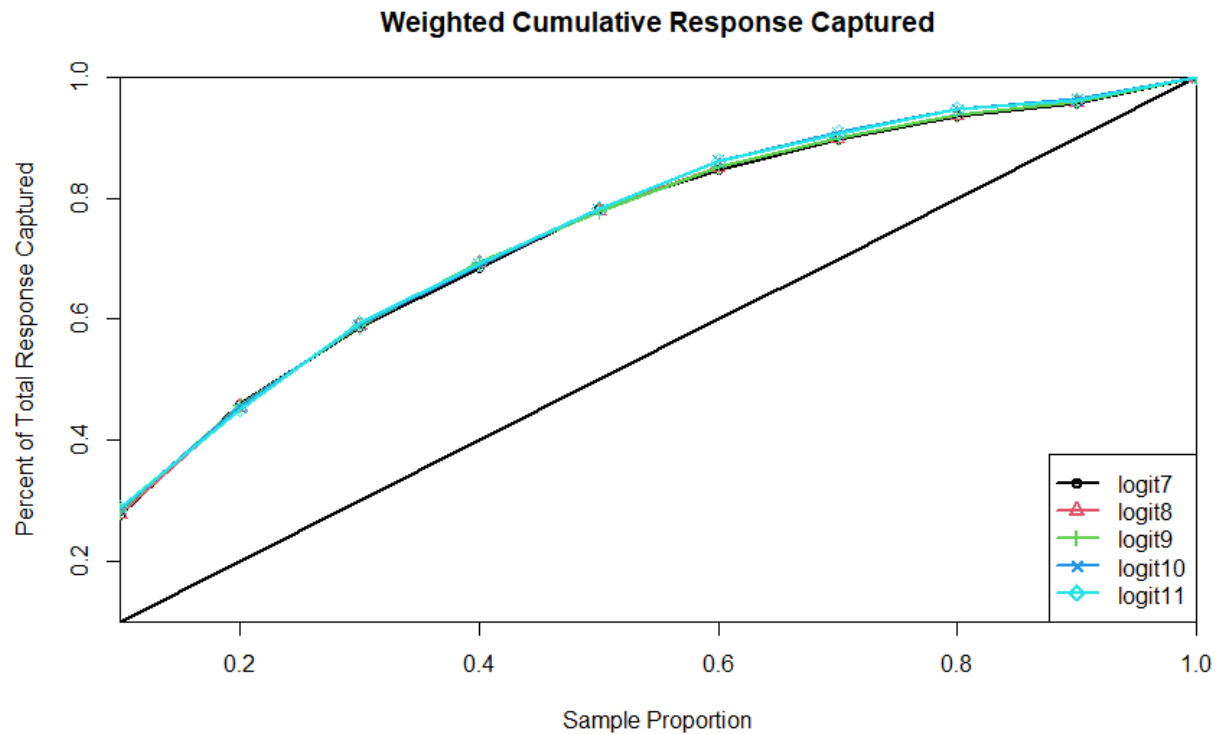
The following charts show non-linear relationship between purchasing a RRSP and the variable age, total dep, txbran, and average income.



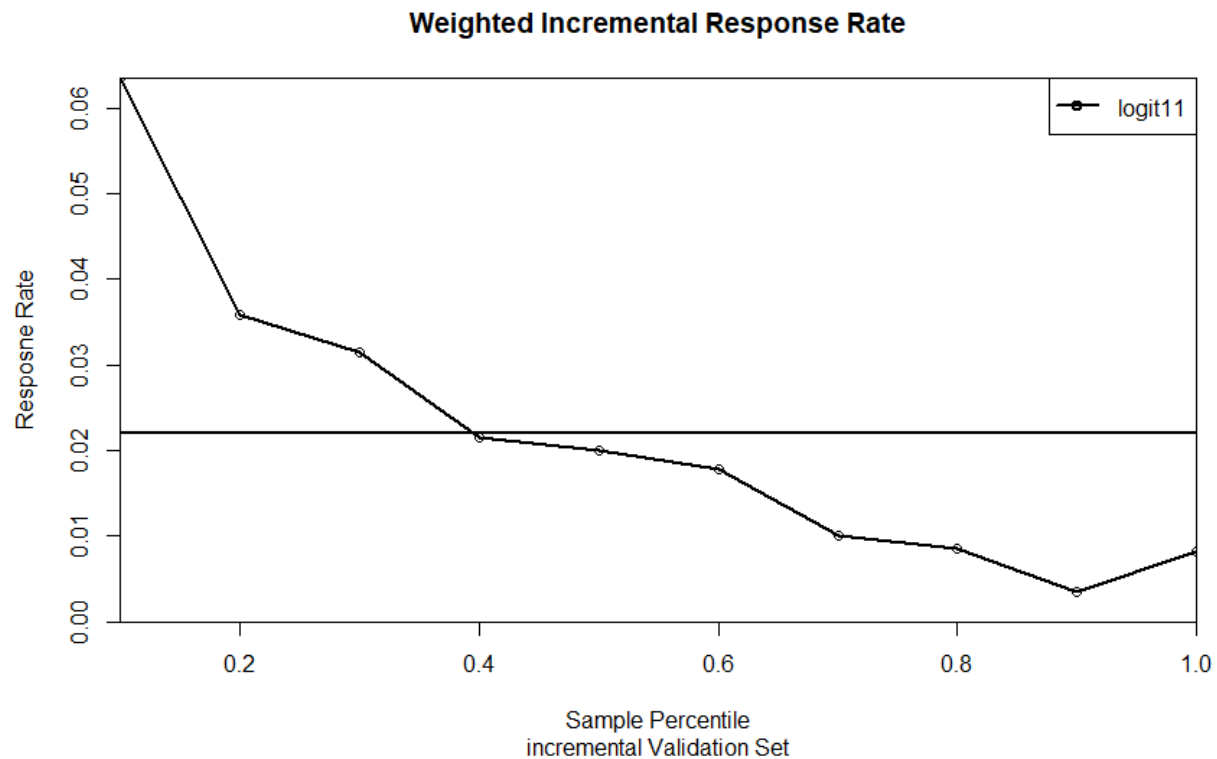
We can log these variables to check if they have a better performance on the models.

Life Charts

Using the models with the best performance I use a lift chart to check which one is better at predicting if someone will purchase a RRSP.



From this we can see that logit11 has the best performance.



The base rate of someone purchasing RRSP at random is 2%. We can see that model(logit11) has a higher response rate than the base when the top 40% of customer are sent ad mail.

Calculation Table

cost per ad mail	revenue per RRSP	sample	total number of Ys		contact	capture	sample	total number of Ys	Revenue	cost	Profit
4.05	200	100000	2200		45%	75%	45000	1650	330000	182250	147750
					60%	85%	60000	1870	374000	243000	131000
					80%	95%	80000	2090	418000	324000	94000
					70%	90%	70000	1980	396000	283500	112500
					40%	70%	40000	1540	308000	162000	146000
					30%	59%	30000	1298	259600	121500	138100
					20%	44%	20000	968	193600	81000	112600
					50%	78%	50000	1716	343200	202500	140700

From the calculation table we see that if we contact the top 45% customer we are able to maximize profits.