

**Project 2:**  
Bank Marketing

**Students**: Matt Farrow, Michael Olheiser, Edward Roske

**Professor**: Dr. Jacob Turner

**Class**: MSDS 6372

**Date**: November 28, 2020

Table of Contents

[Introduction 3](#_Toc57752203)

[Data Description 3](#_Toc57752204)

[Data Cleanup 4](#_Toc57752205)

[Missing Values 4](#_Toc57752206)

[Other Cleanup 4](#_Toc57752207)

[Test & Training Data Sets 4](#_Toc57752208)

[Exploratory Data Analysis 4](#_Toc57752209)

[Model 1 – Interpretable Model 6](#_Toc57752210)

[Feature Selection: Stepwise 7](#_Toc57752211)

[Parameter Interpretation 7](#_Toc57752212)

[Conclusion 8](#_Toc57752213)

[Model 2 – Complicated Model 8](#_Toc57752214)

[Model 3 – Discriminant Analysis Model 9](#_Toc57752215)

[Model 4 – Non-Parameterized Model 10](#_Toc57752216)

[Decision Tree 10](#_Toc57752217)

[Random Forest 10](#_Toc57752218)

[Model 5 – Predicting Duration Model 11](#_Toc57752219)

[Overall Conclusion 12](#_Toc57752220)

[AUC Analysis 12](#_Toc57752221)

[Final Conclusions 12](#_Toc57752222)

[Appendix 13](#_Toc57752223)

[Data Description 13](#_Toc57752224)

[Data Cleanup 16](#_Toc57752225)

[Exploratory Data Analysis 21](#_Toc57752226)

[Logistic Regression (Interpretable) 22](#_Toc57752227)

[Model 2 – Complicated Model 27](#_Toc57752228)

[Model 2 – Complicated Model – Analysis of Maximum Likelihood Estimates 28](#_Toc57752229)

[Discriminant Analysis 41](#_Toc57752230)

[Decision Tree 44](#_Toc57752231)

[`duration` Model 46](#_Toc57752232)

[SAS Code 49](#_Toc57752233)

[Exploratory Data Analysis 49](#_Toc57752234)

[Exploratory Data Analysis 52](#_Toc57752235)

[Model 1 54](#_Toc57752236)

[Model 2 58](#_Toc57752237)

[R Code 61](#_Toc57752238)

# Introduction

The project utilized data from the *UCI Machine Learning Repository* (can be found [here](https://archive.ics.uci.edu/ml/datasets/Bank%20Marketing)). This is data set related to direct marking campaigns of a Portuguese banking institution via phone calls from May 2008 to November 2010. This data provides a wealth of possibly explanatory variables including age, marital status, and job that may correlate to why someone would subscribe to a possible banking offer.

The research team was challenged to develop models to identify which explanatory variables tend to have the biggest impact on whether a subscription was made. The primary purpose of the project was deemed to be explanation not “pre-phone call prediction” so all variables were left in the model to begin with (UCI suggests discarding the duration variable if the primary purpose is prediction).

The team decided to create 5 types of models:

1. Easily interpretable model (in SAS)
2. Complicated model with transformations and interactions (in SAS)
3. Linear/Quadratic Discriminant models (in R)
4. Random Forest model (in R)
5. Model to predict duration then use predicted duration in the final complicated model (in R for prediction then SAS for final model based on #2 above).

# Data Description

The data was provided in multiple files (some of which contained additional explanatory variables). The data set the researchers were given contained 20 explanatory variables related to the response variable, ‘y’ (which we renamed to ‘subscribed’ to be more descriptive). The final data set the team used (after cleaning up data, removing redundancies, and so on) contained 41,188 observations. A table of variables and their descriptions from the UCI Machine Learning Repository (including additional fields not found in the primary data file) can be found in the appendix (Figure 3).

The data contain a fairly even split of continuous (Figure 4) and categorical (Figure 5) variables in addition to the dependent response variable ‘y’ (renamed to ‘subscribed’ and hereafter referred to as such). The response variable has two factor levels, *yes* and *no*, that correspond to whether or not a bank customer signed up for a subscription.

# Data Cleanup

Our first step in cleaning up the data involved replacing placeholder values with *NA;* for most variables this placeholder value was *“unknown”*, but for *pdays* it was *“999”* and for *poutcome* it was *“nonexistent.”* The research team then used R to create a summary statistic table (Figure 6), and the “DataExplorer” and “skimr” packages were used to get a visual summary of the data (Figure 7) (Figure 8) (Figure 9) (Figure 10) (Figure 11) (Figure 12).

## Missing Values

The table reveals that there are many missing values (*NA*) in *job*, *education*, *default*, *housing*, *loan*, *pdays* (the only continuous variable with missing data), and *poutcome*. Once all missing values had been identified, we replaced the missing values in each of the categorical variables with their corresponding mode (Figure 13). We also made the decision to replace missing values of *pday* with the continuous variable’s median, *“6.”*

## Other Cleanup

The summary statistics also revealed that *default* is virtually useless because across the 41,188 observations, only 3 have *default = yes*. Based on that information, the team discarded *default*. Our final cleanup step was to convert all character variables into factors.

## Test & Training Data Sets

In order to build our models off of the same data sets, we built our test and training data sets at the end of our data cleanup process. Two different 75/25 test and training sets were built:

1. The first test/train set was built using all of the available data.
2. The second test/train set was build using a down-sampled set that contained an equal (4,640) number of *“yes”* and *“no”* responses from the *subscribed* variable.

# Exploratory Data Analysis

Overall, the *subscribed* variable is *no =* 88.7% and *yes* = 11.3% (Figure 14). The team looked to the frequencies (PROC FREQ in SAS) where a categorical response revealed percentages substantially different than 88.9% (*no*) and 11.3% (*yes*). For instance, for the categorical variable *job*, at one extreme, 31% of *job = Student* subscribed (substantially more than 11.3%) and at the opposite extreme, 7% of *job = Blue-Collar* subscribed (substantially less than 11.3%). Both of these jobs are far enough away from the expected values that we would include *job* for further evaluation whether or not the Chi-Square test said we should. That said, the team did a Chi-Square test (also from PROC FREQ in SAS) to see which variables correlate to Subscribed (p-value of < 0.05).

Without repeating the entire PROC FREQ tables, the variables that had strong significance levels or seemed to have explanatory values that led to different no/yes percentages were:

* **Categorical**: *job, education, contact, month, marital, day\_of\_week, housing*
* **Continuous**: *duration, campaign, emp\_var\_rate, euribor3m, nr\_employed, previous*

The variables that we rejected (p-value > 0.05 and/or data seemed to be junk/incomplete) were:

* **Categorical**: *default, loan*
* **Continuous**: *age, pdays*

While there was only weak evidence that *cons\_price\_idx* and *cons\_conf\_idx* were significant, for now, we left those variables in. If they turn out to be insignificant, they will hopefully be discarded by the feature selection.

In terms of multicollinearity, *emp\_var\_rate*, *euribor3m*, *nr\_employed*, *cons\_price\_idx*, and *cons\_conf\_idx* are all really just measures of what is happening in the outside world at that moment in time. In other words, they coordinate heavily with “what month is it?” The main purpose of this research is explanation, so we will not be creating a Principle Component with all of these combined, but it is expected that the feature selection in Model 1 (the interpretable model), which will then be used as the starting point for Model 2 (the complex model), will reject the majority of these continuous variables.

Though Model 1 will be using only the non-transformed values (since the goal for that model will be ease of interpretability), the team did add in logs and squares of all possible explanatory variables for Model 2. The team calculated the linear correlation of all explanatory variables (original, log transformed, and squared) to produce the correlation calculations (Figure 16).

The six most negatively statistically correlated variables (meaning, variables that as they increase, the response variable is most significantly decreased) are forms of the *euribor3m* and *nr\_employed* variables. This makes sense for the first because higher interest rates would logically make one less likely to want to deposit money in a Portuguese bank based on a marketing call to solicit deposit subscriptions. It’s less clear why an increase in the number of employed people would make one less likely to subscribe.

Of the top five most positively statistically correlated variables, two are related to the duration of the call and three are related to the consumer price index (inflation, basically). Duration clearly makes sense: the longer the person stays on the phone with the telemarketer, the more likely they are to subscribe. Consumer price index is less clear: if inflation is rising, one would tend to want to save less and spend more. This might possibly be an issue in the data set — the CPI is coming out substantially negative and “negative inflation” is rare and in these large of numbers, extremely rare. We will leave in CPI for now and just point out that the data may be incorrect. Based on this exploratory analysis, it is likely that these key variables listed above will end up being present across the majority of the models developed by the team.[[1]](#footnote-1)

# Model 1 – Interpretable Model

For our interpretable model, we began with 19 statistically relevant variables: *job, marital, education, housing, loan, contact, month, day\_of\_week, poutcome, age, duration, campaign, pdays, previous, emp\_var\_rate, con\_price\_idx, cons\_conf\_idx, euribor3m, and nr\_employed*. We knew that many of these would be discarded based on the EDA above, but we left them in as a sanity check.

We ran a logistic regression (discarding no variables) and, based on the Type 3 Analysis of Effects (Figure 17), several variables had p-values > 0.05: *marital* (0.5086), *education* (0.0591), *housing* (0.8357), *loan* (0.4502), *age* (0.5379), *pdays* (0.6220), and *nr\_employed* (0.1329). Note that many of the more egregiously unimportant variables (*age*, *housing*, *pdays*) were previously identified in the EDA as unimportant. We did leave *education* in the model since it was just outside the alpha of 0.05 and a change in the other variables might make it statistically significant.

## Feature Selection: Stepwise

We then took the remaining variables and tried forward, backward, and stepwise feature selection with stepwise returning the highest model performance. The final variables left in the model (Figure 18) were *job, education, contact, month, day\_of\_week, poutcome, duration, campaign, previous, emp\_var\_rate, cons\_price\_idx,* and *eruibor3m*.

With this model, the *duration* variable had an extremely high Wald Chi-Square score (greater than all the other variables combined) of 4010.6589 with a p-value < 0.0001.

## Parameter Interpretation

Figure 1 - Estimates of Coefficients

Our goal is to produce a formula like the following:

Log(odds) =

A logistic regression of the explanatory variables by the response variable (*subscribed*) produced the coefficients (the Betas in the above formula) shown to the right in the column labeled “Estimate”. The intercept is admittedly extrapolation but it shows that if all other explanatory variables were zero (both the negatively and positively correlated continuous variables) or the reference (unemployed for *job*, Sep for *month*, etc.), the odds of someone subscribing are e-182.9 = 3.7\*10-80 = a really, really, really small number (but still greater than zero… barely).

Many of the most predictive variables are ones we highlighted in our EDA section including *duration, month*, and *euribor3m* are present in this logistic regression formula for the log(odds).

It is easier to interpret this parameter table if we convert it to odds ratios (Figure 19). For odds ratios, the point estimate tells us “for each 1 unit increase in the effect, it is predicted that the odds will increase by a multiplicative factor of the point estimate holding the other variables constant.”

For example, let’s take d*uration*. For each 1 second increase in d*uration*, it is predicted that the odds will increase by a multiplicative factor of 1.005 holding the other variables constant (note that this 95% confidence interval is from 1.005 to 1.005 because of rounding). Confidence intervals for all of the odds ratios can also be found in Figure 19. While a 1.005 multiplicative factor may not seem that great, that’s for only a 1 second increase. If we want to determine the odds ratio for one minute (60 seconds), that would be e60\*0.00471 = 1.327 multiplicative factor increase for each additional one minute the conversation lasts.

## Conclusion

The receiver operating characteristic (ROC) curve for this model produced an area under the curve (AUC) of 0.9358 (Figure 20). Out of interest, we also compared the ROC curve to one containing only the *duration* variable and were able to calculate an AUC value of 0.8184 (Figure 21Figure 21: Logistic Regression (Interpretable) - ROC Curve (just duration)) showing the dramatic significance of the *duration* variable.

Overall, the logistic regression model generated in this case is easy to interpret, easy to explain (telemarketers: call people on their cell phone generally during the cooler months and keep them on the phone for a while), and has a fairly high accuracy rate. Is it possible to do better? Let’s try complicating the model a bit.

# Model 2 – Complicated Model

We began our complicated logistic regression model with our previously identified statistically relevant variables and added transformations from Model 1. For the continuous variables from Model 1 (*duration, campaign, previous, emp\_var\_rate, cons\_price\_idx, and euribor3m*) if the value was greater than 0, we created a log-transformed version[[2]](#endnote-1) of that variable; we also created quadratic versions[[3]](#endnote-2) of those same variables. We did not add any additional predictors beyond the transformations with one major exception: we added in all possible interactions (both two-way and three-way) of the categorical predictors from Model 1.

Table

Description automatically generatedThe final complicated model, based on stepwise selection, returned an AUC of 0.9471 (Figure 22) and was defined by the formula (effects with corresponding p-values listed to the right):

Figure : Logistic Regression (Complicated) – Effects & p-values

For the subscripts listed with a “+” notation, there are actually several coefficients depending on the value of the categorical variable. The model appears fairly simple on the surface (just “nine” effects) but since two of these are categorical — one is a two-way categorical interaction, and another is a three-way categorical interaction — the model is virtually impossible to interpret on an individual coefficient basis (not to mention the square and log of *duration*). To make the model easier to process, it is possible to remove *job\*education\*month*: this does drop the AUC to 0.9401 (Figure 23) which is still remarkably accurate.

The full table of coefficients can be found in the appendix (below).

# Model 3 – Discriminant Analysis Model

The Linear Discriminant Analysis (LDA) model began by selecting the numeric variables along with the *subscribed* variable from the previously down-sampled data, creating training and testing splits, and preprocessing the data using the *“center”* and *“scale”* methods[[4]](#footnote-2). The results of the LDA model showed an accuracy of 82.63% (95% confidence interval: 81.03%, 84.15%) (Figure 25), as well as a sensitivity and specificity of 0.8431 and 0.8095, respectively. The accompanying ROC curve showed an AUC value of 0.9126 (Figure 26).

Interestingly, running the QDA model showed an accuracy of 78.32% (95% confidence interval: 76.59%, 79.98%) (Figure 27), as well as a sensitivity and specificity of 0.9043 and 0.6621, respectively. The accompanying ROC curve showed an AUC value of 0.8997 (Figure 28).

# Model 4 – Non-Parameterized Model

A non-parametric model benefits from the lack of assumptions regarding distributions and linear relationships between explanatory variables and the *log(odds)* of a customer subscribing. We chose to fit decision tree and random forest models using the majority of available predictors except for *pdays* and *outcomes* due to missing data while letting the tree and random forest algorithms determine which variables to split by. The probability threshold for classification was left at the 0.5 default for both models and the test and train data sets were the same down-sampled sets used for other models.

## Decision Tree

In the case of fitting a single decision tree, pruning was not necessary as cross-validation showed no evidence of overfitting. Figure 29 shows the result of a single decision tree resulting in 9 nodes. Similar to logistic regression, the dendrogram indicates that *duration* may be the most important predictor in the model as it's used in the first split and two other subsequent splits. The other predictors included are *month*, *euribor3m*, *nr\_employed*, and *emp\_var\_rate* which all coordinate with time of the year. The single tree resulted in an AUC of 0.914 and accuracy of 0.856 (Figure 30).

## Random Forest

A random forest model leverages bootstrapping techniques to randomize the selected predictors and build multiple trees to avoid overfitting. The final random forest was run using 100 trees and a subset of five random predictors in each bootstrap. Through experimentation, there was no significant change to performance metrics by increasing or decreasing these two parameters. The result of the random forest is an increase in all performance metrics when compared to the single decision tree (Figure 29). The ROC of the single tree (Figure 30) and Random Forest (Figure 31) are fairly similar with the Random Forest having a slight increase in AUC. The Variable Importance plot also shows that duration is the most important variable when it comes to increasing accuracy followed by the same variables selected in the single decision tree in addition to *contact*. (Figure 32).

# Model 5 – Predicting Duration Model

Through EDA and modeling, *duration* is clearly the best predictor in the data set indicating that longer calls are more likely to get a subscription outcome, however it may not be the most practical choice of a predictor depending on the use-case. If the model’s purpose is to guide telemarketers in their pursuit of gaining subscribers, they would have no prior knowledge of what the duration of a call is going to be before making it. Given this, we decided to see if we could build a linear regression model that predicts *duration* which could then be used in our logistic regression model focused on prediction in lieu of the actual *duration* value. Ideally, the result would be a better performing model than leaving out *duration* entirely.

A subset of predictors was used in a candidate regression model based on previous EDA and the reliability of predictors in terms of missing data. Stepwise selection was used to choose the final predictors (Figure 34). The *log(duration)* was used as the response to improve the fit since *duration* is heavily right-skewed. The two absolute largest coefficient estimates were for the Month of November (-0.21, p-value << .001) and December (0.22, p-value = 0.0052). That is, December on average has a median phone call duration that is 1.25 times larger than the month of April (reference) and November has a median phone call duration that is 0.81 times that of April. *Age* was the only numerical predictor that was included in the model making *duration* a difficult response to model since all other numerical variables had a correlation value close to zero with *duration*. The residual plots (Figure 35) show some influential points around *log(duration) = 5* that we did not investigate. Furthermore, there may be evidence of a pattern in the residuals indicating some behavior in *duration* that we are not able to model given the dataset available to us. There is not strong evidence in the QQ plot to suggest the residuals don’t follow a normal distribution. Our attempts to add interaction and complexity did not improve the fit, so we settled on this final model of *duration* to use as a predictor in the complex logistic regression.

# Overall Conclusion

## AUC Analysis

We used the AUC of the ROC curves as our performance metric for evaluating the various model’s performance. The complex logistic regression model, with an AUC of 0.9471, was the best performing model. However, for interpretability, the interpretable logistic regression model offers marginally less performance with a significantly simpler interpretation.

|  |  |
| --- | --- |
| **Model** | **ROC AUC** |
| Interpretable Logistic Regression | 0.9358 |
| **Complex Logistic Regression** | **0.9471** |
| Linear Discriminant Analysis | 0.9126 |
| Quadratic Discriminant Analysis | 0.8997 |
| Decision Tree | 0.914 |
| Random Forest | 0.935 |
| Logistic Regression (Predicting Duration) |  |

## Final Conclusions

The decision to leave *duration* in our models is one that may prove controversial; it is not a variable that can be known beforehand, nor can it be used to predict which clients are most likely prospects. That being said, *duration* was far and away the single greatest predictor of success in our analysis, and that knowledge may prove more crucial to long-term success than client prediction. It is our recommendation that the bank undertake a study of the calls themselves to try and identify whether there are callers, techniques, campaigns, and compounding factors that lend themselves to keeping a prospective client on a call longer.

For the purposes of this analysis, we recommend the interpretable logistic regression model due to its high performance as well as its parameter interpretation.

# Appendix

## Data Description

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Type** | **Description** |
| **age** | *numeric* | Age of the customer in years |
| **job** | *factor* | Type of job the customer holds |
| **marital** | *factor* | Customer’s marital status |
| **education** | *factor* | Customer’s education level |
| **default** | *factor* | Whether the customer has credit in default |
| **housing** | *factor* | Whether the customer has a housing loan |
| **loan** | *factor* | Whether the customer has a personal loan |
| **contact** | *factor* | Method of communication |
| **month** | *factor* | Month of most recent contact |
| **day** | *factor* | Day of the week of most recent contact |
| **duration** | *numeric* | Length of most recent contact |
| **campaign** | *numeric* | Number of customer contacts during the campaign |
| **pdays** | *numeric* | Number of days since the customer’s most recent contact for a previous campaign |
| **previous** | *numeric* | Number of times the customer was contacted prior to this campaign |
| **poutcome** | *numeric* | Outcome of the previous campaign |
| **emp.var.rate** | *numeric* | Employment variation rate (quarterly indicator) |
| **cons.price.idx** | *numeric* | Consumer price index (monthly indicator) |
| **cons.conf.idx** | *numeric* | Consumer confidence index (monthly indicator) |
| **euribor3m** | *numeric* | Euribor (Euro Interbank Offered Rate) 3-month rate (daily indicator) [reference](https://www.euribor-rates.eu/en/) |
| **nr\_employed** | *numeric* | Number of employees (quarterly indicator) |
| **y** | *factor* | Subscription completion |

Figure 3: Data Dictionary

Diagram

Description automatically generated

Figure : Data Profiling Report - Continuous Variables

Graphical user interface, diagram

Description automatically generated

Figure : Data Profiling Report - Categorical Variables

## Data Cleanup

A picture containing text, receipt

Description automatically generated

Figure : Summary Table - All Variables

Chart, scatter chart

Description automatically generated

Figure : Data Profiling Report - Covariance Matrix

A picture containing graphical user interface

Description automatically generated

Figure : Data Profiling Report - Principal Component Analysis

Chart, line chart

Description automatically generated

Figure : Data Profiling Report - QQ Plot

A picture containing graphical user interface, application, table

Description automatically generated

Figure : skimr data summary

Table

Description automatically generated

Figure : skimr factor variable report

Table

Description automatically generated

Figure : skimr numeric variable report

|  |  |
| --- | --- |
| Variable | Mode Value |
| job | admin |
| marital | married |
| housing | yes |
| loan | no |
| contact | cellular |
| education | university degree |
| month | may |
| day\_of\_week | thu |
| poutcome | failure |

Figure : Mode of categorical variables with missing values

## Exploratory Data Analysis

Table

Description automatically generated

Figure : Frequency - Subscribed (Class)

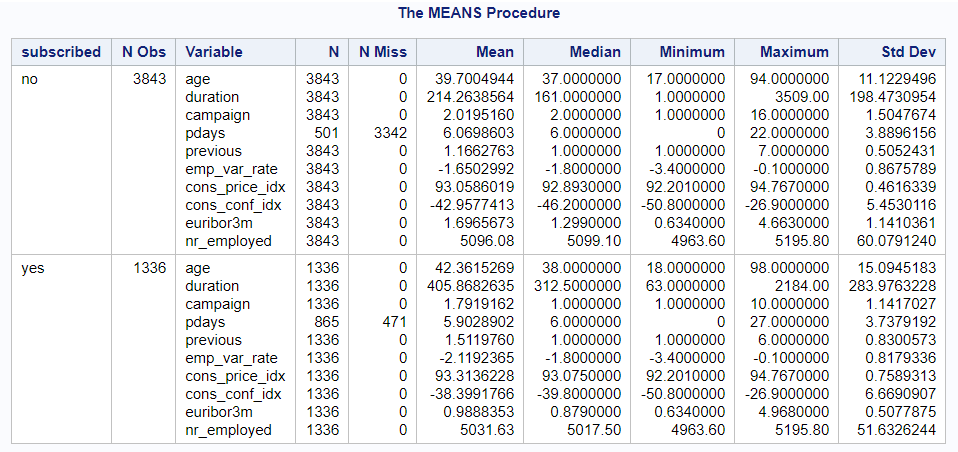


Figure 15: Summary Table - Continuous Variables

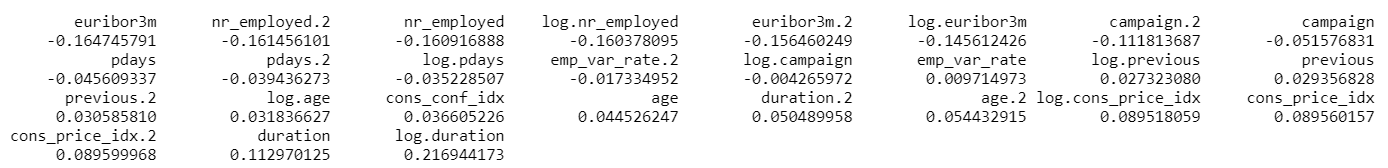


Figure 16: Correlation – Continuous Variables

## Logistic Regression (Interpretable)

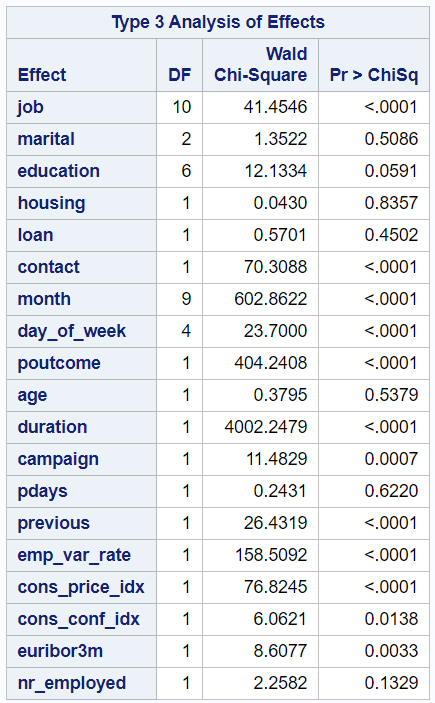


Figure 17: Logistic Regression (All Variables from EDA) - Type 3 Analysis of Effects

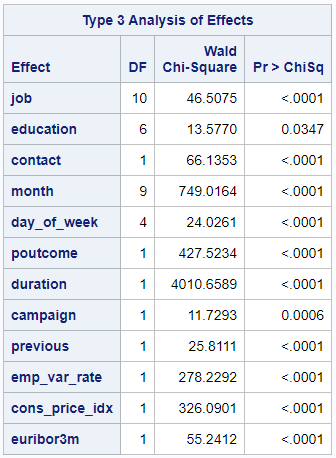


Figure 18: Logistic Regression (Final Interpretable Model) - Type 3 Analysis of Effects

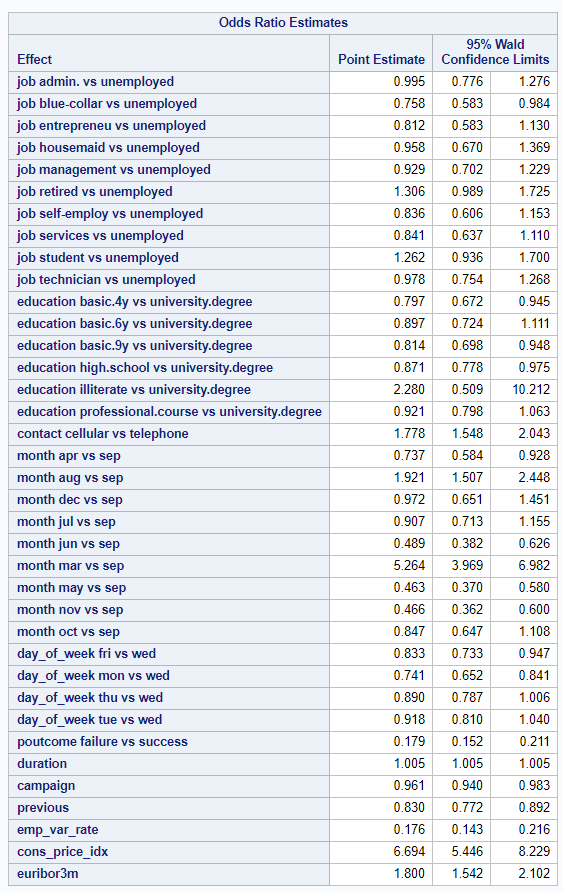


Figure 19: Logistic Regression (Interpretable) – Odds Ratio Estimates

Chart, line chart

Description automatically generated

Figure : Logistic Regression (Interpretable) - ROC Curve

Chart, line chart

Description automatically generated

Figure : Logistic Regression (Interpretable) - ROC Curve (just duration)

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

## Model 2 – Complicated Model

Chart, line chart

Description automatically generated

Figure : Logistic Regression (Complicated) - ROC Curve

Chart, line chart

Description automatically generated

Figure : Logistic Regression (Complicated) - ROC Comparison showing job\*education\*month removed

## Model 2 – Complicated Model – Analysis of Maximum Likelihood Estimates

| **Parameter** | **Factor 1** | **Factor 2** | **Factor 3** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Intercept |  |  |  | -6.0137 | 1.1467 | 27.5024 | <.0001 |
| education | basic.4y |  |  | -0.4253 | 0.3147 | 1.8264 | 0.1766 |
| education | basic.6y |  |  | -0.6366 | 0.8813 | 0.5217 | 0.4701 |
| education | basic.9y |  |  | -0.1431 | 0.2253 | 0.4031 | 0.5255 |
| education | high.school |  |  | -0.0283 | 0.1901 | 0.0221 | 0.8818 |
| education | illiterate |  |  | 0.6689 | 1.381 | 0.2346 | 0.6281 |
| education | professional.course |  |  | 0.1787 | 0.2324 | 0.5909 | 0.4421 |
| month | apr |  |  | 0.5587 | 1.0518 | 0.2822 | 0.5953 |
| month | aug |  |  | 0.9605 | 1.0526 | 0.8327 | 0.3615 |
| month | dec |  |  | -6.307 | 8.5864 | 0.5395 | 0.4626 |
| month | jul |  |  | 0.7678 | 1.0526 | 0.5321 | 0.4657 |
| month | jun |  |  | 0.8378 | 1.0533 | 0.6328 | 0.4263 |
| month | mar |  |  | 1.9212 | 0.3257 | 34.7985 | <.0001 |
| month | may |  |  | -0.2106 | 1.0479 | 0.0404 | 0.8407 |
| month | nov |  |  | 0.7782 | 1.0512 | 0.5481 | 0.4591 |
| month | oct |  |  | 0.582 | 0.3468 | 2.8163 | 0.0933 |
| poutcome | failure |  |  | -0.8641 | 0.0397 | 474.511 | <.0001 |
| job\*education | admin. | basic.4y |  | 5.5734 | 37.6087 | 0.022 | 0.8822 |
| job\*education | admin. | basic.6y |  | -10.8252 | 6.3096 | 2.9435 | 0.0862 |
| job\*education | admin. | basic.9y |  | 1.7441 | 37.4621 | 0.0022 | 0.9629 |
| job\*education | admin. | high.school |  | 1.1086 | 1.0558 | 1.1026 | 0.2937 |
| job\*education | admin. | illiterate |  | 0.3285 | 9.1294 | 0.0013 | 0.9713 |
| job\*education | admin. | professional.course |  | 1.27 | 1.1945 | 1.1305 | 0.2877 |
| job\*education | blue-collar | basic.4y |  | 1.507 | 1.1194 | 1.8126 | 0.1782 |
| job\*education | blue-collar | basic.6y |  | 10.3072 | 1.596 | 41.7089 | <.0001 |
| job\*education | blue-collar | basic.9y |  | 0.4452 | 1.0789 | 0.1703 | 0.6799 |
| job\*education | blue-collar | high.school |  | 2.047 | 1.1534 | 3.1494 | 0.076 |
| job\*education | blue-collar | illiterate |  | -1.5788 | 6.2255 | 0.0643 | 0.7998 |
| job\*education | blue-collar | professional.course |  | 0.1057 | 20.4716 | 0 | 0.9959 |
| job\*education | entrepreneu | basic.4y |  | 0.4614 | 1.918 | 0.0579 | 0.8099 |
| job\*education | entrepreneu | basic.6y |  | 4.7837 | 138.6 | 0.0012 | 0.9725 |
| job\*education | entrepreneu | basic.9y |  | -0.4324 | 2.4952 | 0.03 | 0.8624 |
| job\*education | entrepreneu | high.school |  | -15.1524 | 138.6 | 0.012 | 0.9129 |
| job\*education | entrepreneu | illiterate |  | 3.061 | 4.6578 | 0.4319 | 0.5111 |
| job\*education | entrepreneu | professional.course |  | -2.3706 | 22.3783 | 0.0112 | 0.9156 |
| job\*education | housemaid | basic.4y |  | 1.4617 | 1.1129 | 1.725 | 0.189 |
| job\*education | housemaid | basic.6y |  | -0.6169 | 21.9664 | 0.0008 | 0.9776 |
| job\*education | housemaid | basic.9y |  | -4.2993 | 21.6029 | 0.0396 | 0.8423 |
| job\*education | housemaid | high.school |  | 9.8767 | 5.593 | 3.1184 | 0.0774 |
| job\*education | housemaid | illiterate |  | 1.7293 | 4.8813 | 0.1255 | 0.7231 |
| job\*education | housemaid | professional.course |  | 0.0625 | 3.3398 | 0.0003 | 0.9851 |
| job\*education | management | basic.4y |  | -2.7048 | 111 | 0.0006 | 0.9806 |
| job\*education | management | basic.6y |  | 10.9607 | 138.3 | 0.0063 | 0.9368 |
| job\*education | management | basic.9y |  | 1.8142 | 157.8 | 0.0001 | 0.9908 |
| job\*education | management | high.school |  | -12.3258 | 138.4 | 0.0079 | 0.9291 |
| job\*education | management | illiterate |  | 0.3404 | 2.5549 | 0.0178 | 0.894 |
| job\*education | management | professional.course |  | 1.1592 | 112.5 | 0.0001 | 0.9918 |
| job\*education | retired | basic.4y |  | 1.9603 | 1.0865 | 3.2554 | 0.0712 |
| job\*education | retired | basic.6y |  | -0.8144 | 2.8183 | 0.0835 | 0.7726 |
| job\*education | retired | basic.9y |  | 0.999 | 1.1245 | 0.7893 | 0.3743 |
| job\*education | retired | high.school |  | 0.9549 | 1.0759 | 0.7877 | 0.3748 |
| job\*education | retired | illiterate |  | -4.8246 | 6.2554 | 0.5949 | 0.4405 |
| job\*education | retired | professional.course |  | 0.9989 | 1.0635 | 0.8822 | 0.3476 |
| job\*education | self-employ | basic.4y |  | 9.7595 | 53.7957 | 0.0329 | 0.856 |
| job\*education | self-employ | basic.6y |  | -16.6184 | 57.1962 | 0.0844 | 0.7714 |
| job\*education | self-employ | basic.9y |  | -0.4706 | 18.6253 | 0.0006 | 0.9798 |
| job\*education | self-employ | high.school |  | 2.895 | 2.0795 | 1.9381 | 0.1639 |
| job\*education | self-employ | illiterate |  | 2.8651 | 2.0558 | 1.9425 | 0.1634 |
| job\*education | self-employ | professional.course |  | 1.0256 | 1.5549 | 0.4351 | 0.5095 |
| job\*education | services | basic.4y |  | -5.1804 | 69.745 | 0.0055 | 0.9408 |
| job\*education | services | basic.6y |  | 2.2821 | 58.4899 | 0.0015 | 0.9689 |
| job\*education | services | basic.9y |  | 1.9954 | 1.883 | 1.1229 | 0.2893 |
| job\*education | services | high.school |  | 1.3232 | 1.12 | 1.3957 | 0.2374 |
| job\*education | services | illiterate |  | 1.3293 | 3.6272 | 0.1343 | 0.714 |
| job\*education | services | professional.course |  | -3.604 | 3.2314 | 1.244 | 0.2647 |
| job\*education | student | basic.4y |  | -11.1409 | 23.2235 | 0.2301 | 0.6314 |
| job\*education | student | basic.6y |  | 3.8584 | 1.9406 | 3.9531 | 0.0468 |
| job\*education | student | basic.9y |  | 1.3928 | 1.1014 | 1.599 | 0.206 |
| job\*education | student | high.school |  | 1.4624 | 1.0696 | 1.8694 | 0.1715 |
| job\*education | student | illiterate |  | 1.7139 | 22.2584 | 0.0059 | 0.9386 |
| job\*education | student | professional.course |  | 2.2529 | 1.9417 | 1.3463 | 0.2459 |
| job\*education | technician | basic.4y |  | -2.951 | 137.2 | 0.0005 | 0.9828 |
| job\*education | technician | basic.6y |  | -2.1524 | 15.3339 | 0.0197 | 0.8884 |
| job\*education | technician | basic.9y |  | -0.6207 | 1.7587 | 0.1246 | 0.7241 |
| job\*education | technician | high.school |  | 2.808 | 137.2 | 0.0004 | 0.9837 |
| job\*education | technician | illiterate |  | 1.2585 | 16.5592 | 0.0058 | 0.9394 |
| job\*education | technician | professional.course |  | 0.7471 | 1.042 | 0.5141 | 0.4734 |
| duration |  |  |  | 0.00629 | 0.000396 | 252.1539 | <.0001 |
| duration2 |  |  |  | -1.92E-06 | 1.43E-07 | 181.3026 | <.0001 |
| durationLog |  |  |  | 0.5966 | 0.0974 | 37.5121 | <.0001 |
| euribor3mLog |  |  |  | -1.6176 | 0.0384 | 1777.4226 | <.0001 |
| job\*education\*month | admin. | basic.4y | apr | -5.4031 | 37.6232 | 0.0206 | 0.8858 |
| job\*education\*month | admin. | basic.4y | aug | -4.5551 | 37.6265 | 0.0147 | 0.9036 |
| job\*education\*month | admin. | basic.4y | dec | -64.0778 | 3289.8 | 0.0004 | 0.9845 |
| job\*education\*month | admin. | basic.4y | jul | -5.2004 | 37.6198 | 0.0191 | 0.8901 |
| job\*education\*month | admin. | basic.4y | jun | -4.6955 | 37.624 | 0.0156 | 0.9007 |
| job\*education\*month | admin. | basic.4y | mar | 102.3 | 3272.5 | 0.001 | 0.9751 |
| job\*education\*month | admin. | basic.4y | may | -5.1521 | 37.6148 | 0.0188 | 0.8911 |
| job\*education\*month | admin. | basic.4y | nov | -7.0303 | 37.9052 | 0.0344 | 0.8529 |
| job\*education\*month | admin. | basic.4y | oct | -2.3029 | 37.5324 | 0.0038 | 0.9511 |
| job\*education\*month | admin. | basic.6y | apr | 10.5494 | 6.3023 | 2.8019 | 0.0942 |
| job\*education\*month | admin. | basic.6y | aug | 9.4834 | 7.6414 | 1.5402 | 0.2146 |
| job\*education\*month | admin. | basic.6y | dec | -182 | 3272.5 | 0.0031 | 0.9556 |
| job\*education\*month | admin. | basic.6y | jul | 10.8901 | 6.3193 | 2.9698 | 0.0848 |
| job\*education\*month | admin. | basic.6y | jun | 11.6 | 6.3811 | 3.3046 | 0.0691 |
| job\*education\*month | admin. | basic.6y | mar | 11.7971 | 6.4037 | 3.3938 | 0.0654 |
| job\*education\*month | admin. | basic.6y | may | 11.0231 | 6.2975 | 3.0638 | 0.0801 |
| job\*education\*month | admin. | basic.6y | nov | 10.7107 | 7.9627 | 1.8093 | 0.1786 |
| job\*education\*month | admin. | basic.6y | oct | 0 | . | . | . |
| job\*education\*month | admin. | basic.9y | apr | -1.599 | 37.4667 | 0.0018 | 0.966 |
| job\*education\*month | admin. | basic.9y | aug | -0.8801 | 37.4681 | 0.0006 | 0.9813 |
| job\*education\*month | admin. | basic.9y | dec | 18.9751 | 337.1 | 0.0032 | 0.9551 |
| job\*education\*month | admin. | basic.9y | jul | -1.0416 | 37.4638 | 0.0008 | 0.9778 |
| job\*education\*month | admin. | basic.9y | jun | -1.5916 | 37.4658 | 0.0018 | 0.9661 |
| job\*education\*month | admin. | basic.9y | mar | -3.965 | 37.9578 | 0.0109 | 0.9168 |
| job\*education\*month | admin. | basic.9y | may | -1.3599 | 37.463 | 0.0013 | 0.971 |
| job\*education\*month | admin. | basic.9y | nov | -2.3524 | 37.4682 | 0.0039 | 0.9499 |
| job\*education\*month | admin. | basic.9y | oct | -5.3677 | 37.4912 | 0.0205 | 0.8862 |
| job\*education\*month | admin. | high.school | apr | -1.1782 | 1.0722 | 1.2076 | 0.2718 |
| job\*education\*month | admin. | high.school | aug | -0.7119 | 1.0761 | 0.4376 | 0.5083 |
| job\*education\*month | admin. | high.school | dec | 5.4914 | 8.6108 | 0.4067 | 0.5236 |
| job\*education\*month | admin. | high.school | jul | -0.5548 | 1.0672 | 0.2703 | 0.6032 |
| job\*education\*month | admin. | high.school | jun | -0.8143 | 1.0737 | 0.5752 | 0.4482 |
| job\*education\*month | admin. | high.school | mar | -0.183 | 0.483 | 0.1436 | 0.7047 |
| job\*education\*month | admin. | high.school | may | -1.2916 | 1.0615 | 1.4805 | 0.2237 |
| job\*education\*month | admin. | high.school | nov | -1.5103 | 1.0815 | 1.9502 | 0.1626 |
| job\*education\*month | admin. | high.school | oct | 0.1716 | 0.4695 | 0.1336 | 0.7147 |
| job\*education\*month | admin. | illiterate | apr | -1.2234 | 9.2496 | 0.0175 | 0.8948 |
| job\*education\*month | admin. | illiterate | aug | -0.4765 | 10.2494 | 0.0022 | 0.9629 |
| job\*education\*month | admin. | illiterate | dec | 0 | . | . | . |
| job\*education\*month | admin. | illiterate | jul | -2.2608 | 9.2364 | 0.0599 | 0.8066 |
| job\*education\*month | admin. | illiterate | jun | -2.2127 | 9.326 | 0.0563 | 0.8125 |
| job\*education\*month | admin. | illiterate | mar | -107.2 | 3272.3 | 0.0011 | 0.9739 |
| job\*education\*month | admin. | illiterate | may | -0.6702 | 9.1778 | 0.0053 | 0.9418 |
| job\*education\*month | admin. | illiterate | nov | 1.9589 | 11.434 | 0.0294 | 0.864 |
| job\*education\*month | admin. | illiterate | oct | 0 | . | . | . |
| job\*education\*month | admin. | professional.course | apr | -0.8385 | 1.2891 | 0.4231 | 0.5154 |
| job\*education\*month | admin. | professional.course | aug | -1.8932 | 1.3664 | 1.9197 | 0.1659 |
| job\*education\*month | admin. | professional.course | dec | 8.9684 | 9.9157 | 0.8181 | 0.3658 |
| job\*education\*month | admin. | professional.course | jul | -1.1338 | 1.2748 | 0.791 | 0.3738 |
| job\*education\*month | admin. | professional.course | jun | -1.7068 | 1.3898 | 1.5081 | 0.2194 |
| job\*education\*month | admin. | professional.course | mar | -1.7835 | 1.072 | 2.7682 | 0.0962 |
| job\*education\*month | admin. | professional.course | may | -1.6932 | 1.2657 | 1.7894 | 0.181 |
| job\*education\*month | admin. | professional.course | nov | -0.8892 | 1.346 | 0.4364 | 0.5089 |
| job\*education\*month | admin. | professional.course | oct | -0.5906 | 1.1268 | 0.2747 | 0.6002 |
| job\*education\*month | blue-collar | basic.4y | apr | -2.5616 | 1.1493 | 4.968 | 0.0258 |
| job\*education\*month | blue-collar | basic.4y | aug | -0.8289 | 1.1301 | 0.538 | 0.4633 |
| job\*education\*month | blue-collar | basic.4y | dec | 10.0172 | 8.6846 | 1.3304 | 0.2487 |
| job\*education\*month | blue-collar | basic.4y | jul | -1.2206 | 1.12 | 1.1877 | 0.2758 |
| job\*education\*month | blue-collar | basic.4y | jun | -1.0101 | 1.1216 | 0.811 | 0.3678 |
| job\*education\*month | blue-collar | basic.4y | mar | 0.547 | 0.9527 | 0.3297 | 0.5658 |
| job\*education\*month | blue-collar | basic.4y | may | -1.5559 | 1.1021 | 1.9933 | 0.158 |
| job\*education\*month | blue-collar | basic.4y | nov | -2.0713 | 1.2446 | 2.7696 | 0.0961 |
| job\*education\*month | blue-collar | basic.4y | oct | -0.6872 | 1.5866 | 0.1876 | 0.6649 |
| job\*education\*month | blue-collar | basic.6y | apr | -11.1203 | 1.4467 | 59.0817 | <.0001 |
| job\*education\*month | blue-collar | basic.6y | aug | -9.376 | 1.4805 | 40.1072 | <.0001 |
| job\*education\*month | blue-collar | basic.6y | dec | 89.9836 | 31.8998 | 7.957 | 0.0048 |
| job\*education\*month | blue-collar | basic.6y | jul | -9.2887 | 1.376 | 45.5695 | <.0001 |
| job\*education\*month | blue-collar | basic.6y | jun | -9.6237 | 1.3858 | 48.2286 | <.0001 |
| job\*education\*month | blue-collar | basic.6y | mar | -8.2635 | 1.785 | 21.4321 | <.0001 |
| job\*education\*month | blue-collar | basic.6y | may | -9.3925 | 1.3545 | 48.086 | <.0001 |
| job\*education\*month | blue-collar | basic.6y | nov | -10.1843 | 1.4557 | 48.9489 | <.0001 |
| job\*education\*month | blue-collar | basic.6y | oct | 0 | . | . | . |
| job\*education\*month | blue-collar | basic.9y | apr | -1.396 | 1.1023 | 1.604 | 0.2053 |
| job\*education\*month | blue-collar | basic.9y | aug | -0.2676 | 1.1131 | 0.0578 | 0.81 |
| job\*education\*month | blue-collar | basic.9y | dec | 4.9605 | 8.6688 | 0.3274 | 0.5672 |
| job\*education\*month | blue-collar | basic.9y | jul | -0.0168 | 1.0873 | 0.0002 | 0.9877 |
| job\*education\*month | blue-collar | basic.9y | jun | -0.2401 | 1.0964 | 0.048 | 0.8267 |
| job\*education\*month | blue-collar | basic.9y | mar | -1.7192 | 0.7516 | 5.2319 | 0.0222 |
| job\*education\*month | blue-collar | basic.9y | may | -0.5564 | 1.0778 | 0.2665 | 0.6057 |
| job\*education\*month | blue-collar | basic.9y | nov | -0.343 | 1.108 | 0.0958 | 0.7569 |
| job\*education\*month | blue-collar | basic.9y | oct | 0.6171 | 0.7726 | 0.6381 | 0.4244 |
| job\*education\*month | blue-collar | high.school | apr | -2.6248 | 1.2401 | 4.4801 | 0.0343 |
| job\*education\*month | blue-collar | high.school | aug | -1.6773 | 1.2846 | 1.705 | 0.1916 |
| job\*education\*month | blue-collar | high.school | dec | 7.5528 | 8.7111 | 0.7517 | 0.3859 |
| job\*education\*month | blue-collar | high.school | jul | -1.1672 | 1.1892 | 0.9634 | 0.3263 |
| job\*education\*month | blue-collar | high.school | jun | -1.3797 | 1.1979 | 1.3265 | 0.2494 |
| job\*education\*month | blue-collar | high.school | mar | -1.0159 | 0.8245 | 1.518 | 0.2179 |
| job\*education\*month | blue-collar | high.school | may | -2.1742 | 1.1728 | 3.4367 | 0.0638 |
| job\*education\*month | blue-collar | high.school | nov | -1.3833 | 1.2177 | 1.2906 | 0.2559 |
| job\*education\*month | blue-collar | high.school | oct | 0.4502 | 1.0848 | 0.1723 | 0.6781 |
| job\*education\*month | blue-collar | illiterate | apr | 7.511 | 6.2867 | 1.4274 | 0.2322 |
| job\*education\*month | blue-collar | illiterate | aug | -0.7114 | 6.2863 | 0.0128 | 0.9099 |
| job\*education\*month | blue-collar | illiterate | dec | -13.5761 | 215.2 | 0.004 | 0.9497 |
| job\*education\*month | blue-collar | illiterate | jul | -1.0421 | 6.2147 | 0.0281 | 0.8668 |
| job\*education\*month | blue-collar | illiterate | jun | 0.2891 | 6.2339 | 0.0022 | 0.963 |
| job\*education\*month | blue-collar | illiterate | mar | -2.0852 | 12.098 | 0.0297 | 0.8632 |
| job\*education\*month | blue-collar | illiterate | may | 2.2471 | 6.187 | 0.1319 | 0.7165 |
| job\*education\*month | blue-collar | illiterate | nov | 0.8637 | 6.3269 | 0.0186 | 0.8914 |
| job\*education\*month | blue-collar | illiterate | oct | 0 | . | . | . |
| job\*education\*month | blue-collar | professional.course | apr | -0.7759 | 20.4798 | 0.0014 | 0.9698 |
| job\*education\*month | blue-collar | professional.course | aug | -0.316 | 20.4836 | 0.0002 | 0.9877 |
| job\*education\*month | blue-collar | professional.course | dec | 7.418 | 22.2133 | 0.1115 | 0.7384 |
| job\*education\*month | blue-collar | professional.course | jul | 0.6259 | 20.4773 | 0.0009 | 0.9756 |
| job\*education\*month | blue-collar | professional.course | jun | 0.1654 | 20.4768 | 0.0001 | 0.9936 |
| job\*education\*month | blue-collar | professional.course | mar | 1.8982 | 20.4793 | 0.0086 | 0.9261 |
| job\*education\*month | blue-collar | professional.course | may | -0.7804 | 20.4752 | 0.0015 | 0.9696 |
| job\*education\*month | blue-collar | professional.course | nov | -0.2973 | 20.4805 | 0.0002 | 0.9884 |
| job\*education\*month | blue-collar | professional.course | oct | 0.522 | 20.4641 | 0.0007 | 0.9796 |
| job\*education\*month | entrepreneu | basic.4y | apr | -1.9817 | 2.1871 | 0.821 | 0.3649 |
| job\*education\*month | entrepreneu | basic.4y | aug | -1.1558 | 5.3241 | 0.0471 | 0.8281 |
| job\*education\*month | entrepreneu | basic.4y | dec | 155.7 | 9632.2 | 0.0003 | 0.9871 |
| job\*education\*month | entrepreneu | basic.4y | jul | -4.1997 | 5.6583 | 0.5509 | 0.458 |
| job\*education\*month | entrepreneu | basic.4y | jun | 0.4633 | 2.0973 | 0.0488 | 0.8251 |
| job\*education\*month | entrepreneu | basic.4y | mar | -80.9243 | 2267.2 | 0.0013 | 0.9715 |
| job\*education\*month | entrepreneu | basic.4y | may | -0.2261 | 2.021 | 0.0125 | 0.9109 |
| job\*education\*month | entrepreneu | basic.4y | nov | -2.882 | 4.298 | 0.4496 | 0.5025 |
| job\*education\*month | entrepreneu | basic.4y | oct | 0 | . | . | . |
| job\*education\*month | entrepreneu | basic.6y | apr | 0.4653 | 138.7 | 0 | 0.9973 |
| job\*education\*month | entrepreneu | basic.6y | aug | -7.5312 | 139.2 | 0.0029 | 0.9569 |
| job\*education\*month | entrepreneu | basic.6y | dec | 24.7943 | 970.6 | 0.0007 | 0.9796 |
| job\*education\*month | entrepreneu | basic.6y | jul | -2.825 | 138.6 | 0.0004 | 0.9837 |
| job\*education\*month | entrepreneu | basic.6y | jun | -2.8746 | 138.6 | 0.0004 | 0.9835 |
| job\*education\*month | entrepreneu | basic.6y | mar | 0 | . | . | . |
| job\*education\*month | entrepreneu | basic.6y | may | -7.708 | 138.7 | 0.0031 | 0.9557 |
| job\*education\*month | entrepreneu | basic.6y | nov | -2.0016 | 138.6 | 0.0002 | 0.9885 |
| job\*education\*month | entrepreneu | basic.6y | oct | 0 | . | . | . |
| job\*education\*month | entrepreneu | basic.9y | apr | -1.5648 | 2.7942 | 0.3136 | 0.5755 |
| job\*education\*month | entrepreneu | basic.9y | aug | 1.5346 | 2.9451 | 0.2715 | 0.6023 |
| job\*education\*month | entrepreneu | basic.9y | dec | 0 | . | . | . |
| job\*education\*month | entrepreneu | basic.9y | jul | -0.1678 | 2.8063 | 0.0036 | 0.9523 |
| job\*education\*month | entrepreneu | basic.9y | jun | 0.0417 | 2.6572 | 0.0002 | 0.9875 |
| job\*education\*month | entrepreneu | basic.9y | mar | -0.4567 | 5.7002 | 0.0064 | 0.9361 |
| job\*education\*month | entrepreneu | basic.9y | may | 0.8851 | 2.5428 | 0.1212 | 0.7278 |
| job\*education\*month | entrepreneu | basic.9y | nov | 0.5518 | 2.6605 | 0.043 | 0.8357 |
| job\*education\*month | entrepreneu | basic.9y | oct | 0 | . | . | . |
| job\*education\*month | entrepreneu | high.school | apr | 14.2232 | 138.6 | 0.0105 | 0.9183 |
| job\*education\*month | entrepreneu | high.school | aug | 15.0264 | 138.6 | 0.0118 | 0.9137 |
| job\*education\*month | entrepreneu | high.school | dec | -104.4 | 2516.1 | 0.0017 | 0.9669 |
| job\*education\*month | entrepreneu | high.school | jul | 15.6058 | 138.6 | 0.0127 | 0.9104 |
| job\*education\*month | entrepreneu | high.school | jun | 15.9473 | 138.6 | 0.0132 | 0.9084 |
| job\*education\*month | entrepreneu | high.school | mar | -4.0341 | 2266.5 | 0 | 0.9986 |
| job\*education\*month | entrepreneu | high.school | may | 14.5924 | 138.6 | 0.0111 | 0.9162 |
| job\*education\*month | entrepreneu | high.school | nov | 13.5318 | 138.6 | 0.0095 | 0.9222 |
| job\*education\*month | entrepreneu | high.school | oct | 0 | . | . | . |
| job\*education\*month | entrepreneu | illiterate | apr | -2.1559 | 5.3814 | 0.1605 | 0.6887 |
| job\*education\*month | entrepreneu | illiterate | aug | -1.3892 | 14.584 | 0.0091 | 0.9241 |
| job\*education\*month | entrepreneu | illiterate | dec | 0 | . | . | . |
| job\*education\*month | entrepreneu | illiterate | jul | 3.0508 | 7.7857 | 0.1535 | 0.6952 |
| job\*education\*month | entrepreneu | illiterate | jun | -5.9858 | 5.0257 | 1.4186 | 0.2336 |
| job\*education\*month | entrepreneu | illiterate | mar | 0 | . | . | . |
| job\*education\*month | entrepreneu | illiterate | may | 0.4465 | 5.452 | 0.0067 | 0.9347 |
| job\*education\*month | entrepreneu | illiterate | nov | 0 | . | . | . |
| job\*education\*month | entrepreneu | illiterate | oct | 0 | . | . | . |
| job\*education\*month | entrepreneu | professional.course | apr | 1.4725 | 22.433 | 0.0043 | 0.9477 |
| job\*education\*month | entrepreneu | professional.course | aug | 2.4087 | 22.3974 | 0.0116 | 0.9144 |
| job\*education\*month | entrepreneu | professional.course | dec | -41.845 | 9359.4 | 0 | 0.9964 |
| job\*education\*month | entrepreneu | professional.course | jul | -1.8445 | 22.5793 | 0.0067 | 0.9349 |
| job\*education\*month | entrepreneu | professional.course | jun | 2.6589 | 22.3919 | 0.0141 | 0.9055 |
| job\*education\*month | entrepreneu | professional.course | mar | 7.0069 | 29.9079 | 0.0549 | 0.8148 |
| job\*education\*month | entrepreneu | professional.course | may | 1.2637 | 22.3985 | 0.0032 | 0.955 |
| job\*education\*month | entrepreneu | professional.course | nov | 0.6683 | 22.4158 | 0.0009 | 0.9762 |
| job\*education\*month | entrepreneu | professional.course | oct | 1.1067 | 22.6358 | 0.0024 | 0.961 |
| job\*education\*month | housemaid | basic.4y | apr | 0.2692 | 1.2013 | 0.0502 | 0.8227 |
| job\*education\*month | housemaid | basic.4y | aug | -1.0268 | 1.168 | 0.7729 | 0.3793 |
| job\*education\*month | housemaid | basic.4y | dec | 6.2968 | 8.6483 | 0.5301 | 0.4666 |
| job\*education\*month | housemaid | basic.4y | jul | -1.8083 | 1.1899 | 2.3095 | 0.1286 |
| job\*education\*month | housemaid | basic.4y | jun | 0.1934 | 1.1948 | 0.0262 | 0.8714 |
| job\*education\*month | housemaid | basic.4y | mar | 0.4446 | 1.6386 | 0.0736 | 0.7861 |
| job\*education\*month | housemaid | basic.4y | may | -0.4445 | 1.193 | 0.1388 | 0.7095 |
| job\*education\*month | housemaid | basic.4y | nov | -0.8271 | 1.2466 | 0.4402 | 0.507 |
| job\*education\*month | housemaid | basic.4y | oct | -2.9454 | 1.0689 | 7.5926 | 0.0059 |
| job\*education\*month | housemaid | basic.6y | apr | 1.9052 | 22.0024 | 0.0075 | 0.931 |
| job\*education\*month | housemaid | basic.6y | aug | 1.9114 | 21.9695 | 0.0076 | 0.9307 |
| job\*education\*month | housemaid | basic.6y | dec | -147.1 | 9594.3 | 0.0002 | 0.9878 |
| job\*education\*month | housemaid | basic.6y | jul | 1.7774 | 22.0073 | 0.0065 | 0.9356 |
| job\*education\*month | housemaid | basic.6y | jun | -1.1694 | 22.2474 | 0.0028 | 0.9581 |
| job\*education\*month | housemaid | basic.6y | mar | 101.8 | 1408.9 | 0.0052 | 0.9424 |
| job\*education\*month | housemaid | basic.6y | may | 0.8134 | 21.9769 | 0.0014 | 0.9705 |
| job\*education\*month | housemaid | basic.6y | nov | -2.22 | 21.5985 | 0.0106 | 0.9181 |
| job\*education\*month | housemaid | basic.6y | oct | 0 | . | . | . |
| job\*education\*month | housemaid | basic.9y | apr | 5.4735 | 21.6312 | 0.064 | 0.8002 |
| job\*education\*month | housemaid | basic.9y | aug | 1.1456 | 23.006 | 0.0025 | 0.9603 |
| job\*education\*month | housemaid | basic.9y | dec | 3.7461 | 112.8 | 0.0011 | 0.9735 |
| job\*education\*month | housemaid | basic.9y | jul | 4.5554 | 21.6218 | 0.0444 | 0.8331 |
| job\*education\*month | housemaid | basic.9y | jun | 0.9518 | 21.7542 | 0.0019 | 0.9651 |
| job\*education\*month | housemaid | basic.9y | mar | -15.2424 | 1408 | 0.0001 | 0.9914 |
| job\*education\*month | housemaid | basic.9y | may | 3.8542 | 21.6162 | 0.0318 | 0.8585 |
| job\*education\*month | housemaid | basic.9y | nov | 4.3534 | 21.6622 | 0.0404 | 0.8407 |
| job\*education\*month | housemaid | basic.9y | oct | 0 | . | . | . |
| job\*education\*month | housemaid | high.school | apr | -8.2615 | 5.6874 | 2.11 | 0.1463 |
| job\*education\*month | housemaid | high.school | aug | -10.298 | 5.718 | 3.2435 | 0.0717 |
| job\*education\*month | housemaid | high.school | dec | 96.4215 | 2109 | 0.0021 | 0.9635 |
| job\*education\*month | housemaid | high.school | jul | -12.7539 | 5.8056 | 4.826 | 0.028 |
| job\*education\*month | housemaid | high.school | jun | -9.8892 | 5.6493 | 3.0643 | 0.08 |
| job\*education\*month | housemaid | high.school | mar | -9.1327 | 5.7579 | 2.5158 | 0.1127 |
| job\*education\*month | housemaid | high.school | may | -9.6114 | 5.6314 | 2.913 | 0.0879 |
| job\*education\*month | housemaid | high.school | nov | -9.2896 | 5.6481 | 2.7051 | 0.1 |
| job\*education\*month | housemaid | high.school | oct | 0 | . | . | . |
| job\*education\*month | housemaid | illiterate | apr | -7.311 | 5.5299 | 1.7479 | 0.1861 |
| job\*education\*month | housemaid | illiterate | aug | 0.5352 | 9.6057 | 0.0031 | 0.9556 |
| job\*education\*month | housemaid | illiterate | dec | 0 | . | . | . |
| job\*education\*month | housemaid | illiterate | jul | 0.1003 | 5.322 | 0.0004 | 0.985 |
| job\*education\*month | housemaid | illiterate | jun | 3.0034 | 7.1517 | 0.1764 | 0.6745 |
| job\*education\*month | housemaid | illiterate | mar | 0 | . | . | . |
| job\*education\*month | housemaid | illiterate | may | -3.9903 | 5.2799 | 0.5712 | 0.4498 |
| job\*education\*month | housemaid | illiterate | nov | 0 | . | . | . |
| job\*education\*month | housemaid | illiterate | oct | 0 | . | . | . |
| job\*education\*month | housemaid | professional.course | apr | 0.7317 | 3.4852 | 0.0441 | 0.8337 |
| job\*education\*month | housemaid | professional.course | aug | -0.5699 | 3.6554 | 0.0243 | 0.8761 |
| job\*education\*month | housemaid | professional.course | dec | 26.8863 | 9359 | 0 | 0.9977 |
| job\*education\*month | housemaid | professional.course | jul | -0.2733 | 3.5645 | 0.0059 | 0.9389 |
| job\*education\*month | housemaid | professional.course | jun | -1.2217 | 3.8413 | 0.1011 | 0.7505 |
| job\*education\*month | housemaid | professional.course | mar | 0 | . | . | . |
| job\*education\*month | housemaid | professional.course | may | 0.8707 | 3.5233 | 0.0611 | 0.8048 |
| job\*education\*month | housemaid | professional.course | nov | -0.4817 | 3.6938 | 0.017 | 0.8962 |
| job\*education\*month | housemaid | professional.course | oct | 0 | . | . | . |
| job\*education\*month | management | basic.4y | apr | 1.7962 | 111.1 | 0.0003 | 0.9871 |
| job\*education\*month | management | basic.4y | aug | -0.967 | 122.7 | 0.0001 | 0.9937 |
| job\*education\*month | management | basic.4y | dec | -11.8865 | 997.4 | 0.0001 | 0.9905 |
| job\*education\*month | management | basic.4y | jul | 4.1814 | 111.1 | 0.0014 | 0.97 |
| job\*education\*month | management | basic.4y | jun | 0.6925 | 111.1 | 0 | 0.995 |
| job\*education\*month | management | basic.4y | mar | -4.7021 | 115.4 | 0.0017 | 0.9675 |
| job\*education\*month | management | basic.4y | may | 2.458 | 111.1 | 0.0005 | 0.9823 |
| job\*education\*month | management | basic.4y | nov | 3.0829 | 111.1 | 0.0008 | 0.9779 |
| job\*education\*month | management | basic.4y | oct | 2.5356 | 111.1 | 0.0005 | 0.9818 |
| job\*education\*month | management | basic.6y | apr | -10.0847 | 138.3 | 0.0053 | 0.9419 |
| job\*education\*month | management | basic.6y | aug | -2.9749 | 142.8 | 0.0004 | 0.9834 |
| job\*education\*month | management | basic.6y | dec | 72.9485 | 1106.9 | 0.0043 | 0.9475 |
| job\*education\*month | management | basic.6y | jul | -11.9792 | 138.5 | 0.0075 | 0.9311 |
| job\*education\*month | management | basic.6y | jun | -10.5194 | 138.3 | 0.0058 | 0.9394 |
| job\*education\*month | management | basic.6y | mar | -10.6453 | 138.3 | 0.0059 | 0.9386 |
| job\*education\*month | management | basic.6y | may | -9.4618 | 138.3 | 0.0047 | 0.9455 |
| job\*education\*month | management | basic.6y | nov | -9.5843 | 138.3 | 0.0048 | 0.9447 |
| job\*education\*month | management | basic.6y | oct | 0 | . | . | . |
| job\*education\*month | management | basic.9y | apr | -2.8595 | 157.8 | 0.0003 | 0.9855 |
| job\*education\*month | management | basic.9y | aug | -5.9364 | 158 | 0.0014 | 0.97 |
| job\*education\*month | management | basic.9y | dec | 3.708 | 158 | 0.0006 | 0.9813 |
| job\*education\*month | management | basic.9y | jul | -1.822 | 157.8 | 0.0001 | 0.9908 |
| job\*education\*month | management | basic.9y | jun | -1.3668 | 157.8 | 0.0001 | 0.9931 |
| job\*education\*month | management | basic.9y | mar | 14.9679 | 1416 | 0.0001 | 0.9916 |
| job\*education\*month | management | basic.9y | may | -1.6718 | 157.8 | 0.0001 | 0.9915 |
| job\*education\*month | management | basic.9y | nov | -3.3314 | 157.8 | 0.0004 | 0.9832 |
| job\*education\*month | management | basic.9y | oct | -0.2057 | 157.8 | 0 | 0.999 |
| job\*education\*month | management | high.school | apr | 11.4959 | 138.5 | 0.0069 | 0.9338 |
| job\*education\*month | management | high.school | aug | 12.8705 | 138.4 | 0.0086 | 0.9259 |
| job\*education\*month | management | high.school | dec | 0 | . | . | . |
| job\*education\*month | management | high.school | jul | 13.1184 | 138.4 | 0.009 | 0.9245 |
| job\*education\*month | management | high.school | jun | 11.3975 | 138.5 | 0.0068 | 0.9344 |
| job\*education\*month | management | high.school | mar | -14.0167 | 1803.6 | 0.0001 | 0.9938 |
| job\*education\*month | management | high.school | may | 11.3346 | 138.4 | 0.0067 | 0.9348 |
| job\*education\*month | management | high.school | nov | 11.6615 | 138.4 | 0.0071 | 0.9329 |
| job\*education\*month | management | high.school | oct | 0 | . | . | . |
| job\*education\*month | management | illiterate | apr | 2.6942 | 4.0402 | 0.4447 | 0.5049 |
| job\*education\*month | management | illiterate | aug | 0.4401 | 69.1886 | 0 | 0.9949 |
| job\*education\*month | management | illiterate | dec | 0 | . | . | . |
| job\*education\*month | management | illiterate | jul | -2.0753 | 7.8876 | 0.0692 | 0.7925 |
| job\*education\*month | management | illiterate | jun | 0.0481 | 5.8758 | 0.0001 | 0.9935 |
| job\*education\*month | management | illiterate | mar | 0 | . | . | . |
| job\*education\*month | management | illiterate | may | -0.8842 | 2.9609 | 0.0892 | 0.7652 |
| job\*education\*month | management | illiterate | nov | 0 | . | . | . |
| job\*education\*month | management | illiterate | oct | 0 | . | . | . |
| job\*education\*month | management | professional.course | apr | -1.5849 | 112.5 | 0.0002 | 0.9888 |
| job\*education\*month | management | professional.course | aug | -2.5234 | 112.6 | 0.0005 | 0.9821 |
| job\*education\*month | management | professional.course | dec | 0.2626 | 118.6 | 0 | 0.9982 |
| job\*education\*month | management | professional.course | jul | -0.6473 | 112.5 | 0 | 0.9954 |
| job\*education\*month | management | professional.course | jun | 0.1796 | 112.5 | 0 | 0.9987 |
| job\*education\*month | management | professional.course | mar | 14.372 | 1011.8 | 0.0002 | 0.9887 |
| job\*education\*month | management | professional.course | may | -0.79 | 112.5 | 0 | 0.9944 |
| job\*education\*month | management | professional.course | nov | -1.1226 | 112.5 | 0.0001 | 0.992 |
| job\*education\*month | management | professional.course | oct | -2.1648 | 112.5 | 0.0004 | 0.9847 |
| job\*education\*month | retired | basic.4y | apr | 0.0894 | 1.0924 | 0.0067 | 0.9347 |
| job\*education\*month | retired | basic.4y | aug | -1.01 | 1.0916 | 0.8561 | 0.3548 |
| job\*education\*month | retired | basic.4y | dec | 5.6666 | 8.6145 | 0.4327 | 0.5107 |
| job\*education\*month | retired | basic.4y | jul | -1.071 | 1.1055 | 0.9386 | 0.3326 |
| job\*education\*month | retired | basic.4y | jun | -1.0525 | 1.1513 | 0.8357 | 0.3606 |
| job\*education\*month | retired | basic.4y | mar | -0.5525 | 0.5196 | 1.1306 | 0.2876 |
| job\*education\*month | retired | basic.4y | may | -0.6625 | 1.126 | 0.3462 | 0.5563 |
| job\*education\*month | retired | basic.4y | nov | -0.5202 | 1.1414 | 0.2077 | 0.6486 |
| job\*education\*month | retired | basic.4y | oct | -0.6785 | 0.4583 | 2.1912 | 0.1388 |
| job\*education\*month | retired | basic.6y | apr | -3.6274 | 9.0996 | 0.1589 | 0.6902 |
| job\*education\*month | retired | basic.6y | aug | 0.7829 | 2.8099 | 0.0776 | 0.7805 |
| job\*education\*month | retired | basic.6y | dec | -31.3188 | 43.0267 | 0.5298 | 0.4667 |
| job\*education\*month | retired | basic.6y | jul | 0.9718 | 2.9101 | 0.1115 | 0.7384 |
| job\*education\*month | retired | basic.6y | jun | -0.5411 | 7.2059 | 0.0056 | 0.9401 |
| job\*education\*month | retired | basic.6y | mar | -2.6844 | 29.1145 | 0.0085 | 0.9265 |
| job\*education\*month | retired | basic.6y | may | 1.2123 | 2.9354 | 0.1706 | 0.6796 |
| job\*education\*month | retired | basic.6y | nov | 8.9825 | 16.7767 | 0.2867 | 0.5924 |
| job\*education\*month | retired | basic.6y | oct | 2.3781 | 2.2594 | 1.1079 | 0.2925 |
| job\*education\*month | retired | basic.9y | apr | 0.9524 | 1.5008 | 0.4027 | 0.5257 |
| job\*education\*month | retired | basic.9y | aug | -1.2037 | 1.281 | 0.883 | 0.3474 |
| job\*education\*month | retired | basic.9y | dec | 5.9122 | 8.8795 | 0.4433 | 0.5055 |
| job\*education\*month | retired | basic.9y | jul | -0.0744 | 1.2848 | 0.0034 | 0.9538 |
| job\*education\*month | retired | basic.9y | jun | -0.2947 | 1.5217 | 0.0375 | 0.8464 |
| job\*education\*month | retired | basic.9y | mar | -2.1019 | 1.3721 | 2.3467 | 0.1255 |
| job\*education\*month | retired | basic.9y | may | -2.2507 | 1.6372 | 1.8899 | 0.1692 |
| job\*education\*month | retired | basic.9y | nov | 0.1205 | 1.4464 | 0.0069 | 0.9336 |
| job\*education\*month | retired | basic.9y | oct | -0.4046 | 1.0072 | 0.1613 | 0.6879 |
| job\*education\*month | retired | high.school | apr | -0.6268 | 1.2078 | 0.2693 | 0.6038 |
| job\*education\*month | retired | high.school | aug | -1.1325 | 1.1594 | 0.9542 | 0.3286 |
| job\*education\*month | retired | high.school | dec | 7.9184 | 8.6253 | 0.8428 | 0.3586 |
| job\*education\*month | retired | high.school | jul | -0.2274 | 1.1758 | 0.0374 | 0.8467 |
| job\*education\*month | retired | high.school | jun | -0.7125 | 1.2072 | 0.3484 | 0.555 |
| job\*education\*month | retired | high.school | mar | -4.921 | 1.4387 | 11.6999 | 0.0006 |
| job\*education\*month | retired | high.school | may | -0.959 | 1.1696 | 0.6723 | 0.4122 |
| job\*education\*month | retired | high.school | nov | 0.4508 | 1.217 | 0.1372 | 0.7111 |
| job\*education\*month | retired | high.school | oct | -0.1403 | 0.6848 | 0.042 | 0.8377 |
| job\*education\*month | retired | illiterate | apr | 1.4918 | 10.7933 | 0.0191 | 0.8901 |
| job\*education\*month | retired | illiterate | aug | 4.2896 | 6.3225 | 0.4603 | 0.4975 |
| job\*education\*month | retired | illiterate | dec | 0 | . | . | . |
| job\*education\*month | retired | illiterate | jul | 1.4966 | 6.4418 | 0.054 | 0.8163 |
| job\*education\*month | retired | illiterate | jun | 3.7668 | 9.2702 | 0.1651 | 0.6845 |
| job\*education\*month | retired | illiterate | mar | 12.9532 | 29.286 | 0.1956 | 0.6583 |
| job\*education\*month | retired | illiterate | may | 4.0268 | 6.5315 | 0.3801 | 0.5376 |
| job\*education\*month | retired | illiterate | nov | -8.2063 | 17.7519 | 0.2137 | 0.6439 |
| job\*education\*month | retired | illiterate | oct | 0 | . | . | . |
| job\*education\*month | retired | professional.course | apr | 0.9816 | 1.2312 | 0.6357 | 0.4253 |
| job\*education\*month | retired | professional.course | aug | -0.4434 | 1.1504 | 0.1485 | 0.6999 |
| job\*education\*month | retired | professional.course | dec | 5.726 | 8.6546 | 0.4377 | 0.5082 |
| job\*education\*month | retired | professional.course | jul | -1.1433 | 1.2212 | 0.8765 | 0.3492 |
| job\*education\*month | retired | professional.course | jun | -0.4126 | 1.2025 | 0.1177 | 0.7315 |
| job\*education\*month | retired | professional.course | mar | -1.7162 | 0.8706 | 3.8857 | 0.0487 |
| job\*education\*month | retired | professional.course | may | -0.8293 | 1.2199 | 0.4621 | 0.4966 |
| job\*education\*month | retired | professional.course | nov | -0.5726 | 1.2194 | 0.2205 | 0.6387 |
| job\*education\*month | retired | professional.course | oct | 0.5286 | 0.7184 | 0.5414 | 0.4619 |
| job\*education\*month | self-employ | basic.4y | apr | -12.0283 | 53.8358 | 0.0499 | 0.8232 |
| job\*education\*month | self-employ | basic.4y | aug | -8.4109 | 53.807 | 0.0244 | 0.8758 |
| job\*education\*month | self-employ | basic.4y | dec | -26.9584 | 22.1886 | 1.4761 | 0.2244 |
| job\*education\*month | self-employ | basic.4y | jul | -13.0526 | 55.69 | 0.0549 | 0.8147 |
| job\*education\*month | self-employ | basic.4y | jun | -8.72 | 53.808 | 0.0263 | 0.8713 |
| job\*education\*month | self-employ | basic.4y | mar | -5.2088 | 18.7168 | 0.0774 | 0.7808 |
| job\*education\*month | self-employ | basic.4y | may | -10.7159 | 53.8471 | 0.0396 | 0.8423 |
| job\*education\*month | self-employ | basic.4y | nov | -10.8475 | 53.8486 | 0.0406 | 0.8404 |
| job\*education\*month | self-employ | basic.4y | oct | 1.3009 | 0.9021 | 2.0795 | 0.1493 |
| job\*education\*month | self-employ | basic.6y | apr | 14.6286 | 58.5703 | 0.0624 | 0.8028 |
| job\*education\*month | self-employ | basic.6y | aug | 15.2187 | 58.1083 | 0.0686 | 0.7934 |
| job\*education\*month | self-employ | basic.6y | dec | 0 | . | . | . |
| job\*education\*month | self-employ | basic.6y | jul | 16.8888 | 59.1291 | 0.0816 | 0.7752 |
| job\*education\*month | self-employ | basic.6y | jun | 16.1991 | 58.0478 | 0.0779 | 0.7802 |
| job\*education\*month | self-employ | basic.6y | mar | 0 | . | . | . |
| job\*education\*month | self-employ | basic.6y | may | 17.0128 | 57.2268 | 0.0884 | 0.7662 |
| job\*education\*month | self-employ | basic.6y | nov | 15.8181 | 57.2144 | 0.0764 | 0.7822 |
| job\*education\*month | self-employ | basic.6y | oct | 0 | . | . | . |
| job\*education\*month | self-employ | basic.9y | apr | 0.1694 | 18.6565 | 0.0001 | 0.9928 |
| job\*education\*month | self-employ | basic.9y | aug | 2.4454 | 18.6355 | 0.0172 | 0.8956 |
| job\*education\*month | self-employ | basic.9y | dec | 0 | . | . | . |
| job\*education\*month | self-employ | basic.9y | jul | -1.2258 | 18.6693 | 0.0043 | 0.9477 |
| job\*education\*month | self-employ | basic.9y | jun | 0.3226 | 18.6629 | 0.0003 | 0.9862 |
| job\*education\*month | self-employ | basic.9y | mar | 0 | . | . | . |
| job\*education\*month | self-employ | basic.9y | may | 0.4682 | 18.6318 | 0.0006 | 0.98 |
| job\*education\*month | self-employ | basic.9y | nov | 0.9753 | 18.6478 | 0.0027 | 0.9583 |
| job\*education\*month | self-employ | basic.9y | oct | 0 | . | . | . |
| job\*education\*month | self-employ | high.school | apr | -5.0442 | 2.5152 | 4.0219 | 0.0449 |
| job\*education\*month | self-employ | high.school | aug | -2.4393 | 2.2919 | 1.1327 | 0.2872 |
| job\*education\*month | self-employ | high.school | dec | 20.1158 | 16.8687 | 1.422 | 0.2331 |
| job\*education\*month | self-employ | high.school | jul | -3.8239 | 2.7001 | 2.0057 | 0.1567 |
| job\*education\*month | self-employ | high.school | jun | -4.2335 | 2.7932 | 2.2973 | 0.1296 |
| job\*education\*month | self-employ | high.school | mar | 0 | . | . | . |
| job\*education\*month | self-employ | high.school | may | -3.1585 | 2.23 | 2.0061 | 0.1567 |
| job\*education\*month | self-employ | high.school | nov | -4.3298 | 2.8144 | 2.3668 | 0.1239 |
| job\*education\*month | self-employ | high.school | oct | 0 | . | . | . |
| job\*education\*month | self-employ | illiterate | apr | 2.9768 | 13.1022 | 0.0516 | 0.8203 |
| job\*education\*month | self-employ | illiterate | aug | -2.6416 | 11.2324 | 0.0553 | 0.8141 |
| job\*education\*month | self-employ | illiterate | dec | 0 | . | . | . |
| job\*education\*month | self-employ | illiterate | jul | 2.4549 | 21.021 | 0.0136 | 0.907 |
| job\*education\*month | self-employ | illiterate | jun | -2.6186 | 10.4313 | 0.063 | 0.8018 |
| job\*education\*month | self-employ | illiterate | mar | 0 | . | . | . |
| job\*education\*month | self-employ | illiterate | may | -1.4128 | 3.8342 | 0.1358 | 0.7125 |
| job\*education\*month | self-employ | illiterate | nov | 0 | . | . | . |
| job\*education\*month | self-employ | illiterate | oct | 0 | . | . | . |
| job\*education\*month | self-employ | professional.course | apr | -0.722 | 1.8569 | 0.1512 | 0.6974 |
| job\*education\*month | self-employ | professional.course | aug | -2.6816 | 4.0566 | 0.437 | 0.5086 |
| job\*education\*month | self-employ | professional.course | dec | 0 | . | . | . |
| job\*education\*month | self-employ | professional.course | jul | -0.4641 | 1.717 | 0.073 | 0.7869 |
| job\*education\*month | self-employ | professional.course | jun | -0.7301 | 1.6857 | 0.1876 | 0.665 |
| job\*education\*month | self-employ | professional.course | mar | 5.8396 | 18.7009 | 0.0975 | 0.7548 |
| job\*education\*month | self-employ | professional.course | may | -1.2857 | 1.6851 | 0.5821 | 0.4455 |
| job\*education\*month | self-employ | professional.course | nov | -0.2577 | 1.7166 | 0.0225 | 0.8807 |
| job\*education\*month | self-employ | professional.course | oct | 0 | . | . | . |
| job\*education\*month | services | basic.4y | apr | 4.285 | 69.7595 | 0.0038 | 0.951 |
| job\*education\*month | services | basic.4y | aug | 5.5027 | 69.7501 | 0.0062 | 0.9371 |
| job\*education\*month | services | basic.4y | dec | -48.0232 | 532.4 | 0.0081 | 0.9281 |
| job\*education\*month | services | basic.4y | jul | 7.2481 | 69.7521 | 0.0108 | 0.9172 |
| job\*education\*month | services | basic.4y | jun | 2.5114 | 69.852 | 0.0013 | 0.9713 |
| job\*education\*month | services | basic.4y | mar | 1.9771 | 37.8948 | 0.0027 | 0.9584 |
| job\*education\*month | services | basic.4y | may | 5.3257 | 69.7508 | 0.0058 | 0.9391 |
| job\*education\*month | services | basic.4y | nov | 4.0484 | 69.6946 | 0.0034 | 0.9537 |
| job\*education\*month | services | basic.4y | oct | 2.1396 | 37.6573 | 0.0032 | 0.9547 |
| job\*education\*month | services | basic.6y | apr | -3.1144 | 58.6356 | 0.0028 | 0.9576 |
| job\*education\*month | services | basic.6y | aug | -1.6518 | 58.6352 | 0.0008 | 0.9775 |
| job\*education\*month | services | basic.6y | dec | 11.2152 | 410.4 | 0.0007 | 0.9782 |
| job\*education\*month | services | basic.6y | jul | -1.4248 | 58.63 | 0.0006 | 0.9806 |
| job\*education\*month | services | basic.6y | jun | -2.3451 | 58.6275 | 0.0016 | 0.9681 |
| job\*education\*month | services | basic.6y | mar | 0 | . | . | . |
| job\*education\*month | services | basic.6y | may | -1.2954 | 58.6233 | 0.0005 | 0.9824 |
| job\*education\*month | services | basic.6y | nov | -2.3507 | 58.6391 | 0.0016 | 0.968 |
| job\*education\*month | services | basic.6y | oct | 0 | . | . | . |
| job\*education\*month | services | basic.9y | apr | -2.9123 | 2.072 | 1.9756 | 0.1599 |
| job\*education\*month | services | basic.9y | aug | -0.4885 | 2.0167 | 0.0587 | 0.8086 |
| job\*education\*month | services | basic.9y | dec | 0 | . | . | . |
| job\*education\*month | services | basic.9y | jul | -1.9377 | 1.9488 | 0.9887 | 0.3201 |
| job\*education\*month | services | basic.9y | jun | -1.3486 | 1.9705 | 0.4684 | 0.4937 |
| job\*education\*month | services | basic.9y | mar | 0 | . | . | . |
| job\*education\*month | services | basic.9y | may | -1.774 | 1.914 | 0.8591 | 0.354 |
| job\*education\*month | services | basic.9y | nov | -1.7736 | 2.0829 | 0.725 | 0.3945 |
| job\*education\*month | services | basic.9y | oct | 0 | . | . | . |
| job\*education\*month | services | high.school | apr | -2.5598 | 1.153 | 4.9289 | 0.0264 |
| job\*education\*month | services | high.school | aug | -1.3417 | 1.1716 | 1.3115 | 0.2521 |
| job\*education\*month | services | high.school | dec | 11.496 | 9.2258 | 1.5527 | 0.2127 |
| job\*education\*month | services | high.school | jul | -0.9161 | 1.1385 | 0.6474 | 0.421 |
| job\*education\*month | services | high.school | jun | -1.2223 | 1.1534 | 1.1232 | 0.2892 |
| job\*education\*month | services | high.school | mar | -1.9157 | 0.8968 | 4.5632 | 0.0327 |
| job\*education\*month | services | high.school | may | -1.3504 | 1.1271 | 1.4355 | 0.2309 |
| job\*education\*month | services | high.school | nov | -1.6807 | 1.1752 | 2.0454 | 0.1527 |
| job\*education\*month | services | high.school | oct | 0 | . | . | . |
| job\*education\*month | services | illiterate | apr | 3.079 | 4.2529 | 0.5241 | 0.4691 |
| job\*education\*month | services | illiterate | aug | -5.5523 | 4.06 | 1.8702 | 0.1714 |
| job\*education\*month | services | illiterate | dec | 0 | . | . | . |
| job\*education\*month | services | illiterate | jul | -7.1575 | 3.8793 | 3.4043 | 0.065 |
| job\*education\*month | services | illiterate | jun | 1.2 | 5.5685 | 0.0464 | 0.8294 |
| job\*education\*month | services | illiterate | mar | 0 | . | . | . |
| job\*education\*month | services | illiterate | may | -2.1939 | 3.7508 | 0.3421 | 0.5586 |
| job\*education\*month | services | illiterate | nov | 0 | . | . | . |
| job\*education\*month | services | illiterate | oct | 0 | . | . | . |
| job\*education\*month | services | professional.course | apr | 3.6952 | 3.3848 | 1.1918 | 0.275 |
| job\*education\*month | services | professional.course | aug | 4.4617 | 3.3865 | 1.7358 | 0.1877 |
| job\*education\*month | services | professional.course | dec | -30.2329 | 26.4112 | 1.3103 | 0.2523 |
| job\*education\*month | services | professional.course | jul | 5.1991 | 3.285 | 2.5048 | 0.1135 |
| job\*education\*month | services | professional.course | jun | 2.9734 | 3.4061 | 0.7621 | 0.3827 |
| job\*education\*month | services | professional.course | mar | 2.508 | 4.0848 | 0.377 | 0.5392 |
| job\*education\*month | services | professional.course | may | 3.3264 | 3.2839 | 1.0261 | 0.3111 |
| job\*education\*month | services | professional.course | nov | 3.5078 | 3.3978 | 1.0658 | 0.3019 |
| job\*education\*month | services | professional.course | oct | 0 | . | . | . |
| job\*education\*month | student | basic.4y | apr | 12.5705 | 23.2589 | 0.2921 | 0.5889 |
| job\*education\*month | student | basic.4y | aug | 21.1551 | 45.1455 | 0.2196 | 0.6394 |
| job\*education\*month | student | basic.4y | dec | -102.8 | 197.5 | 0.2707 | 0.6029 |
| job\*education\*month | student | basic.4y | jul | 8.3 | 26.8943 | 0.0952 | 0.7576 |
| job\*education\*month | student | basic.4y | jun | 9.1484 | 23.6915 | 0.1491 | 0.6994 |
| job\*education\*month | student | basic.4y | mar | 5.3036 | 59.0181 | 0.0081 | 0.9284 |
| job\*education\*month | student | basic.4y | may | 11.5902 | 23.2495 | 0.2485 | 0.6181 |
| job\*education\*month | student | basic.4y | nov | 16.1832 | 21.9423 | 0.544 | 0.4608 |
| job\*education\*month | student | basic.4y | oct | 1.4247 | 1.7572 | 0.6574 | 0.4175 |
| job\*education\*month | student | basic.6y | apr | -4.2746 | 20.175 | 0.0449 | 0.8322 |
| job\*education\*month | student | basic.6y | aug | -5.0589 | 2.2878 | 4.8897 | 0.027 |
| job\*education\*month | student | basic.6y | dec | -0.7068 | 194.5 | 0 | 0.9971 |
| job\*education\*month | student | basic.6y | jul | -10.8516 | 19.4951 | 0.3098 | 0.5778 |
| job\*education\*month | student | basic.6y | jun | -5.9121 | 2.381 | 6.1653 | 0.013 |
| job\*education\*month | student | basic.6y | mar | 20.675 | 59.0529 | 0.1226 | 0.7263 |
| job\*education\*month | student | basic.6y | may | -5.2603 | 9.1606 | 0.3297 | 0.5658 |
| job\*education\*month | student | basic.6y | nov | -13.2319 | 21.27 | 0.387 | 0.5339 |
| job\*education\*month | student | basic.6y | oct | 0 | . | . | . |
| job\*education\*month | student | basic.9y | apr | 0.0945 | 1.4362 | 0.0043 | 0.9475 |
| job\*education\*month | student | basic.9y | aug | -1.1289 | 1.2197 | 0.8566 | 0.3547 |
| job\*education\*month | student | basic.9y | dec | 6.2799 | 8.6285 | 0.5297 | 0.4667 |
| job\*education\*month | student | basic.9y | jul | -1.1792 | 1.4258 | 0.684 | 0.4082 |
| job\*education\*month | student | basic.9y | jun | -1.2097 | 1.6651 | 0.5278 | 0.4675 |
| job\*education\*month | student | basic.9y | mar | 0.0701 | 0.7109 | 0.0097 | 0.9214 |
| job\*education\*month | student | basic.9y | may | -1.4936 | 1.3186 | 1.283 | 0.2573 |
| job\*education\*month | student | basic.9y | nov | -0.9544 | 1.8898 | 0.255 | 0.6135 |
| job\*education\*month | student | basic.9y | oct | -0.7091 | 1.2216 | 0.3369 | 0.5616 |
| job\*education\*month | student | high.school | apr | -0.3176 | 1.1445 | 0.077 | 0.7814 |
| job\*education\*month | student | high.school | aug | -1.8158 | 1.121 | 2.6238 | 0.1053 |
| job\*education\*month | student | high.school | dec | 4.0732 | 8.6576 | 0.2214 | 0.638 |
| job\*education\*month | student | high.school | jul | -1.2522 | 1.1676 | 1.1502 | 0.2835 |
| job\*education\*month | student | high.school | jun | -0.4786 | 1.1228 | 0.1817 | 0.6699 |
| job\*education\*month | student | high.school | mar | 0.9451 | 0.8513 | 1.2325 | 0.2669 |
| job\*education\*month | student | high.school | may | -0.6359 | 1.1033 | 0.3322 | 0.5644 |
| job\*education\*month | student | high.school | nov | -0.4693 | 1.1841 | 0.1571 | 0.6919 |
| job\*education\*month | student | high.school | oct | 0.078 | 0.5841 | 0.0178 | 0.8938 |
| job\*education\*month | student | illiterate | apr | -8.0623 | 27.1172 | 0.0884 | 0.7662 |
| job\*education\*month | student | illiterate | aug | -11.8472 | 44.8784 | 0.0697 | 0.7918 |
| job\*education\*month | student | illiterate | dec | 0 | . | . | . |
| job\*education\*month | student | illiterate | jul | 6.7464 | 29.9202 | 0.0508 | 0.8216 |
| job\*education\*month | student | illiterate | jun | 0.2045 | 23.2858 | 0.0001 | 0.993 |
| job\*education\*month | student | illiterate | mar | 0 | . | . | . |
| job\*education\*month | student | illiterate | may | -2.9277 | 21.2002 | 0.0191 | 0.8902 |
| job\*education\*month | student | illiterate | nov | 0 | . | . | . |
| job\*education\*month | student | illiterate | oct | 0 | . | . | . |
| job\*education\*month | student | professional.course | apr | -0.1435 | 2.3113 | 0.0039 | 0.9505 |
| job\*education\*month | student | professional.course | aug | -0.504 | 2.139 | 0.0555 | 0.8137 |
| job\*education\*month | student | professional.course | dec | 9.5051 | 14.7038 | 0.4179 | 0.518 |
| job\*education\*month | student | professional.course | jul | -1.7044 | 2.3088 | 0.5449 | 0.4604 |
| job\*education\*month | student | professional.course | jun | -2.1301 | 3.0792 | 0.4785 | 0.4891 |
| job\*education\*month | student | professional.course | mar | 0 | . | . | . |
| job\*education\*month | student | professional.course | may | -1.094 | 2.4155 | 0.2051 | 0.6506 |
| job\*education\*month | student | professional.course | nov | -1.2214 | 2.0934 | 0.3404 | 0.5596 |
| job\*education\*month | student | professional.course | oct | 0 | . | . | . |
| job\*education\*month | technician | basic.4y | apr | 4.7934 | 137.3 | 0.0012 | 0.9721 |
| job\*education\*month | technician | basic.4y | aug | -0.1704 | 138.2 | 0 | 0.999 |
| job\*education\*month | technician | basic.4y | dec | -15.5315 | 1334.6 | 0.0001 | 0.9907 |
| job\*education\*month | technician | basic.4y | jul | 5.7396 | 137.3 | 0.0017 | 0.9666 |
| job\*education\*month | technician | basic.4y | jun | 2.7479 | 137.3 | 0.0004 | 0.984 |
| job\*education\*month | technician | basic.4y | mar | -0.1432 | 137.4 | 0 | 0.9992 |
| job\*education\*month | technician | basic.4y | may | 1.79 | 137.3 | 0.0002 | 0.9896 |
| job\*education\*month | technician | basic.4y | nov | 4.34 | 137.3 | 0.001 | 0.9748 |
| job\*education\*month | technician | basic.4y | oct | -0.0935 | 137.2 | 0 | 0.9995 |
| job\*education\*month | technician | basic.6y | apr | 2.9223 | 15.3584 | 0.0362 | 0.8491 |
| job\*education\*month | technician | basic.6y | aug | 2.6362 | 15.3968 | 0.0293 | 0.8641 |
| job\*education\*month | technician | basic.6y | dec | -5.3995 | 503.7 | 0.0001 | 0.9914 |
| job\*education\*month | technician | basic.6y | jul | 3.6086 | 15.3367 | 0.0554 | 0.814 |
| job\*education\*month | technician | basic.6y | jun | 1.1989 | 15.7217 | 0.0058 | 0.9392 |
| job\*education\*month | technician | basic.6y | mar | 0 | . | . | . |
| job\*education\*month | technician | basic.6y | may | 2.1904 | 15.3515 | 0.0204 | 0.8865 |
| job\*education\*month | technician | basic.6y | nov | 1.6388 | 15.9491 | 0.0106 | 0.9182 |
| job\*education\*month | technician | basic.6y | oct | 0 | . | . | . |
| job\*education\*month | technician | basic.9y | apr | 0.5949 | 1.8488 | 0.1035 | 0.7476 |
| job\*education\*month | technician | basic.9y | aug | 1.507 | 1.9613 | 0.5904 | 0.4423 |
| job\*education\*month | technician | basic.9y | dec | -11.5451 | 15.5568 | 0.5508 | 0.458 |
| job\*education\*month | technician | basic.9y | jul | 1.0111 | 1.8294 | 0.3055 | 0.5805 |
| job\*education\*month | technician | basic.9y | jun | 0.7752 | 1.862 | 0.1733 | 0.6772 |
| job\*education\*month | technician | basic.9y | mar | 3.687 | 6.3736 | 0.3346 | 0.5629 |
| job\*education\*month | technician | basic.9y | may | 0.929 | 1.8009 | 0.2661 | 0.6059 |
| job\*education\*month | technician | basic.9y | nov | 0.6023 | 1.8798 | 0.1027 | 0.7487 |
| job\*education\*month | technician | basic.9y | oct | 0 | . | . | . |
| job\*education\*month | technician | high.school | apr | -2.4563 | 137.2 | 0.0003 | 0.9857 |
| job\*education\*month | technician | high.school | aug | -2.2307 | 137.2 | 0.0003 | 0.987 |
| job\*education\*month | technician | high.school | dec | 20.4455 | 1235.1 | 0.0003 | 0.9868 |
| job\*education\*month | technician | high.school | jul | -3.1602 | 137.2 | 0.0005 | 0.9816 |
| job\*education\*month | technician | high.school | jun | -2.1663 | 137.2 | 0.0002 | 0.9874 |
| job\*education\*month | technician | high.school | mar | -2.2952 | 137.2 | 0.0003 | 0.9867 |
| job\*education\*month | technician | high.school | may | -2.7425 | 137.2 | 0.0004 | 0.9841 |
| job\*education\*month | technician | high.school | nov | -2.3575 | 137.2 | 0.0003 | 0.9863 |
| job\*education\*month | technician | high.school | oct | 0.2454 | 137.2 | 0 | 0.9986 |
| job\*education\*month | technician | illiterate | apr | -4.0195 | 16.7317 | 0.0577 | 0.8102 |
| job\*education\*month | technician | illiterate | aug | 0 | . | . | . |
| job\*education\*month | technician | illiterate | dec | 0 | . | . | . |
| job\*education\*month | technician | illiterate | jul | -4.9583 | 16.638 | 0.0888 | 0.7657 |
| job\*education\*month | technician | illiterate | jun | -1.301 | 17.1066 | 0.0058 | 0.9394 |
| job\*education\*month | technician | illiterate | mar | 0 | . | . | . |
| job\*education\*month | technician | illiterate | may | -0.4273 | 16.6347 | 0.0007 | 0.9795 |
| job\*education\*month | technician | illiterate | nov | -2.165 | 17.1977 | 0.0158 | 0.8998 |
| job\*education\*month | technician | illiterate | oct | 0 | . | . | . |
| job\*education\*month | technician | professional.course | apr | -0.8188 | 1.0774 | 0.5775 | 0.4473 |
| job\*education\*month | technician | professional.course | aug | -0.6619 | 1.0685 | 0.3837 | 0.5356 |
| job\*education\*month | technician | professional.course | dec | 4.7184 | 8.5279 | 0.3061 | 0.5801 |
| job\*education\*month | technician | professional.course | jul | -0.7261 | 1.0775 | 0.4542 | 0.5004 |
| job\*education\*month | technician | professional.course | jun | -0.486 | 1.0795 | 0.2026 | 0.6526 |
| job\*education\*month | technician | professional.course | mar | 0 | . | . | . |
| job\*education\*month | technician | professional.course | may | -0.8601 | 1.0635 | 0.6541 | 0.4186 |
| job\*education\*month | technician | professional.course | nov | -0.8668 | 1.0859 | 0.6372 | 0.4247 |
| job\*education\*month | technician | professional.course | oct | 0 | . | . | . |

## Discriminant Analysis

Graphical user interface, text

Description automatically generated

Figure 24: LDA model summary

Graphical user interface, text, application

Description automatically generated

Figure 25: LDA confusion matrix

Chart, scatter chart

Description automatically generated

Figure 26: LDA ROC Curve

Graphical user interface, text, application

Description automatically generated

Figure 27: QDA confusion matrix

Chart, scatter chart

Description automatically generated

Figure 28: QDA ROC Curve

## Decision Tree

*A picture containing table

Description automatically generated*

Figure 29: Decision Tree Dendrogram

Chart

Description automatically generated

Figure 30: ROC for Single Decision Tree

Chart

Description automatically generated

Figure 31ROC for Random Forest, mtry = 5, ntrees = 100

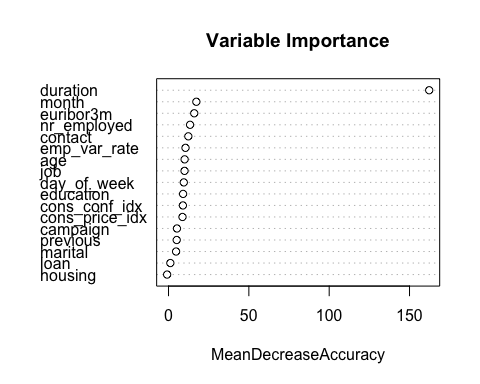


Figure 32: Variable Importance Plot for Random Forest of 100 Trees

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Model* | *Specificity* | *Sensitivity* | *Accuracy* | *AUC* |
| *Tree* | *0.791* | *0.921* | *0.856* | *0.914* |
| *Random Forest* | *0.826* | *0.934* | *0.880* | *0.935* |

Figure 33: Non-Parametric Model Comparison

## `duration` Model

##   
## Selection Summary   
## --------------------------------------------------------------------------  
## Variable AIC Sum Sq RSS R-Sq Adj. R-Sq   
## --------------------------------------------------------------------------  
## campaign 109319.983 1351.852 34270.843 0.03795 0.03793   
## month 109053.291 1587.916 34034.778 0.04458 0.04434   
## day\_of\_week 108963.736 1668.433 33954.262 0.04684 0.04651   
## contact 108896.063 1725.821 33896.874 0.04845 0.04810   
## previous 108876.933 1743.206 33879.489 0.04894 0.04857   
## housing 108866.533 1753.404 33869.291 0.04922 0.04883   
## age 108859.396 1760.917 33861.778 0.04943 0.04902   
## loan 108857.834 1763.845 33858.850 0.04951 0.04908   
## job 108857.174 1780.825 33841.870 0.04999 0.04932   
## --------------------------------------------------------------------------

##   
## Call:  
## lm(formula = log(duration) ~ education + campaign + month + day\_of\_week +   
## housing + previous + job + loan + contact + age, data = bank\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.5020 -0.5431 0.0054 0.5810 3.8447   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.3508519 0.0420524 127.242 < 2e-16 \*\*\*  
## educationbasic.6y -0.0257410 0.0278649 -0.924 0.355607   
## educationbasic.9y -0.0316826 0.0218508 -1.450 0.147082   
## educationhigh.school -0.0146059 0.0224421 -0.651 0.515165   
## educationilliterate -0.1258995 0.2628717 -0.479 0.631985   
## educationprofessional.course -0.0373782 0.0252974 -1.478 0.139538   
## educationuniversity.degree -0.0098018 0.0220095 -0.445 0.656075   
## campaign -0.0628795 0.0019014 -33.070 < 2e-16 \*\*\*  
## monthaug -0.1886039 0.0248075 -7.603 2.98e-14 \*\*\*  
## monthdec 0.2205021 0.0789634 2.792 0.005234 \*\*   
## monthjul -0.0215968 0.0243183 -0.888 0.374501   
## monthjun -0.1187685 0.0272027 -4.366 1.27e-05 \*\*\*  
## monthmar -0.0382149 0.0496438 -0.770 0.441435   
## monthmay -0.0370079 0.0235713 -1.570 0.116416   
## monthnov -0.2083387 0.0263909 -7.894 3.01e-15 \*\*\*  
## monthoct -0.0660363 0.0452403 -1.460 0.144389   
## monthsep -0.0359120 0.0485747 -0.739 0.459721   
## day\_of\_weekmon 0.0260554 0.0164378 1.585 0.112956   
## day\_of\_weekthu 0.0572566 0.0164192 3.487 0.000489 \*\*\*  
## day\_of\_weektue 0.0929925 0.0166859 5.573 2.52e-08 \*\*\*  
## day\_of\_weekwed 0.1181446 0.0166711 7.087 1.40e-12 \*\*\*  
## housingyes -0.0293846 0.0104315 -2.817 0.004852 \*\*   
## previous 0.0419413 0.0111593 3.758 0.000171 \*\*\*  
## jobblue-collar 0.0213907 0.0185718 1.152 0.249418   
## jobentrepreneur 0.0193468 0.0297560 0.650 0.515581   
## jobhousemaid -0.0019219 0.0348875 -0.055 0.956069   
## jobmanagement 0.0045535 0.0224813 0.203 0.839491   
## jobretired 0.0696342 0.0312935 2.225 0.026075 \*   
## jobself-employed 0.0311845 0.0298703 1.044 0.296496   
## jobservices 0.0097098 0.0205817 0.472 0.637095   
## jobstudent 0.0855856 0.0384006 2.229 0.025837 \*   
## jobtechnician 0.0085579 0.0182799 0.468 0.639674   
## jobunemployed -0.0073504 0.0346073 -0.212 0.831800   
## loanyes -0.0229029 0.0144699 -1.583 0.113478   
## contacttelephone -0.0956898 0.0139985 -6.836 8.31e-12 \*\*\*  
## age 0.0011278 0.0005856 1.926 0.054115 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.9083 on 30855 degrees of freedom  
## Multiple R-squared: 0.04927, Adjusted R-squared: 0.04819   
## F-statistic: 45.69 on 35 and 30855 DF, p-value: < 2.2e-16

Figure 34: Model and Selection Summary for `duration`

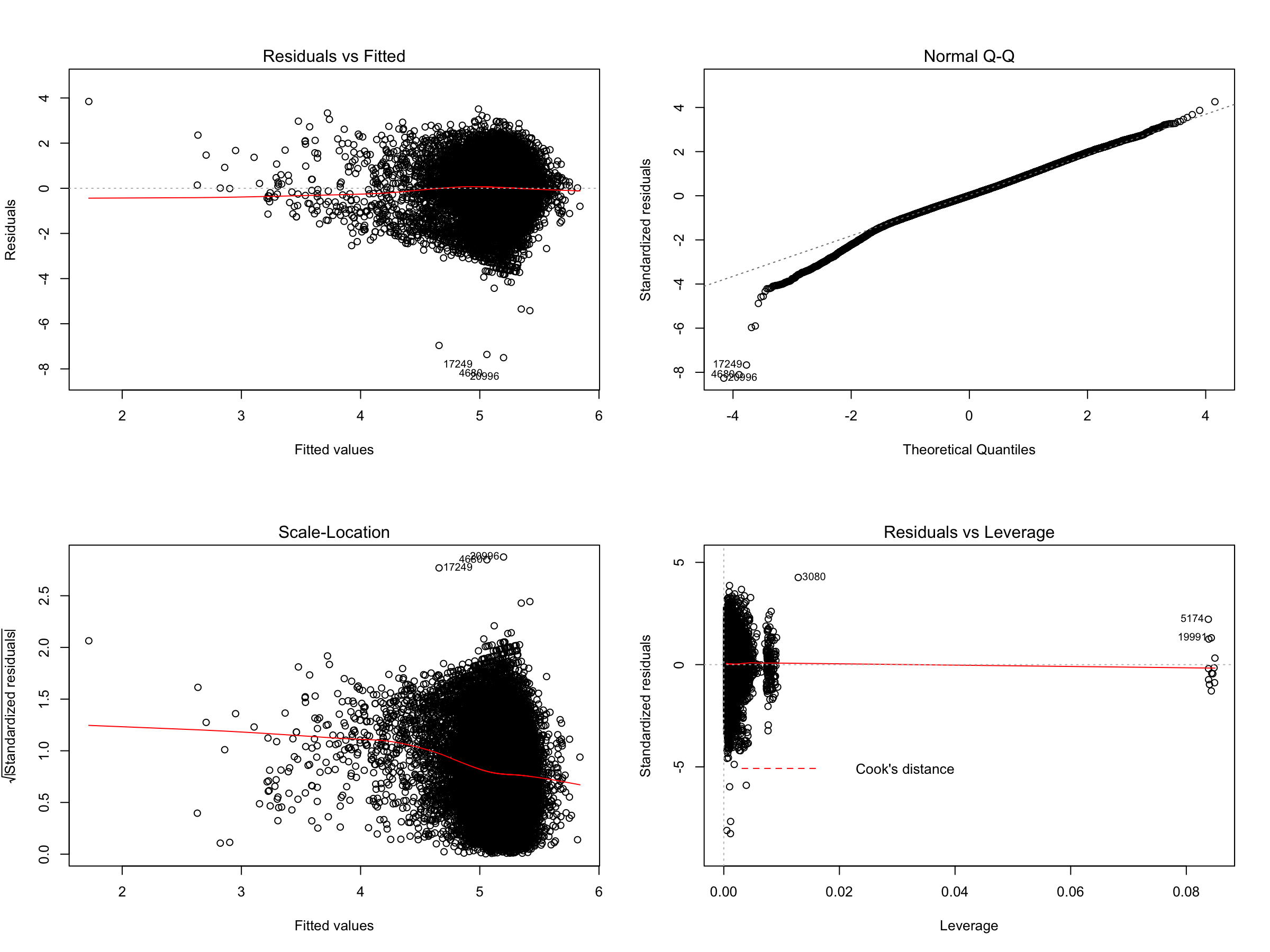
**

Figure 35: Residual Diagnostics for `duration` Model

**Chart, line chart

Description automatically generatedChart, scatter chart

Description automatically generated**

## SAS Code

### Exploratory Data Analysis

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* STUDENT: Edward Roske \*

\* DATE: November 8, 2020 \*

\* CLASS: MSDS6372 - Applied Statistics \*

\* PROJECT: 2 - Logistic Regression \*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* Header Information;

LIBNAME STATS "/home/u43010517/sasuser.v94/STATS/DS6372";

**RUN**;

TITLE "Logistic Regression";

\* Explore data;

**PROC** **FREQ** DATA=STATS.BANK;

\* default\*subscribed;

TABLES

job\*subscribed

marital\*subscribed

education\*subscribed

housing\*subscribed

loan\*subscribed

contact\*subscribed

month\*subscribed

day\_of\_week\*subscribed

poutcome\*subscribed

/ chisq relrisk;

**RUN**;**QUIT**;

**PROC** **MEANS** DATA=STATS.BANK N NMISS MEAN MEDIAN MIN MAX STD;

\* default;

CLASS subscribed

job

marital

education

housing

loan

contact

month

day\_of\_week

poutcome

;

\* default\*subscribed;

TYPES subscribed

job\*subscribed

marital\*subscribed

education\*subscribed

housing\*subscribed

loan\*subscribed

contact\*subscribed

month\*subscribed

day\_of\_week\*subscribed

poutcome\*subscribed

;

VAR age duration campaign pdays previous emp\_var\_rate cons\_price\_idx

cons\_conf\_idx euribor3m nr\_employed ;

**RUN**;

**PROC** **MEANS** DATA=STATS.BANK N NMISS MEAN MEDIAN MIN MAX STD;

CLASS subscribed

month

;

TYPES subscribed

month\*subscribed

;

VAR Duration Campaign Emp\_var\_rate Euribor3m Nr\_employed Previous ;

**RUN**;

/\*

PROC MEANS DATA=STATS.BANK N NMISS MEAN MEDIAN MIN MAX STD;

CLASS subscribed

month

;

TYPES subscribed

month\*subscribed

;

VAR nr\_employed ;

RUN;

\*/

### Exploratory Data Analysis

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* STUDENT: Edward Roske \*

\* DATE: November 8, 2020 \*

\* CLASS: MSDS6372 - Applied Statistics \*

\* PROJECT: 2 - Logistic Regression \*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* Header Information;

LIBNAME STATS "/home/u43010517/sasuser.v94/STATS/DS6372";

**RUN**;

TITLE "Logistic Regression";

\* Explore data;

**PROC** **FREQ** DATA=STATS.BANK;

\* default\*subscribed;

TABLES

job\*subscribed

marital\*subscribed

education\*subscribed

housing\*subscribed

loan\*subscribed

contact\*subscribed

month\*subscribed

day\_of\_week\*subscribed

poutcome\*subscribed

/ chisq relrisk;

**RUN**;**QUIT**;

**PROC** **MEANS** DATA=STATS.BANK N NMISS MEAN MEDIAN MIN MAX STD;

\* default;

CLASS subscribed

job

marital

education

housing

loan

contact

month

day\_of\_week

poutcome

;

\* default\*subscribed;

TYPES subscribed

job\*subscribed

marital\*subscribed

education\*subscribed

housing\*subscribed

loan\*subscribed

contact\*subscribed

month\*subscribed

day\_of\_week\*subscribed

poutcome\*subscribed

;

VAR age duration campaign pdays previous emp\_var\_rate cons\_price\_idx

cons\_conf\_idx euribor3m nr\_employed ;

**RUN**;

**PROC** **MEANS** DATA=STATS.BANK N NMISS MEAN MEDIAN MIN MAX STD;

CLASS subscribed

month

;

TYPES subscribed

month\*subscribed

;

VAR Duration Campaign Emp\_var\_rate Euribor3m Nr\_employed Previous ;

**RUN**;

/\*

PROC MEANS DATA=STATS.BANK N NMISS MEAN MEDIAN MIN MAX STD;

CLASS subscribed

month

;

TYPES subscribed

month\*subscribed

;

VAR nr\_employed ;

RUN;

\*/

### Model 1

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* STUDENT: Edward Roske \*

\* DATE: November 8, 2020 \*

\* CLASS: MSDS6372 - Applied Statistics \*

\* PROJECT: 2 - Logistic Regression \*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* Header Information;

LIBNAME STATS "/home/u43010517/sasuser.v94/STATS/DS6372";

**RUN**;

TITLE "Logistic Regression";

\* Logistic Regression;

\* - All Continuous Variables;

**PROC** **LOGISTIC** data=STATS.bank ;

TITLE2 "All Continuous Variables";

MODEL subscribed(event='ye') =

age duration campaign pdays previous emp\_var\_rate cons\_price\_idx

cons\_conf\_idx euribor3m nr\_employed

/ SCALE=none aggregate lackfit; \* Can add "influence" but it does too many regression diagnostics;

**RUN**;

\* Logistic Regression;

\* - All Categorical Variables;

**PROC** **LOGISTIC** data=STATS.bank ;

TITLE2 "All Categorical Variables";

CLASS subscribed

job

marital

education

housing

loan

contact

month

day\_of\_week

poutcome

;

MODEL subscribed(event='ye') =

job marital education housing loan contact month day\_of\_week poutcome

/ SCALE=none aggregate lackfit; \* Can add "influence" but it does too many regression diagnostics;

**RUN**;

\* Logistic Regression;

\* - All Variables;

**PROC** **LOGISTIC** data=STATS.bank ;

TITLE2 "All Continuous & Categorical Variables";

CLASS subscribed

job

marital

education

housing

loan

contact

month

day\_of\_week

poutcome

;

MODEL subscribed(event='ye') =

job marital education housing loan contact month day\_of\_week poutcome

age duration campaign pdays previous emp\_var\_rate cons\_price\_idx

cons\_conf\_idx euribor3m nr\_employed

/ SCALE=none aggregate lackfit; \* Can add "influence" but it does too many regression diagnostics;

**RUN**;

\* Logistic Regression;

\* - Statistically Relevant Variables (based on alpha<0.5 p-values from step above);

**PROC** **LOGISTIC** data=STATS.bank ;

TITLE2 "Statistically Relevant Variables";

CLASS subscribed

job

contact

month

day\_of\_week

poutcome

;

MODEL subscribed(event='ye') =

job contact month day\_of\_week poutcome

duration campaign previous emp\_var\_rate cons\_price\_idx cons\_conf\_idx euribor3m

/ SCALE=none aggregate lackfit; \* Can add "influence" but it does too many regression diagnostics;

**RUN**;

\* Logistic Regression with Effect Plots;

\* - Statistically Relevant Variables;

**PROC** **LOGISTIC** data=STATS.bank ;

TITLE2 "Statistically Relevant Variables";

CLASS subscribed job marital contact month day\_of\_week poutcome ;

MODEL subscribed(event='ye') =

job marital contact month day\_of\_week poutcome

duration previous emp\_var\_rate cons\_price\_idx cons\_conf\_idx euribor3m

/ SCALE=none;

EFFECTPLOT slicefit(sliceby=job plotby=duration) / noobs;

EFFECTPLOT slicefit(sliceby=marital plotby=duration) / noobs;

EFFECTPLOT slicefit(sliceby=contact plotby=duration) / noobs;

EFFECTPLOT slicefit(sliceby=month plotby=duration) / noobs;

EFFECTPLOT slicefit(sliceby=day\_of\_week plotby=duration) / noobs;

EFFECTPLOT slicefit(sliceby=poutcome plotby=duration) / noobs;

EFFECTPLOT slicefit(sliceby=previous plotby=duration) / noobs;

EFFECTPLOT slicefit(sliceby=cons\_price\_idx plotby=duration) / noobs;

EFFECTPLOT slicefit(sliceby=cons\_conf\_idx plotby=duration) / noobs;

EFFECTPLOT slicefit(sliceby=euribor3m plotby=duration) / noobs;

\* And a few interesting plots;

EFFECTPLOT slicefit(sliceby=month plotby=cons\_price\_idx) / noobs;

EFFECTPLOT slicefit(sliceby=duration plotby=emp\_var\_rate) / noobs;

**RUN**;

\* Logistic Regression;

\* - All Variables

\* - Feature Selection: Forward;

**PROC** **LOGISTIC** data=STATS.bank ;

TITLE2 "Feature Selection: Forward";

CLASS subscribed

job marital education housing loan contact month day\_of\_week poutcome;

MODEL subscribed(event='ye') =

job marital education housing loan contact month day\_of\_week poutcome

age duration campaign pdays previous emp\_var\_rate cons\_price\_idx

cons\_conf\_idx euribor3m nr\_employed

/ SELECTION=FORWARD start=**3**

SCALE=none aggregate lackfit;

**RUN**;

\* Logistic Regression;

\* - All Variables

\* - Feature Selection: Backward;

**PROC** **LOGISTIC** data=STATS.bank ;

TITLE2 "Feature Selection: Backward";

CLASS subscribed

job marital education housing loan contact month day\_of\_week poutcome;

MODEL subscribed(event='ye') =

job marital education housing loan contact month day\_of\_week poutcome

age duration campaign pdays previous emp\_var\_rate cons\_price\_idx

cons\_conf\_idx euribor3m nr\_employed

/ SELECTION=BACKWARD start=**3**

SCALE=none aggregate lackfit;

**RUN**;

\* Logistic Regression;

\* - All Variables

\* - Feature Selection: Stepwise;

**PROC** **LOGISTIC** data=STATS.bank ;

TITLE2 "Feature Selection: Stepwise";

CLASS subscribed

job marital education housing loan contact month day\_of\_week poutcome;

MODEL subscribed(event='ye') =

job marital education housing loan contact month day\_of\_week poutcome

age duration campaign pdays previous emp\_var\_rate cons\_price\_idx

cons\_conf\_idx euribor3m nr\_employed

/ SELECTION=STEPWISE start=**3**

SCALE=none aggregate lackfit;

**RUN**;

\* Logistic Regression;

\* - Statistically Relevant Variables (based on results of STEPWISE);

**PROC** **LOGISTIC** data=STATS.bank ;

TITLE2 "Statistically Relevant Variables";

CLASS subscribed

job education contact month day\_of\_week poutcome ;

MODEL subscribed(event='ye') =

job education contact month day\_of\_week poutcome

duration campaign previous emp\_var\_rate cons\_price\_idx euribor3m

/ SCALE=none aggregate lackfit;

**RUN**;

\* Logistic Regression;

\* - Statistically Relevant Variables (based on results of STEPWISE)

\* - ROC Curves;

**PROC** **LOGISTIC** data=STATS.bank ;

TITLE2 "Statistically Relevant Variables";

CLASS subscribed

job education contact month day\_of\_week poutcome ;

MODEL subscribed(event='ye') =

job education contact month day\_of\_week poutcome

duration campaign previous emp\_var\_rate cons\_price\_idx euribor3m

/ SCALE=none aggregate lackfit;

ROC 'Just Duration'

duration

;

ROC 'Just Education'

education

;

ROC 'Random Chance'

;

ROCCONTRAST / estimate e;

**RUN**;

\* Logistic Regression;

\* - Statistically Relevant Variables (based on results of STEPWISE)

\* - ROC Curve;

**PROC** **LOGISTIC** data=STATS.bank plots(only)=roc;

TITLE2 "Model 1: ROC Curve";

CLASS subscribed

job education contact month day\_of\_week poutcome ;

LogisticModel: MODEL subscribed(event='ye') =

job education contact month day\_of\_week poutcome

duration campaign previous emp\_var\_rate cons\_price\_idx euribor3m;

\*/ SCALE=none aggregate lackfit;

OUTPUT out=LogiOut predicted=LogiPred; /\* output predicted value, to be used later if we want to see the predictions \*/

**RUN**;

### Model 2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* STUDENT: Edward Roske \*

\* DATE: November 23, 2020 \*

\* CLASS: MSDS6372 - Applied Statistics \*

\* PROJECT: 2 - Logistic Regression \*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* Header Information;

LIBNAME STATS "/home/u43010517/sasuser.v94/STATS/DS6372";

**RUN**;

TITLE "Model 2 - More Complicated Model";

\* Add transformations;

\* - Begin with previously relevant variables;

**DATA** STATS.BANKnew

(DROP=age marital housing loan pdays cons\_conf\_idx nr\_employed);

SET STATS.BANK;

\* Logarithms;

if duration gt **0** then durationLog=log(duration);

if campaign gt **0** then campaignLog=log(campaign);

if previous gt **0** then previousLog=log(previous);

if emp\_var\_rate gt **0** then emp\_var\_rateLog=log(emp\_var\_rate);

if cons\_price\_idx gt **0** then cons\_price\_idxLog=log(cons\_price\_idx);

if euribor3m gt **0** then euribor3mLog=log(euribor3m);

\* Quadratics;

duration2=duration\*\***2**;

campaign2=campaign\*\***2**;

previous2=previous\*\***2**;

emp\_var\_rate2=emp\_var\_rate\*\***2**;

cons\_price\_idx2=cons\_price\_idx\*\***2**;

euribor3m2=euribor3m\*\***2**;

\*\*\* Will add interactions later in model code;

**RUN**;

\* Logistic Regression;

\* - All Variables including new transformations;

\* - Had to leave out previousLog and emp\_var\_rateLog because of too many NAs;

**PROC** **LOGISTIC** data=STATS.BANKnew ;

TITLE2 "All Continuous & Categorical Variables";

CLASS subscribed

job education contact month day\_of\_week poutcome;

MODEL subscribed(event='ye') =

job education contact month day\_of\_week poutcome

duration campaign previous emp\_var\_rate cons\_price\_idx euribor3m

duration2 campaign2 previous2 emp\_var\_rate2 cons\_price\_idx2 euribor3m2

durationLog campaignLog cons\_price\_idxLog euribor3mLog

/ SCALE=none aggregate lackfit; \* Can add "influence" but it does too many regression diagnostics;

**RUN**;

\* All variables are still statistically impactful to the model;

\* - Campaign is less relevant but still alpha < 0.05;

\* Logistic Regression;

\* - Add interactions;

\* - All Variables including new transformations;

\* - Had to leave out previousLog and emp\_var\_rateLog because of too many NAs;

**PROC** **LOGISTIC** data=STATS.BANKnew plots(only)=roc;

TITLE2 "All Continuous & Categorical Variables with Interactions";

CLASS subscribed

job education contact month day\_of\_week poutcome

;

MODEL subscribed(event='ye') =

job education contact month day\_of\_week month day\_of\_week poutcome

job\*education job\*education\*month month\*day\_of\_week month\*poutcome

campaign | previous | emp\_var\_rate | cons\_price\_idx | euribor3m

duration campaign previous emp\_var\_rate cons\_price\_idx euribor3m

duration2 campaign2 previous2 emp\_var\_rate2 cons\_price\_idx2 euribor3m2

durationLog campaignLog cons\_price\_idxLog euribor3mLog

/ SCALE=none aggregate lackfit; \* Can add "influence" but it does too many regression diagnostics;

ROC 'Random Chance'

;

ROCCONTRAST / estimate e;

**RUN**;

\* Logistic Regression;

\* - All Variables p-value > 0.05 (from last step)

\* - Feature Selection: Stepwise;

**PROC** **LOGISTIC** data=STATS.BANKnew ;

TITLE2 "Variables: Stepwise Selection";

CLASS subscribed

job education contact month day\_of\_week poutcome

;

StepwiseModel: MODEL subscribed(event='ye') =

education contact month day\_of\_week month day\_of\_week poutcome

job\*education job\*education\*month month\*day\_of\_week month\*poutcome

previous\*cons\_price\_idx emp\_var\_rate\*euribor3m cons\_price\_idx\*euribor3m

emp\_var\_rate\*cons\_price\_idx\*euribor3m

duration previous cons\_price\_idx euribor3m

duration2 campaign2 previous2 emp\_var\_rate2 cons\_price\_idx2 euribor3m2

durationLog cons\_price\_idxLog euribor3mLog

/ SELECTION=STEPWISE start=**3**

SCALE=none aggregate lackfit;

ROC 'All variables'

education contact month day\_of\_week month day\_of\_week poutcome

job\*education job\*education\*month month\*day\_of\_week month\*poutcome

previous\*cons\_price\_idx emp\_var\_rate\*euribor3m cons\_price\_idx\*euribor3m

emp\_var\_rate\*cons\_price\_idx\*euribor3m

duration previous cons\_price\_idx euribor3m

duration2 campaign2 previous2 emp\_var\_rate2 cons\_price\_idx2 euribor3m2

durationLog cons\_price\_idxLog euribor3mLog

;

ROCCONTRAST / estimate e;

**RUN**;

\* ROC for stepwise variables: 0.9349

\* ROC for stepwise if I leave in job\*education\*month: 0.9401

\* ROC for all statistically relevant variables: 0.9471

\* Logistic Regression;

\* - Statistically Relevant Variables (based on results of STEPWISE);

**PROC** **LOGISTIC** data=STATS.BANKnew plots(only)=roc;

TITLE2 "Final Stepwise Model";

CLASS subscribed

job education month poutcome ;

FullModel: MODEL subscribed(event='ye') =

education month poutcome job\*education duration duration2 durationLog euribor3mLog

job\*education\*month

/ SCALE=none aggregate lackfit;

ROC 'Stepwise (no job\*education\*month)'

education month poutcome job\*education duration duration2 durationLog euribor3mLog

job\*education\*month

;

ROC 'Only job\*education\*month'

job\*education\*month

;

ROC 'Random Chance'

;

ROCCONTRAST / estimate e;

**RUN**;

\* Final variables in the more complex model:

education month poutcome duration

duration2 durationLog euribor3mLog

job\*education

job\*education\*month

## R Code

### Data Loading, Cleanup, Train/Test Splits, LDA/QDA

knitr::opts\_chunk$set(echo = TRUE)

knitr::opts\_chunk$set(**warning** = FALSE)

knitr::opts\_chunk$set(message = FALSE)

**library**(tidyverse) *# general data processing & plotting*

**library**(here) *# relative location references*

**library**(janitor) *# data cleanup tools*

**library**(naniar) *# dealing with missing values*

**library**(caret) *# misc functions for training and plotting classification and regression models*

**library**(tidymodels)

**library**(GGally) *# for ggpairs*

**library**(MASS) *# for LDA/QDA*

*# Read data*

bank <- read\_delim(here("data - raw", "bank-additional", "bank-additional-full.csv"),

";", escape\_double = FALSE, trim\_ws = TRUE)

*# Clean column names*

bank <- clean\_names(bank)

*# Quick examination of the data*

glimpse(bank)

head(bank)

*# Begin by replacing missing value placeholders with NA*

bank\_clean <- bank %>%

replace\_with\_na\_all(condition = ~ .x == "unknown") %>%

replace\_with\_na(replace = list(pdays = 999)) %>%

replace\_with\_na(replace = list(poutcome = "nonexistent"))

*# Replace each missing value in a categorical variable with the mode as*

*# determined during SAS EDA*

bank\_clean <- bank\_clean %>%

mutate(job = replace\_na(job, "admin."),

marital = replace\_na(marital, "married"),

housing = replace\_na(housing, "yes"),

loan = replace\_na(loan, "no"),

contact = replace\_na(contact, "cellular"),

education = replace\_na(education, "university.degree"),

month = replace\_na(month, "may"),

day\_of\_week = replace\_na(day\_of\_week, "thu"),

poutcome = replace\_na(poutcome, "failure"))

*# Drop `default` column since there are only 3 "Yes" values*

bank\_clean <- bank\_clean %>%

dplyr::select(-default)

*# Only Pdays is missing, and we should delete it, but for completeness, replace*

*# with median (6)*

bank\_clean <- bank\_clean %>%

mutate(pdays = replace\_na(pdays, 6))

*# Rename response variable*

bank\_clean <- bank\_clean %>%

rename(subscribed = y)

*# Convert all character variables to factors*

bank\_clean <- bank\_clean %>%

mutate(across(where(is\_character),as\_factor))

*# Re-run reports*

DataExplorer::create\_report(bank\_clean)

Hmisc::describe(bank\_clean)

psych::describe(bank\_clean)

skimr::skim(bank\_clean)

*# Reports from inspectdf (categorical variables and Pearson correlation*

*# coefficients)*

bank\_cat <- inspectdf::inspect\_cat(bank\_clean)

bank\_pear <- inspectdf::inspect\_cor(bank\_clean)

*# Set seed*

set.seed(123)

*# Save full bank\_clean data set*

*# write\_csv(bank\_clean, here("data - output", "bank\_clean.csv"))*

*# Create test/train data sets of full data*

inTraining <- createDataPartition(bank\_clean$subscribed, p = .75, list = FALSE)

training <- bank\_clean[ inTraining,]

testing <- bank\_clean[-inTraining,]

*# Save bank\_clean test/train sets*

*# write\_csv(training, here("data - output", "bank\_clean\_train.csv"))*

*# write\_csv(testing, here("data - output", "bank\_clean\_test.csv"))*

*# Down-sample the data to get an equal number of yes and no responses*

bank\_clean\_ds <- downSample(x = bank\_clean[, 1:19],

y = bank\_clean$subscribed)

*# Rename `Class` to `subscribed`*

bank\_clean\_ds <- bank\_clean\_ds %>%

rename(subscribed = Class)

*# Double-check the number of yes/no responses*

bank\_clean\_ds %>%

count(subscribed)

*# Save downsampled data set*

*# write\_csv(bank\_clean\_ds, here("data - output", "bank\_clean\_downsampled.csv"))*

*# Create test/train data sets of downsampled data*

inTraining <- createDataPartition(bank\_clean\_ds$subscribed, p = .75, list = FALSE)

training <- bank\_clean\_ds[ inTraining,]

testing <- bank\_clean\_ds[-inTraining,]

*# Save downsampled data test/train sets*

*# write\_csv(training, here("data - output", "bank\_clean\_ds\_train.csv"))*

*# write\_csv(testing, here("data - output", "bank\_clean\_ds\_test.csv"))*

set.seed(123)

*# Create LDA data set*

lda\_data <- bank\_clean\_ds %>%

dplyr::select(c(1, 10:13, 15:17, 19:20)) *# class + numeric columns + drop euribor3m*

*# Check for collinearity*

view(inspectdf::inspect\_cor(lda\_data))

*# Create test/train data sets of downsampled data*

inTraining <- createDataPartition(lda\_data$subscribed, p = .75, list = FALSE)

training <- lda\_data[ inTraining,]

testing <- lda\_data[-inTraining,]

*# Estimate preprocessing parameters*

preproc\_parameter <- training %>%

preProcess(method = c("center", "scale"))

*# Transform the data using the estimated parameters*

train\_transform <- preproc\_parameter %>% predict(training)

test\_transform <- preproc\_parameter %>% predict(testing)

*# Fit the model*

lda\_model <- lda(subscribed ~ ., data = train\_transform)

*# Look at the model*

lda\_model

*# Make predictions*

lda\_predictions <- lda\_model %>% predict(test\_transform)

*# Model accuracy*

mean(lda\_predictions$class == test\_transform$subscribed)

*# Generate confusion matrix*

confusionMatrix(predict(lda\_model, newdata = test\_transform)$class,

test\_transform$subscribed)

*# Get the posteriors as a tibble*

lda\_posteriors <- tibble(predict(lda\_model, newdata = test\_transform)$posterior)

*# Create prediction and performance values*

**library**(ROCR)

lda\_predict <- prediction(lda\_predictions$posterior[,2], test\_transform$subscribed)

lda\_perf <- performance(lda\_predict, "tpr", "fpr")

*# Calculate AUC*

lda\_auc <- performance(lda\_predict, measure = "auc")

lda\_auc <- lda\_auc@y.values[[1]]

lda\_auc

*# Create and plot ROC curve for LDA model*

**library**(ggfortify)

autoplot(lda\_perf) +

labs(title = "LDA Model",

subtitle = "True Positive Rate vs. False Positive Rate",

x = "False Positive Rate",

y = "True Positive Rate") +

annotate(

geom = "text", x = 0.50, y = 0.50,

label = paste("AUC =", round(lda\_auc, 4)), size = 6

) +

theme\_minimal()

*# QDA*

qda\_model <- qda(subscribed ~ ., data = train\_transform)

*# Make predictions*

qda\_predictions <- qda\_model %>% predict(test\_transform)

*# Model accuracy*

mean(qda\_predictions$class == test\_transform$subscribed)

*# Generate confusion matrix*

confusionMatrix(predict(qda\_model, newdata = test\_transform)$class,

test\_transform$subscribed)

*# Get the posteriors as a tibble*

qda\_posteriors <- tibble(predict(qda\_model, newdata = test\_transform)$posterior)

*# Create prediction and performance values*

qda\_predict <- prediction(qda\_predictions$posterior[,2], test\_transform$subscribed)

qda\_perf <- performance(qda\_predict, "tpr", "fpr")

*# Calculate AUC*

qda\_auc <- performance(qda\_predict, measure = "auc")

qda\_auc <- qda\_auc@y.values[[1]]

qda\_auc

*# Create and plot ROC curve for LDA model*

autoplot(qda\_perf) +

labs(title = "QDA Model",

subtitle = "True Positive Rate vs. False Positive Rate",

x = "False Positive Rate",

y = "True Positive Rate") +

annotate(

geom = "text", x = 0.50, y = 0.50,

label = paste("AUC =", round(qda\_auc, 4)), size = 6

) +

theme\_minimal()

### Decision Trees

*# Read in Data (locally)*

bank\_clean <- read.csv("/Users/molheiser/Downloads/bank\_clean.csv")

bank\_ds <- read.csv("/Users/molheiser/Downloads/bank\_clean\_downsampled.csv")

bank\_train <- read.csv("/Users/molheiser/Downloads/bank\_clean\_train.csv")

bank\_test <- read.csv("/Users/molheiser/Downloads/bank\_clean\_test.csv")

bank\_train\_ds <- read.csv("/Users/molheiser/Downloads/bank\_clean\_ds\_train.csv")

bank\_test\_ds <- read.csv("/Users/molheiser/Downloads/bank\_clean\_ds\_test.csv")

*### Decision Tree model*

**library**(tree)

mytree <- tree(subscribed ~ job + marital + day\_of\_week + housing + duration + education + contact + month + campaign + emp\_var\_rate + euribor3m + nr\_employed +

cons\_price\_idx + cons\_conf\_idx, bank\_ds)

summary(mytree)

plot(mytree, main = "Decision Tree For Subscribed")

text(mytree, pretty = 0)

*# CV for Pruning tree*

set.seed(123)

cv.tree <- cv.tree(mytree, FUN = prune.tree, method = "deviance")

names(cv.tree)

cv.tree

plot(cv.tree)

par(mfrow = c(1, 1))

plot(cv.tree$size, cv.tree$dev, type = "b")

*# The result is essnetially no pruning.*

pruned.tree <- prune.tree(mytree, best = 9)

plot(pruned.tree)

text(pruned.tree, pretty = 0)

*## test train split for tree model above*

train.tree <- tree(subscribed ~ job + marital + day\_of\_week + housing + education + contact + month + duration + campaign + emp\_var\_rate + euribor3m + nr\_employed

+ cons\_price\_idx + cons\_conf\_idx, bank\_train\_ds)

pruned.train.tree <- prune.tree(train.tree, best = 9)

summary(pruned.train.tree)

plot(pruned.train.tree)

text(pruned.train.tree, pretty = 0, cex = 1)

*# Make predictions on testing set, calculate perforance metrics*

tree.pred <- predict(pruned.train.tree, bank\_test\_ds, type = "class")

cm.tree.pruned <- table(tree.pred, bank\_test\_ds$subscribed)

*# sens (recall) = TP / TP + FN*

*# Spec = TN / TN + FP*

*# accuracy = TP + TN / total*

sensitivity\_tree1 <- cm.tree.pruned[2, 2] / (cm.tree.pruned[2, 2] + cm.tree.pruned[1, 2])

specificity\_tree1 <- cm.tree.pruned[1, 1] / (cm.tree.pruned[1, 1] + cm.tree.pruned[2, 1])

accuracy\_tree1 <- (cm.tree.pruned[2, 2] + cm.tree.pruned[1, 1]) / length(bank\_test\_ds$subscribed)

*# ROC curve for single Decision tree*

**library**(ROCR)

tree.pred <- predict(pruned.train.tree, bank\_test\_ds, type = "vector")

pred <- prediction(tree.pred[, 2], bank\_test\_ds$subscribed)

roc.perf.tree <- performance(pred, measure = "tpr", x.measure = "fpr")

auc.tree <- performance(pred, measure = "auc")

auc.tree <- auc.tree@y.values

plot(roc.perf.tree, main = "ROC of Single Tree for Subscribed", xlab = "1-Specificity", ylab = "Sensitivity")

abline(a = 0, b = 1)

text(x = .40, y = .6, paste("AUC = ", round(auc.tree[[1]], 3), sep = ""))

*# Random Forest*

**library**(randomForest)

*# model removing poutcome and pdays*

rf.bank <- randomForest(subscribed ~ . - subscribed - pdays - poutcome, bank\_train\_ds, mtry = 5, importance = T, ntree = 100)

*# Making predictions on test and then observing performnace metrics*

fit.pred <- predict(rf.bank, newdata = bank\_test\_ds, type = "response")

cm.rf <- table(fit.pred, bank\_test\_ds$subscribed)

sensitivity\_rf1 <- cm.rf[2, 2] / (cm.rf[2, 2] + cm.rf[1, 2])

specificity\_rf1 <- cm.rf[1, 1] / (cm.rf[1, 1] + cm.rf[2, 1])

accuracy\_rf1 <- (cm.rf[2, 2] + cm.rf[1, 1]) / length(bank\_test\_ds$subscribed)

*# Plot ROC*

rf.pred <- predict(rf.bank, newdata = bank\_test\_ds, type = "prob")

pred <- prediction(rf.pred[, 2], bank\_test\_ds$subscribed)

roc.perf.rf <- performance(pred, measure = "tpr", x.measure = "fpr")

auc.rf <- performance(pred, measure = "auc")

auc.rf <- auc.rf@y.values

plot(roc.perf.rf, main = "ROC of Random Forest of for Subscribed - mtry=5", xlab = "1-Specificity", ylab = "Sensitivity")

abline(a = 0, b = 1)

text(x = .40, y = .6, paste("AUC = ", round(auc.rf[[1]], 3), sep = ""))

*# Which variables are important. We can use variable importance.*

*# 1=mean decrease in accuracy, 2=mean decrease in node impurity*

varImpPlot(rf.bank, type = 1, main = "Variable Importance")

varImpPlot(rf.bank, type = 2, main = "Variable Importance")

*# compare models*

model\_results <- data.frame(

Model = c("Tree", "Random Forest"),

Specificity = c(round(specificity\_tree1, 3), round(specificity\_rf1, 3)),

Sensitivity = c(round(sensitivity\_tree1, 3), round(sensitivity\_rf1, 3)),

Accuracy = c(round(accuracy\_tree1, 3), round(accuracy\_rf1, 3)),

AUC = c(round(auc.tree[[1]], 3), round(auc.rf[[1]], 3))

)

### Linear Regression for Duration

*#### read in data from Matts Github (downloaded locally because of git issues on work computer)*

bank\_clean <- read.csv("/Users/molheiser/Downloads/bank\_clean.csv")

bank\_train <- read.csv("/Users/molheiser/Downloads/bank\_clean\_train.csv")

bank\_test <- read.csv("/Users/molheiser/Downloads/bank\_clean\_test.csv")

*# Take log transform of Duration due to skewness*

bank\_clean$duration[bank\_clean$duration == 0] <- 0.1

bank\_train$duration[bank\_train$duration == 0] <- 0.1

bank\_test$duration[bank\_test$duration == 0] <- 0.1

bank\_clean$log\_duration <- log(bank\_clean$duration)

*# bit of EDA*

**library**(tidyverse)

b1 <- bank\_clean %>% ggplot(aes(x = job, y = log\_duration)) +

geom\_boxplot() +

theme(legend.position = "none")

b2 <- bank\_clean %>% ggplot(aes(x = marital, y = log\_duration)) +

geom\_boxplot() +

theme(legend.position = "none")

b3 <- bank\_clean %>% ggplot(aes(x = education, y = log\_duration)) +

geom\_boxplot() +

theme(legend.position = "none")

b4 <- bank\_clean %>% ggplot(aes(x = housing, y = log\_duration)) +

geom\_boxplot() +

theme(legend.position = "none")

b5 <- bank\_clean %>% ggplot(aes(x = loan, y = log\_duration)) +

geom\_boxplot() +

theme(legend.position = "none")

b6 <- bank\_clean %>% ggplot(aes(x = contact, y = log\_duration)) +

geom\_boxplot() +

theme(legend.position = "none")

b7 <- bank\_clean %>% ggplot(aes(x = month, y = log\_duration)) +

geom\_boxplot() +

theme(legend.position = "none")

b8 <- bank\_clean %>% ggplot(aes(x = day\_of\_week, y = log\_duration)) +

geom\_boxplot() +

theme(legend.position = "none")

b9 <- bank\_clean %>% ggplot(aes(x = as.factor(previous), y = log\_duration)) +

geom\_boxplot() +

theme(legend.position = "none")

**library**(gridExtra)

grid.arrange(b1, b2, b3, b4, b5, b6, b7, b8, b9)

numericals <- seq(1, length(bank\_clean))[unlist(lapply(bank\_clean, is.numeric))]

*# correlation plot for numericals*

**library**("corrplot")

**source**("http://www.sthda.com/upload/rquery\_cormat.r")

rquery.cormat(bank\_clean[, sapply(bank\_clean, is.numeric)])

**library**(GGally)

ggpairs(bank\_clean, columns = numericals)

*### Modeling ##########*

*# Fwd, Bwd, Step selection kept same variables*

*# Candidate model*

duration\_lm\_1 <- lm(log\_duration ~ education + age + job + marital + education + housing +

loan + contact + day\_of\_week + campaign + previous + month, data = bank\_clean)

*# stepwise selection*

**library**(olsrr)

step\_model <- ols\_step\_forward\_aic(duration\_lm\_1, details = FALSE)

step\_model

*# fit final model after stepwise selection*

duration\_lm\_1.fit <- lm(log(duration) ~ education + campaign + month + day\_of\_week + housing + previous + job + loan + contact + age, data = bank\_train)

summary(duration\_lm\_1.fit)

*# predict*

step\_pred <- predict(duration\_lm\_1.fit, newdata = bank\_test)

AIC(duration\_lm\_1.fit)

mean((step\_pred - log(bank\_test$duration))^2, na.rm = TRUE)

*# residual diagnostics*

par(mfrow = c(2, 2))

plot(duration\_lm\_1.fit)

*# use model to predict duration for entire data set and export*

duration\_lm\_1.fit <- lm(log(duration) ~ campaign + month + day\_of\_week + housing + previous + job + loan + contact + age, data = bank\_clean)

predicted\_log\_duration <- predict(duration\_lm\_1.fit, newdata = bank\_clean)

Duration\_Predicted <- merge(bank\_clean, predicted\_log\_duration, by = "row.names")

colnames(Duration\_Predicted)[22] <- "predicted\_log\_duration"

Duration\_Predicted$predicted\_duration <- exp(Duration\_Predicted$predicted\_log\_duration)

write.csv(Duration\_Predicted, "/Users/molheiser/Downloads/Duration\_Predicted.csv")

1. **Note**: though these variables are correlated statistically, we cannot say that an increase in any of these variables causes an increase in subscriptions (or vice versa) because this is an observational study. The team will not be calling this out during the rest of the paper, but it’s important to remember. [↑](#footnote-ref-1)
2. [↑](#endnote-ref-1)
3. [↑](#endnote-ref-2)
4. The methods in this case came from the `caret` package in R. *“Center”* subtracts the mean of the predictor’s data from the predictor values, and *“scale”* divides by the standard deviation. [↑](#footnote-ref-2)