Predictive Modelling DS 432 - Assignment 1

Saha Debanshee Gopal

28th September 2016

A) Predicting Bank Loan Defaulters

Introduction

A bank is a place where people go for loans. A problem every bank encounters is the the repayment of loans by a customer. A bank wants to weed out such loan applicants on the basis of previous data and successfully avoid dealing with such defaulters.

Challenge

Whenever an applicant applies for a loan, based on the history of customers, we should be able to predict whether the applicant will be successfully be able to pay the loan or not.

Business Understanding

Background Information

The bank judges the applicant based on the 5 C's of credit.Its used to gauge the credibility of the potential borrowers. The system judges the applicants on 5 factors. They are -
- Character or credit history
- Capacity
- Capital (Down Payment)
- Collateral
- Conditions

In addition to examining income, lenders look at the number of years an applicant has been at his job and job stability.

Problem description

The bank wants to reduce the "default" and try and examine if an applicant will be able to repay the loan or not.The bank thus, wants to generate a profile for the defaulters or parameters that results in applicants not repaying the loan.

Success Criteria

To model the profile of a defaulter based on certain parameters and past data

Data Mining Goals

With the given data,
- Determine what kind of candidate would or would not return the loan.
- Profile defaulters
- Predict, given the data, whether an applicant would default or not.

Data Understanding

Source of data : BankBayesLaon.txt

Data Description Report

The data set contains 850 observations of 9 variables.
The target variable is "defualt".

Data Quantity

Characteristics -

```
## 'data.frame':    850 obs. of  9 variables:
##  $ age     : num  41 27 40 41 24 41 39 43 24 36 ...
##  $ ed      : num  3 1 1 1 2 2 1 1 1 1 ...
##  $ employ  : num  17 10 15 15 2 5 20 12 3 0 ...
##  $ address : num  12 6 14 14 0 5 9 11 4 13 ...
##  $ income  : num  176 31 55 120 28 25 67 38 19 25 ...
##  $ debtinc : num  9.3 17.3 5.5 2.9 17.3 10.2 30.6 3.6 24.4 19.7 ...
##  $ creddebt: num  11.359 1.362 0.856 2.659 1.787 ...
##  $ othdebt : num  5.009 4.001 2.169 0.821 3.057 ...
##  $ default : num  1 0 0 0 1 0 0 0 1 0 ...
```

Data quality

Statistics -

```
##       age        ed          employ          address           income      
##  Min.   :20.00   1:372   Min.   : 0.000   Min.   : 0.000   Min.   : 14.0  
##  1st Qu.:29.00   2:198   1st Qu.: 3.000   1st Qu.: 3.000   1st Qu.: 24.0  
##  Median :34.00   3: 87   Median : 7.000   Median : 7.000   Median : 34.0  
##  Mean   :34.86   4: 38   Mean   : 8.389   Mean   : 8.279   Mean   : 45.6  
##  3rd Qu.:40.00   5:  5   3rd Qu.:12.000   3rd Qu.:12.000   3rd Qu.: 55.0  
##  Max.   :56.00           Max.   :31.000   Max.   :34.000   Max.   :446.0  
##     debtinc         creddebt          othdebt            default      
##  Min.   : 0.40   Min.   : 0.0117   Min.   : 0.04558   Min.   :0.0000  
##  1st Qu.: 5.00   1st Qu.: 0.3691   1st Qu.: 1.04418   1st Qu.:0.0000  
##  Median : 8.60   Median : 0.8549   Median : 1.98757   Median :0.0000  
##  Mean   :10.26   Mean   : 1.5536   Mean   : 3.05821   Mean   :0.2614  
##  3rd Qu.:14.12   3rd Qu.: 1.9020   3rd Qu.: 3.92306   3rd Qu.:1.0000  
##  Max.   :41.30   Max.   :20.5613   Max.   :27.03360   Max.   :1.0000
```

```
## [1] 517
```

```
## [1] 183
```

```
## [1] "formula"
```

```
## default ~ age + ed + employ + address + income + debtinc + creddebt +
##     othdebt
```

Logistic Regression

```
##
## Call:
## glm(formula = frml, family = "binomial", data = Trngdt)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6876  -0.7018  -0.3118   0.6095   2.8130
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.293906   0.718430  -1.801 0.071700 .
## age           0.033808   0.020076   1.684 0.092177 .
## ed2           0.158343   0.293803   0.539 0.589927
## ed3           0.337954   0.398677   0.848 0.396612
## ed4          -0.256491   0.542943  -0.472 0.636635
## ed5         -12.967965 680.501346  -0.019 0.984796
## employ       -0.226124   0.038449  -5.881 4.07e-09 ***
## address      -0.094722   0.026738  -3.543 0.000396 ***
## income       -0.009357   0.013958  -0.670 0.502633
## debtinc       0.049013   0.039376   1.245 0.213224
## creddebt      0.605106   0.145914   4.147 3.37e-05 ***
## othdebt       0.052981   0.097409   0.544 0.586511
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 571.62  on 490  degrees of freedom
## Residual deviance: 410.81  on 479  degrees of freedom
## AIC: 434.81
##
## Number of Fisher Scoring iterations: 14
```

```
## Start:  AIC=434.81
## default ~ age + ed + employ + address + income + debtinc + creddebt +
##     othdebt
##
##            Df Deviance    AIC
## - ed        4   412.91 428.91
## - othdebt   1   411.11 433.11
## - income    1   411.29 433.29
## - debtinc   1   412.34 434.34
## <none>          410.81 434.81
## - age       1   413.62 435.62
## - address   1   424.26 446.26
## - creddebt  1   438.07 460.07
## - employ    1   456.69 478.69
##
## Step:  AIC=428.91
## default ~ age + employ + address + income + debtinc + creddebt +
##     othdebt
##
```

```
##            Df Deviance    AIC
## - othdebt   1   413.45 427.45
## - income    1   413.61 427.61
## - debtinc   1   414.16 428.16
## <none>          412.91 428.91
## - age       1   415.48 429.48
## - address   1   426.88 440.88
## - creddebt  1   442.44 456.44
## - employ    1   468.50 482.50
##
## Step:  AIC=427.45
## default ~ age + employ + address + income + debtinc + creddebt
##
##            Df Deviance    AIC
## - income    1   413.64 425.64
## <none>          413.45 427.45
## - age       1   416.00 428.00
## - debtinc   1   421.13 433.13
## - address   1   427.55 439.55
## - creddebt  1   443.30 455.30
## - employ    1   468.81 480.81
##
## Step:  AIC=425.64
## default ~ age + employ + address + debtinc + creddebt
##
##            Df Deviance    AIC
## <none>          413.64 425.64
## - age       1   416.06 426.06
## - debtinc   1   423.32 433.32
## - address   1   428.13 438.13
## - creddebt  1   449.45 459.45
## - employ    1   480.99 490.99
```

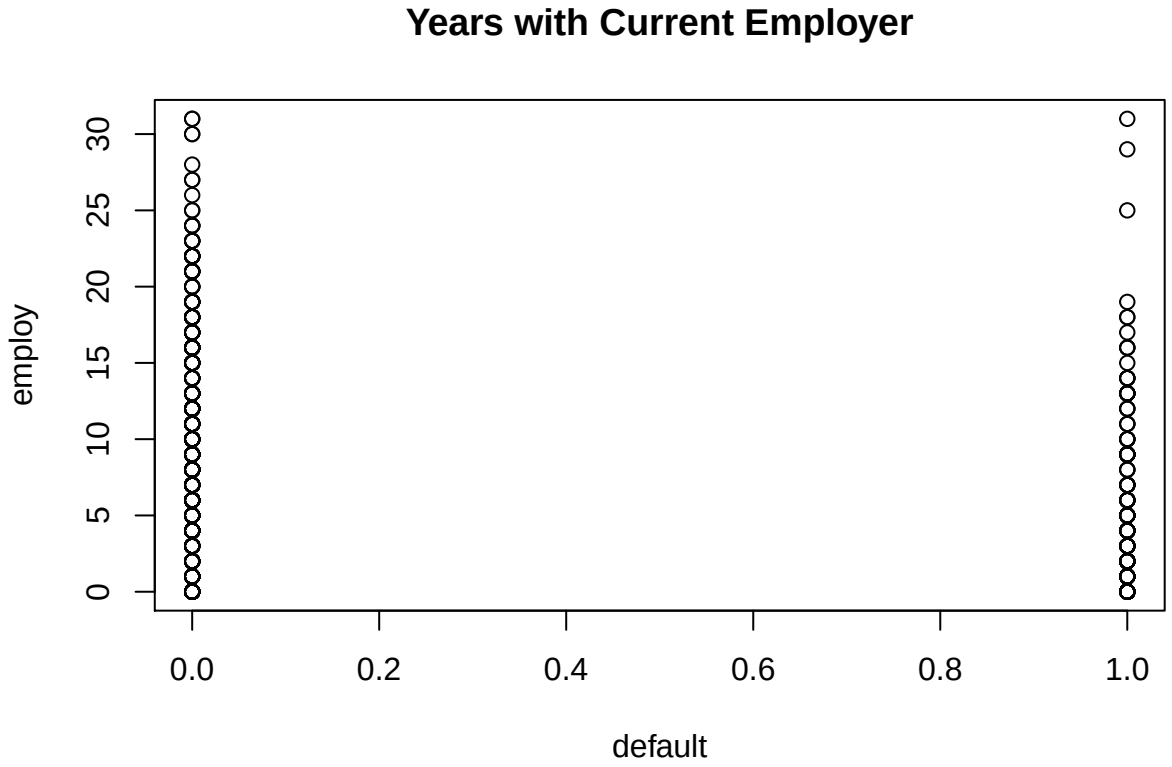
Data Exploration Report

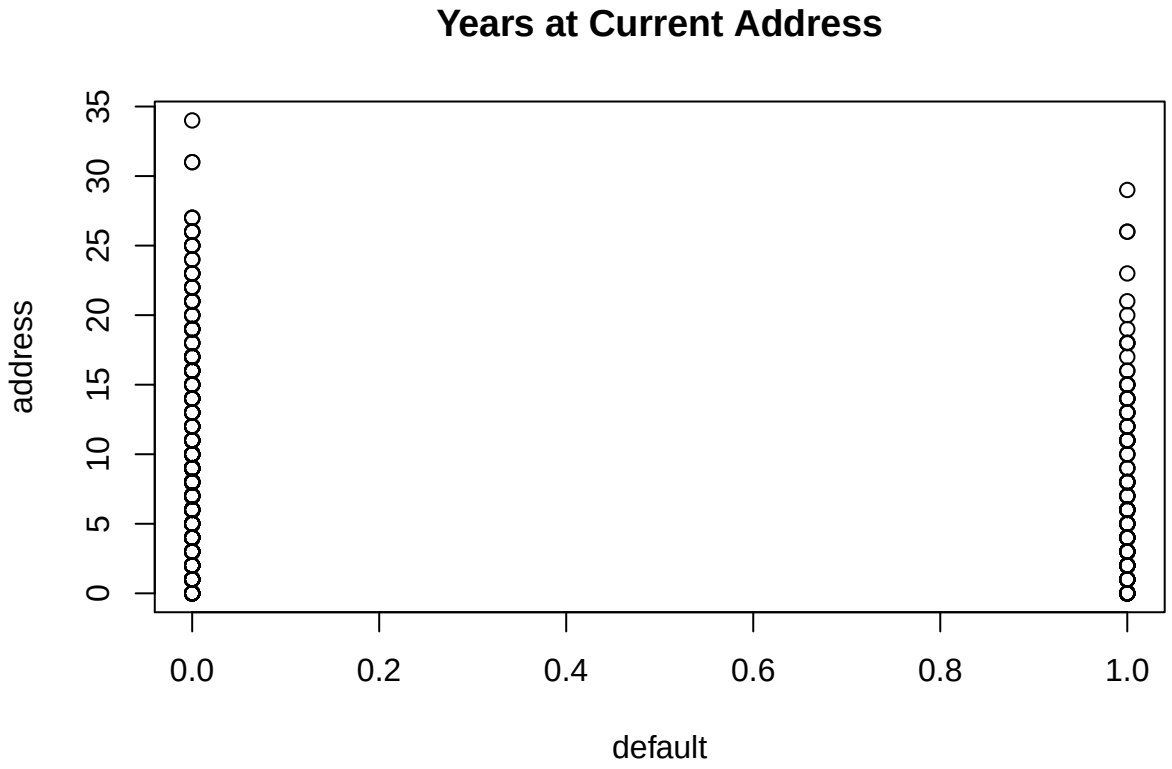
The above data gives us the relation between the target variable and the independent variables. The data shows that defaulters can be profile with the following formula.
default ~ employ + address + creddebt + debtinc

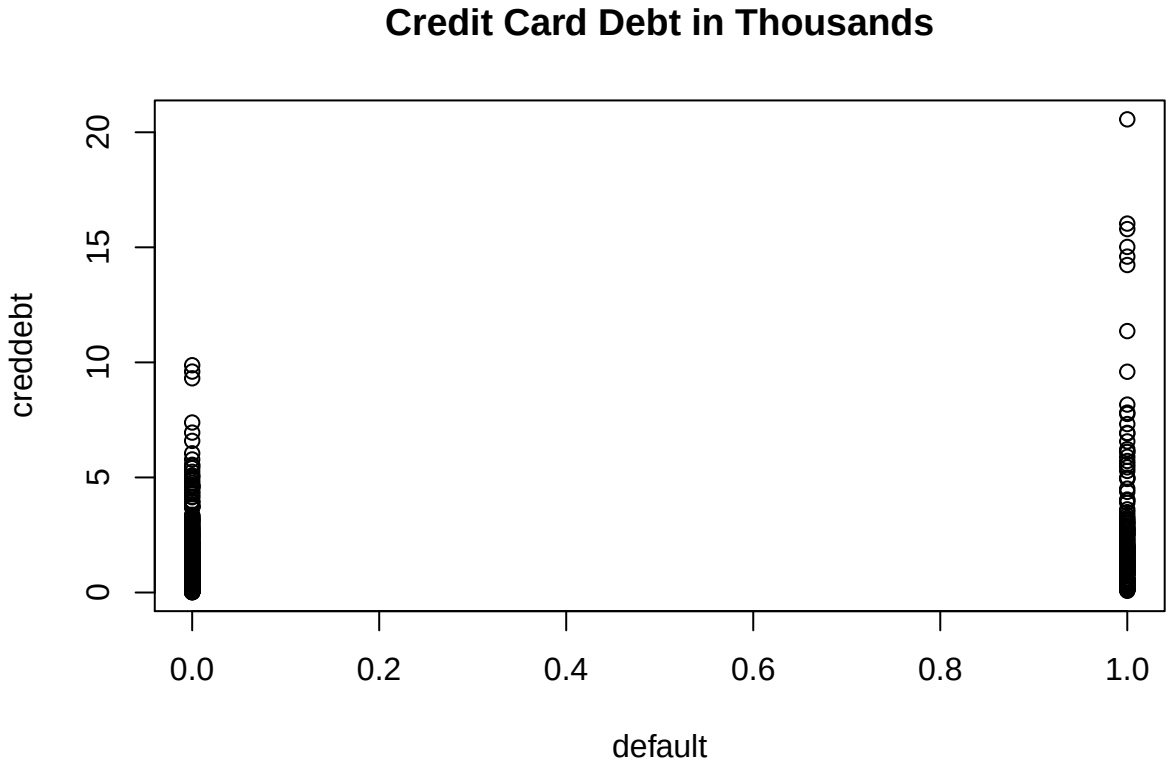
Hypothesis formed

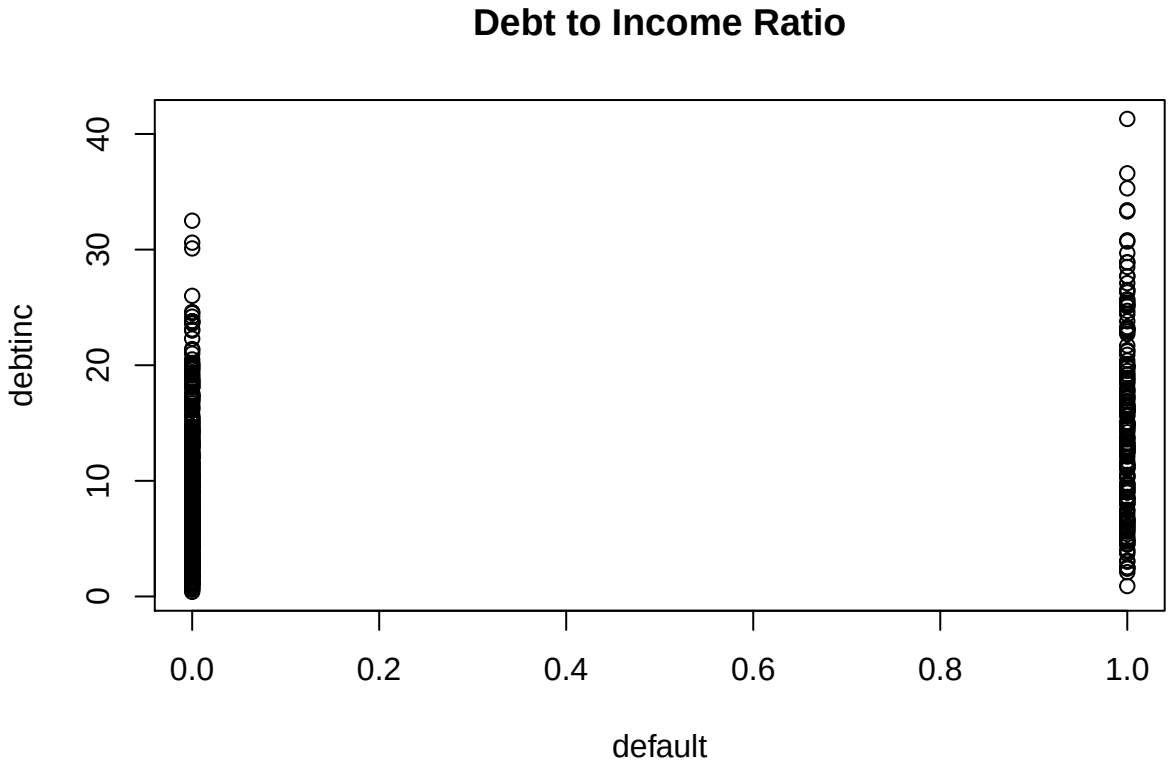
H0:Employment and Creddebt will have a significant effect on deafault.
Other characteristics about the data -
Address : The number of years at the current address plays a part in profiling a defaulter.
Debt to income ratio : The existing debt of the applicant to the ratio of his income is also plays a significant role in the profiling of a defaulter.
Age does not particularly have much impact on the fact that whether an applicant will churn or not. The credit card debt and the number of years the applicant has been working with the current employer are the significant attributes.

Graphs


Years with Current Employer
employ
0
5
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0
default


Years at Current Address
address
0
5
10
15
20
25
30
35
0.0
0.2
0.4
0.6
0.8
1.0
default





Verifying Hypothesis

```
## Groups for Hosmer-Lemeshow C statistic:
## cutfit
```

```
## [0.00105,0.0141]  (0.0141,0.0376]  (0.0376,0.0759]   (0.0759,0.139]
##               50                49                49                49
##    (0.139,0.205]    (0.205,0.282]    (0.282,0.384]    (0.384,0.492]
##               49                49                49                49
##    (0.492,0.652]    (0.652,0.999]
##               49                49
## Groups for Hosmer-Lemeshow H statistic:
## cutfit1
## [5.35e-05,0.101]    (0.101,0.201]      (0.201,0.3]        (0.3,0.4]
##              172               72               65               36
##        (0.4,0.5]        (0.5,0.6]        (0.6,0.7]      (0.7,0.799]
##               48               39               29               16
##    (0.799,0.899]        (0.899,1]
##                6                8
## $C
##
##  Hosmer-Lemeshow C statistic
##
## data:  Training_model$fitted.values and Training_model$y
## X-squared = 3.3346, df = 8, p-value = 0.9116
##
##
## $H
##
##  Hosmer-Lemeshow H statistic
##
## data:  Training_model$fitted.values and Training_model$y
## X-squared = 3.18, df = 8, p-value = 0.9226
```

H0 (Initial hypotheses): The model is a good fit.

Since the p value is greater than 0.5, it proves that our initial hypotheses is right and that our model is a good fit.

B) Campaign Offers

Introduction

Banks like every other industry in the world today has competition. A bank wants to solve the problem of customer churn by offering the right offer on financial services to the customers.

Challenge

The challenge is to identify the characteristic of customers who are most likely to respond to an offer based on the past data.

Business Understanding

Background Information

The current business scenario needs to be analyzed in this situation :
Marketing campaigns are strategically important to any firm. However, satisfactory response rates from such marketing strategies have traditionally been a challenge.
Given that there are time and budget constraints on marketing strategies, it becomes a key focus area for any business wanting to improve its efficiency and market position.
Another challenge is the ever changing consumer behavior to such advertisements.
Customers want more out of the advertisements such as personalised greetingse etc.. The fact that today there are newer and more sophisticated ways of reaching out to the customer can add a significant burden on the marketing dollar.
The marketing campaigns need to be :
-Relevant : Making the right offer to the right person.
-Timely.

Problem description

The bank wants to simply identify what offers to advertise to which customer based on data analytics.

Business Success Criteria

To match campaign offers with the right customer profile to get the best response.

Data Mining Goals

With the help of the data -
- Profile and classify the customer as a right or a wrong candidate for an offer. - Predicting which profile of customer will result in a positive response for the offer.

Data Understanding

Data Description Report

Data Quantity

Characteristics -

```
## 'data.frame':    22465 obs. of  31 variables:
##  $ customer_id                   : num  7 13 15 16 23 24 30 30 33 42 ...
##  $ campaign                      : num  2 2 2 2 2 2 2 3 2 3 ...
##  $ response                      : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ response_date                 : num  NA NA NA 38903 NA ...
##  $ purchase                      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ purchase_date                 : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ product_id                    : num  NA NA NA 183 NA NA NA NA NA NA ...
##  $ Rowid                         : num  1 2 3 761 4 5 6 7 8 9 ...
##  $ age                           : num  19 44 45 43 42 39 23 23 24 35 ...
##  $ age_youngest_child            : num  0 12 12 12 11 7 0 0 0 8 ...
##  $ average_balance_feed_index    : num  0 0 75 161 0 521 0 0 0 0 ...
##  $ branch                        : chr  "Ravensville" "Catburg" "Ravensville" "Catburg" ...
##  $ debt_equity                   : num  19 44 45 43 42 39 23 23 24 35 ...
##  $ gender                        : chr  "M" "M" "F" "F" ...
##  $ bad_payment                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gold_card                     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pension_plan                  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ household_debt_to_equity_ratio: num  64 65 65 65 65 65 65 65 65 65 ...
##  $ income                        : num  13362 13453 13453 13453 13453 ...
##  $ marital                       : chr  "S" "S" "U" "U" ...
##  $ members_in_household          : num  2 2 2 2 2 2 2 2 2 2 ...
##  $ months_current_account        : num  -4 17 33 34 22 26 30 30 -2 22 ...
##  $ months_customer               : num  0 24 36 36 24 36 36 36 0 24 ...
##  $ call_center_contacts          : num  1 0 0 0 0 0 0 0 0 0 ...
##  $ loan_accounts                 : num  1 0 4 0 0 0 1 1 0 1 ...
##  $ number_products               : num  1 0 2 3 0 0 0 0 0 0 ...
##  $ number_transactions           : num  0 0 1 2 0 3 0 0 0 0 ...
##  $ non_worker_percentage         : num  9 14 14 14 14 14 14 14 14 14 ...
##  $ white_collar_percentage       : num  26 19 19 19 19 19 19 19 19 19 ...
##  $ rfm_score                     : num  0 0 7.6 10.1 0 ...
##  $ X_random                      : num  3 1 1 1 3 2 1 1 1 3 ...
```

Summary after cleaning data

```
##   customer_id     campaign     response      response_date
##  Min.   :     7   1: 1838   Min.   :0.0000   Min.   :38808
##  1st Qu.: 17414   2:13951   1st Qu.:0.0000   1st Qu.:38902
##  Median : 34074   3: 5285   Median :0.0000   Median :38902
##  Mean   : 36724   4: 1391   Mean   :0.1276   Mean   :38902
##  3rd Qu.: 50720             3rd Qu.:0.0000   3rd Qu.:38902
##  Max.   :140097             Max.   :1.0000   Max.   :38929
##     purchase      purchase_date     product_id      Rowid
##  Min.   :0.0000   Min.   :38808   Min.   :183   Min.   :    1
##  1st Qu.:0.0000   1st Qu.:38840   1st Qu.:311   1st Qu.: 2809
##  Median :0.0000   Median :38840   Median :311   Median : 8367
```

```
##  Mean   :0.0884   Mean   :38840   Mean   :311   Mean   : 8733
##  3rd Qu.:0.0000   3rd Qu.:38840   3rd Qu.:311   3rd Qu.:13983
##  Max.   :1.0000   Max.   :38868   Max.   :421   Max.   :19599
##       age        age_youngest_child average_balance_feed_index
##  Min.   :10.00   Min.   : 0.00      Min.   :   0.0
##  1st Qu.:32.00   1st Qu.: 2.00      1st Qu.:   0.0
##  Median :39.00   Median : 9.00      Median :  58.0
##  Mean   :39.51   Mean   :11.18      Mean   : 138.7
##  3rd Qu.:46.00   3rd Qu.:17.00      3rd Qu.: 202.0
##  Max.   :96.00   Max.   :66.00      Max.   :3322.0
##     branch           debt_equity    gender     bad_payment
##  Length:22465       Min.   :10.00   F:11348   Min.   :0.000000
##  Class :character   1st Qu.:32.00   M:11117   1st Qu.:0.000000
##  Mode  :character   Median :39.00             Median :0.000000
##                     Mean   :39.51             Mean   :0.002982
##                     3rd Qu.:46.00             3rd Qu.:0.000000
##                     Max.   :96.00             Max.   :1.000000
##    gold_card       pension_plan      household_debt_to_equity_ratio
##  Min.   :0.0000   Min.   :0.000000   Min.   :10.0
##  1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.:49.0
##  Median :0.0000   Median :0.000000   Median :55.0
##  Mean   :0.0312   Mean   :0.005965   Mean   :54.5
##  3rd Qu.:0.0000   3rd Qu.:0.000000   3rd Qu.:61.0
##  Max.   :1.0000   Max.   :1.000000   Max.   :84.0
##      income      marital   members_in_household months_current_account
##  Min.   :13362   M: 5950   Min.   :1.000        Min.   : 0.00
##  1st Qu.:35367   S:10518   1st Qu.:3.000        1st Qu.:11.00
##  Median :42678   U: 5997   Median :4.000        Median :22.00
##  Mean   :43707             Mean   :4.301        Mean   :21.48
##  3rd Qu.:50954             3rd Qu.:6.000        3rd Qu.:32.00
##  Max.   :88016             Max.   :7.000        Max.   :41.00
##  months_customer call_center_contacts loan_accounts   number_products
##  Min.   : 0.00   Min.   : 0.000       Min.   :0.000   Min.   :0.000
##  1st Qu.:24.00   1st Qu.: 1.000       1st Qu.:0.000   1st Qu.:0.000
##  Median :24.00   Median : 2.000       Median :1.000   Median :1.000
##  Mean   :27.51   Mean   : 2.249       Mean   :1.096   Mean   :1.643
##  3rd Qu.:36.00   3rd Qu.: 3.000       3rd Qu.:2.000   3rd Qu.:3.000
##  Max.   :48.00   Max.   :10.000       Max.   :7.000   Max.   :6.000
##  number_transactions non_worker_percentage white_collar_percentage
##  Min.   : 0.000      Min.   : 0            Min.   : 5
##  1st Qu.: 0.000      1st Qu.: 8            1st Qu.:27
##  Median : 1.000      Median :11            Median :33
##  Mean   : 2.085      Mean   :11            Mean   :33
##  3rd Qu.: 2.000      3rd Qu.:13            3rd Qu.:39
##  Max.   :45.000      Max.   :37            Max.   :73
##    rfm_score         X_random
##  Min.   : 0.000   Min.   :1.000
##  1st Qu.: 0.000   1st Qu.:1.000
##  Median : 7.732   Median :2.000
##  Mean   : 6.431   Mean   :2.006
##  3rd Qu.:10.483   3rd Qu.:3.000
##  Max.   :52.568   Max.   :3.000

## [1] "formula"
```

```
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + age_youngest_child +
##     average_balance_feed_index + branch + debt_equity + gender +
##     bad_payment + gold_card + pension_plan + household_debt_to_equity_ratio +
##     income + marital + members_in_household + months_current_account +
##     months_customer + call_center_contacts + loan_accounts +
##     number_products + number_transactions + non_worker_percentage +
##     white_collar_percentage + rfm_score + X_random
```

Data quality

```
##
## Call:
## glm(formula = frmla, data = Trngdt)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.44420  -0.05454  -0.01781   0.02362   0.96345
##
## Coefficients: (1 not defined because of singularities)
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -3.378e+01  1.660e+01  -2.035 0.041818 *
## customer_id                    1.444e-06  3.390e-07   4.258 2.07e-05 ***
## campaign2                      2.094e-02  4.787e-03   4.375 1.22e-05 ***
## campaign3                     -1.338e-02  4.956e-03  -2.699 0.006961 **
## campaign4                     -4.745e-02  6.005e-03  -7.902 2.94e-15 ***
## response_date                  2.602e-03  1.236e-04  21.054  < 2e-16 ***
## purchase                       5.334e-01  5.940e-03  89.796  < 2e-16 ***
## purchase_date                 -1.735e-03  4.427e-04  -3.920 8.91e-05 ***
## product_id                    -4.583e-05  3.062e-05  -1.497 0.134479
## Rowid                         -2.493e-05  2.772e-07 -89.947  < 2e-16 ***
## age                            1.232e-03  3.088e-04   3.991 6.62e-05 ***
## age_youngest_child            -2.936e-04  3.436e-04  -0.854 0.392842
## average_balance_feed_index    -2.581e-05  7.606e-06  -3.393 0.000692 ***
## branchFoxton                  -4.284e-03  3.692e-03  -1.160 0.245893
## branchKingsville              -6.616e-03  3.667e-03  -1.804 0.071266 .
## branchPrinceton               -2.374e-04  3.662e-03  -0.065 0.948306
## branchQueensbury               7.449e-04  3.694e-03   0.202 0.840217
## branchRavensville             -4.822e-03  3.643e-03  -1.324 0.185650
## debt_equity                           NA         NA      NA       NA
## genderM                       -1.481e-02  2.134e-03  -6.942 4.03e-12 ***
## bad_payment                   -2.197e-02  1.996e-02  -1.101 0.270997
## gold_card                      1.995e-02  6.180e-03   3.228 0.001248 **
## pension_plan                   2.680e-05  1.397e-02   0.002 0.998469
## household_debt_to_equity_ratio -2.739e-03  2.147e-03  -1.276 0.201979
## income                         9.479e-06  7.405e-07  12.802  < 2e-16 ***
## maritalS                      -2.991e-03  2.792e-03  -1.071 0.284078
## maritalU                      -3.933e-03  2.976e-03  -1.322 0.186334
## members_in_household          -1.091e-03  6.227e-04  -1.752 0.079746 .
## months_current_account        -2.398e-04  2.945e-04  -0.814 0.415427
## months_customer                2.223e-04  2.476e-04   0.898 0.369254
## call_center_contacts          -3.652e-03  2.816e-03  -1.297 0.194712
## loan_accounts                  1.489e-03  7.816e-04   1.905 0.056761 .
```

```
## number_products                 3.122e-03  9.399e-04   3.322 0.000897 ***
## number_transactions            -1.069e-02  5.208e-04 -20.533  < 2e-16 ***
## non_worker_percentage          -2.311e-03  2.143e-03  -1.078 0.281028
## white_collar_percentage        -2.896e-03  2.145e-03  -1.350 0.177053
## rfm_score                       7.640e-03  4.441e-04  17.203  < 2e-16 ***
## X_random                       -9.527e-04  1.308e-03  -0.729 0.466296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.01775382)
##
##     Null deviance: 1755.92  on 15716  degrees of freedom
## Residual deviance:  278.38  on 15680  degrees of freedom
## AIC: -18716
##
## Number of Fisher Scoring iterations: 2
```

Data Mining Success Criteria

```
## Start:  AIC=-18715.79
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + age_youngest_child +
##     average_balance_feed_index + branch + debt_equity + gender +
##     bad_payment + gold_card + pension_plan + household_debt_to_equity_ratio +
##     income + marital + members_in_household + months_current_account +
##     months_customer + call_center_contacts + loan_accounts +
##     number_products + number_transactions + non_worker_percentage +
##     white_collar_percentage + rfm_score + X_random
##
##
## Step:  AIC=-18715.79
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + age_youngest_child +
##     average_balance_feed_index + branch + gender + bad_payment +
##     gold_card + pension_plan + household_debt_to_equity_ratio +
##     income + marital + members_in_household + months_current_account +
##     months_customer + call_center_contacts + loan_accounts +
##     number_products + number_transactions + non_worker_percentage +
##     white_collar_percentage + rfm_score + X_random
##
##                                 Df Deviance    AIC
## - branch                         5   278.50 -18719
## - marital                        2   278.41 -18718
## - pension_plan                   1   278.38 -18718
## - X_random                       1   278.39 -18717
## - months_current_account         1   278.39 -18717
## - age_youngest_child             1   278.39 -18717
## - months_customer                1   278.39 -18717
## - non_worker_percentage          1   278.40 -18717
## - bad_payment                    1   278.40 -18717
## - household_debt_to_equity_ratio 1   278.41 -18716
## - call_center_contacts           1   278.41 -18716
## - white_collar_percentage        1   278.41 -18716
```

```
## <none>                                278.38 -18716
## - product_id                       1   278.42 -18716
## - members_in_household             1   278.43 -18715
## - loan_accounts                    1   278.44 -18714
## - gold_card                        1   278.56 -18707
## - number_products                  1   278.58 -18707
## - average_balance_feed_index       1   278.58 -18706
## - purchase_date                    1   278.65 -18702
## - age                              1   278.66 -18702
## - customer_id                      1   278.70 -18700
## - gender                           1   279.24 -18670
## - income                           1   281.29 -18554
## - rfm_score                        1   283.63 -18424
## - campaign                         3   283.76 -18421
## - number_transactions              1   285.86 -18301
## - response_date                    1   286.25 -18280
## - purchase                         1   421.54 -12196
## - Rowid                            1   422.02 -12179
##
## Step:  AIC=-18718.81
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + age_youngest_child +
##     average_balance_feed_index + gender + bad_payment + gold_card +
##     pension_plan + household_debt_to_equity_ratio + income +
##     marital + members_in_household + months_current_account +
##     months_customer + call_center_contacts + loan_accounts +
##     number_products + number_transactions + non_worker_percentage +
##     white_collar_percentage + rfm_score + X_random
##
##                                  Df Deviance    AIC
## - marital                         2   278.54 -18721
## - pension_plan                    1   278.50 -18721
## - X_random                        1   278.51 -18720
## - months_current_account          1   278.52 -18720
## - age_youngest_child              1   278.52 -18720
## - months_customer                 1   278.52 -18720
## - bad_payment                     1   278.53 -18720
## - non_worker_percentage           1   278.53 -18720
## - call_center_contacts            1   278.53 -18719
## - household_debt_to_equity_ratio  1   278.53 -18719
## - white_collar_percentage         1   278.54 -18719
## <none>                                278.50 -18719
## - product_id                      1   278.54 -18719
## - members_in_household            1   278.56 -18718
## - loan_accounts                   1   278.57 -18717
## - gold_card                       1   278.69 -18710
## - number_products                 1   278.70 -18710
## - average_balance_feed_index      1   278.71 -18709
## - purchase_date                   1   278.77 -18706
## - age                             1   278.78 -18705
## - customer_id                     1   278.82 -18703
## - gender                          1   279.36 -18672
## - income                          1   281.43 -18556
## - rfm_score                       1   283.74 -18428
```

```
## - campaign                        3   283.88 -18424
## - number_transactions             1   285.97 -18305
## - response_date                   1   286.37 -18283
## - purchase                        1   421.66 -12202
## - Rowid                           1   422.27 -12179
##
## Step:  AIC=-18720.93
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + age_youngest_child +
##     average_balance_feed_index + gender + bad_payment + gold_card +
##     pension_plan + household_debt_to_equity_ratio + income +
##     members_in_household + months_current_account + months_customer +
##     call_center_contacts + loan_accounts + number_products +
##     number_transactions + non_worker_percentage + white_collar_percentage +
##     rfm_score + X_random
##
##                                  Df Deviance    AIC
## - pension_plan                    1   278.54 -18723
## - X_random                        1   278.55 -18722
## - months_current_account          1   278.55 -18722
## - months_customer                 1   278.55 -18722
## - age_youngest_child              1   278.55 -18722
## - bad_payment                     1   278.56 -18722
## - non_worker_percentage           1   278.56 -18722
## - call_center_contacts            1   278.57 -18721
## - household_debt_to_equity_ratio  1   278.57 -18721
## <none>                                278.54 -18721
## - white_collar_percentage         1   278.57 -18721
## - product_id                      1   278.58 -18721
## - members_in_household            1   278.59 -18720
## - loan_accounts                   1   278.62 -18718
## - gold_card                       1   278.73 -18712
## - number_products                 1   278.73 -18712
## - average_balance_feed_index      1   278.74 -18711
## - purchase_date                   1   278.81 -18708
## - age                             1   278.85 -18705
## - customer_id                     1   278.85 -18705
## - gender                          1   279.40 -18674
## - income                          1   281.47 -18558
## - rfm_score                       1   283.77 -18430
## - campaign                        3   283.91 -18426
## - number_transactions             1   286.00 -18307
## - response_date                   1   286.40 -18286
## - purchase                        1   421.67 -12205
## - Rowid                           1   422.40 -12178
##
## Step:  AIC=-18722.93
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + age_youngest_child +
##     average_balance_feed_index + gender + bad_payment + gold_card +
##     household_debt_to_equity_ratio + income + members_in_household +
##     months_current_account + months_customer + call_center_contacts +
##     loan_accounts + number_products + number_transactions + non_worker_percentage +
##     white_collar_percentage + rfm_score + X_random
```

```
##
##                                  Df Deviance    AIC
## - X_random                        1   278.55 -18724
## - months_current_account          1   278.55 -18724
## - months_customer                 1   278.55 -18724
## - age_youngest_child              1   278.55 -18724
## - bad_payment                     1   278.56 -18724
## - non_worker_percentage           1   278.56 -18724
## - call_center_contacts            1   278.57 -18723
## - household_debt_to_equity_ratio  1   278.57 -18723
## <none>                                278.54 -18723
## - white_collar_percentage         1   278.57 -18723
## - product_id                      1   278.58 -18723
## - members_in_household            1   278.59 -18722
## - loan_accounts                   1   278.62 -18720
## - gold_card                       1   278.73 -18714
## - number_products                 1   278.73 -18714
## - average_balance_feed_index      1   278.74 -18713
## - purchase_date                   1   278.81 -18710
## - age                             1   278.85 -18707
## - customer_id                     1   278.85 -18707
## - gender                          1   279.40 -18676
## - income                          1   281.47 -18560
## - rfm_score                       1   283.77 -18432
## - campaign                        3   283.91 -18428
## - number_transactions             1   286.05 -18307
## - response_date                   1   286.40 -18287
## - purchase                        1   421.70 -12206
## - Rowid                           1   422.41 -12180
##
## Step:  AIC=-18724.36
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + age_youngest_child +
##     average_balance_feed_index + gender + bad_payment + gold_card +
##     household_debt_to_equity_ratio + income + members_in_household +
##     months_current_account + months_customer + call_center_contacts +
##     loan_accounts + number_products + number_transactions + non_worker_percentage +
##     white_collar_percentage + rfm_score
##
##                                  Df Deviance    AIC
## - months_current_account          1   278.56 -18726
## - months_customer                 1   278.56 -18726
## - age_youngest_child              1   278.56 -18726
## - bad_payment                     1   278.57 -18725
## - non_worker_percentage           1   278.57 -18725
## - call_center_contacts            1   278.58 -18725
## - household_debt_to_equity_ratio  1   278.58 -18724
## <none>                                278.55 -18724
## - white_collar_percentage         1   278.58 -18724
## - product_id                      1   278.59 -18724
## - members_in_household            1   278.60 -18723
## - loan_accounts                   1   278.63 -18722
## - gold_card                       1   278.74 -18716
## - number_products                 1   278.74 -18715
```

```
## - average_balance_feed_index      1   278.75 -18715
## - purchase_date                   1   278.81 -18711
## - age                             1   278.86 -18709
## - customer_id                     1   278.86 -18708
## - gender                          1   279.41 -18678
## - income                          1   281.48 -18562
## - rfm_score                       1   283.79 -18434
## - campaign                        3   283.92 -18430
## - number_transactions             1   286.07 -18308
## - response_date                   1   286.41 -18289
## - purchase                        1   421.73 -12207
## - Rowid                           1   422.41 -12182
##
## Step:  AIC=-18725.73
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + age_youngest_child +
##     average_balance_feed_index + gender + bad_payment + gold_card +
##     household_debt_to_equity_ratio + income + members_in_household +
##     months_customer + call_center_contacts + loan_accounts +
##     number_products + number_transactions + non_worker_percentage +
##     white_collar_percentage + rfm_score
##
##                                  Df Deviance    AIC
## - months_customer                 1   278.56 -18728
## - age_youngest_child              1   278.57 -18727
## - bad_payment                     1   278.58 -18726
## - non_worker_percentage           1   278.58 -18726
## - call_center_contacts            1   278.59 -18726
## - household_debt_to_equity_ratio  1   278.59 -18726
## <none>                                278.56 -18726
## - white_collar_percentage         1   278.59 -18726
## - product_id                      1   278.60 -18726
## - members_in_household            1   278.61 -18725
## - loan_accounts                   1   278.64 -18723
## - gold_card                       1   278.75 -18717
## - number_products                 1   278.75 -18717
## - average_balance_feed_index      1   278.76 -18716
## - purchase_date                   1   278.82 -18713
## - age                             1   278.87 -18710
## - customer_id                     1   278.87 -18710
## - gender                          1   279.42 -18679
## - income                          1   281.49 -18563
## - rfm_score                       1   283.85 -18432
## - campaign                        3   283.97 -18429
## - number_transactions             1   286.12 -18306
## - response_date                   1   286.41 -18291
## - purchase                        1   421.81 -12206
## - Rowid                           1   422.42 -12184
##
## Step:  AIC=-18727.56
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + age_youngest_child +
##     average_balance_feed_index + gender + bad_payment + gold_card +
##     household_debt_to_equity_ratio + income + members_in_household +
```

```
##     call_center_contacts + loan_accounts + number_products +
##     number_transactions + non_worker_percentage + white_collar_percentage +
##     rfm_score
##
##                                  Df Deviance    AIC
## - age_youngest_child              1   278.58 -18729
## - bad_payment                     1   278.58 -18728
## - non_worker_percentage           1   278.58 -18728
## - call_center_contacts            1   278.59 -18728
## - household_debt_to_equity_ratio  1   278.59 -18728
## <none>                                278.56 -18728
## - white_collar_percentage         1   278.60 -18728
## - product_id                      1   278.60 -18727
## - members_in_household            1   278.62 -18726
## - loan_accounts                   1   278.64 -18725
## - gold_card                       1   278.75 -18719
## - number_products                 1   278.75 -18719
## - average_balance_feed_index      1   278.77 -18718
## - purchase_date                   1   278.83 -18714
## - age                             1   278.87 -18712
## - customer_id                     1   278.88 -18712
## - gender                          1   279.42 -18681
## - income                          1   281.49 -18565
## - campaign                        3   284.00 -18430
## - rfm_score                       1   284.04 -18423
## - number_transactions             1   286.29 -18300
## - response_date                   1   286.44 -18291
## - purchase                        1   421.81 -12208
## - Rowid                           1   422.42 -12186
##
## Step:  AIC=-18728.77
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + average_balance_feed_index +
##     gender + bad_payment + gold_card + household_debt_to_equity_ratio +
##     income + members_in_household + call_center_contacts + loan_accounts +
##     number_products + number_transactions + non_worker_percentage +
##     white_collar_percentage + rfm_score
##
##                                  Df Deviance    AIC
## - bad_payment                     1   278.60 -18730
## - non_worker_percentage           1   278.60 -18730
## - call_center_contacts            1   278.61 -18729
## - household_debt_to_equity_ratio  1   278.61 -18729
## <none>                                278.58 -18729
## - white_collar_percentage         1   278.61 -18729
## - product_id                      1   278.61 -18729
## - members_in_household            1   278.63 -18728
## - loan_accounts                   1   278.66 -18726
## - gold_card                       1   278.76 -18720
## - number_products                 1   278.77 -18720
## - average_balance_feed_index      1   278.78 -18719
## - purchase_date                   1   278.84 -18716
## - customer_id                     1   278.89 -18713
## - gender                          1   279.44 -18682
```

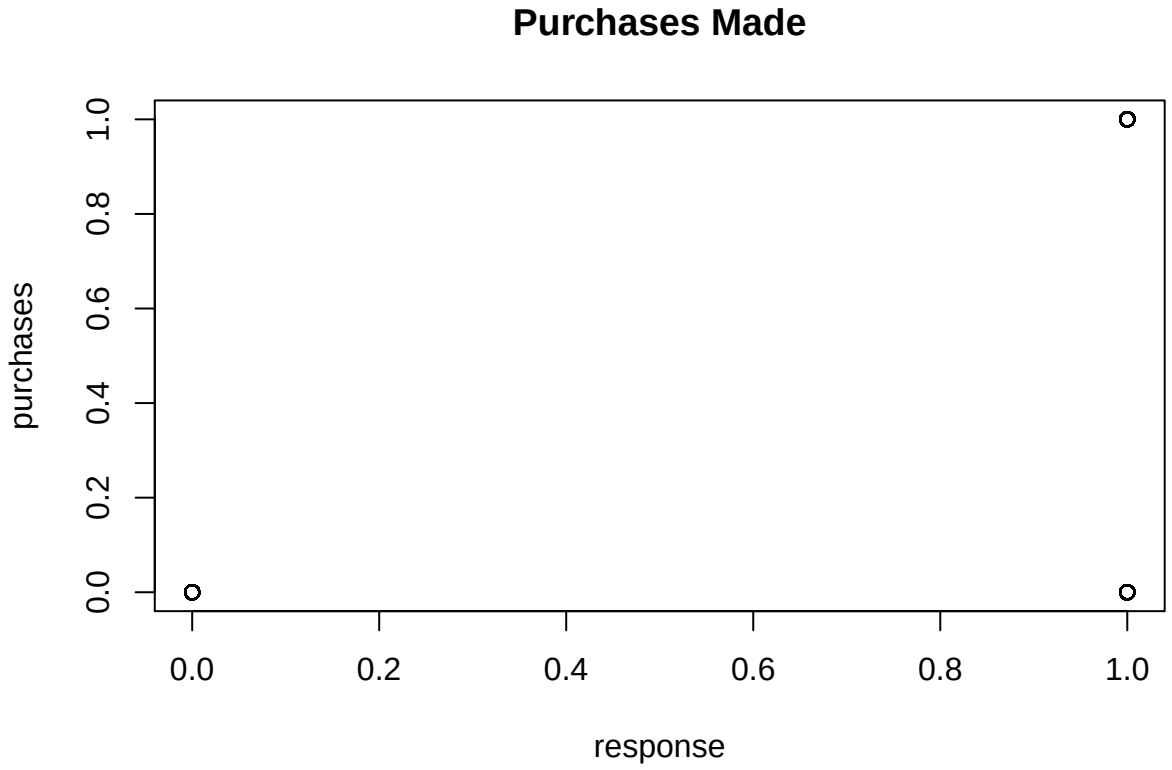
```
## - age                              1   280.63 -18615
## - income                           1   281.51 -18566
## - campaign                         3   284.01 -18431
## - rfm_score                        1   284.05 -18425
## - number_transactions              1   286.30 -18301
## - response_date                    1   286.47 -18292
## - purchase                         1   422.00 -12203
## - Rowid                            1   422.42 -12188
##
## Step:  AIC=-18729.5
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + average_balance_feed_index +
##     gender + gold_card + household_debt_to_equity_ratio + income +
##     members_in_household + call_center_contacts + loan_accounts +
##     number_products + number_transactions + non_worker_percentage +
##     white_collar_percentage + rfm_score
##
##                                   Df Deviance    AIC
## - non_worker_percentage            1   278.62 -18730
## - white_collar_percentage          1   278.63 -18730
## - household_debt_to_equity_ratio   1   278.63 -18730
## <none>                                 278.60 -18730
## - product_id                       1   278.64 -18729
## - call_center_contacts             1   278.65 -18729
## - members_in_household             1   278.65 -18729
## - loan_accounts                    1   278.68 -18727
## - gold_card                        1   278.78 -18721
## - number_products                  1   278.79 -18720
## - average_balance_feed_index       1   278.80 -18720
## - purchase_date                    1   278.87 -18716
## - customer_id                      1   278.91 -18714
## - gender                           1   279.45 -18683
## - age                              1   280.65 -18616
## - income                           1   281.53 -18567
## - campaign                         3   284.03 -18432
## - rfm_score                        1   284.07 -18426
## - number_transactions              1   286.32 -18302
## - response_date                    1   286.49 -18292
## - purchase                         1   422.00 -12205
## - Rowid                            1   422.59 -12183
##
## Step:  AIC=-18730.22
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + average_balance_feed_index +
##     gender + gold_card + household_debt_to_equity_ratio + income +
##     members_in_household + call_center_contacts + loan_accounts +
##     number_products + number_transactions + white_collar_percentage +
##     rfm_score
##
##                                   Df Deviance    AIC
## - household_debt_to_equity_ratio   1   278.65 -18731
## <none>                                 278.62 -18730
## - product_id                       1   278.66 -18730
## - members_in_household             1   278.67 -18729
```

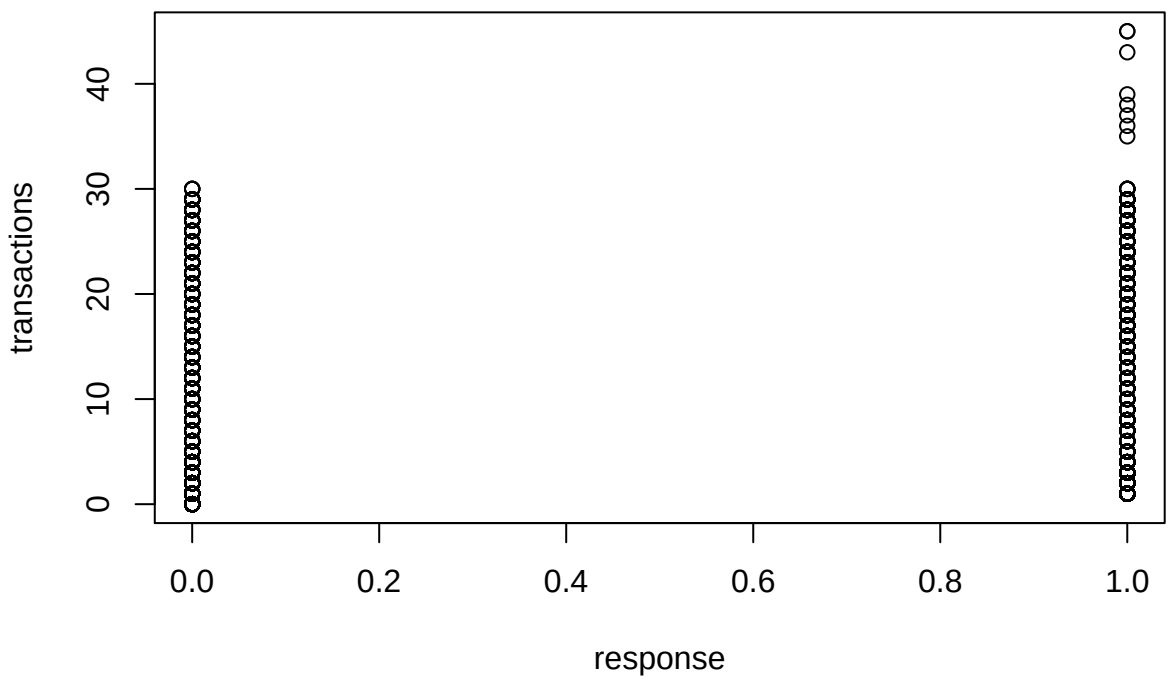
```
## - call_center_contacts           1   278.68 -18729
## - white_collar_percentage        1   278.69 -18728
## - loan_accounts                  1   278.70 -18727
## - gold_card                      1   278.81 -18722
## - number_products                1   278.81 -18721
## - average_balance_feed_index     1   278.82 -18721
## - purchase_date                  1   278.89 -18717
## - customer_id                    1   278.93 -18715
## - gender                         1   279.48 -18684
## - age                            1   280.68 -18617
## - income                         1   281.57 -18567
## - campaign                       3   284.04 -18433
## - rfm_score                      1   284.09 -18426
## - number_transactions            1   286.35 -18302
## - response_date                  1   286.52 -18293
## - purchase                       1   422.13 -12202
## - Rowid                          1   422.59 -12185
##
## Step:  AIC=-18730.66
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + average_balance_feed_index +
##     gender + gold_card + income + members_in_household + call_center_contacts +
##     loan_accounts + number_products + number_transactions + white_collar_percentage +
##     rfm_score
##
##                              Df Deviance    AIC
## - call_center_contacts        1   278.68 -18731
## <none>                            278.65 -18731
## - product_id                  1   278.69 -18730
## - white_collar_percentage     1   278.69 -18730
## - members_in_household        1   278.70 -18730
## - loan_accounts               1   278.73 -18728
## - gold_card                   1   278.83 -18722
## - number_products             1   278.84 -18722
## - average_balance_feed_index  1   278.85 -18721
## - purchase_date               1   278.92 -18718
## - customer_id                 1   278.96 -18715
## - gender                      1   279.51 -18684
## - age                         1   280.71 -18617
## - income                      1   281.59 -18567
## - campaign                    3   284.06 -18434
## - rfm_score                   1   284.11 -18428
## - number_transactions         1   286.36 -18304
## - response_date               1   286.54 -18294
## - purchase                    1   422.14 -12204
## - Rowid                       1   422.65 -12185
##
## Step:  AIC=-18730.76
## response ~ customer_id + campaign + response_date + purchase +
##     purchase_date + product_id + Rowid + age + average_balance_feed_index +
##     gender + gold_card + income + members_in_household + loan_accounts +
##     number_products + number_transactions + white_collar_percentage +
##     rfm_score
##
```

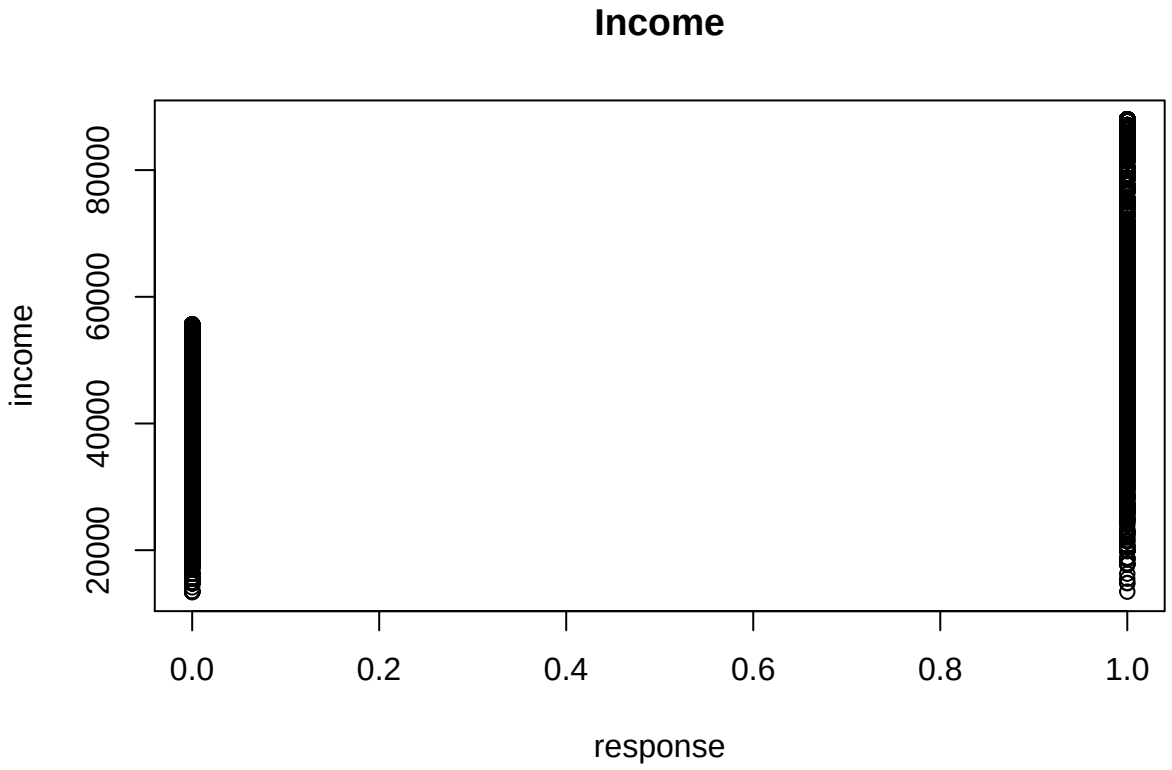
```
##                               Df Deviance    AIC
## <none>                             278.68 -18731
## - product_id                   1   278.72 -18731
## - members_in_household         1   278.73 -18730
## - loan_accounts                1   278.76 -18728
## - gold_card                    1   278.87 -18722
## - number_products              1   278.88 -18722
## - average_balance_feed_index   1   278.89 -18721
## - purchase_date                1   278.95 -18718
## - customer_id                  1   278.99 -18715
## - white_collar_percentage      1   279.28 -18699
## - gender                       1   279.54 -18685
## - age                          1   280.75 -18617
## - income                       1   281.62 -18568
## - campaign                     3   284.09 -18434
## - rfm_score                    1   284.15 -18428
## - number_transactions          1   286.39 -18304
## - response_date                1   286.57 -18294
## - purchase                     1   422.26 -12202
## - Rowid                        1   422.65 -12187
```

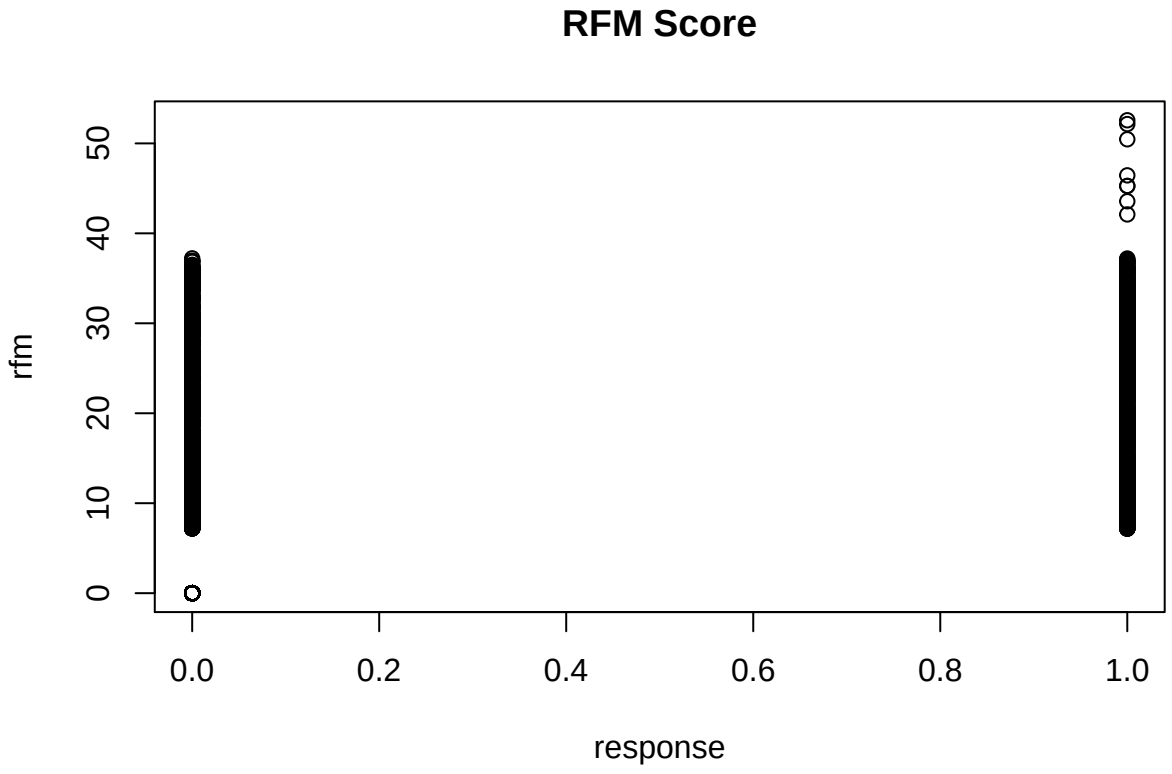
Data Exploration Report

The data shows us that the chances of a customer purchasing an offer is significantly affected by RFM score of the customer.
If a customer has purchased an offer before, he's more likely to purchase another one again.
The number of transactions of a customer show how active he/she is and this variable also plays a role in the sale of an offer. The more transactions there are, the more is the likelihood of the customer going for the offer.
Other than this, the income and the offer itself affect the response variable.


Purchases Made
purchases
0.0
0.2
0.4
0.6
0.8
1.0
response


Number of transactions
transactions
0
10
20
30
40
0.0
0.2
0.4
0.6
0.8
1.0
response





Verifying Data Quality

```
## Groups for Hosmer-Lemeshow C statistic:
## cutfit
```

```
##   [-0.151,-0.0403]  (-0.0403,-0.0178] (-0.0178,0.000314]
##               1572               1572               1571
##  (0.000314,0.0156]    (0.0156,0.0303]    (0.0303,0.0444]
##               1572               1572               1571
##    (0.0444,0.0624]    (0.0624,0.0906]     (0.0906,0.631]
##               1572               1571               1572
##       (0.631,1.44]
##               1572
## Groups for Hosmer-Lemeshow H statistic:
## cutfit1
## [-0.153,0.00815]  (0.00815,0.167]    (0.167,0.327]    (0.327,0.486]
##             5510             8224              119              115
##    (0.486,0.645]    (0.645,0.804]    (0.804,0.963]     (0.963,1.12]
##              203              360              384              328
##      (1.12,1.28]      (1.28,1.44]
##              281              193

## $C
##
##  Hosmer-Lemeshow C statistic
##
## data:  Training_model$fitted.values and Training_model$y
## X-squared = 237.68, df = 8, p-value < 2.2e-16
##
##
## $H
##
##  Hosmer-Lemeshow H statistic
##
## data:  Training_model$fitted.values and Training_model$y
## X-squared = 937.48, df = 8, p-value < 2.2e-16
```

The model is not a good fit.

C) Retail Forecasting

Introduction

A Retail company wants to forecast for the next four months the number of items that will be consumed by males. They wish to forecast the growth of their sales for Men's items into the future so that the company can get an idea of the future demand and use that for better production planning of its merchandize. In doing so, it can have a chance to both minimize the errors due to underproduction and having product shortages, or overproduction which will lead to excess stock that would lead to an additional cost. For this, they have engaged on a forecasting project based on time series analysis techniques.

Challenge.

The objective of the catalog prediction exercise is to build a forecast for the number of items that will be consumed by men (male customers) in the next four months. The data is historic data for ten years, from 1989 to 1998.

Business Objectives:

Background Information

In retail management, forecasting services are used to predict and meet the demands of consumers in retail establishments while controlling pricing and inventory. Holding excess inventory adds to overhead costs for a business. When forecasting helps the retailer to meet the demands of the customer by understanding consumer purchase patterns better, more efficient use of shelf and display space within the retail establishment is the result, in addition to optimal use of inventory space.
Close examination of the data will uncover trends in the months, days and times of day. The most obvious peaks and troughs are usually the seasons. Slighter trends can be caused by things like local events, weather, promotions or government announcements

Problem description

The retailer wants to forecast the nature and magnitude of demand in male clothing to optimize sales and maximize profit.
The retailer also wants to identify the factors that can help increase demand, and customize them accordingly.

Business Success Criteria

To forecast demand for men's merchandise based on certain parameters and past data.

Data Mining Goals

With the help of the data, we translate the retailer's business objectives into data mining terms.

The goals for the initial study to be completed are:
a. Profiling: Use the data given to determine products that are likely to sell the most based on past customer behaviour.
b. Predicting: Approximate sale of different families of products.
c. Documenting of technical goals using specific units of time, i.e predictions with a 4-month validity.
d. Attempting to provide actual numbers for desired outcomes, such as increase or decrease in demand in percent of various products for men.

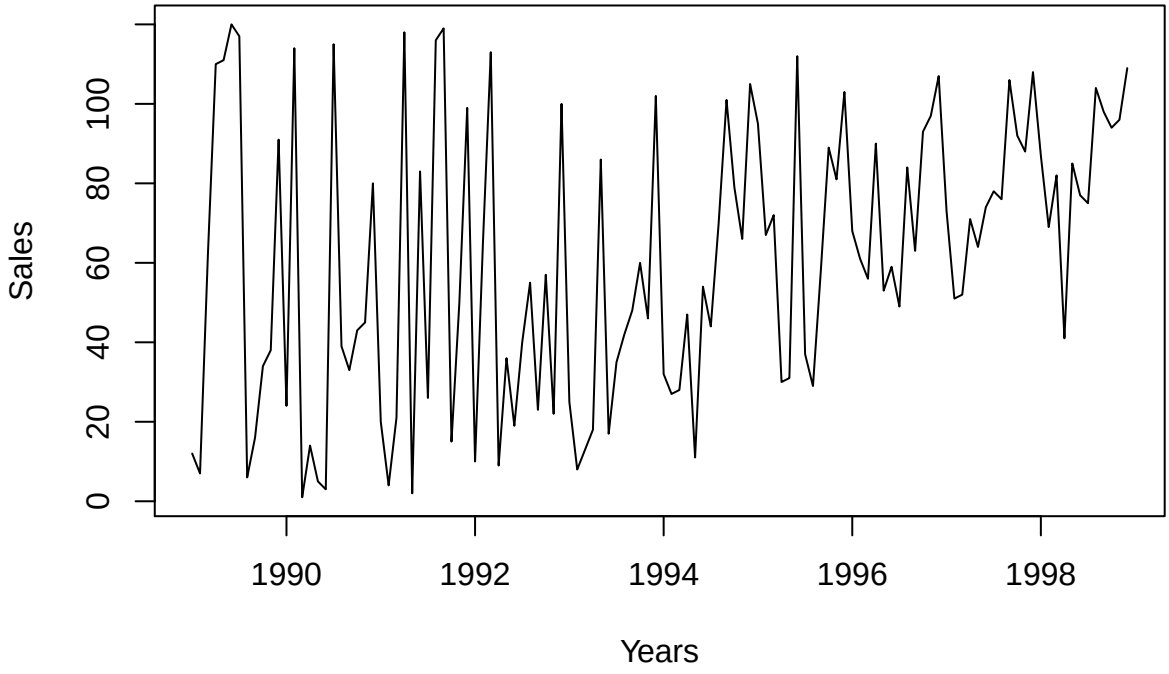
Data Understanding

```
## 'data.frame':    120 obs. of  16 variables:
##  $ date                  : Factor w/ 120 levels "32509","32540",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ men                   : Factor w/ 120 levels "10061.19","10074.24",..: 12 7 62 110 111 120 117 6
##  $ women                 : Factor w/ 120 levels "16578.93","18103.06",..: 1 3 73 23 19 20 24 25 22 5
##  $ jewel                 : Factor w/ 120 levels "10388.66","10732.06",..: 3 4 90 9 71 8 10 73 57 70
##  $ mail                  : Factor w/ 117 levels "10047","10065",..: 68 73 69 64 82 66 67 71 65 84 .
##  $ page                  : Factor w/ 48 levels "101","102","104",..: 25 40 17 37 26 39 31 24 35 7 .
##  $ phone                 : Factor w/ 36 levels "17","20","21",..: 16 11 6 2 1 12 10 9 17 7 ...
##  $ print                 : Factor w/ 120 levels "18061.2","20911.52",..: 4 45 51 71 5 43 29 89 32 10
##  $ service               : Factor w/ 43 levels "15","16","18",..: 5 5 11 7 6 8 7 5 1 5 ...
##  $ YEAR_                 : Factor w/ 10 levels "1989","1990",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ MONTH_                : Factor w/ 12 levels "1","10","11",..: 1 5 6 7 8 9 10 11 12 2 ...
```

```
##  $ DATE_                  : Factor w/ 120 levels "32509","32540",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Seasonal_Err_Men       : Factor w/ 120 levels "0.235599","0.621646",..: 31 17 120 3 6 107 70 47 9
##  $ Seasonal_AdjSer_Men    : Factor w/ 120 levels "10129.166245",..: 14 20 82 114 115 9 6 5 19 8 ...
##  $ Seasonal_Factors_Men   : Factor w/ 12 levels "0.845035","0.849452",..: 9 1 3 2 4 6 5 8 7 11 ...
##  $ Seasonal_TrendCycle_Men: Factor w/ 120 levels "10019.529191",..: 29 54 36 6 2 1 3 5 8 10 ...
```

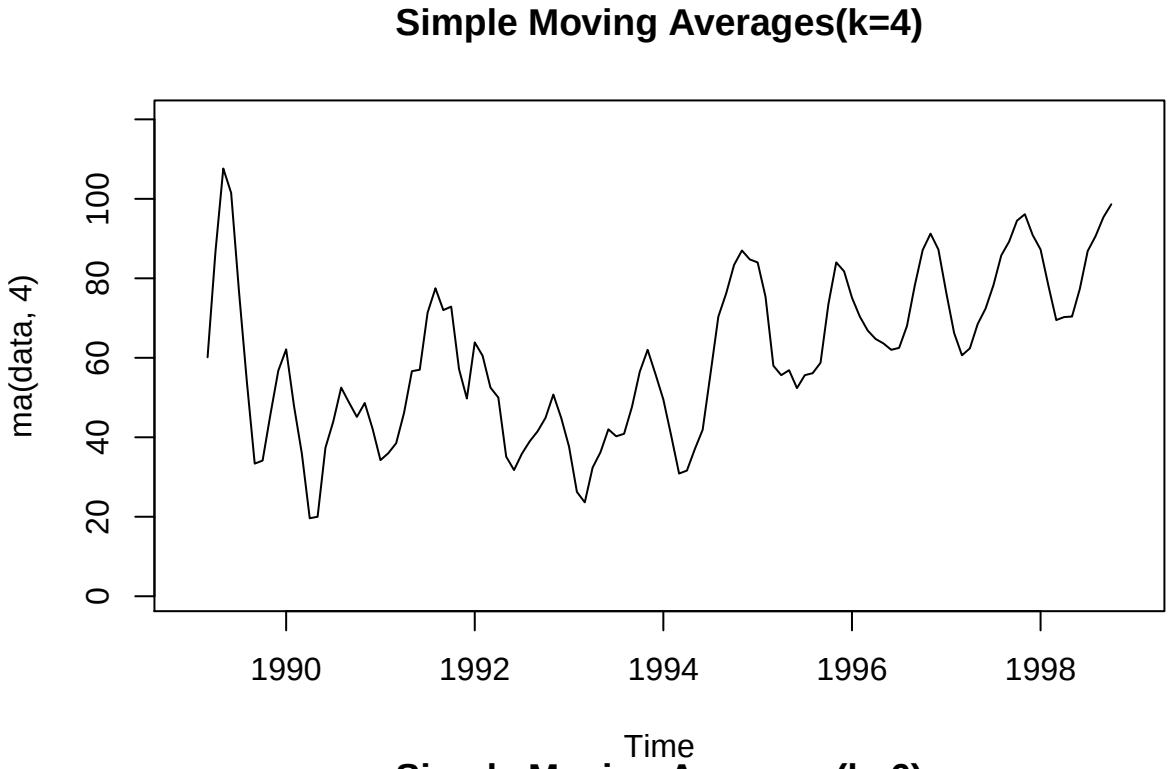
Data Description Report

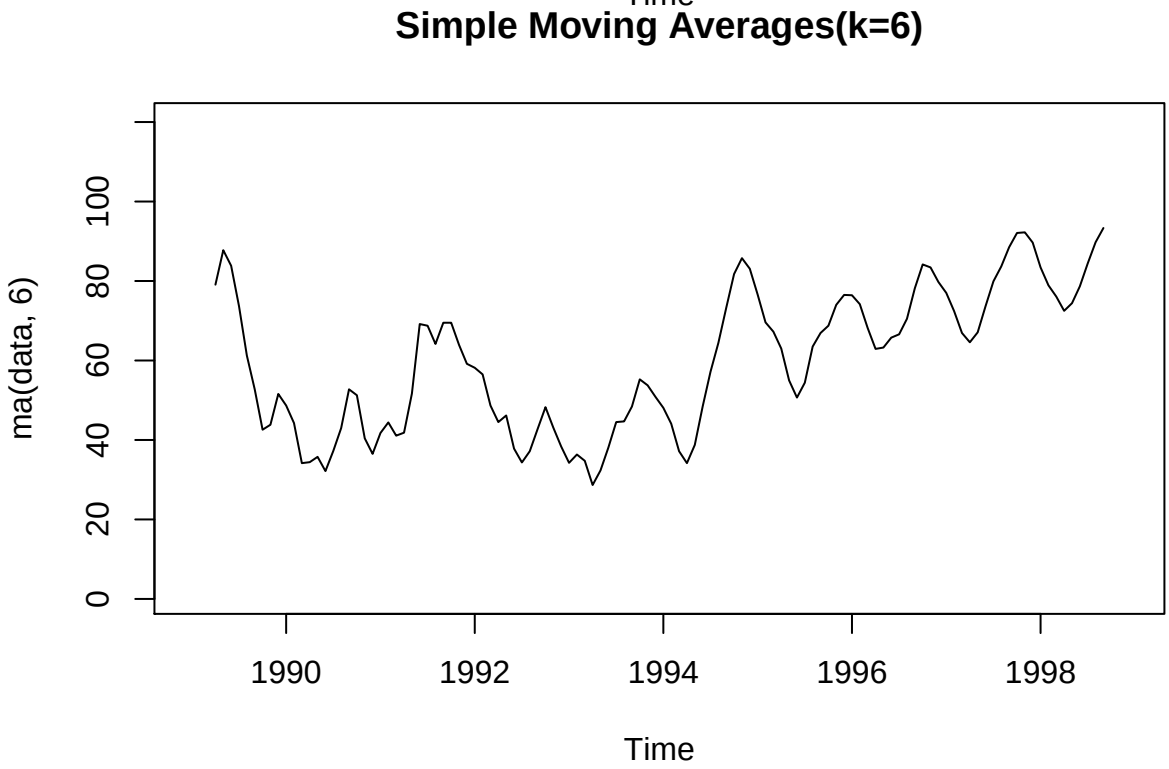
There are 2 ways to describe the data, but most descriptions focus on the quantity and quality of data. Data Quantity KEY CHARACTERISTICS -

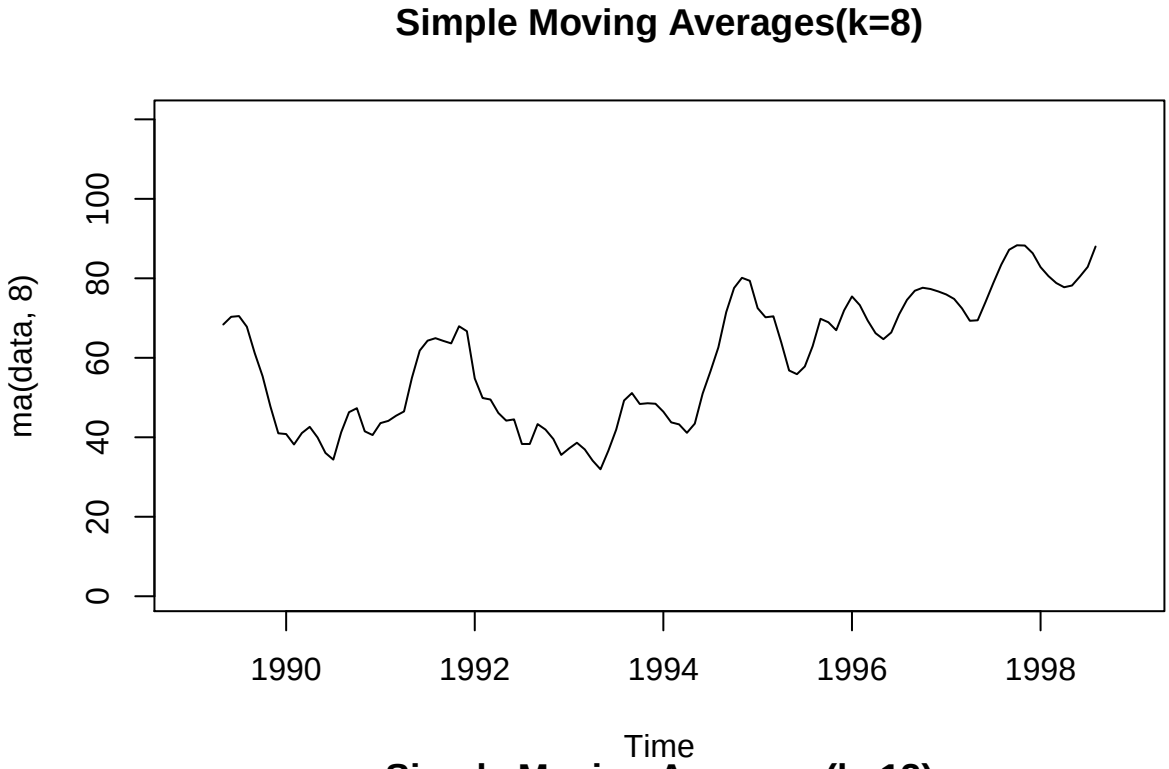


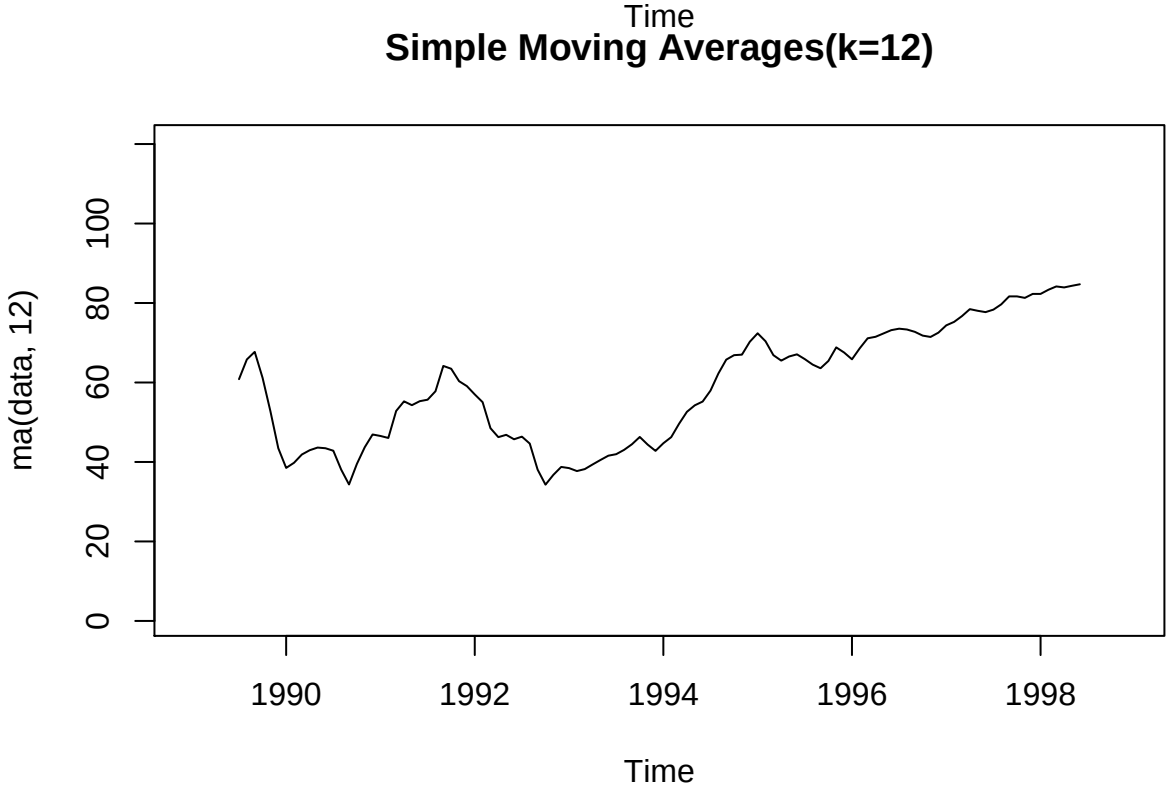
Data quality

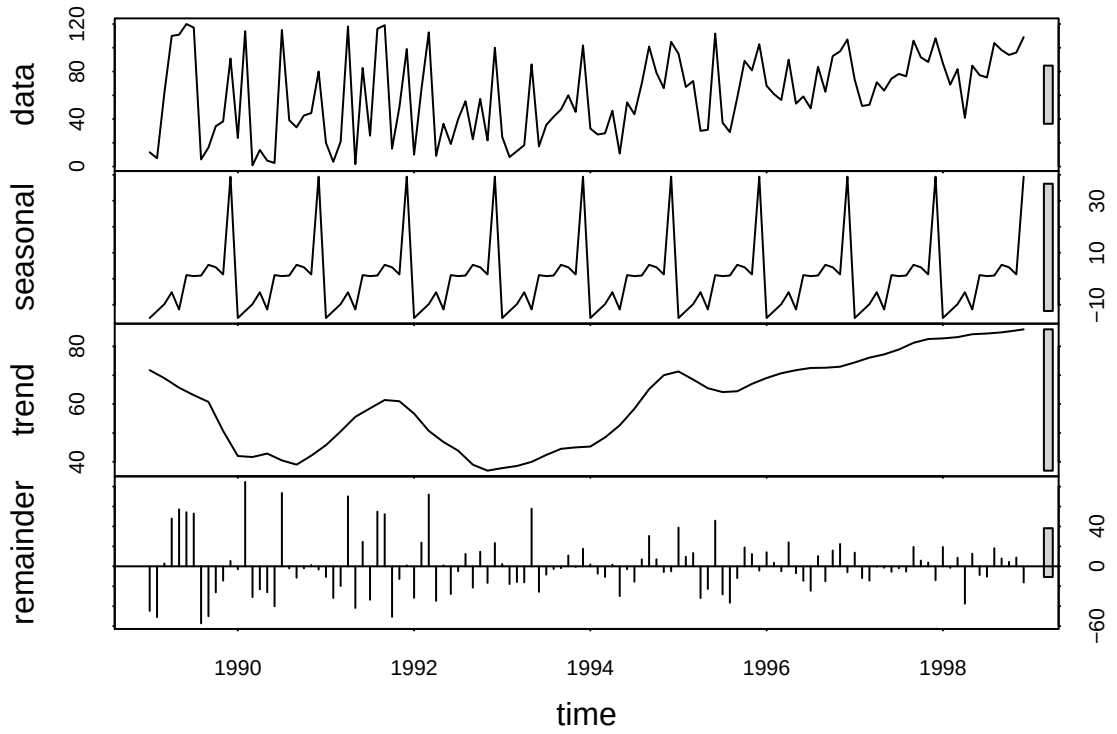
Does the data include characteristics relevant to the business question? Basic statistics / descriptive statistics for the data need to be computed and any insights need to be written here. Prioritizing the relevant attributes.


Simple Moving Averages(k=4)
ma(data, 4)
0
20
40
60
80
100
1990
1992
1994
1996
1998
Time


Simple Moving Averages(k=6)
ma(data, 6)
0
20
40
60
80
100
1990
1992
1994
1996
1998
Time


Simple Moving Averages(k=8)
ma(data, 8)
0
20
40
60
80
100
1990
1992
1994
1996
1998
Time

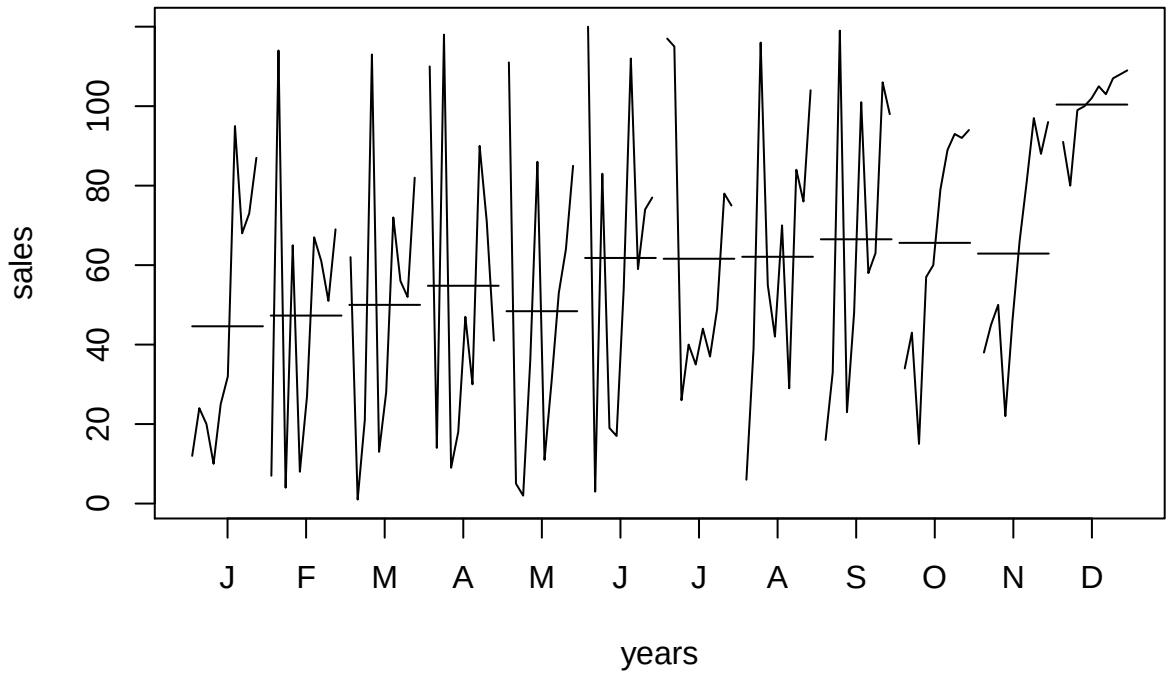

Simple Moving Averages(k=12)
ma(data, 12)
0
20
40
60
80
100
1990
1992
1994
1996
1998
Time



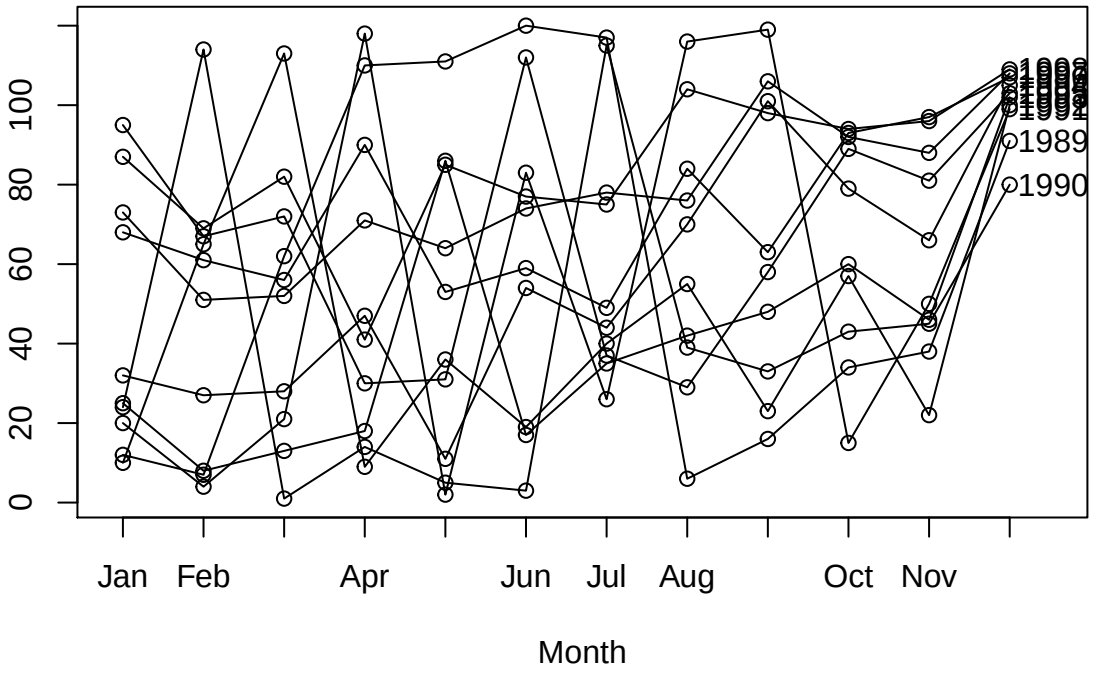

```
##            seasonal    trend   remainder
## Jan 1989 -15.107616 71.75595 -44.6483372
## Feb 1989 -12.443113 70.36411 -50.9209931
## Mar 1989  -9.778609 68.97226   2.8063495
## Apr 1989  -5.200350 67.32274  47.8776055
## May 1989 -11.822086 65.67323  57.1488565
## Jun 1989   1.392935 64.38724  54.2198222
## Jul 1989   1.007935 63.10126  52.8908083
## Aug 1989   1.231619 61.92883 -57.1604466
## Sep 1989   5.355271 60.75640 -50.1116691
## Oct 1989   4.414771 55.73193 -26.1467009
## Nov 1989   1.674262 50.70746 -14.3817241
## Dec 1989  39.274981 46.35642   5.3685955
## Jan 1990 -15.107616 42.00539  -2.8977692
## Feb 1990 -12.443113 41.83578  84.6073337
## Mar 1990  -9.778609 41.66617 -30.8875649
## Apr 1990  -5.200350 42.26026 -23.0599091
## May 1990 -11.822086 42.85434 -26.0322584
## Jun 1990   1.392935 41.67348 -40.0664124
## Jul 1990   1.007935 40.49261  73.4994538
## Aug 1990   1.231619 39.75256  -1.9841839
## Sep 1990   5.355271 39.01252 -11.3677894
## Oct 1990   4.414771 40.57676  -1.9915301
## Nov 1990   1.674262 42.14100   1.1847379
## Dec 1990  39.274981 43.95714  -3.2321234
## Jan 1991 -15.107616 45.77329 -10.6656691
## Feb 1991 -12.443113 48.14851 -31.7053990
## Mar 1991  -9.778609 50.52374 -19.7451303
## Apr 1991  -5.200350 53.05537  70.1449828
## May 1991 -11.822086 55.58699 -41.7649092
## Jun 1991   1.392935 57.06019  24.5468789
```

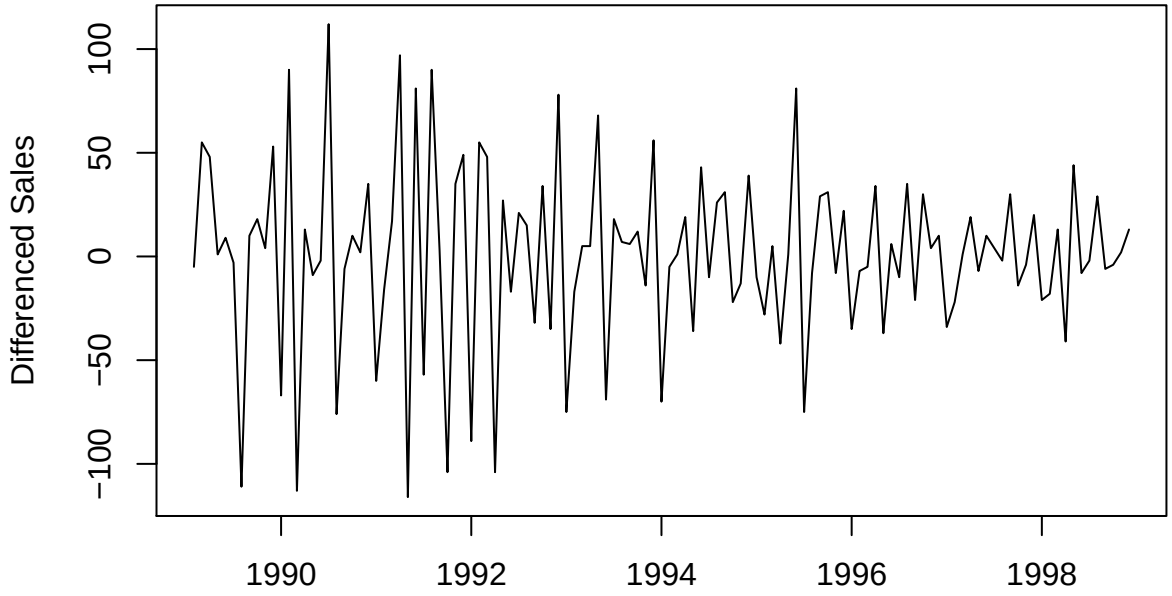
```
## Jul 1991   1.007935 58.53338 -33.5413127
## Aug 1991   1.231619 59.96965  54.7987301
## Sep 1991   5.355271 61.40592  52.2388051
## Oct 1991   4.414771 61.19323 -50.6080048
## Nov 1991   1.674262 60.98054 -12.6548061
## Dec 1991  39.274981 58.83713   0.8878854
## Jan 1992 -15.107616 56.69372 -31.5861075
## Feb 1992 -12.443113 53.71102  23.7320969
## Mar 1992  -9.778609 50.72831  72.0502998
## Apr 1992  -5.200350 48.79527 -34.5949156
## May 1992 -11.822086 46.86222   0.9598640
## Jun 1992   1.392935 45.35941 -27.7523400
## Jul 1992   1.007935 43.85659  -4.8645237
## Aug 1992   1.231619 41.43169  12.3366901
## Sep 1992   5.355271 39.00679 -21.3620638
## Oct 1992   4.414771 37.97325  14.6119823
## Nov 1992   1.674262 36.93970 -16.6139631
## Dec 1992  39.274981 37.38070  23.3443142
## Jan 1993 -15.107616 37.82171   2.2859070
## Feb 1993 -12.443113 38.19152 -17.7484028
## Mar 1993  -9.778609 38.56132 -15.7827141
## Apr 1993  -5.200350 39.27013 -16.0697821
## May 1993 -11.822086 39.97894  57.8431447
## Jun 1993   1.392935 41.18114 -25.5740736
## Jul 1993   1.007935 42.38334  -8.3912717
## Aug 1993   1.231619 43.44381  -2.6754291
## Sep 1993   5.355271 44.50428  -1.8595543
## Oct 1993   4.414771 44.75274  10.8324860
## Nov 1993   1.674262 45.00120  -0.6754651
## Dec 1993  39.274981 45.15061  17.5744109
## Jan 1994 -15.107616 45.30001   1.8076026
## Feb 1994 -12.443113 46.87508  -7.4319631
## Mar 1994  -9.778609 48.45014 -10.6715302
## Apr 1994  -5.200350 50.54921   1.6511428
## May 1994 -11.822086 52.64828 -29.8261893
## Jun 1994   1.392935 55.52424  -2.9171707
## Jul 1994   1.007935 58.40020 -15.4081318
## Aug 1994   1.231619 61.78776   6.9806209
## Sep 1994   5.355271 65.17532  30.4694059
## Oct 1994   4.414771 67.59060   6.9946249
## Nov 1994   1.674262 70.00589  -5.6801475
## Dec 1994  39.274981 70.65565  -4.9306296
## Jan 1995 -15.107616 71.30541  38.8022039
## Feb 1995 -12.443113 69.90133   9.5417820
## Mar 1995  -9.778609 68.49725  13.2813586
## Apr 1995  -5.200350 66.99222 -31.7918719
## May 1995 -11.822086 65.48719 -22.6651076
## Jun 1995   1.392935 64.82572  45.7813422
## Jul 1995   1.007935 64.16425 -28.1721876
## Aug 1995   1.231619 64.30115 -36.5327714
## Sep 1995   5.355271 64.43805 -11.7933228
## Oct 1995   4.414771 65.71709  18.8681364
## Nov 1995   1.674262 66.99613  12.3296043
## Dec 1995  39.274981 68.01513  -4.2901077
```

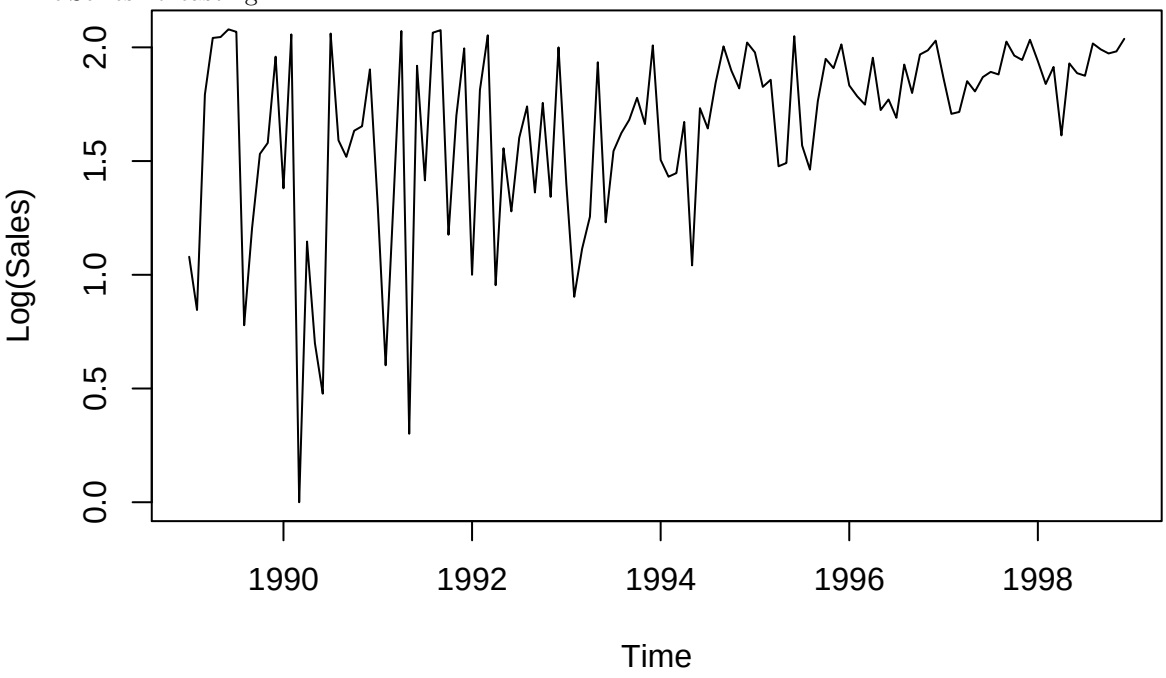
```
## Jan 1996 -15.107616 69.03412  14.0734959
## Feb 1996 -12.443113 69.86006   3.5830519
## Mar 1996  -9.778609 70.68600  -4.9073936
## Apr 1996  -5.200350 71.21545  23.9848997
## May 1996 -11.822086 71.74490  -6.9228120
## Jun 1996   1.392935 72.12970 -14.5226387
## Jul 1996   1.007935 72.51451 -24.5224450
## Aug 1996   1.231619 72.55322  10.2151578
## Sep 1996   5.355271 72.59194 -14.9472072
## Oct 1996   4.414771 72.76684  15.8183850
## Nov 1996   1.674262 72.94175  22.3839858
## Dec 1996  39.274981 73.68334  -5.9583182
## Jan 1997 -15.107616 74.42492  13.6826935
## Feb 1997 -12.443113 75.25688 -11.8137685
## Mar 1997  -9.778609 76.08884 -14.3102318
## Apr 1997  -5.200350 76.65113  -0.4507774
## May 1997 -11.822086 77.21341  -1.3913281
## Jun 1997   1.392935 78.05495  -5.4478892
## Jul 1997   1.007935 78.89650  -1.9044301
## Aug 1997   1.231619 80.05966  -5.2912760
## Sep 1997   5.355271 81.22282  19.4219102
## Oct 1997   4.414771 81.87923   5.7059992
## Nov 1997   1.674262 82.53564   3.7900969
## Dec 1997  39.274981 82.64819 -13.9231732
## Jan 1998 -15.107616 82.76074  19.3468723
## Feb 1998 -12.443113 82.97794  -1.5348302
## Mar 1998  -9.778609 83.19514   8.5834659
## Apr 1998  -5.200350 83.69559 -37.4952378
## May 1998 -11.822086 84.19603  12.6260535
## Jun 1998   1.392935 84.31995  -8.7128891
## Jul 1998   1.007935 84.44388 -10.4518115
## Aug 1998   1.231619 84.64399  18.1243885
## Sep 1998   5.355271 84.84411   7.8006207
## Oct 1998   4.414771 85.18009   4.4051347
## Nov 1998   1.674262 85.51608   8.8096572
## Dec 1998  39.274981 85.90161 -16.1765921
```

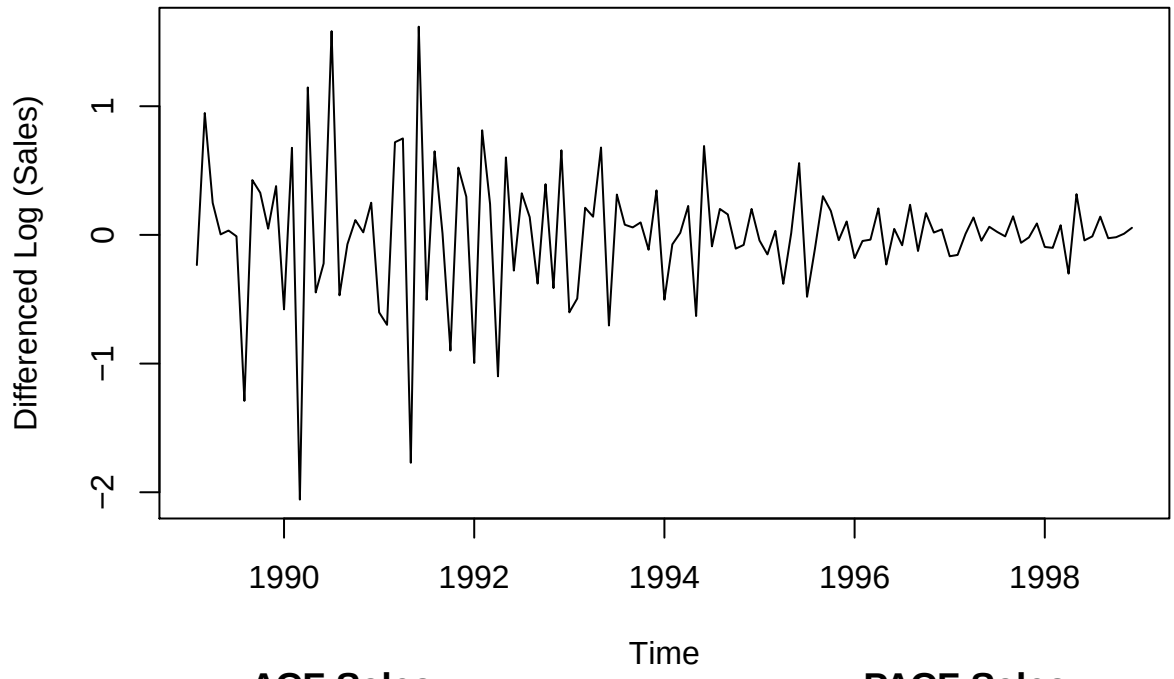

sales
0
20
40
60
80
100
J
F
M
A
M
J
J
A
S
O
N
D
years

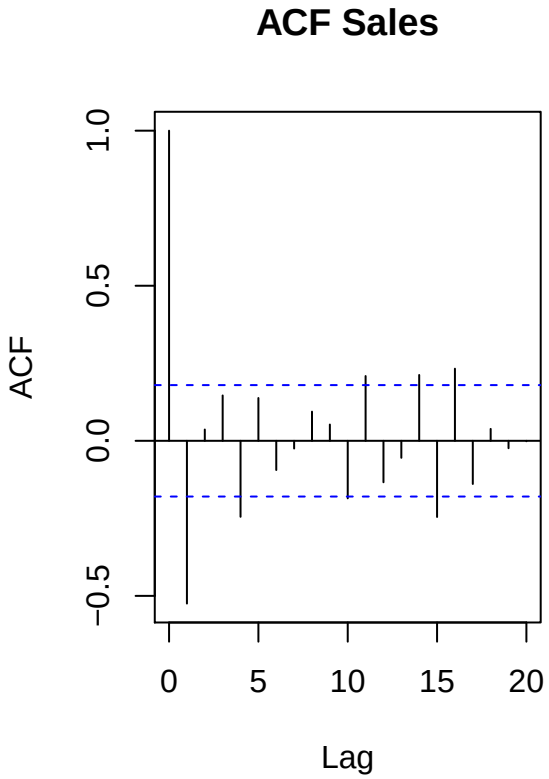
season plot

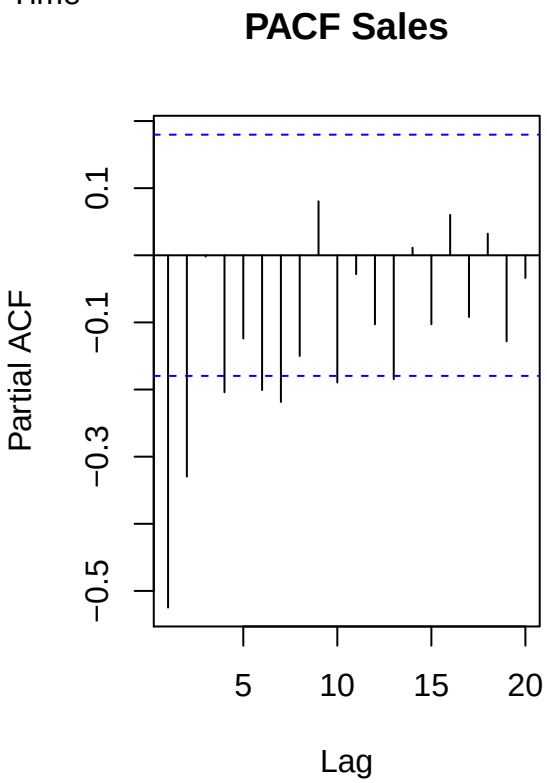

0
20
40
60
80
100
1989
1990
Jan
Feb
Apr
Jun
Jul
Aug
Oct
Nov
Month


Differenced Sales
100
50
0
−50
−100
1990
1992
1994
1996
1998
Time


Log(Sales)
0.0
0.5
1.0
1.5
2.0
1990
1992
1994
1996
1998
Time







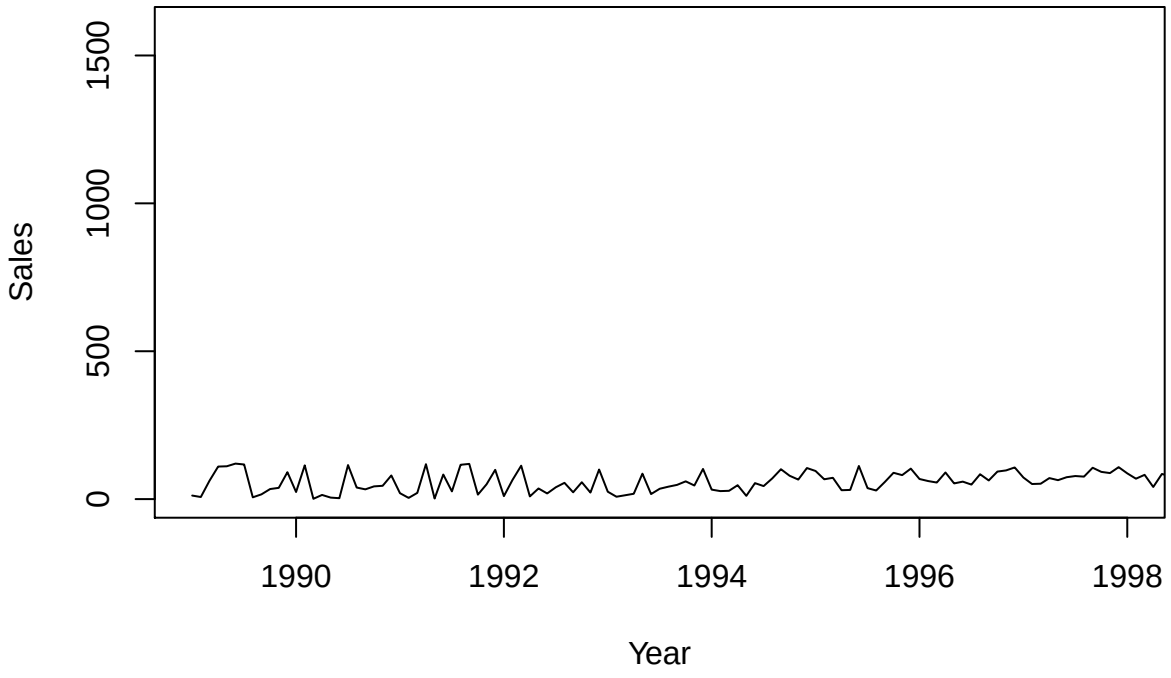
Data Exploration Report

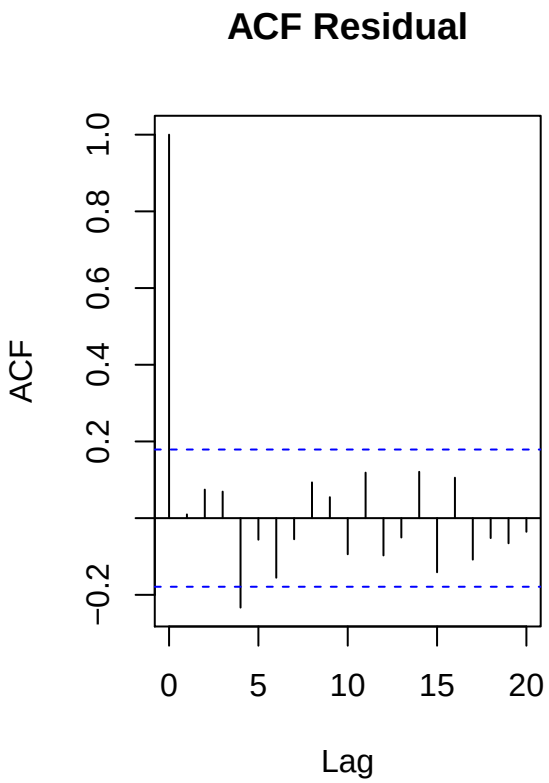
```
## Series: log10(data)
## ARIMA(0,1,1)
##
## Coefficients:
##          ma1
```

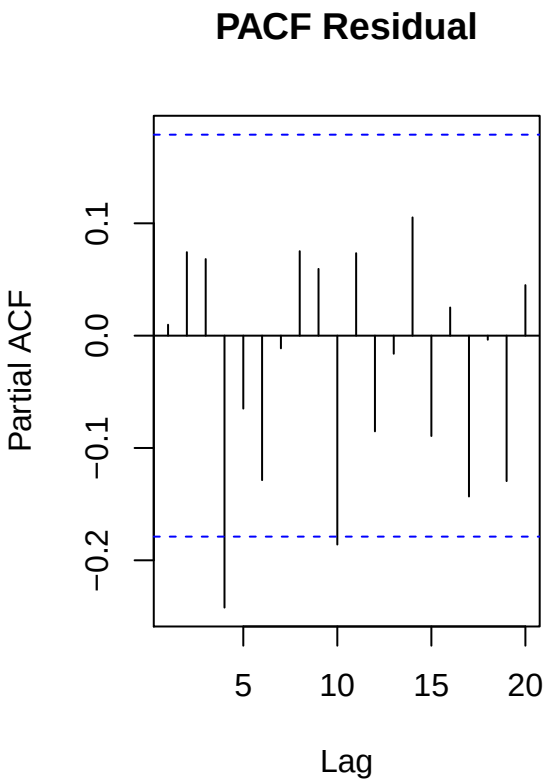
```
##       -0.9221
## s.e.   0.0308
##
## sigma^2 estimated as 0.1455:  log likelihood=-54.62
## AIC=113.24   AICc=113.34   BIC=118.8
##
## Training set error measures:
##                      ME      RMSE       MAE  MPE MAPE      MASE
## Training set 0.05294276 0.3782851 0.2756294 -Inf  Inf 0.7744427
##                     ACF1
## Training set 0.009724099
```

Verifying Data Quality

```
## $pred
##           Jan      Feb      Mar      Apr
## 1999 1.910576 1.910576 1.910576 1.910576
##
## $se
##            Jan       Feb       Mar       Apr
## 1999 0.3814774 0.3826341 0.3837872 0.3849369
```







The residuals are random implying that the model is fine.

D) Telecom Churn

Introduction

Every firm has competition, and the market of telecommunications faces a huge problem regarding the churn of customers. Customers needs and behavior changes a lot and because of that fact, a customer might leave a company for another. A telecom firm wants to resolve this issue of telecom churn.

Challenge

To analyse the data and profile probable “churners”.

Business Understanding

Background information:

The telecommunication companies face this problem of churning where a percentage of subscribers or customers switch to some other service and stop using the services of previous company after a certain period of time.

A company spends a lot of resources on each of its customers and employee, and hence wants to minimise the loss of customers and employee. This affects the overall profit of the company.

Problem Description

The company wants to reduce the number of churners and examine if a customer will churn or not.The company thus, wants to identify the profile of the churner or parameters that result in a defaulter.

Business Success Criteria

The model the profile of a churner based on certain parameters and past data.

```
## 'data.frame':    1000 obs. of  42 variables:
##  $ region  : num  2 3 3 2 2 2 3 2 3 1 ...
##  $ tenure  : num  13 11 68 33 23 41 45 38 45 68 ...
##  $ age     : num  44 33 52 33 30 39 22 35 59 41 ...
##  $ marital : num  1 1 1 0 1 0 1 0 1 1 ...
##  $ address : num  9 7 24 12 9 17 2 5 7 21 ...
##  $ income  : num  64 136 116 33 30 78 19 76 166 72 ...
##  $ ed      : num  4 5 1 2 1 2 2 2 4 1 ...
##  $ employ  : num  5 5 29 0 2 16 4 10 31 22 ...
##  $ retire  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gender  : num  0 0 1 1 0 1 1 0 0 0 ...
##  $ reside  : num  2 6 2 1 4 1 5 3 5 3 ...
##  $ tollfree: num  0 1 1 0 0 1 0 1 1 0 ...
##  $ equip   : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ callcard: num  1 1 1 0 0 1 1 1 1 1 ...
##  $ wireless: num  0 1 0 0 0 0 0 1 0 0 ...
##  $ longmon : num  3.7 4.4 18.15 9.45 6.3 ...
##  $ tollmon : num  0 20.8 18 0 0 ...
##  $ equipmon: num  0 0 0 0 0 0 0 50.1 0 0 ...
##  $ cardmon : num  7.5 15.2 30.2 0 0 ...
##  $ wiremon : num  0 35.7 0 0 0 0 0 64.9 0 0 ...
##  $ longten : num  37.5 42 1300.6 288.8 157.1 ...
##  $ tollten : num  0 211 1247 0 0 ...
##  $ equipten: num  0 0 0 0 0 ...
##  $ cardten : num  110 125 2150 0 0 ...
##  $ wireten : num  0 380 0 0 0 ...
##  $ multline: num  0 0 0 0 0 0 1 1 1 1 ...
##  $ voice   : num  0 1 0 0 0 0 0 1 0 0 ...
##  $ pager   : num  0 1 0 0 0 0 0 1 0 0 ...
##  $ internet: num  0 0 0 0 0 0 1 1 0 0 ...
##  $ callid  : num  0 1 1 0 1 1 0 1 1 0 ...
##  $ callwait: num  0 1 1 0 0 1 1 1 1 0 ...
```

```
##  $ forward : num  1 1 0 0 1 0 0 1 1 0 ...
##  $ confer  : num  0 1 1 0 1 0 0 1 1 0 ...
##  $ ebill   : num  0 0 0 0 0 0 1 1 0 0 ...
##  $ loglong : num  1.31 1.48 2.9 2.25 1.84 ...
##  $ logtoll : num  2.01 3.03 2.89 3.5 3.4 ...
##  $ logequi : num  4.16 2.72 3.41 1 3 ...
##  $ logcard : num  1 3.58 4.75 1 0 ...
##  $ logwire : num  1 4.91 3 NA NA ...
##  $ lninc   : num  NA 4 0 NA NA ...
##  $ custcat : num  NA 1 NA NA NA NA NA 4 NA NA ...
##  $ churn   : num  NA NA NA NA NA NA NA 0 NA NA ...
```

Summary after data cleaning

```
## [1] 869
```

```
##      region          tenure           age           marital
##  Min.   :1.000   Min.   : 1.00   Min.   :18.00   Min.   :0.000
##  1st Qu.:1.000   1st Qu.:17.00   1st Qu.:32.00   1st Qu.:0.000
##  Median :2.000   Median :34.00   Median :40.00   Median :0.000
##  Mean   :2.022   Mean   :35.53   Mean   :41.68   Mean   :0.495
##  3rd Qu.:3.000   3rd Qu.:54.00   3rd Qu.:51.00   3rd Qu.:1.000
##  Max.   :3.000   Max.   :72.00   Max.   :77.00   Max.   :1.000
##     address          income             ed            employ
##  Min.   : 0.00   Min.   :   9.00   Min.   :1.000   Min.   : 0.00
##  1st Qu.: 3.00   1st Qu.:  29.00   1st Qu.:2.000   1st Qu.: 3.00
##  Median : 9.00   Median :  47.00   Median :3.000   Median : 8.00
##  Mean   :11.55   Mean   :  77.53   Mean   :2.671   Mean   :10.99
##  3rd Qu.:18.00   3rd Qu.:  83.00   3rd Qu.:4.000   3rd Qu.:17.00
##  Max.   :55.00   Max.   :1668.00   Max.   :5.000   Max.   :47.00
##      retire          gender          reside         tollfree
##  Min.   :0.000   Min.   :0.000   Min.   :1.000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:1.000   1st Qu.:0.000
##  Median :0.000   Median :1.000   Median :2.000   Median :0.000
##  Mean   :0.047   Mean   :0.517   Mean   :2.331   Mean   :0.474
##  3rd Qu.:0.000   3rd Qu.:1.000   3rd Qu.:3.000   3rd Qu.:1.000
##  Max.   :1.000   Max.   :1.000   Max.   :8.000   Max.   :1.000
##      equip          callcard        wireless        longmon
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   : 0.900
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.: 5.200
##  Median :0.000   Median :1.000   Median :0.000   Median : 8.525
##  Mean   :0.386   Mean   :0.678   Mean   :0.296   Mean   :11.723
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:14.412
##  Max.   :1.000   Max.   :1.000   Max.   :1.000   Max.   :99.950
##     tollmon         equipmon        cardmon         wiremon
##  Min.   :  0.00   Min.   : 0.00   Min.   :  0.00   Min.   :  0.00
##  1st Qu.:  0.00   1st Qu.: 0.00   1st Qu.:  0.00   1st Qu.:  0.00
##  Median :  0.00   Median : 0.00   Median : 12.00   Median :  0.00
##  Mean   : 13.27   Mean   :14.22   Mean   : 13.78   Mean   : 11.58
##  3rd Qu.: 24.25   3rd Qu.:31.48   3rd Qu.: 20.50   3rd Qu.: 24.71
##  Max.   :173.00   Max.   :77.70   Max.   :109.25   Max.   :111.95
##     longten          tollten         equipten         cardten
##  Min.   :   0.90   Min.   :   0.0   Min.   :   0.0   Min.   :   0.0
##  1st Qu.:  90.14   1st Qu.:   0.0   1st Qu.:   0.0   1st Qu.:   0.0
```

```
##  Median : 285.48   Median :   0.0   Median :   0.0   Median : 332.5
##  Mean   : 574.05   Mean   : 551.3   Mean   : 465.6   Mean   : 605.8
##  3rd Qu.: 755.02   3rd Qu.: 846.9   3rd Qu.: 579.5   3rd Qu.: 910.0
##  Max.   :7257.60   Max.   :5916.0   Max.   :5028.6   Max.   :7515.0
##     wireten          multline         voice           pager
##  Min.   :   0.0   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:   0.0   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000
##  Median :   0.0   Median :0.000   Median :0.000   Median :0.000
##  Mean   : 442.7   Mean   :0.475   Mean   :0.304   Mean   :0.261
##  3rd Qu.: 316.5   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :7856.9   Max.   :1.000   Max.   :1.000   Max.   :1.000
##     internet         callid         callwait        forward
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000
##  Median :0.000   Median :0.000   Median :0.000   Median :0.000
##  Mean   :0.368   Mean   :0.481   Mean   :0.485   Mean   :0.493
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :1.000   Max.   :1.000   Max.   :1.000   Max.   :1.000
##      confer          ebill          loglong           logtoll
##  Min.   :0.000   Min.   :0.000   Min.   :-0.1054   Min.   :1.504
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.: 1.6487   1st Qu.:2.970
##  Median :1.000   Median :0.000   Median : 2.1430   Median :3.258
##  Mean   :0.502   Mean   :0.371   Mean   : 2.1821   Mean   :3.253
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.: 2.6681   3rd Qu.:3.549
##  Max.   :1.000   Max.   :1.000   Max.   : 4.6047   Max.   :6.375
##     logequi         logcard         logwire          lninc
##  Min.   :1.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:2.657   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :3.264   Median :2.998   Median :2.257   Median :1.985
##  Mean   :3.137   Mean   :2.592   Mean   :2.257   Mean   :1.985
##  3rd Qu.:3.761   3rd Qu.:3.657   3rd Qu.:3.466   3rd Qu.:3.000
##  Max.   :7.419   Max.   :6.833   Max.   :7.031   Max.   :6.850
##     custcat         churn
##  Min.   :0.00   Min.   :0.000
##  1st Qu.:1.94   1st Qu.:0.374
##  Median :1.94   Median :0.374
##  Mean   :1.94   Mean   :0.374
##  3rd Qu.:1.94   3rd Qu.:0.374
##  Max.   :4.00   Max.   :1.000
```

```
## [1] "formula"
```

```
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + reside + tollfree + equip +
##     callcard + wireless + longmon + tollmon + equipmon + cardmon +
##     wiremon + longten + tollten + equipten + cardten + wireten +
##     multline + voice + pager + internet + callid + callwait +
##     forward + confer + ebill + loglong + logtoll + logequi +
##     logcard + logwire + lninc + custcat
```

```
##
## Call:
## glm(formula = frmla, family = "binomial", data = Trngdt)
##
## Deviance Residuals:
```

```
##      Min        1Q    Median        3Q       Max
## -1.38551  -0.06725  -0.00454   0.06121   1.86458
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.136e-01  1.389e+00  -0.514    0.608
## region      -4.416e-02  9.959e-02  -0.443    0.657
## tenure       6.438e-03  1.425e-02   0.452    0.651
## age          1.902e-03  1.116e-02   0.170    0.865
## marital      8.346e-03  2.147e-01   0.039    0.969
## address     -6.401e-03  1.142e-02  -0.561    0.575
## income       3.204e-04  1.156e-03   0.277    0.782
## ed          -1.221e-02  8.428e-02  -0.145    0.885
## employ      -3.876e-03  1.402e-02  -0.276    0.782
## retire       1.280e-01  4.861e-01   0.263    0.792
## gender       1.059e-01  1.634e-01   0.648    0.517
## reside       2.859e-03  7.402e-02   0.039    0.969
## tollfree    -4.724e-02  3.722e-01  -0.127    0.899
## equip        2.798e-01  7.723e-01   0.362    0.717
## callcard     9.469e-02  3.375e-01   0.281    0.779
## wireless    -2.374e-01  5.346e-01  -0.444    0.657
## longmon     -3.179e-02  1.009e-01  -0.315    0.753
## tollmon     -5.647e-04  1.193e-02  -0.047    0.962
## equipmon     5.279e-03  2.695e-02   0.196    0.845
## cardmon      1.534e-04  2.081e-02   0.007    0.994
## wiremon      1.566e-02  2.142e-02   0.731    0.465
## longten      3.899e-04  1.156e-03   0.337    0.736
## tollten     -6.744e-05  2.243e-04  -0.301    0.764
## equipten    -4.576e-04  2.820e-04  -1.623    0.105
## cardten     -3.969e-05  3.494e-04  -0.114    0.910
## wireten     -2.757e-04  2.986e-04  -0.923    0.356
## multline    -5.713e-04  2.101e-01  -0.003    0.998
## voice        5.697e-03  2.493e-01   0.023    0.982
## pager        9.012e-02  2.747e-01   0.328    0.743
## internet    -8.047e-02  2.352e-01  -0.342    0.732
## callid       8.829e-02  2.491e-01   0.354    0.723
## callwait    -1.060e-02  2.460e-01  -0.043    0.966
## forward      5.062e-02  2.398e-01   0.211    0.833
## confer      -3.134e-02  2.342e-01  -0.134    0.894
## ebill        7.829e-02  2.300e-01   0.340    0.734
## loglong      2.618e-02  5.637e-01   0.046    0.963
## logtoll      7.183e-03  2.367e-01   0.030    0.976
## logequi     -3.950e-02  1.626e-01  -0.243    0.808
## logcard      8.545e-03  1.433e-01   0.060    0.952
## logwire      2.484e-02  1.409e-01   0.176    0.860
## lninc        1.600e-02  1.328e-01   0.121    0.904
## custcat      4.884e-02  1.323e-01   0.369    0.712
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 109.691  on 697  degrees of freedom
## Residual deviance:  96.242  on 656  degrees of freedom
## AIC: 752.6
##
```

```
## Number of Fisher Scoring iterations: 4
```

```
## Start:  AIC=752.6
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + reside + tollfree + equip +
##     callcard + wireless + longmon + tollmon + equipmon + cardmon +
##     wiremon + longten + tollten + equipten + cardten + wireten +
##     multline + voice + pager + internet + callid + callwait +
##     forward + confer + ebill + loglong + logtoll + logequi +
##     logcard + logwire + lninc + custcat
##
##             Df Deviance    AIC
## - multline  1   96.242 750.60
## - cardmon   1   96.242 750.60
## - voice     1   96.242 750.60
## - logtoll   1   96.243 750.60
## - reside    1   96.243 750.60
## - marital   1   96.243 750.60
## - callwait  1   96.244 750.60
## - loglong   1   96.244 750.61
## - tollmon   1   96.244 750.61
## - logcard   1   96.245 750.61
## - cardten   1   96.255 750.62
## - lninc     1   96.256 750.62
## - tollfree  1   96.258 750.62
## - confer    1   96.260 750.62
## - ed        1   96.263 750.62
## - age       1   96.271 750.63
## - logwire   1   96.273 750.63
## - equipmon  1   96.280 750.64
## - forward   1   96.286 750.65
## - logequi   1   96.301 750.66
## - retire    1   96.311 750.67
## - income    1   96.318 750.68
## - employ    1   96.318 750.68
## - callcard  1   96.321 750.68
## - tollten   1   96.332 750.69
## - longmon   1   96.341 750.70
## - pager     1   96.349 750.71
## - longten   1   96.356 750.72
## - ebill     1   96.358 750.72
## - internet  1   96.359 750.72
## - callid    1   96.368 750.73
## - equip     1   96.373 750.73
## - custcat   1   96.378 750.74
## - region    1   96.439 750.80
## - wireless  1   96.440 750.80
## - tenure    1   96.446 750.81
## - address   1   96.556 750.92
## - gender    1   96.663 751.02
## - wiremon   1   96.776 751.14
## - wireten   1   97.098 751.46
## <none>          96.242 752.60
## - equipten  1   98.901 753.26
```

```
##
## Step:  AIC=750.6
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + reside + tollfree + equip +
##     callcard + wireless + longmon + tollmon + equipmon + cardmon +
##     wiremon + longten + tollten + equipten + cardten + wireten +
##     voice + pager + internet + callid + callwait + forward +
##     confer + ebill + loglong + logtoll + logequi + logcard +
##     logwire + lninc + custcat
##
##            Df Deviance    AIC
## - cardmon   1   96.242 748.60
## - voice     1   96.242 748.60
## - logtoll   1   96.243 748.60
## - reside    1   96.243 748.60
## - marital   1   96.243 748.60
## - callwait  1   96.244 748.60
## - loglong   1   96.244 748.61
## - tollmon   1   96.244 748.61
## - logcard   1   96.246 748.61
## - cardten   1   96.255 748.62
## - lninc     1   96.257 748.62
## - tollfree  1   96.258 748.62
## - confer    1   96.260 748.62
## - ed        1   96.263 748.62
## - age       1   96.271 748.63
## - logwire   1   96.275 748.64
## - equipmon  1   96.280 748.64
## - forward   1   96.287 748.65
## - logequi   1   96.305 748.67
## - retire    1   96.312 748.67
## - employ    1   96.319 748.68
## - income    1   96.320 748.68
## - callcard  1   96.321 748.68
## - tollten   1   96.332 748.69
## - longmon   1   96.341 748.70
## - pager     1   96.349 748.71
## - longten   1   96.356 748.72
## - ebill     1   96.358 748.72
## - internet  1   96.359 748.72
## - callid    1   96.370 748.73
## - equip     1   96.374 748.74
## - custcat   1   96.381 748.74
## - region    1   96.439 748.80
## - wireless  1   96.440 748.80
## - tenure    1   96.449 748.81
## - address   1   96.556 748.92
## - gender    1   96.663 749.02
## - wiremon   1   96.778 749.14
## - wireten   1   97.099 749.46
## <none>          96.242 750.60
## - equipten  1   98.906 751.27
##
## Step:  AIC=748.6
```

```
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + reside + tollfree + equip +
##     callcard + wireless + longmon + tollmon + equipmon + wiremon +
##     longten + tollten + equipten + cardten + wireten + voice +
##     pager + internet + callid + callwait + forward + confer +
##     ebill + loglong + logtoll + logequi + logcard + logwire +
##     lninc + custcat
##
##            Df Deviance    AIC
## - voice     1   96.243 746.60
## - logtoll   1   96.243 746.60
## - reside    1   96.243 746.60
## - marital   1   96.244 746.60
## - callwait  1   96.244 746.61
## - loglong   1   96.244 746.61
## - tollmon   1   96.244 746.61
## - logcard   1   96.246 746.61
## - lninc     1   96.258 746.62
## - tollfree  1   96.258 746.62
## - confer    1   96.260 746.62
## - ed        1   96.263 746.62
## - age       1   96.271 746.63
## - logwire   1   96.277 746.64
## - equipmon  1   96.281 746.64
## - forward   1   96.287 746.65
## - cardten   1   96.289 746.65
## - logequi   1   96.310 746.67
## - retire    1   96.312 746.67
## - income    1   96.320 746.68
## - employ    1   96.321 746.68
## - tollten   1   96.333 746.69
## - longmon   1   96.342 746.70
## - pager     1   96.350 746.71
## - longten   1   96.357 746.72
## - ebill     1   96.358 746.72
## - callcard  1   96.360 746.72
## - internet  1   96.360 746.72
## - callid    1   96.370 746.73
## - equip     1   96.374 746.74
## - custcat   1   96.388 746.75
## - region    1   96.439 746.80
## - wireless  1   96.440 746.80
## - tenure    1   96.460 746.82
## - address   1   96.557 746.92
## - gender    1   96.666 747.03
## - wiremon   1   96.778 747.14
## - wireten   1   97.106 747.47
## <none>          96.242 748.60
## - equipten  1   98.925 749.29
##
## Step:  AIC=746.6
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + reside + tollfree + equip +
##     callcard + wireless + longmon + tollmon + equipmon + wiremon +
```

```
##     longten + tollten + equipten + cardten + wireten + pager +
##     internet + callid + callwait + forward + confer + ebill +
##     loglong + logtoll + logequi + logcard + logwire + lninc +
##     custcat
##
##            Df Deviance    AIC
## - logtoll   1   96.244 744.60
## - marital   1   96.244 744.60
## - reside    1   96.244 744.60
## - callwait  1   96.244 744.60
## - loglong   1   96.245 744.60
## - tollmon   1   96.245 744.60
## - logcard   1   96.247 744.60
## - tollfree  1   96.259 744.61
## - lninc     1   96.259 744.61
## - confer    1   96.260 744.61
## - ed        1   96.264 744.62
## - age       1   96.272 744.62
## - logwire   1   96.279 744.63
## - equipmon  1   96.282 744.63
## - forward   1   96.288 744.64
## - cardten   1   96.290 744.64
## - logequi   1   96.310 744.66
## - retire    1   96.313 744.67
## - income    1   96.321 744.67
## - employ    1   96.324 744.68
## - tollten   1   96.334 744.69
## - longmon   1   96.343 744.70
## - pager     1   96.351 744.70
## - longten   1   96.358 744.71
## - ebill     1   96.361 744.71
## - callcard  1   96.361 744.71
## - internet  1   96.361 744.71
## - callid    1   96.371 744.72
## - equip     1   96.374 744.73
## - custcat   1   96.390 744.74
## - region    1   96.440 744.79
## - wireless  1   96.440 744.79
## - tenure    1   96.460 744.81
## - address   1   96.557 744.91
## - gender    1   96.668 745.02
## - wiremon   1   96.783 745.14
## - wireten   1   97.106 745.46
## <none>          96.243 746.60
## - equipten  1   98.938 747.29
##
## Step:  AIC=744.6
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + reside + tollfree + equip +
##     callcard + wireless + longmon + tollmon + equipmon + wiremon +
##     longten + tollten + equipten + cardten + wireten + pager +
##     internet + callid + callwait + forward + confer + ebill +
##     loglong + logequi + logcard + logwire + lninc + custcat
##
```

```
##            Df Deviance    AIC
## - tollmon   1   96.245 742.60
## - callwait  1   96.245 742.60
## - reside    1   96.245 742.60
## - marital   1   96.245 742.60
## - loglong   1   96.246 742.60
## - logcard   1   96.247 742.60
## - lninc     1   96.260 742.61
## - confer    1   96.260 742.61
## - ed        1   96.264 742.62
## - tollfree  1   96.266 742.62
## - age       1   96.274 742.63
## - logwire   1   96.280 742.63
## - equipmon  1   96.285 742.64
## - forward   1   96.291 742.64
## - cardten   1   96.294 742.65
## - retire    1   96.313 742.66
## - employ    1   96.324 742.68
## - income    1   96.326 742.68
## - logequi   1   96.334 742.69
## - tollten   1   96.335 742.69
## - longmon   1   96.343 742.70
## - pager     1   96.353 742.70
## - longten   1   96.358 742.71
## - internet  1   96.361 742.71
## - ebill     1   96.363 742.71
## - callcard  1   96.368 742.72
## - equip     1   96.376 742.73
## - callid    1   96.376 742.73
## - custcat   1   96.391 742.74
## - wireless  1   96.441 742.79
## - region    1   96.446 742.80
## - tenure    1   96.463 742.81
## - address   1   96.561 742.91
## - gender    1   96.671 743.02
## - wiremon   1   96.794 743.15
## - wireten   1   97.133 743.48
## <none>          96.244 744.60
## - equipten  1   98.946 745.30
##
## Step:  AIC=742.6
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + reside + tollfree + equip +
##     callcard + wireless + longmon + equipmon + wiremon + longten +
##     tollten + equipten + cardten + wireten + pager + internet +
##     callid + callwait + forward + confer + ebill + loglong +
##     logequi + logcard + logwire + lninc + custcat
##
##            Df Deviance    AIC
## - reside    1   96.247 740.60
## - marital   1   96.247 740.60
## - callwait  1   96.247 740.60
## - loglong   1   96.248 740.60
## - logcard   1   96.249 740.60
```

```
## - lninc     1   96.261 740.61
## - confer    1   96.262 740.61
## - ed        1   96.265 740.62
## - tollfree  1   96.273 740.62
## - age       1   96.275 740.63
## - logwire   1   96.282 740.63
## - equipmon  1   96.287 740.64
## - forward   1   96.292 740.64
## - cardten   1   96.295 740.65
## - retire    1   96.314 740.66
## - employ    1   96.325 740.68
## - income    1   96.327 740.68
## - logequi   1   96.336 740.69
## - longmon   1   96.346 740.70
## - pager     1   96.354 740.70
## - longten   1   96.361 740.71
## - internet  1   96.363 740.71
## - ebill     1   96.366 740.72
## - callcard  1   96.368 740.72
## - equip     1   96.376 740.73
## - callid    1   96.377 740.73
## - custcat   1   96.393 740.74
## - tollten   1   96.431 740.78
## - wireless  1   96.442 740.79
## - region    1   96.449 740.80
## - tenure    1   96.467 740.82
## - address   1   96.561 740.91
## - gender    1   96.671 741.02
## - wiremon   1   96.797 741.15
## - wireten   1   97.144 741.49
## <none>          96.245 742.60
## - equipten  1   98.985 743.34
##
## Step:  AIC=740.6
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + tollfree + equip + callcard +
##     wireless + longmon + equipmon + wiremon + longten + tollten +
##     equipten + cardten + wireten + pager + internet + callid +
##     callwait + forward + confer + ebill + loglong + logequi +
##     logcard + logwire + lninc + custcat
##
##            Df Deviance    AIC
## - callwait  1   96.248 738.60
## - loglong   1   96.249 738.60
## - logcard   1   96.250 738.61
## - marital   1   96.253 738.61
## - lninc     1   96.263 738.62
## - confer    1   96.263 738.62
## - ed        1   96.267 738.62
## - tollfree  1   96.275 738.63
## - age       1   96.275 738.63
## - logwire   1   96.283 738.64
## - equipmon  1   96.289 738.64
## - forward   1   96.293 738.65
```

```
## - cardten   1   96.296 738.65
## - retire    1   96.317 738.67
## - employ    1   96.326 738.68
## - income    1   96.329 738.68
## - logequi   1   96.337 738.69
## - longmon   1   96.347 738.70
## - pager     1   96.355 738.71
## - longten   1   96.362 738.72
## - internet  1   96.364 738.72
## - ebill     1   96.368 738.72
## - callcard  1   96.369 738.72
## - equip     1   96.377 738.73
## - callid    1   96.378 738.73
## - custcat   1   96.393 738.75
## - tollten   1   96.431 738.79
## - wireless  1   96.445 738.80
## - region    1   96.452 738.81
## - tenure    1   96.470 738.82
## - address   1   96.568 738.92
## - gender    1   96.673 739.03
## - wiremon   1   96.805 739.16
## - wireten   1   97.163 739.52
## <none>          96.247 740.60
## - equipten  1   98.987 741.34
##
## Step:  AIC=738.6
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + tollfree + equip + callcard +
##     wireless + longmon + equipmon + wiremon + longten + tollten +
##     equipten + cardten + wireten + pager + internet + callid +
##     forward + confer + ebill + loglong + logequi + logcard +
##     logwire + lninc + custcat
##
##            Df Deviance    AIC
## - loglong   1   96.251 736.61
## - logcard   1   96.251 736.61
## - marital   1   96.255 736.61
## - lninc     1   96.264 736.62
## - confer    1   96.266 736.62
## - ed        1   96.268 736.62
## - age       1   96.278 736.63
## - tollfree  1   96.278 736.63
## - logwire   1   96.283 736.64
## - equipmon  1   96.290 736.65
## - forward   1   96.293 736.65
## - cardten   1   96.296 736.65
## - retire    1   96.317 736.67
## - employ    1   96.328 736.68
## - income    1   96.331 736.69
## - logequi   1   96.344 736.70
## - longmon   1   96.351 736.71
## - pager     1   96.356 736.71
## - internet  1   96.365 736.72
## - longten   1   96.366 736.72
```

```
## - callcard  1   96.372 736.73
## - ebill     1   96.372 736.73
## - callid    1   96.378 736.73
## - equip     1   96.381 736.74
## - custcat   1   96.393 736.75
## - tollten   1   96.433 736.79
## - wireless  1   96.447 736.80
## - region    1   96.455 736.81
## - tenure    1   96.470 736.83
## - address   1   96.571 736.93
## - gender    1   96.677 737.03
## - wiremon   1   96.808 737.16
## - wireten   1   97.169 737.52
## <none>          96.248 738.60
## - equipten  1   98.988 739.34
##
## Step:  AIC=736.6
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + tollfree + equip + callcard +
##     wireless + longmon + equipmon + wiremon + longten + tollten +
##     equipten + cardten + wireten + pager + internet + callid +
##     forward + confer + ebill + logequi + logcard + logwire +
##     lninc + custcat
##
##            Df Deviance    AIC
## - logcard   1   96.254 734.61
## - marital   1   96.257 734.61
## - lninc     1   96.266 734.62
## - confer    1   96.268 734.62
## - ed        1   96.269 734.62
## - age       1   96.280 734.63
## - tollfree  1   96.280 734.63
## - logwire   1   96.286 734.64
## - equipmon  1   96.294 734.65
## - forward   1   96.295 734.65
## - cardten   1   96.297 734.65
## - retire    1   96.319 734.67
## - employ    1   96.329 734.68
## - income    1   96.334 734.69
## - logequi   1   96.347 734.70
## - pager     1   96.357 734.71
## - internet  1   96.367 734.72
## - callcard  1   96.372 734.73
## - ebill     1   96.374 734.73
## - callid    1   96.381 734.73
## - equip     1   96.381 734.74
## - custcat   1   96.397 734.75
## - tollten   1   96.435 734.79
## - wireless  1   96.447 734.80
## - region    1   96.457 734.81
## - longten   1   96.556 734.91
## - address   1   96.571 734.93
## - longmon   1   96.654 735.01
## - gender    1   96.679 735.03
```

```
## - wiremon   1   96.810 735.16
## - tenure    1   96.992 735.35
## - wireten   1   97.171 735.53
## <none>          96.251 736.60
## - equipten  1   99.009 737.36
##
## Step:  AIC=734.56
## churn ~ region + tenure + age + marital + address + income +
##     ed + employ + retire + gender + tollfree + equip + callcard +
##     wireless + longmon + equipmon + wiremon + longten + tollten +
##     equipten + cardten + wireten + pager + internet + callid +
##     forward + confer + ebill + logequi + logwire + lninc + custcat
##
##            Df Deviance    AIC
## - marital   1   96.261 732.56
## - lninc     1   96.268 732.57
## - confer    1   96.270 732.57
## - ed        1   96.271 732.58
## - tollfree  1   96.280 732.58
## - age       1   96.283 732.59
## - logwire   1   96.286 732.59
## - equipmon  1   96.296 732.60
## - cardten   1   96.297 732.60
## - forward   1   96.304 732.61
## - retire    1   96.319 732.62
## - employ    1   96.329 732.63
## - income    1   96.352 732.66
## - logequi   1   96.366 732.67
## - internet  1   96.368 732.67
## - pager     1   96.371 732.68
## - ebill     1   96.380 732.68
## - equip     1   96.398 732.70
## - callid    1   96.401 732.70
## - custcat   1   96.408 732.71
## - callcard  1   96.416 732.72
## - tollten   1   96.437 732.74
## - wireless  1   96.448 732.75
## - region    1   96.469 732.77
## - longten   1   96.557 732.86
## - address   1   96.571 732.88
## - longmon   1   96.654 732.96
## - gender    1   96.681 732.99
## - wiremon   1   96.845 733.15
## - tenure    1   96.996 733.30
## - wireten   1   97.192 733.50
## <none>          96.254 734.56
## - equipten  1   99.009 735.31
##
## Step:  AIC=732.55
## churn ~ region + tenure + age + address + income + ed + employ +
##     retire + gender + tollfree + equip + callcard + wireless +
##     longmon + equipmon + wiremon + longten + tollten + equipten +
##     cardten + wireten + pager + internet + callid + forward +
##     confer + ebill + logequi + logwire + lninc + custcat
```

```
##
##            Df Deviance    AIC
## - lninc     1   96.275 730.57
## - confer    1   96.277 730.57
## - ed        1   96.278 730.57
## - tollfree  1   96.288 730.58
## - age       1   96.290 730.58
## - logwire   1   96.294 730.59
## - cardten   1   96.302 730.59
## - equipmon  1   96.303 730.59
## - forward   1   96.312 730.60
## - retire    1   96.322 730.61
## - employ    1   96.338 730.63
## - income    1   96.355 730.65
## - logequi   1   96.372 730.66
## - internet  1   96.372 730.66
## - ebill     1   96.383 730.67
## - pager     1   96.384 730.68
## - equip     1   96.404 730.70
## - callid    1   96.410 730.70
## - custcat   1   96.416 730.71
## - callcard  1   96.420 730.71
## - tollten   1   96.447 730.74
## - wireless  1   96.454 730.75
## - region    1   96.482 730.77
## - longten   1   96.563 730.85
## - address   1   96.588 730.88
## - longmon   1   96.657 730.95
## - gender    1   96.694 730.99
## - wiremon   1   96.850 731.14
## - tenure    1   97.010 731.30
## - wireten   1   97.194 731.49
## <none>          96.261 732.55
## - equipten  1   99.013 733.30
##
## Step:  AIC=730.91
## churn ~ region + tenure + age + address + income + ed + employ +
##     retire + gender + tollfree + equip + callcard + wireless +
##     longmon + equipmon + wiremon + longten + tollten + equipten +
##     cardten + wireten + pager + internet + callid + forward +
##     confer + ebill + logequi + logwire + custcat
##
##            Df Deviance    AIC
## - confer    1   96.292 728.92
## - ed        1   96.295 728.93
## - age       1   96.303 728.94
## - logwire   1   96.304 728.94
## - tollfree  1   96.306 728.94
## - cardten   1   96.315 728.95
## - equipmon  1   96.320 728.95
## - forward   1   96.327 728.96
## - retire    1   96.340 728.97
## - employ    1   96.354 728.99
## - income    1   96.378 729.01
```

```
## - logequi   1   96.380 729.01
## - internet  1   96.388 729.02
## - ebill     1   96.390 729.02
## - pager     1   96.404 729.04
## - equip     1   96.414 729.05
## - callcard  1   96.424 729.06
## - callid    1   96.425 729.06
## - custcat   1   96.429 729.06
## - wireless  1   96.457 729.09
## - tollten   1   96.469 729.10
## - region    1   96.496 729.13
## - longten   1   96.571 729.20
## - address   1   96.606 729.24
## - longmon   1   96.666 729.30
## - gender    1   96.711 729.34
## - wiremon   1   96.898 729.53
## - tenure    1   97.015 729.65
## - wireten   1   97.195 729.83
## <none>          96.275 730.91
## - equipten  1   99.044 731.68
##
## Step:  AIC=728.96
## churn ~ region + tenure + age + address + income + ed + employ +
##     retire + gender + tollfree + equip + callcard + wireless +
##     longmon + equipmon + wiremon + longten + tollten + equipten +
##     cardten + wireten + pager + internet + callid + forward +
##     ebill + logequi + logwire + custcat
##
##            Df Deviance    AIC
## - ed        1   96.313 726.98
## - logwire   1   96.317 726.98
## - age       1   96.320 726.98
## - tollfree  1   96.324 726.99
## - cardten   1   96.330 726.99
## - forward   1   96.335 727.00
## - equipmon  1   96.335 727.00
## - retire    1   96.354 727.02
## - employ    1   96.369 727.03
## - income    1   96.395 727.06
## - internet  1   96.408 727.07
## - logequi   1   96.408 727.07
## - ebill     1   96.410 727.07
## - callid    1   96.425 727.09
## - pager     1   96.427 727.09
## - callcard  1   96.438 727.10
## - equip     1   96.440 727.10
## - custcat   1   96.444 727.11
## - wireless  1   96.478 727.14
## - tollten   1   96.503 727.17
## - region    1   96.509 727.17
## - longten   1   96.582 727.25
## - address   1   96.618 727.28
## - longmon   1   96.675 727.34
## - gender    1   96.726 727.39
```

```
## - wiremon   1   96.912 727.58
## - tenure    1   97.024 727.69
## - wireten   1   97.202 727.87
## <none>          96.292 728.96
## - equipten  1   99.057 729.72
##
## Step:  AIC=727.04
## churn ~ region + tenure + age + address + income + employ + retire +
##     gender + tollfree + equip + callcard + wireless + longmon +
##     equipmon + wiremon + longten + tollten + equipten + cardten +
##     wireten + pager + internet + callid + forward + ebill + logequi +
##     logwire + custcat
##
##            Df Deviance    AIC
## - logwire   1   96.335 725.07
## - age       1   96.341 725.07
## - tollfree  1   96.344 725.07
## - cardten   1   96.348 725.08
## - forward   1   96.357 725.09
## - equipmon  1   96.357 725.09
## - retire    1   96.369 725.10
## - employ    1   96.375 725.11
## - income    1   96.399 725.13
## - ebill     1   96.422 725.15
## - logequi   1   96.437 725.17
## - pager     1   96.446 725.18
## - callid    1   96.446 725.18
## - internet  1   96.447 725.18
## - equip     1   96.457 725.19
## - callcard  1   96.466 725.20
## - custcat   1   96.470 725.20
## - wireless  1   96.495 725.23
## - tollten   1   96.524 725.25
## - region    1   96.525 725.25
## - longten   1   96.610 725.34
## - address   1   96.633 725.36
## - longmon   1   96.705 725.43
## - gender    1   96.751 725.48
## - wiremon   1   96.918 725.65
## - tenure    1   97.032 725.76
## - wireten   1   97.207 725.94
## <none>          96.313 727.04
## - equipten  1   99.131 727.86
##
## Step:  AIC=725.11
## churn ~ region + tenure + age + address + income + employ + retire +
##     gender + tollfree + equip + callcard + wireless + longmon +
##     equipmon + wiremon + longten + tollten + equipten + cardten +
##     wireten + pager + internet + callid + forward + ebill + logequi +
##     custcat
##
##            Df Deviance    AIC
## - tollfree  1   96.353 723.12
## - age       1   96.364 723.14
```

```
## - cardten   1   96.365 723.14
## - equipmon  1   96.383 723.16
## - retire    1   96.385 723.16
## - employ    1   96.392 723.16
## - forward   1   96.396 723.17
## - ebill     1   96.448 723.22
## - income    1   96.450 723.22
## - internet  1   96.470 723.24
## - custcat   1   96.471 723.24
## - pager     1   96.475 723.25
## - equip     1   96.481 723.25
## - callid    1   96.489 723.26
## - callcard  1   96.502 723.27
## - wireless  1   96.510 723.28
## - tollten   1   96.537 723.31
## - region    1   96.541 723.31
## - longten   1   96.631 723.40
## - address   1   96.669 723.44
## - logequi   1   96.698 723.47
## - longmon   1   96.723 723.50
## - gender    1   96.778 723.55
## - wiremon   1   96.996 723.77
## - tenure    1   97.041 723.81
## - wireten   1   97.267 724.04
## <none>          96.335 725.11
## - equipten  1   99.133 725.91
##
## Step:  AIC=722.98
## churn ~ region + tenure + age + address + income + employ + retire +
##     gender + equip + callcard + wireless + longmon + equipmon +
##     wiremon + longten + tollten + equipten + cardten + wireten +
##     pager + internet + callid + forward + ebill + logequi + custcat
##
##            Df Deviance    AIC
## - age       1   96.379 721.01
## - cardten   1   96.379 721.01
## - retire    1   96.403 721.03
## - forward   1   96.403 721.03
## - equipmon  1   96.403 721.03
## - employ    1   96.407 721.03
## - ebill     1   96.470 721.10
## - income    1   96.470 721.10
## - custcat   1   96.481 721.11
## - internet  1   96.487 721.11
## - pager     1   96.489 721.12
## - callid    1   96.493 721.12
## - equip     1   96.499 721.13
## - callcard  1   96.505 721.13
## - wireless  1   96.535 721.16
## - region    1   96.562 721.19
## - longten   1   96.660 721.29
## - address   1   96.681 721.31
## - logequi   1   96.731 721.36
## - longmon   1   96.750 721.38
```

```
## - tollten   1   96.764 721.39
## - gender    1   96.790 721.42
## - wiremon   1   97.000 721.63
## - tenure    1   97.114 721.74
## - wireten   1   97.272 721.90
## <none>          96.353 722.98
## - equipten  1   99.212 723.84
##
## Step:  AIC=720.98
## churn ~ region + tenure + address + income + employ + retire +
##     gender + equip + callcard + wireless + longmon + equipmon +
##     wiremon + longten + tollten + equipten + cardten + wireten +
##     pager + internet + callid + forward + ebill + logequi + custcat
##
##            Df Deviance    AIC
## - cardten   1   96.404 719.00
## - employ    1   96.412 719.01
## - equipmon  1   96.427 719.03
## - forward   1   96.428 719.03
## - retire    1   96.469 719.07
## - ebill     1   96.494 719.09
## - custcat   1   96.504 719.10
## - income    1   96.506 719.11
## - pager     1   96.509 719.11
## - internet  1   96.510 719.11
## - callid    1   96.516 719.12
## - equip     1   96.526 719.13
## - callcard  1   96.533 719.13
## - wireless  1   96.558 719.16
## - region    1   96.584 719.18
## - longten   1   96.670 719.27
## - address   1   96.694 719.29
## - logequi   1   96.743 719.34
## - longmon   1   96.764 719.36
## - tollten   1   96.790 719.39
## - gender    1   96.799 719.40
## - wiremon   1   97.026 719.63
## - tenure    1   97.163 719.76
## - wireten   1   97.296 719.90
## <none>          96.379 720.98
## - equipten  1   99.214 721.81
##
## Step:  AIC=719.03
## churn ~ region + tenure + address + income + employ + retire +
##     gender + equip + callcard + wireless + longmon + equipmon +
##     wiremon + longten + tollten + equipten + wireten + pager +
##     internet + callid + forward + ebill + logequi + custcat
##
##            Df Deviance    AIC
## - employ    1   96.438 717.06
## - equipmon  1   96.454 717.08
## - forward   1   96.456 717.08
## - retire    1   96.491 717.12
## - ebill     1   96.520 717.15
```

```
## - pager     1   96.531 717.16
## - internet  1   96.532 717.16
## - custcat   1   96.533 717.16
## - callcard  1   96.537 717.16
## - income    1   96.540 717.17
## - callid    1   96.542 717.17
## - equip     1   96.552 717.18
## - wireless  1   96.576 717.20
## - region    1   96.609 717.23
## - longten   1   96.672 717.30
## - address   1   96.737 717.36
## - longmon   1   96.769 717.39
## - logequi   1   96.811 717.44
## - tollten   1   96.812 717.44
## - gender    1   96.823 717.45
## - wiremon   1   97.041 717.67
## - tenure    1   97.165 717.79
## - wireten   1   97.326 717.95
## <none>          96.404 719.03
## - equipten  1   99.276 719.90
##
## Step:  AIC=717.06
## churn ~ region + tenure + address + income + retire + gender +
##     equip + callcard + wireless + longmon + equipmon + wiremon +
##     longten + tollten + equipten + wireten + pager + internet +
##     callid + forward + ebill + logequi + custcat
##
##            Df Deviance    AIC
## - equipmon  1   96.491 715.11
## - forward   1   96.491 715.11
## - retire    1   96.504 715.13
## - income    1   96.540 715.16
## - ebill     1   96.555 715.18
## - internet  1   96.559 715.18
## - custcat   1   96.564 715.19
## - pager     1   96.565 715.19
## - callcard  1   96.569 715.19
## - callid    1   96.570 715.19
## - equip     1   96.582 715.21
## - wireless  1   96.610 715.23
## - region    1   96.641 715.26
## - longten   1   96.693 715.32
## - longmon   1   96.789 715.41
## - address   1   96.812 715.43
## - gender    1   96.845 715.47
## - tollten   1   96.868 715.49
## - logequi   1   96.874 715.50
## - wiremon   1   97.070 715.69
## - tenure    1   97.170 715.79
## - wireten   1   97.346 715.97
## <none>          96.438 717.06
## - equipten  1   99.311 717.93
##
## Step:  AIC=715.2
```

```
## churn ~ region + tenure + address + income + retire + gender +
##     equip + callcard + wireless + longmon + wiremon + longten +
##     tollten + equipten + wireten + pager + internet + callid +
##     forward + ebill + logequi + custcat
##
##            Df Deviance    AIC
## - forward   1   96.550 713.26
## - retire    1   96.568 713.28
## - income    1   96.590 713.30
## - internet  1   96.605 713.31
## - callcard  1   96.616 713.32
## - ebill     1   96.616 713.32
## - callid    1   96.620 713.33
## - pager     1   96.632 713.34
## - wireless  1   96.682 713.39
## - region    1   96.693 713.40
## - custcat   1   96.733 713.44
## - longten   1   96.751 713.46
## - longmon   1   96.852 713.56
## - address   1   96.875 713.58
## - gender    1   96.901 713.61
## - logequi   1   96.907 713.61
## - tollten   1   96.911 713.62
## - tenure    1   97.209 713.92
## - wiremon   1   97.402 714.11
## - wireten   1   97.649 714.36
## - equip     1   98.075 714.78
## <none>          96.491 715.20
## - equipten  1  100.345 717.05
##
## Step:  AIC=713.35
## churn ~ region + tenure + address + income + retire + gender +
##     equip + callcard + wireless + longmon + wiremon + longten +
##     tollten + equipten + wireten + pager + internet + callid +
##     ebill + logequi + custcat
##
##            Df Deviance    AIC
## - retire    1   96.630 711.43
## - income    1   96.650 711.45
## - ebill     1   96.667 711.47
## - internet  1   96.670 711.47
## - callcard  1   96.707 711.51
## - pager     1   96.712 711.51
## - wireless  1   96.731 711.53
## - region    1   96.759 711.56
## - callid    1   96.779 711.58
## - longten   1   96.793 711.59
## - custcat   1   96.801 711.60
## - longmon   1   96.893 711.69
## - address   1   96.920 711.72
## - tollten   1   96.926 711.72
## - gender    1   96.945 711.74
## - logequi   1   96.954 711.75
## - tenure    1   97.254 712.05
```

```
## - wiremon   1   97.471 712.27
## - wireten   1   97.730 712.53
## - equip     1   98.123 712.92
## <none>          96.550 713.35
## - equipten  1  100.422 715.22
##
## Step:  AIC=711.5
## churn ~ region + tenure + address + income + gender + equip +
##     callcard + wireless + longmon + wiremon + longten + tollten +
##     equipten + wireten + pager + internet + callid + ebill +
##     logequi + custcat
##
##            Df Deviance    AIC
## - income    1   96.701 709.57
## - ebill     1   96.741 709.61
## - internet  1   96.760 709.63
## - callcard  1   96.797 709.67
## - pager     1   96.799 709.67
## - wireless  1   96.816 709.69
## - region    1   96.828 709.70
## - callid    1   96.860 709.73
## - custcat   1   96.883 709.76
## - longten   1   96.915 709.79
## - address   1   96.939 709.81
## - tollten   1   96.994 709.87
## - longmon   1   96.997 709.87
## - gender    1   97.033 709.91
## - logequi   1   97.072 709.95
## - tenure    1   97.295 710.17
## - wiremon   1   97.555 710.43
## - wireten   1   97.804 710.68
## - equip     1   98.250 711.12
## <none>          96.630 711.50
## - equipten  1  100.600 713.47
##
## Step:  AIC=709.6
## churn ~ region + tenure + address + gender + equip + callcard +
##     wireless + longmon + wiremon + longten + tollten + equipten +
##     wireten + pager + internet + callid + ebill + logequi + custcat
##
##            Df Deviance    AIC
## - ebill     1   96.818 707.72
## - internet  1   96.836 707.74
## - callcard  1   96.862 707.77
## - wireless  1   96.882 707.79
## - pager     1   96.894 707.80
## - region    1   96.904 707.81
## - callid    1   96.949 707.85
## - custcat   1   96.950 707.85
## - address   1   96.980 707.88
## - longten   1   97.000 707.90
## - tollten   1   97.048 707.95
## - longmon   1   97.087 707.99
## - logequi   1   97.094 708.00
```

```
## - gender    1   97.108 708.01
## - tenure    1   97.377 708.28
## - wiremon   1   97.594 708.50
## - wireten   1   97.827 708.73
## - equip     1   98.291 709.20
## <none>          96.701 709.60
## - equipten  1  100.667 711.57
##
## Step:  AIC=707.52
## churn ~ region + tenure + address + gender + equip + callcard +
##     wireless + longmon + wiremon + longten + tollten + equipten +
##     wireten + pager + internet + callid + logequi + custcat
##
##            Df Deviance    AIC
## - internet  1   96.906 705.61
## - callcard  1   96.968 705.67
## - wireless  1   96.995 705.69
## - region    1   97.002 705.70
## - pager     1   97.023 705.72
## - callid    1   97.043 705.74
## - custcat   1   97.082 705.78
## - address   1   97.099 705.80
## - longten   1   97.123 705.82
## - tollten   1   97.183 705.88
## - longmon   1   97.207 705.91
## - gender    1   97.210 705.91
## - logequi   1   97.222 705.92
## - tenure    1   97.478 706.18
## - wiremon   1   97.737 706.44
## - wireten   1   97.954 706.65
## - equip     1   98.745 707.44
## <none>          96.818 707.52
## - equipten  1  100.796 709.50
##
## Step:  AIC=705.76
## churn ~ region + tenure + address + gender + equip + callcard +
##     wireless + longmon + wiremon + longten + tollten + equipten +
##     wireten + pager + callid + logequi + custcat
##
##            Df Deviance    AIC
## - callcard  1   97.050 703.91
## - wireless  1   97.078 703.94
## - pager     1   97.090 703.95
## - region    1   97.092 703.95
## - callid    1   97.149 704.01
## - address   1   97.179 704.04
## - custcat   1   97.190 704.05
## - longten   1   97.210 704.07
## - tollten   1   97.253 704.11
## - gender    1   97.285 704.14
## - longmon   1   97.291 704.15
## - logequi   1   97.292 704.15
## - tenure    1   97.582 704.44
## - wiremon   1   97.780 704.64
```

```
## - wireten   1   98.015 704.87
## - equip     1   98.772 705.63
## <none>          96.906 705.76
## - equipten  1  100.915 707.77
##
## Step:  AIC=704.48
## churn ~ region + tenure + address + gender + equip + wireless +
##     longmon + wiremon + longten + tollten + equipten + wireten +
##     pager + callid + logequi + custcat
##
##            Df Deviance    AIC
## - wireless  1   97.207 702.64
## - region    1   97.219 702.65
## - pager     1   97.244 702.67
## - address   1   97.307 702.74
## - custcat   1   97.329 702.76
## - longten   1   97.333 702.76
## - logequi   1   97.350 702.78
## - tollten   1   97.388 702.82
## - callid    1   97.406 702.83
## - longmon   1   97.413 702.84
## - gender    1   97.427 702.85
## - tenure    1   97.958 703.39
## - wiremon   1   98.022 703.45
## - wireten   1   98.379 703.81
## - equip     1   98.793 704.22
## <none>          97.050 704.48
## - equipten  1  100.952 706.38
##
## Step:  AIC=701.99
## churn ~ region + tenure + address + gender + equip + longmon +
##     wiremon + longten + tollten + equipten + wireten + pager +
##     callid + logequi + custcat
##
##            Df Deviance    AIC
## - region    1   97.366 700.15
## - pager     1   97.405 700.19
## - address   1   97.467 700.25
## - logequi   1   97.488 700.27
## - longten   1   97.491 700.27
## - callid    1   97.517 700.30
## - tollten   1   97.519 700.30
## - longmon   1   97.572 700.35
## - gender    1   97.600 700.38
## - custcat   1   97.651 700.43
## - tenure    1   98.098 700.88
## - wireten   1   98.379 701.16
## - wiremon   1   98.407 701.19
## - equip     1   98.841 701.62
## <none>          97.207 701.99
## - equipten  1  101.065 703.85
##
## Step:  AIC=700.2
## churn ~ tenure + address + gender + equip + longmon + wiremon +
```

```
##     longten + tollten + equipten + wireten + pager + callid +
##     logequi + custcat
##
##            Df Deviance    AIC
## - pager     1   97.553 698.39
## - address   1   97.625 698.46
## - longten   1   97.633 698.47
## - logequi   1   97.672 698.50
## - callid    1   97.672 698.50
## - tollten   1   97.679 698.51
## - longmon   1   97.711 698.54
## - gender    1   97.723 698.56
## - custcat   1   97.830 698.66
## - tenure    1   98.255 699.09
## - wireten   1   98.566 699.40
## - wiremon   1   98.598 699.43
## - equip     1   98.998 699.83
## <none>          97.366 700.20
## - equipten  1  101.215 702.05
##
## Step:  AIC=698.29
## churn ~ tenure + address + gender + equip + longmon + wiremon +
##     longten + tollten + equipten + wireten + callid + logequi +
##     custcat
##
##            Df Deviance    AIC
## - tollten   1   97.814 696.55
## - longten   1   97.827 696.57
## - address   1   97.830 696.57
## - logequi   1   97.849 696.59
## - gender    1   97.880 696.62
## - callid    1   97.900 696.64
## - longmon   1   97.911 696.65
## - custcat   1   98.062 696.80
## - tenure    1   98.431 697.17
## - wireten   1   98.917 697.66
## - equip     1   99.275 698.01
## <none>          97.553 698.29
## - wiremon   1   99.577 698.32
## - equipten  1  101.393 700.13
##
## Step:  AIC=696.2
## churn ~ tenure + address + gender + equip + longmon + wiremon +
##     longten + equipten + wireten + callid + logequi + custcat
##
##            Df Deviance    AIC
## - callid    1   97.985 694.38
## - logequi   1   98.032 694.42
## - longten   1   98.048 694.44
## - gender    1   98.109 694.50
## - address   1   98.127 694.52
## - longmon   1   98.161 694.55
## - custcat   1   98.279 694.67
## - tenure    1   98.545 694.94
```

```
## - equip     1   99.390 695.78
## <none>          97.814 696.20
## - wireten   1   99.888 696.28
## - wiremon   1  100.027 696.42
## - equipten  1  101.394 697.78
##
## Step:  AIC=694.76
## churn ~ tenure + address + gender + equip + longmon + wiremon +
##     longten + equipten + wireten + logequi + custcat
##
##            Df Deviance    AIC
## - logequi   1   98.176 692.95
## - longten   1   98.225 693.00
## - gender    1   98.291 693.06
## - address   1   98.300 693.07
## - longmon   1   98.325 693.10
## - custcat   1   98.437 693.21
## - tenure    1   98.665 693.44
## - equip     1   99.423 694.20
## <none>          97.985 694.76
## - wireten   1  100.164 694.94
## - wiremon   1  100.756 695.53
## - equipten  1  101.464 696.24
##
## Step:  AIC=692.94
## churn ~ tenure + address + gender + equip + longmon + wiremon +
##     longten + equipten + wireten + custcat
##
##            Df Deviance    AIC
## - longten   1   98.412 691.17
## - address   1   98.489 691.25
## - gender    1   98.493 691.25
## - longmon   1   98.508 691.27
## - custcat   1   98.552 691.31
## - tenure    1   98.705 691.46
## - equip     1   99.428 692.19
## <none>          98.176 692.94
## - wireten   1  100.261 693.02
## - wiremon   1  100.796 693.56
## - equipten  1  101.466 694.23
##
## Step:  AIC=691.11
## churn ~ tenure + address + gender + equip + longmon + wiremon +
##     equipten + wireten + custcat
##
##            Df Deviance    AIC
## - longmon   1   98.563 689.27
## - address   1   98.677 689.38
## - gender    1   98.752 689.45
## - custcat   1   98.802 689.50
## - tenure    1   99.288 689.99
## - equip     1   99.825 690.53
## - wireten   1  100.396 691.10
## <none>          98.412 691.11
```

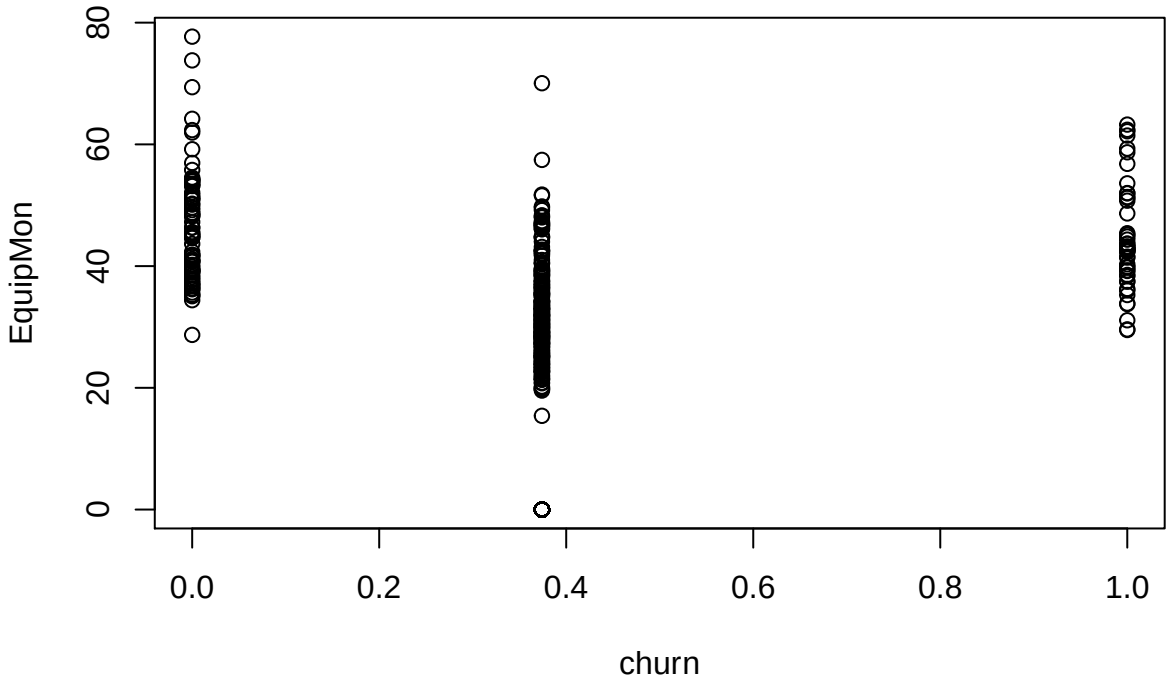
```
## - wiremon   1  100.946 691.65
## - equipten  1  101.950 692.65
##
## Step:  AIC=689.29
## churn ~ tenure + address + gender + equip + wiremon + equipten +
##     wireten + custcat
##
##            Df Deviance    AIC
## - gender    1   98.898 687.63
## - address   1   98.907 687.64
## - custcat   1   98.954 687.68
## - tenure    1   99.324 688.05
## - equip     1   99.918 688.65
## <none>          98.563 689.29
## - wireten   1  100.592 689.32
## - wiremon   1  101.132 689.86
## - equipten  1  102.013 690.74
##
## Step:  AIC=687.85
## churn ~ tenure + address + equip + wiremon + equipten + wireten +
##     custcat
##
##            Df Deviance    AIC
## - address   1   99.257 686.21
## - custcat   1   99.306 686.26
## - tenure    1   99.626 686.58
## - equip     1  100.223 687.17
## - wireten   1  100.868 687.82
## <none>          98.898 687.85
## - wiremon   1  101.403 688.35
## - equipten  1  102.260 689.21
##
## Step:  AIC=686.34
## churn ~ tenure + equip + wiremon + equipten + wireten + custcat
##
##            Df Deviance    AIC
## - custcat   1   99.620 684.70
## - tenure    1   99.680 684.76
## - equip     1  100.521 685.61
## <none>          99.257 686.34
## - wireten   1  101.302 686.39
## - wiremon   1  101.792 686.88
## - equipten  1  102.407 687.49
##
## Step:  AIC=689.47
## churn ~ tenure + equip + wiremon + equipten + wireten
```

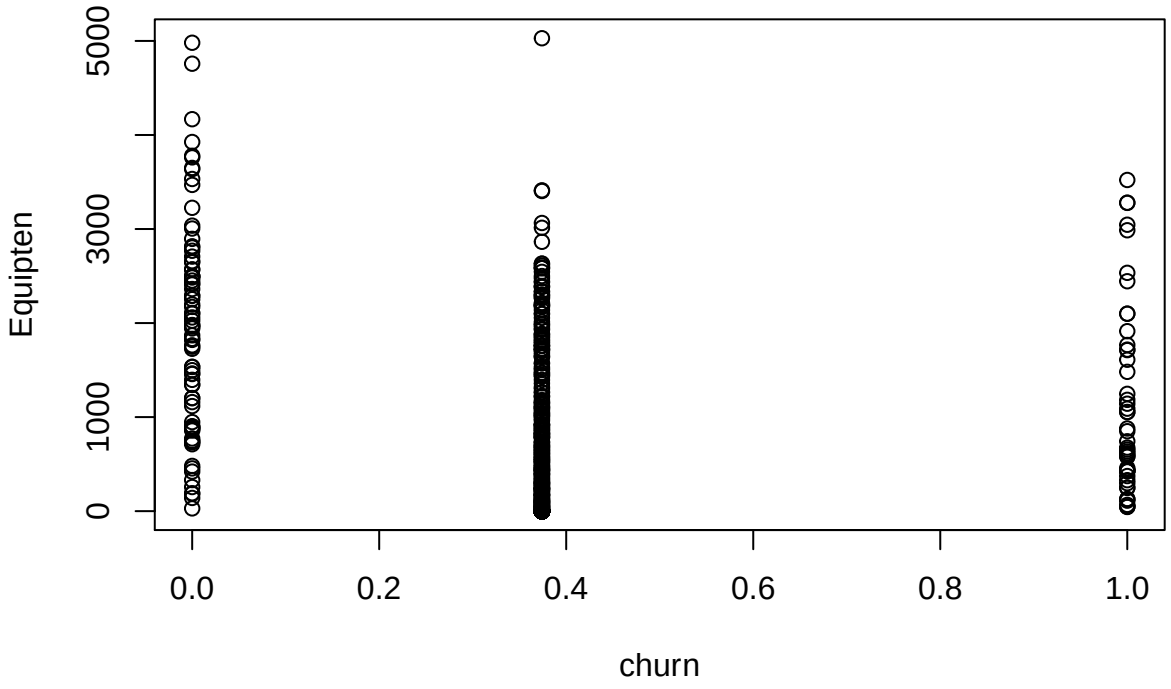
Data mining goals:

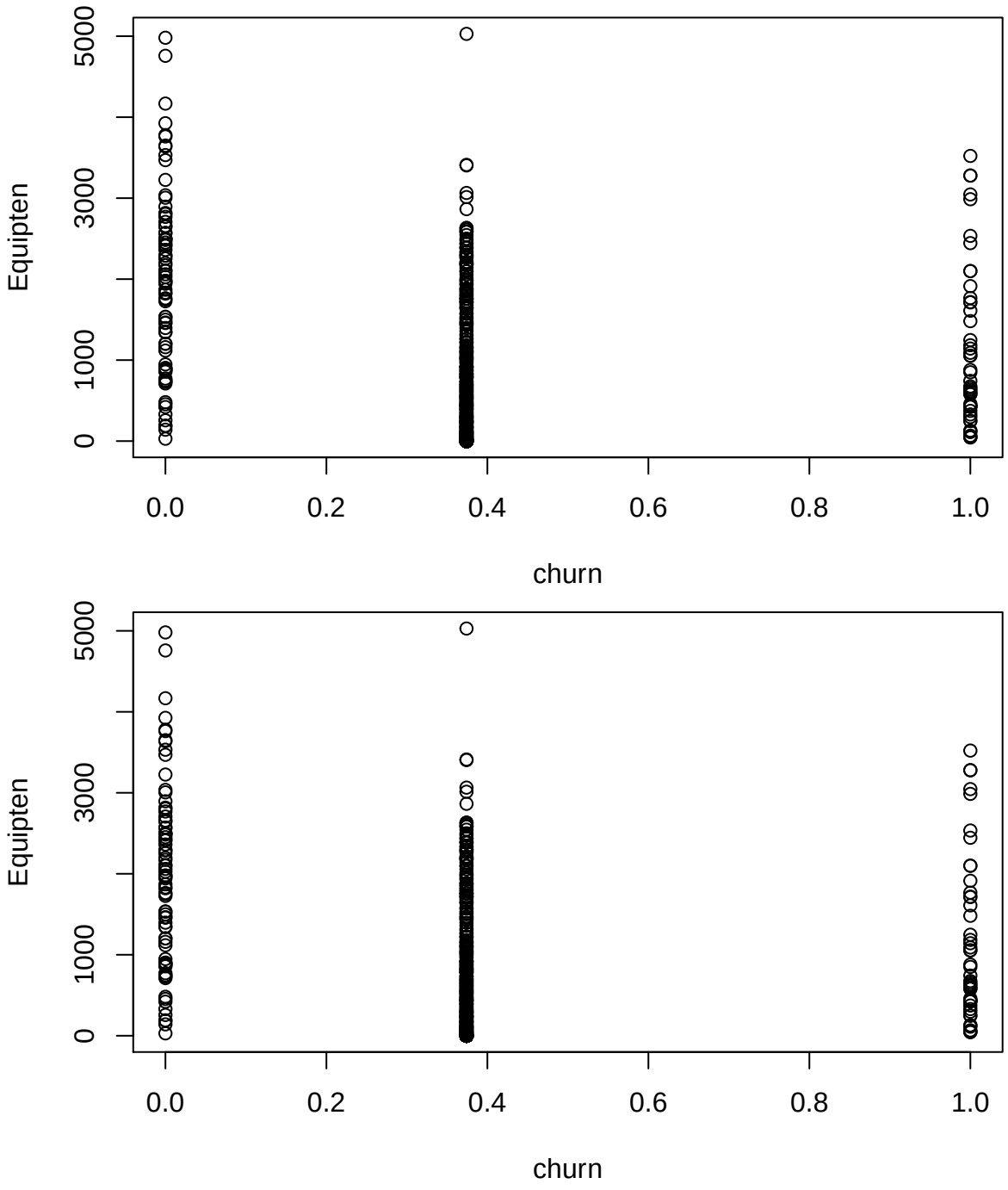
Retention is one of the major goals of any telecom industry and therefore it is necessary to profile the "churners".
We analyze the data to determine what customer might churn and which customer will retain and be loyal to the company.

H0 : The churn is affected by the following variables - equipmon,equipten,wiremon,wireten.

Graphs







Verifying Data Quality :

```
## Groups for Hosmer-Lemeshow C statistic:
## cutfit
## [0.103,0.318] (0.318,0.351] (0.351,0.359] (0.359,0.367] (0.367,0.374]
##            70            70            72            68            69
## (0.374,0.382]  (0.382,0.39]  (0.39,0.399] (0.399,0.429] (0.429,0.595]
##            70            69            70            70            70
## Groups for Hosmer-Lemeshow H statistic:
```

```
## cutfit1
## [0.103,0.152] (0.152,0.202] (0.202,0.251]   (0.251,0.3]   (0.3,0.349]
##             7             5             8            29            70
## (0.349,0.398] (0.398,0.447] (0.447,0.496] (0.496,0.546] (0.546,0.595]
##           437            89            33            12             8
```

```
## $C
##
##  Hosmer-Lemeshow C statistic
##
## data:  Training_model$fitted.values and Training_model$y
## X-squared = 1.4946, df = 8, p-value = 0.9928
##
##
## $H
##
##  Hosmer-Lemeshow H statistic
##
## data:  Training_model$fitted.values and Training_model$y
## X-squared = 5.223, df = 8, p-value = 0.7335
```

The value of p is greater than 0.5,the model is a good fit