

DATA 621: BUSINESS ANALYTICS AND DATA MINING

HOMEWORK#5 Assignment Requirements

Group 2 - Gabriel Campos, Melissa Bowman, Alexander Khaykin, & Jennifer Abinette

Last edited November 30, 2023

Contents

1 Overview	1
1.1 Deliverables	2
1.2 Write Up:	2
1.2.1 1. DATA EXPLORATION (25 Points)	2
1.2.2 2. DATA PREPARATION (25 Points)	3
1.2.3 3. BUILD MODELS (25 Points)	3
1.2.4 4. SELECT MODELS (25 Points)	3
2 Import Data	4
2.1 Data Summary	5
2.1.1 df_wine_eval	5
2.1.2 df_wine_train	7

1 Overview

In this homework assignment, you will explore, analyze and model a data set containing information on approximately 12,000 commercially available wines. The variables are mostly related to the chemical properties of the wine being sold. The response variable is the number of sample cases of wine that were purchased by wine distribution companies after sampling a wine. These cases would be used to provide tasting samples to restaurants and wine stores around the United States. The more sample cases purchased, the more likely is a wine to be sold at a high end restaurant. A large wine manufacturer is studying the data in order to predict the number of wine cases ordered based upon the wine characteristics. If the wine manufacturer can predict the number of cases, then that manufacturer will be able to adjust their wine offering to maximize sales.

Your objective is to build a count regression model to predict the number of cases of wine that will be sold given certain properties of the wine. HINT: Sometimes, the fact that a variable is missing is actually predictive of the target. You can only use the variables given to you (or variables that you derive from the variables provided). Below is a short description of the variables of interest in the data set:

VARIABLE NAME	DEFINITION	THEORETICAL EFFECT
INDEX	Identification Variable (do not use)	None
TARGET	Number of Cases Purchased	None
AcidIndex	Proprietary method of testing total acidity of wine by using a weighted average	
Alcohol	Alcohol Content	
Chlorides	Chloride content of wine	
CitricAcid	Citric Acid Content	
Density	Density of Wine	
FixedAcidity	Fixed Acidity of Wine	
FreeSulfurDioxide	Sulfur Dioxide content of wine	
LabelAppeal	Marketing Score indicating the appeal of label design for consumers. High numbers suggest customers like the label design. Negative numbers suggest customers don't like the design.	Many consumers purchase based on the visual appeal of the wine label design. Higher numbers suggest better sales.
ResidualSugar	Residual Sugar of wine	
STARS	Wine rating by a team of experts. 4 Stars = Excellent, 1 Star = Poor	A high number of stars suggests high sales
Sulphates	Sulfate content of wine	
TotalSulfurDioxide	Total Sulfur Dioxide of Wine	
VolatileAcidity	Volatile Acid content of wine	
pH	pH of wine	

1.1 Deliverables

- A write-up submitted in PDF format. Your write-up should have four sections. Each one is described below. You may assume you are addressing me as a fellow data scientist, so do not need to shy away from technical details.
- Assigned predictions (number of cases of wine sold) for the evaluation data set.
- Include your R statistical programming code in an Appendix.

1.2 Write Up:

1.2.1 1. DATA EXPLORATION (25 Points)

Describe the size and the variables in the wine training data set. Consider that too much detail will cause a manager to lose interest while too little detail will make the manager consider that you aren't doing your job. Some suggestions are given below. Please do NOT treat this as a check list of things to do to complete the assignment. You should have your own thoughts on what to tell the boss. These are just ideas.

- Mean / Standard Deviation / Median
- Bar Chart or Box Plot of the data
- Is the data correlated to the target variable (or to other variables?)
- Are any of the variables missing and need to be imputed "fixed"?

1.2.2 2. DATA PREPARATION (25 Points)

Describe how you have transformed the data by changing the original variables or creating new variables. If you did transform the data or create new variables, discuss why you did this. Here are some possible transformations.

- a. Fix missing values (maybe with a Mean or Median value)
- b. Create flags to suggest if a variable was missing
- c. Transform data by putting it into buckets
- d. Mathematical transforms such as log or square root (or use Box-Cox)
- e. Combine variables (such as ratios or adding or multiplying) to create new variables

1.2.3 3. BUILD MODELS (25 Points)

Using the training data set, build at least two different poisson regression models, at least two different negative binomial regression models, and at least two multiple linear regression models, using different variables (or the same variables with different transformations). Sometimes poisson and negative binomial regression models give the same results. If that is the case, comment on that. Consider changing the input variables if that occurs so that you get different models. Although not covered in class, you may also want to consider building zero-inflated poisson and negative binomial regression models. You may select the variables manually, use an approach such as Forward or Stepwise, use a different approach such as trees, or use a combination of techniques. Describe the techniques you used. If you manually selected a variable for inclusion into the model or exclusion into the model, indicate why this was done

Discuss the coefficients in the models, do they make sense? In this case, about the only thing you can comment on is the number of stars and the wine label appeal. However, you might comment on the coefficient and magnitude of variables and how they are similar or different from model to model. For example, you might say “pH seems to have a major positive impact in my poisson regression model, but a negative effect in my multiple linear regression model”. Are you keeping the model even though it is counter intuitive? Why? The boss needs to know.

1.2.4 4. SELECT MODELS (25 Points)

Decide on the criteria for selecting the best count regression model. Will you select models with slightly worse performance if it makes more sense or is more parsimonious? Discuss why you selected your models.

For the count regression model, will you use a metric such as AIC, average squared error, etc.? Be sure to explain how you can make inferences from the model, and discuss other relevant model output. If you like the multiple linear regression model the best, please say why. However, you must select a count regression model for model deployment. Using the training data set, evaluate the performance of the count regression model. Make predictions using the evaluation data set.

2 Import Data

```
df_wine_eval <-  
  read.csv(paste0(url_git,"wine-evaluation-data.csv"))  
  
head(df_wine_eval)
```

```
##   IN TARGET FixedAcidity VolatileAcidity CitricAcid ResidualSugar Chlorides  
## 1  3      NA          5.4         -0.860         0.27         -10.7         0.092  
## 2  9      NA          12.4         0.385         -0.76         -19.7         1.169  
## 3 10      NA          7.2          1.750         0.17         -33.0         0.065  
## 4 18      NA          6.2          0.100         1.80          1.0        -0.179  
## 5 21      NA          11.4         0.210         0.28          1.2         0.038  
## 6 30      NA          17.6         0.040        -1.15          1.4         0.535  
##   FreeSulfurDioxide TotalSulfurDioxide Density    pH Sulphates Alcohol  
## 1                   23                   398 0.98527 5.02     0.64    12.30  
## 2                   -37                   68 0.99048 3.37     1.09    16.00  
## 3                     9                   76 1.04641 4.61     0.68     8.55  
## 4                   104                   89 0.98877 3.20     2.11    12.30  
## 5                     70                   53 1.02899 2.54    -0.07     4.80  
## 6                  -250                   140 0.95028 3.06    -0.02    11.40  
##   LabelAppeal AcidIndex STARS  
## 1            -1         6    NA  
## 2             0         6     2  
## 3             0         8     1  
## 4            -1         8     1  
## 5             0        10    NA  
## 6             1         8     4
```

```
df_wine_train <-  
  read.csv(paste0(url_git,"wine-training-data.csv"))  
  
head(df_wine_train)
```

```
##   INDEX TARGET FixedAcidity VolatileAcidity CitricAcid ResidualSugar Chlorides  
## 1     1     3           3.2           1.160        -0.98          54.2        -0.567  
## 2     2     3           4.5           0.160        -0.81          26.1        -0.425  
## 3     4     5           7.1           2.640        -0.88          14.8         0.037  
## 4     5     3           5.7           0.385         0.04          18.8        -0.425  
## 5     6     4           8.0           0.330        -1.26           9.4         NA  
## 6     7     0          11.3           0.320         0.59           2.2         0.556  
##   FreeSulfurDioxide TotalSulfurDioxide Density    pH Sulphates Alcohol  
## 1                   NA                   268 0.99280 3.33    -0.59     9.9  
## 2                   15                  -327 1.02792 3.38     0.70     NA  
## 3                   214                   142 0.99518 3.12     0.48    22.0  
## 4                   22                   115 0.99640 2.24     1.83     6.2  
## 5                  -167                   108 0.99457 3.12     1.77    13.7  
## 6                  -37                   15 0.99940 3.20     1.29    15.4  
##   LabelAppeal AcidIndex STARS  
## 1             0         8     2  
## 2            -1         7     3  
## 3            -1         8     3
```

```
## 4      -1      6      1
## 5      0      9      2
## 6      0     11     NA
```

2.1 Data Summary

2.1.1 df_wine_eval

```
dim(df_wine_eval)
```

```
## [1] 3335  16
```

```
describe(df_wine_eval)
```

```
##          vars      n    mean      sd  median trimmed      mad      min
## IN              1 3335 8048.31 4655.48 7906.00 8044.28 5960.05      3.00
## TARGET          2   0      NaN      NA      NA      NaN      NA      Inf
## FixedAcidity    3 3335   6.86   6.32   6.90   6.91   2.82 -18.20
## VolatileAcidity 4 3335   0.31   0.81   0.28   0.31   0.46  -2.83
## CitricAcid      5 3335   0.31   0.87   0.31   0.31   0.44  -3.12
## ResidualSugar   6 3167   5.32  34.37   3.60   5.46  16.90 -128.30
## Chlorides       7 3197   0.06   0.31   0.05   0.06   0.12  -1.15
## FreeSulfurDioxide 8 3183  34.95 149.63  30.00  34.26  57.82 -563.00
## TotalSulfurDioxide 9 3178 123.41 225.80 124.00 124.00 137.88 -769.00
## Density        10 3335   0.99   0.03   0.99   0.99   0.01   0.89
## pH             11 3231   3.24   0.68   3.21   3.23   0.37   0.60
## Sulphates      12 3025   0.53   0.91   0.50   0.53   0.39  -3.07
## Alcohol       13 3150  10.58   3.76  10.40  10.58   2.52  -4.20
## LabelAppeal   14 3335   0.01   0.89   0.00   0.01   1.48  -2.00
## AcidIndex     15 3335   7.75   1.32   8.00   7.62   1.48   5.00
## STARS         16 2494   2.04   0.91   2.00   1.97   1.48   1.00
##          max    range  skew kurtosis      se
## IN        16130.00 16127.00  0.01    -1.20 80.62
## TARGET      -Inf    -Inf    NA      NA    NA
## FixedAcidity    33.50   51.70 -0.12    2.04  0.11
## VolatileAcidity   3.61    6.44 -0.04    1.62  0.01
## CitricAcid       3.76    6.88 -0.03    1.66  0.02
## ResidualSugar   145.40  273.70 -0.06    1.97  0.61
## Chlorides        1.26    2.41 -0.04    1.74  0.01
## FreeSulfurDioxide 617.00 1180.00  0.07    1.88  2.65
## TotalSulfurDioxide 1004.00 1773.00 -0.05    1.50  4.01
## Density          1.10    0.21 -0.03    1.94  0.00
## pH               6.21    5.61  0.12    1.69  0.01
## Sulphates        4.18    7.25  0.01    1.83  0.02
## Alcohol          25.60   29.80  0.05    1.54  0.07
## LabelAppeal      2.00    4.00  0.05   -0.26  0.02
## AcidIndex        17.00   12.00  1.51    4.28  0.02
## STARS            4.00    3.00  0.44   -0.75  0.02
```

```
summary(df_wine_eval)
```

```
##           IN           TARGET      FixedAcidity      VolatileAcidity
##  Min.      :    3      Mode:logical    Min.      :-18.200    Min.      :-2.8300
##  1st Qu.: 4018      NA's:3335      1st Qu.:  5.200    1st Qu.:  0.0800
##  Median : 7906                                Median :  6.900    Median :  0.2800
##  Mean   : 8048                                Mean   :  6.864    Mean   :  0.3103
##  3rd Qu.:12061                                3rd Qu.:  9.000    3rd Qu.:  0.6300
##  Max.    :16130                                Max.    : 33.500    Max.    :  3.6100
##
##  CitricAcid      ResidualSugar      Chlorides      FreeSulfurDioxide
##  Min.      :-3.1200    Min.      :-128.300    Min.      :-1.15000    Min.      :-563.00
##  1st Qu.:  0.0000    1st Qu.:  -2.600    1st Qu.:  0.01600    1st Qu.:   3.00
##  Median :  0.3100    Median :   3.600    Median :  0.04700    Median :  30.00
##  Mean   :  0.3124    Mean   :   5.319    Mean   :  0.06143    Mean   :  34.95
##  3rd Qu.:  0.6050    3rd Qu.:  17.200    3rd Qu.:  0.17100    3rd Qu.:  79.25
##  Max.    :  3.7600    Max.    : 145.400    Max.    :  1.26300    Max.    : 617.00
##                      NA's      :168      NA's      :138      NA's      :152
##  TotalSulfurDioxide      Density      pH      Sulphates
##  Min.      :-769.00    Min.      :0.8898    Min.      :0.600    Min.      :-3.0700
##  1st Qu.:  27.25    1st Qu.:0.9883    1st Qu.:2.980    1st Qu.:  0.3300
##  Median : 124.00    Median :0.9946    Median :3.210    Median :  0.5000
##  Mean   : 123.41    Mean   :0.9947    Mean   :3.237    Mean   :  0.5346
##  3rd Qu.: 210.00    3rd Qu.:1.0005    3rd Qu.:3.490    3rd Qu.:  0.8200
##  Max.    :1004.00    Max.    :1.0998    Max.    :6.210    Max.    :  4.1800
##  NA's      :157                                NA's      :104    NA's      :310
##  Alcohol      LabelAppeal      AcidIndex      STARS
##  Min.      :-4.20    Min.      :-2.00000    Min.      : 5.000    Min.      :1.00
##  1st Qu.:  9.00    1st Qu.: -1.00000    1st Qu.: 7.000    1st Qu.:1.00
##  Median :10.40    Median : 0.00000    Median : 8.000    Median :2.00
##  Mean   :10.58    Mean   : 0.01349    Mean   : 7.748    Mean   :2.04
##  3rd Qu.:12.50    3rd Qu.: 1.00000    3rd Qu.: 8.000    3rd Qu.:3.00
##  Max.    :25.60    Max.    : 2.00000    Max.    :17.000    Max.    :4.00
##  NA's      :185                                NA's      :841
```

```
str(df_wine_eval)
```

```
## 'data.frame': 3335 obs. of 16 variables:
## $ IN : int 3 9 10 18 21 30 31 37 39 47 ...
## $ TARGET : logi NA NA NA NA NA NA ...
## $ FixedAcidity : num 5.4 12.4 7.2 6.2 11.4 17.6 15.5 15.9 11.6 3.8 ...
## $ VolatileAcidity : num -0.86 0.385 1.75 0.1 0.21 0.04 0.53 1.19 0.32 0.22 ...
## $ CitricAcid : num 0.27 -0.76 0.17 1.8 0.28 -1.15 -0.53 1.14 0.55 0.31 ...
## $ ResidualSugar : num -10.7 -19.7 -33 1 1.2 1.4 4.6 31.9 -50.9 -7.7 ...
## $ Chlorides : num 0.092 1.169 0.065 -0.179 0.038 ...
## $ FreeSulfurDioxide : num 23 -37 9 104 70 -250 10 115 35 40 ...
## $ TotalSulfurDioxide: num 398 68 76 89 53 140 17 381 83 129 ...
## $ Density : num 0.985 0.99 1.046 0.989 1.029 ...
## $ pH : num 5.02 3.37 4.61 3.2 2.54 3.06 3.07 2.99 3.32 4.72 ...
## $ Sulphates : num 0.64 1.09 0.68 2.11 -0.07 -0.02 0.75 0.31 2.18 -0.64 ...
## $ Alcohol : num 12.3 16 8.55 12.3 4.8 11.4 8.5 11.4 -0.5 10.9 ...
## $ LabelAppeal : int -1 0 0 -1 0 1 0 1 0 0 ...
```

```
## $ AcidIndex      : int   6 6 8 8 10 8 12 7 12 7 ...
## $ STARS          : int   NA 2 1 1 NA 4 3 NA NA NA ...
```

Missing Data

```
for (i in colnames(df_wine_eval)){
  print(paste(i, " ", sum(is.na(df_wine_eval[,i])), sep = " "))
}
```

```
## [1] "IN 0"
## [1] "TARGET 3335"
## [1] "FixedAcidity 0"
## [1] "VolatileAcidity 0"
## [1] "CitricAcid 0"
## [1] "ResidualSugar 168"
## [1] "Chlorides 138"
## [1] "FreeSulfurDioxide 152"
## [1] "TotalSulfurDioxide 157"
## [1] "Density 0"
## [1] "pH 104"
## [1] "Sulphates 310"
## [1] "Alcohol 185"
## [1] "LabelAppeal 0"
## [1] "AcidIndex 0"
## [1] "STARS 841"
```

2.1.2 df_wine_train

```
describe(df_wine_train)
```

```
##          vars      n   mean      sd median trimmed   mad    min
## INDEX          1 12795 8069.98 4656.91 8110.00 8071.03 5977.84    1.00
## TARGET          2 12795   3.03   1.93   3.00   3.05   1.48    0.00
## FixedAcidity    3 12795   7.08   6.32   6.90   7.07   3.26  -18.10
## VolatileAcidity  4 12795   0.32   0.78   0.28   0.32   0.43   -2.79
## CitricAcid      5 12795   0.31   0.86   0.31   0.31   0.42   -3.24
## ResidualSugar   6 12179   5.42  33.75   3.90   5.58  15.72 -127.80
## Chlorides       7 12157   0.05   0.32   0.05   0.05   0.13   -1.17
## FreeSulfurDioxide 8 12148  30.85 148.71  30.00  30.93  56.34 -555.00
## TotalSulfurDioxide 9 12113 120.71 231.91 123.00 120.89 134.92 -823.00
## Density        10 12795   0.99   0.03   0.99   0.99   0.01    0.89
## pH             11 12400   3.21   0.68   3.20   3.21   0.39    0.48
## Sulphates      12 11585   0.53   0.93   0.50   0.53   0.44   -3.13
## Alcohol        13 12142  10.49   3.73  10.40  10.50   2.37   -4.70
## LabelAppeal    14 12795  -0.01   0.89   0.00  -0.01   1.48   -2.00
## AcidIndex      15 12795   7.77   1.32   8.00   7.64   1.48    4.00
## STARS          16  9436   2.04   0.90   2.00   1.97   1.48    1.00
##          max    range skew kurtosis    se
## INDEX    16129.00 16128.00  0.00    -1.20 41.17
## TARGET      8.00     8.00 -0.33    -0.88  0.02
## FixedAcidity 34.40    52.50 -0.02     1.67  0.06
```

```
## VolatileAcidity      3.68      6.47  0.02      1.83  0.01
## CitricAcid           3.86      7.10 -0.05      1.84  0.01
## ResidualSugar       141.15    268.95 -0.05      1.88  0.31
## Chlorides            1.35      2.52  0.03      1.79  0.00
## FreeSulfurDioxide    623.00   1178.00  0.01      1.84  1.35
## TotalSulfurDioxide  1057.00   1880.00 -0.01      1.67  2.11
## Density              1.10      0.21 -0.02      1.90  0.00
## pH                   6.13      5.65  0.04      1.65  0.01
## Sulphates            4.24      7.37  0.01      1.75  0.01
## Alcohol              26.50     31.20 -0.03      1.54  0.03
## LabelAppeal          2.00      4.00  0.01     -0.26  0.01
## AcidIndex            17.00     13.00  1.65      5.19  0.01
## STARS                 4.00      3.00  0.45     -0.69  0.01
```

```
summary(df_wine_train)
```

```
##      INDEX      TARGET      FixedAcidity      VolatileAcidity
## Min.   :    1  Min.   :0.000  Min.   : -18.100  Min.   : -2.7900
## 1st Qu.: 4038  1st Qu.:2.000  1st Qu.:  5.200  1st Qu.:  0.1300
## Median : 8110  Median :3.000  Median :  6.900  Median :  0.2800
## Mean   : 8070  Mean   :3.029  Mean   :  7.076  Mean   :  0.3241
## 3rd Qu.:12106  3rd Qu.:4.000  3rd Qu.:  9.500  3rd Qu.:  0.6400
## Max.   :16129  Max.   :8.000  Max.   : 34.400  Max.   :  3.6800
##
##      CitricAcid      ResidualSugar      Chlorides      FreeSulfurDioxide
## Min.   : -3.2400  Min.   : -127.800  Min.   : -1.1710  Min.   : -555.00
## 1st Qu.:  0.0300  1st Qu.:  -2.000  1st Qu.: -0.0310  1st Qu.:   0.00
## Median :  0.3100  Median :   3.900  Median :  0.0460  Median :  30.00
## Mean   :  0.3084  Mean   :   5.419  Mean   :  0.0548  Mean   :  30.85
## 3rd Qu.:  0.5800  3rd Qu.: 15.900  3rd Qu.:  0.1530  3rd Qu.:  70.00
## Max.   :  3.8600  Max.   :141.150  Max.   :  1.3510  Max.   : 623.00
##      NA's :616      NA's :638      NA's :647
## TotalSulfurDioxide      Density      pH      Sulphates
## Min.   : -823.0  Min.   :0.8881  Min.   :0.480  Min.   : -3.1300
## 1st Qu.:  27.0  1st Qu.:0.9877  1st Qu.:2.960  1st Qu.:  0.2800
## Median : 123.0  Median :0.9945  Median :3.200  Median :  0.5000
## Mean   : 120.7  Mean   :0.9942  Mean   :3.208  Mean   :  0.5271
## 3rd Qu.: 208.0  3rd Qu.:1.0005  3rd Qu.:3.470  3rd Qu.:  0.8600
## Max.   :1057.0  Max.   :1.0992  Max.   :6.130  Max.   :  4.2400
## NA's   :682      NA's :395      NA's :1210
##      Alcohol      LabelAppeal      AcidIndex      STARS
## Min.   : -4.70  Min.   : -2.000000  Min.   :  4.000  Min.   :  1.000
## 1st Qu.:  9.00  1st Qu.: -1.000000  1st Qu.:  7.000  1st Qu.:  1.000
## Median :10.40  Median :  0.000000  Median :  8.000  Median :  2.000
## Mean   :10.49  Mean   : -0.009066  Mean   :  7.773  Mean   :  2.042
## 3rd Qu.:12.40  3rd Qu.:  1.000000  3rd Qu.:  8.000  3rd Qu.:  3.000
## Max.   :26.50  Max.   :  2.000000  Max.   :17.000  Max.   :  4.000
## NA's   :653      NA's :3359
```

```
str(df_wine_train)
```

```
## 'data.frame':    12795 obs. of  16 variables:
## $ INDEX          : int  1 2 4 5 6 7 8 11 12 13 ...
```



```
## $ TARGET          : int  3 3 5 3 4 0 0 4 3 6 ...
## $ FixedAcidity    : num  3.2 4.5 7.1 5.7 8 11.3 7.7 6.5 14.8 5.5 ...
## $ VolatileAcidity : num  1.16 0.16 2.64 0.385 0.33 0.32 0.29 -1.22 0.27 -0.22 ...
## $ CitricAcid      : num  -0.98 -0.81 -0.88 0.04 -1.26 0.59 -0.4 0.34 1.05 0.39 ...
## $ ResidualSugar   : num  54.2 26.1 14.8 18.8 9.4 ...
## $ Chlorides       : num  -0.567 -0.425 0.037 -0.425 NA 0.556 0.06 0.04 -0.007 -0.277 ...
## $ FreeSulfurDioxide : num  NA 15 214 22 -167 -37 287 523 -213 62 ...
## $ TotalSulfurDioxide: num  268 -327 142 115 108 15 156 551 NA 180 ...
## $ Density         : num  0.993 1.028 0.995 0.996 0.995 ...
## $ pH              : num  3.33 3.38 3.12 2.24 3.12 3.2 3.49 3.2 4.93 3.09 ...
## $ Sulphates       : num  -0.59 0.7 0.48 1.83 1.77 1.29 1.21 NA 0.26 0.75 ...
## $ Alcohol         : num  9.9 NA 22 6.2 13.7 15.4 10.3 11.6 15 12.6 ...
## $ LabelAppeal     : int  0 -1 -1 -1 0 0 0 1 0 0 ...
## $ AcidIndex       : int  8 7 8 6 9 11 8 7 6 8 ...
## $ STARS           : int  2 3 3 1 2 NA NA 3 NA 4 ...
```

Missing Data

```
for (i in colnames(df_wine_train)){
  print(paste(i, " ", sum(is.na(df_wine_train[,i])), sep = " "))
}
```

```
## [1] "INDEX 0"
## [1] "TARGET 0"
## [1] "FixedAcidity 0"
## [1] "VolatileAcidity 0"
## [1] "CitricAcid 0"
## [1] "ResidualSugar 616"
## [1] "Chlorides 638"
## [1] "FreeSulfurDioxide 647"
## [1] "TotalSulfurDioxide 682"
## [1] "Density 0"
## [1] "pH 395"
## [1] "Sulphates 1210"
## [1] "Alcohol 653"
## [1] "LabelAppeal 0"
## [1] "AcidIndex 0"
## [1] "STARS 3359"
```