Critical Thinking Group 4 - HW5 - Wine

Sreejaya, Suman, Vuthy

November 15, 2016

Overview

The objective of this assignment is to predict the number of sample cases of wine that will be sold based on the properties of the wine. A *count regression* model will be used to predict wine sales of sample cases.

Dataset
Wine - Training data
Wine - Evaluation Data

Data Exploration

Looks like the INDEX column name need to be corrected.

```
## Observations: 12,795
## Variables: 16
## $ INDEX              <int> 1, 2, 4, 5, 6, 7, 8, 11, 12, 13, 14, 15, 16...
## $ TARGET             <int> 3, 3, 5, 3, 4, 0, 0, 4, 3, 6, 0, 4, 3, 7, 4...
## $ FixedAcidity       <dbl> 3.2, 4.5, 7.1, 5.7, 8.0, 11.3, 7.7, 6.5, 14...
## $ VolatileAcidity    <dbl> 1.160, 0.160, 2.640, 0.385, 0.330, 0.320, 0...
## $ CitricAcid         <dbl> -0.98, -0.81, -0.88, 0.04, -1.26, 0.59, -0....
## $ ResidualSugar      <dbl> 54.20, 26.10, 14.80, 18.80, 9.40, 2.20, 21....
## $ Chlorides          <dbl> -0.567, -0.425, 0.037, -0.425, NA, 0.556, 0...
## $ FreeSulfurDioxide  <dbl> NA, 15, 214, 22, -167, -37, 287, 523, -213,...
## $ TotalSulfurDioxide <dbl> 268, -327, 142, 115, 108, 15, 156, 551, NA,...
## $ Density            <dbl> 0.99280, 1.02792, 0.99518, 0.99640, 0.99457...
## $ pH                 <dbl> 3.33, 3.38, 3.12, 2.24, 3.12, 3.20, 3.49, 3...
## $ Sulphates          <dbl> -0.59, 0.70, 0.48, 1.83, 1.77, 1.29, 1.21, ...
## $ Alcohol            <dbl> 9.9, NA, 22.0, 6.2, 13.7, 15.4, 10.3, 11.6,...
## $ LabelAppeal        <int> 0, -1, -1, -1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 2...
## $ AcidIndex          <int> 8, 7, 8, 6, 9, 11, 8, 7, 6, 8, 5, 10, 7, 8,...
## $ STARS              <int> 2, 3, 3, 1, 2, NA, NA, 3, NA, 4, 1, 2, 2, 3...
```

Missing Data

Eight of the variables have missing data.

ColName	NA_Count	NA_Percent
ResidualSugar	616	4.81
Chlorides	638	4.99
FreeSulfurDioxide	647	5.06
TotalSulfurDioxide	682	5.33
pH	395	3.09
Sulphates	1210	9.46
Alcohol	653	5.10
STARS	3359	26.25

Lets explore more on the missing values here:

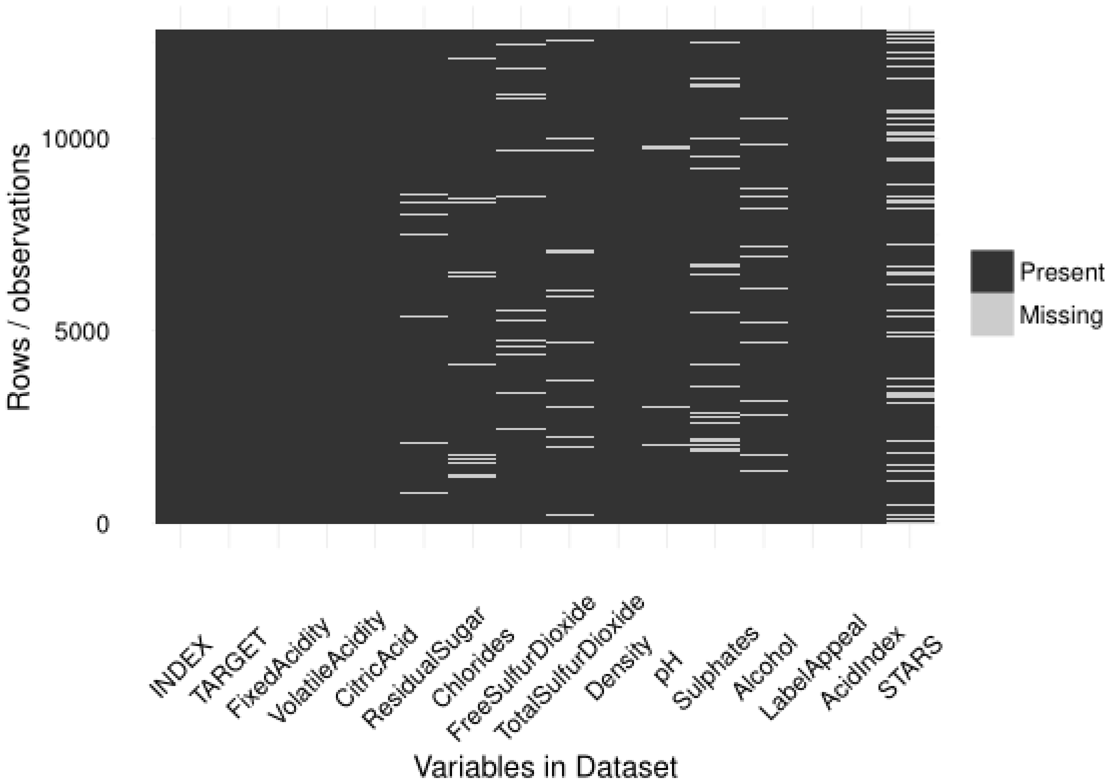
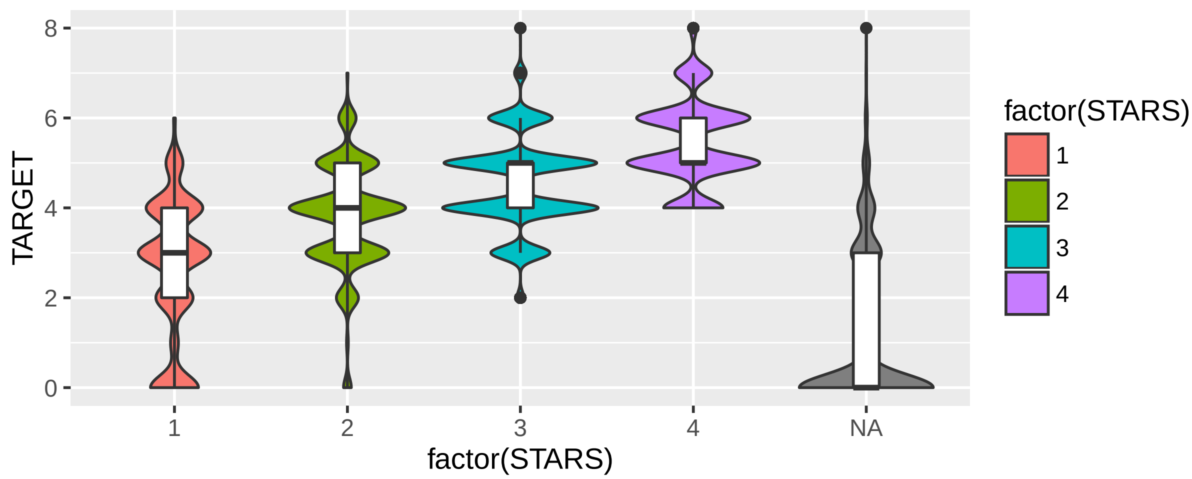
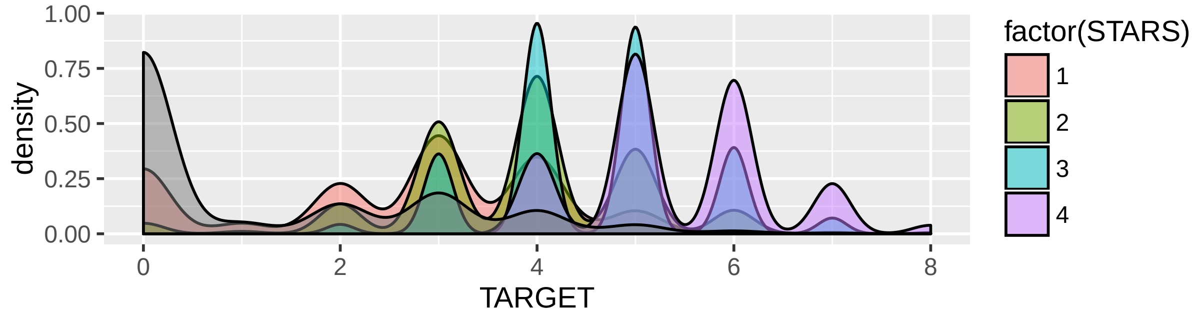


Figure 1:

Though there are lot of missing values, we could not see a definite pattern here, but we difinitely notice that there are highest number of missing values for *STARS* variable.

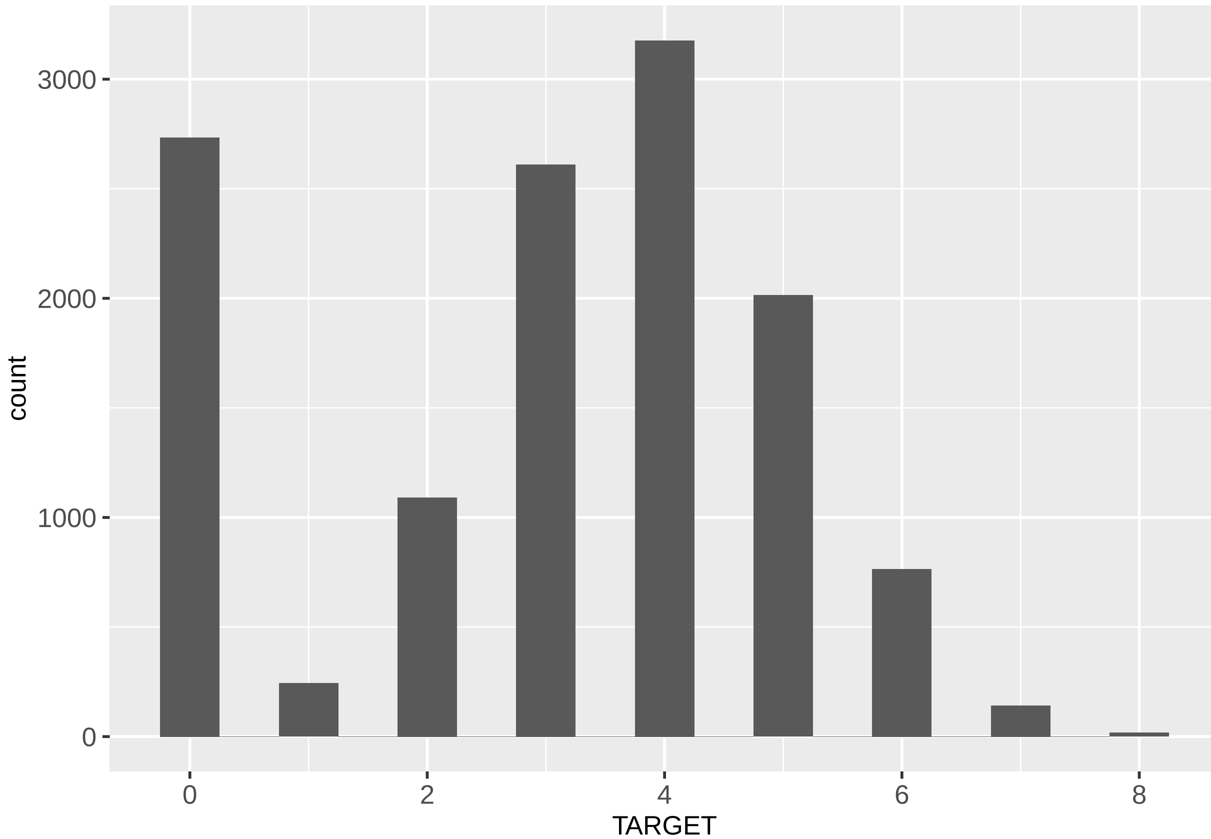




From the above diagrams, we notice that the NAs for STARS showing us a different distribution. So, we have to take care of this in the data preparation. (NA is valid category here)

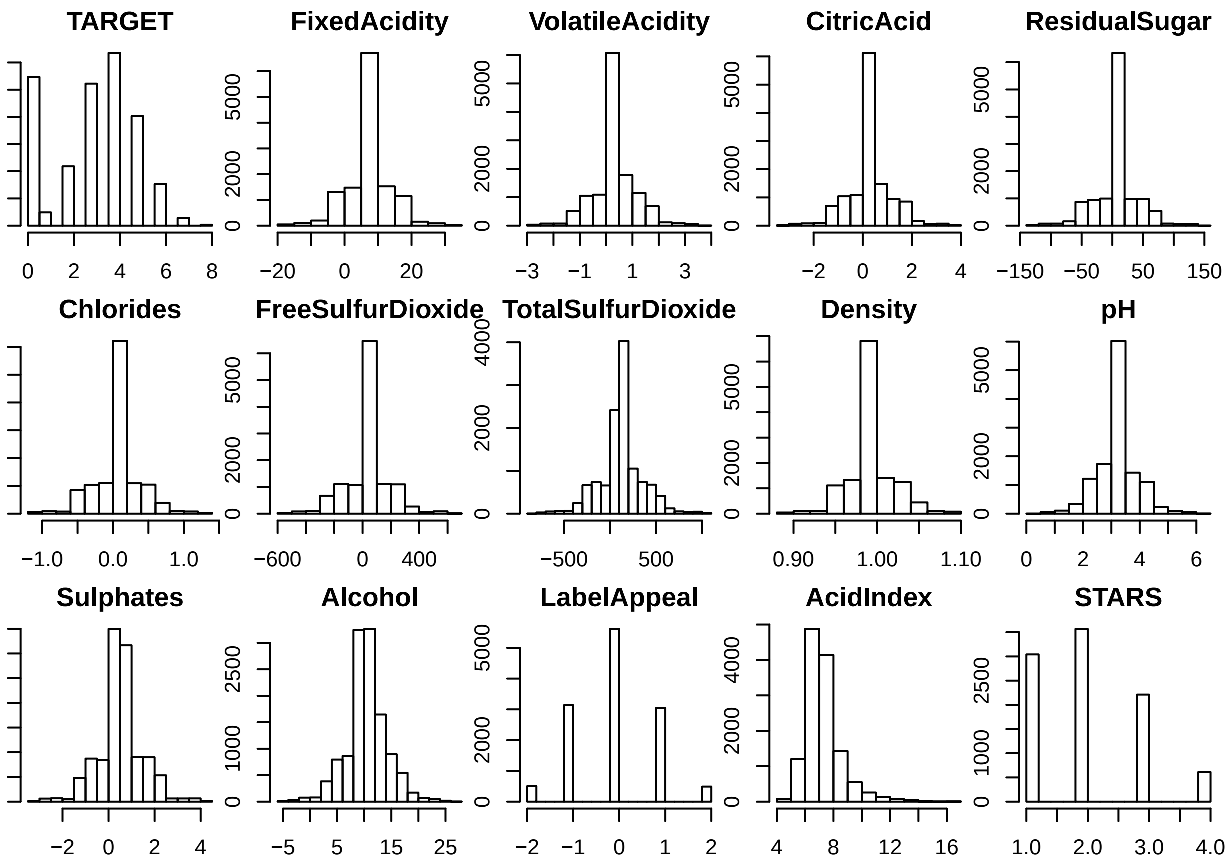
Data Distributions

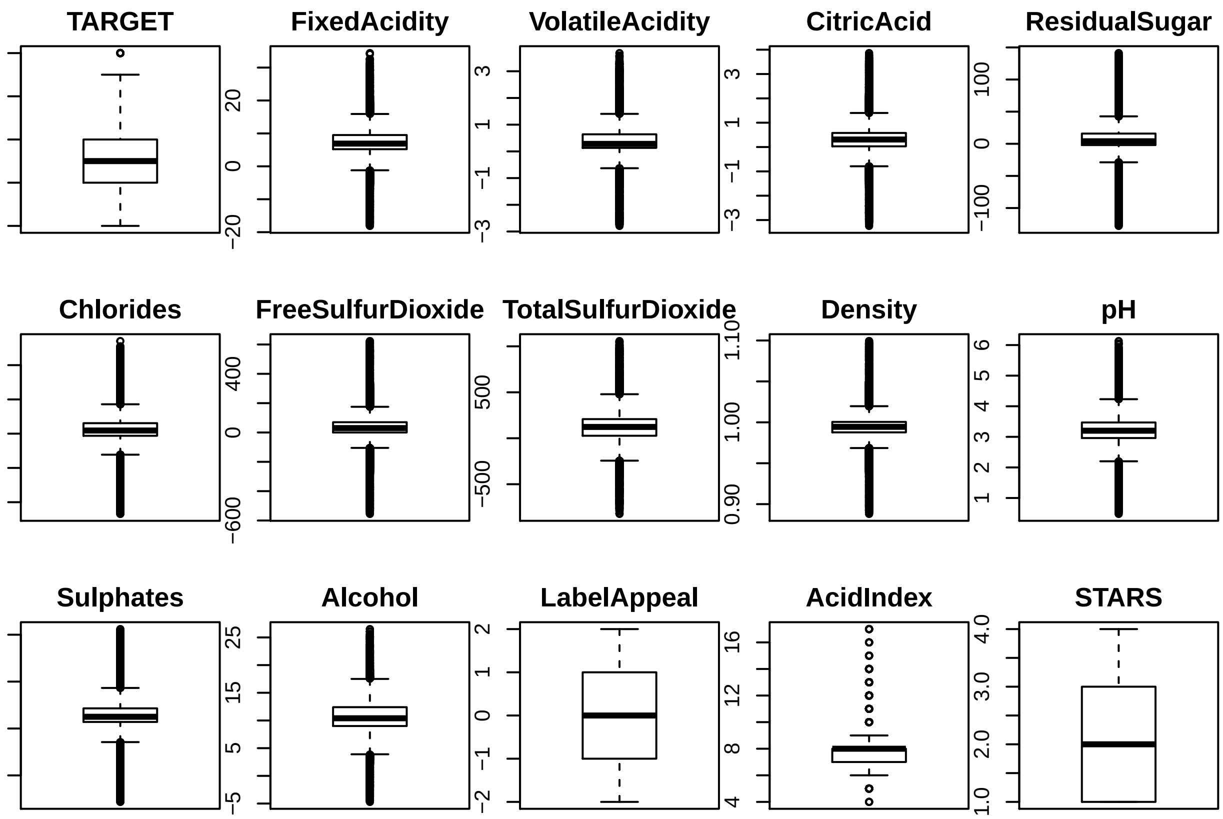
Lets check the overall distribution of the TAGET variable (which is a *count variable* indicating the number of sample cases):



The above *TARGET* distribution has lot of *ZERO* values, which would indicate the *no sample cases purchased*, which could be due to NA values presence Or, some business reasons. But overall this appears close to *Poisson* distribution.

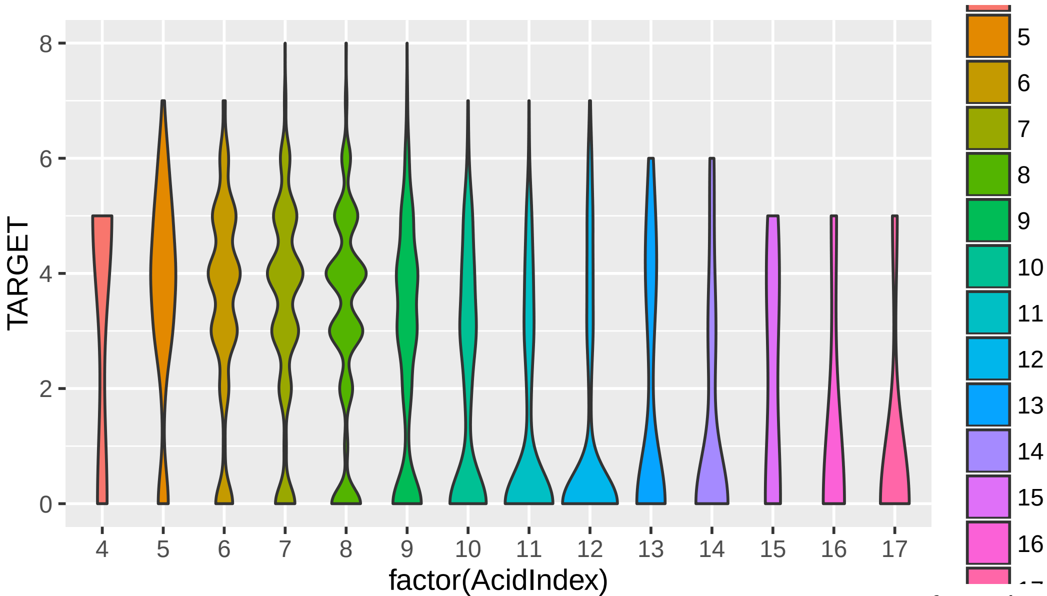
Lets check other variables distributions:

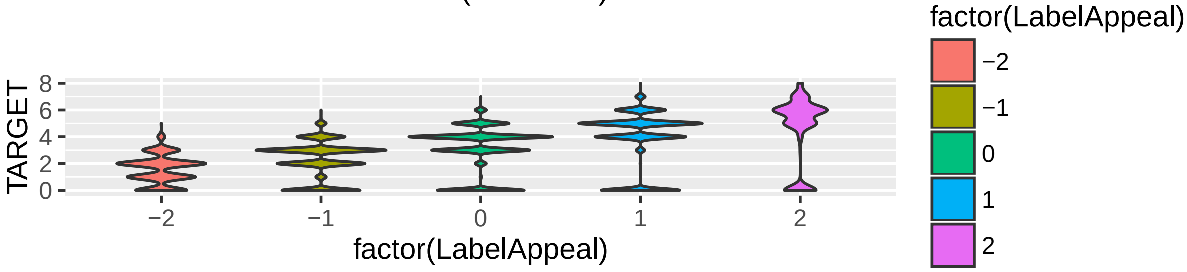

TARGET
FixedAcidity
VolatileAcidity
CitricAcid
ResidualSugar
Chlorides
FreeSulfurDioxide
TotalSulfurDioxide
Density
pH
Sulphates
Alcohol
LabelAppeal
AcidIndex
STARS



Majority of the variables appears to be *numerical and normally distributed.* Lets also review the *Ordinal* variables here:

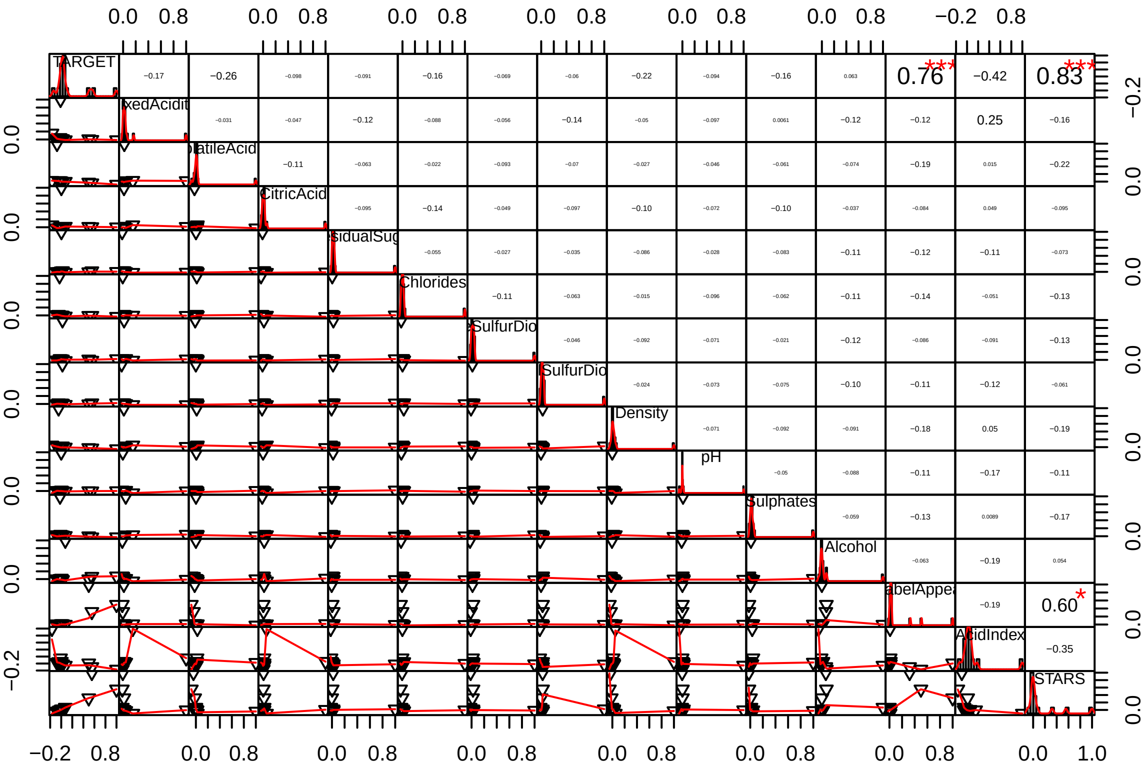
We have seen the *STARS* distribution previosly in *Missing Data* section, lets now review the *Acid Index*, and *LabelAppeal*, which can be treated as categorical similar to *STARS*:





Correlations

Lets visualize the correlation graph:



The above indicates the *STARS* and *LabelAppeal* are significant variables from correlation perspective. And *AcidIndex* and *VolatileAcidity* also got moderately correlated with the TARGET variable.

Data Preparation

Transform NAs

We will be modeling based on 2 different dataframes. One with *AcidIndex and LabelAppeal* as factor variables and the second one with numeric *AcidIndex and LabelAppeal*

Factorize

Lets factorize the *STARS* , *AcidIndex* and *LabelAppeal* for our first data frame.

lets take complete cases only in both cases, as we have got sufficient number of observations after we took care of the NAs for STARS and Alcohol variables.

Multicollinearity

Lets check for Multicollinearity in the predictors:

Numerica DataFrame:

	sort(vifFit1.numeric, decreasing = T)
STARS1	0.8910983
STARS2	0.8596886
STARS3	0.6821156
AcidIndex	0.4697698
LabelAppeal	0.3788666
FixedAcidity	0.3423291
STARS4	0.3376683
VolatileAcidity	0.3354500
Alcohol	0.3347811
TotalSulfurDioxide	0.3345273
Chlorides	0.3343219
ResidualSugar	0.3340220
pH	0.3339136
Sulphates	0.3333092
Density	0.3330000
FreeSulfurDioxide	0.3321442
CitricAcid	0.3306037

Categorical DataFrame:

	Multicolinearity score
AcidIndex7	216.4659703
AcidIndex8	201.6610282
AcidIndex9	92.2911942
AcidIndex6	77.1844205
AcidIndex10	39.3035671
AcidIndex11	19.5030397
AcidIndex12	10.5988711
AcidIndex5	5.5846113
AcidIndex13	5.2930521
AcidIndex14	4.5463148
LabelAppeal0	4.2704399
LabelAppeal-1	3.3390954
LabelAppeal1	3.2363811
AcidIndex17	1.5074415
AcidIndex16	1.2042507
AcidIndex15	1.1046126
STARS1	0.8934462
STARS2	0.8645223
LabelAppeal2	0.8417245
STARS3	0.6851982
FixedAcidity	0.3430885
STARS4	0.3388301
VolatileAcidity	0.3361709
TotalSulfurDioxide	0.3355376
Alcohol	0.3353972
ResidualSugar	0.3353051
Chlorides	0.3347751
pH	0.3346397
Sulphates	0.3338405
Density	0.3331687

	Multicolinearity score
FreeSulfurDioxide	0.3319512
CitricAcid	0.3305677

Multicollinearity noticed for AcidIndex dummy variables AcidIndex values 6, 7 , 8, 9, 10, 11, 12, for the data frame where the *AcidIndex and LabelAppeal* are *categorical*.

Lets try consolidating those rows and retry the vif again.

But there is no Multicollinearity noticed for any of the variables in our numeric dataframe. Therefore we will keep all the variables for modelling for the dataframe where the *AcidIndex and LabelAppeal* are *numerical*.

	Multicolinearity score
AcidIndex5	10.1576511
AcidIndex13	5.2928877
AcidIndex14	4.5462571
LabelAppeal0	4.2669710
LabelAppeal-1	3.3370975
LabelAppeal1	3.2315314
AcidIndex17	1.5074156
AcidIndex16	1.2042184
AcidIndex15	1.1046116
STARS1	0.8876861
STARS2	0.8534218
LabelAppeal2	0.8404632
STARS3	0.6715474
VolatileAcidity	0.3363584
FixedAcidity	0.3359405
Alcohol	0.3348208
ResidualSugar	0.3344179
TotalSulfurDioxide	0.3342609
STARS4	0.3340824
Chlorides	0.3339860
pH	0.3332575
Sulphates	0.3327584
FreeSulfurDioxide	0.3318813
Density	0.3317041
CitricAcid	0.3299254

The above variables looks good enough to proceed with model building.

Split the dataset into training and test:

We will randomly split our dataset into training (80%) and test (20%).

```
set.seed(3)

s0 = sample(1:nrow(wine.trn1.numeric.omit.na), 0.8 * nrow(wine.trn1.numeric.omit.na))
wine.training0 = wine.trn1.numeric.omit.na[s0, ]
wine.test0 = wine.trn1.numeric.omit.na[-s0, ]
```

```
s = sample(1:nrow(wine.trn.omit.na), 0.8 * nrow(wine.trn.omit.na))
wine.training = wine.trn.omit.na[s, ]
wine.test = wine.trn.omit.na[-s, ]
```

Number of observations in *training* dataset for categorical is 7306

Number of observations in *test* dataset for categorical is 1827

Number of observations in *training* dataset for numerical is 7306

Number of observations in *test* dataset for numerical is 1827

Build Models

Poisson Model - Stepwise Backward

First, Include all variables and build the model. And then use the stepwise backward.

```
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)              0.25       0.51    0.49     0.62
## FixedAcidity             0.00       0.00   -1.75     0.08
## VolatileAcidity         -0.03       0.01   -3.31     0.00
## Chlorides               -0.05       0.02   -2.33     0.02
## FreeSulfurDioxide        0.00       0.00    2.74     0.01
## TotalSulfurDioxide       0.00       0.00    3.72     0.00
## Density                 -0.58       0.25   -2.28     0.02
## Sulphates               -0.02       0.01   -2.29     0.02
## Alcohol                  0.00       0.00    1.45     0.15
## LabelAppeal-1            0.24       0.05    4.88     0.00
## LabelAppeal0             0.42       0.05    8.84     0.00
## LabelAppeal1             0.55       0.05   11.24     0.00
## LabelAppeal2             0.68       0.06   12.23     0.00
## AcidIndex5               0.11       0.45    0.25     0.80
## AcidIndex13             -0.44       0.47   -0.94     0.35
## AcidIndex14             -0.20       0.47   -0.43     0.67
## AcidIndex15              0.11       0.53    0.21     0.83
## AcidIndex16            -12.73     162.49   -0.08     0.94
## AcidIndex17             -0.28       0.63   -0.44     0.66
## STARS1                   0.79       0.03   30.48     0.00
## STARS2                   1.12       0.02   46.37     0.00
## STARS3                   1.24       0.03   49.06     0.00
## STARS4                   1.37       0.03   42.43     0.00
```

```
## TARGET ~ FixedAcidity + VolatileAcidity + Chlorides + FreeSulfurDioxide +
##     TotalSulfurDioxide + Density + Sulphates + Alcohol + LabelAppeal +
##     AcidIndex + STARS
```

Numerical:

```
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)              1.21       0.26    4.69     0.00
## VolatileAcidity         -0.03       0.01   -3.61     0.00
```

```
## Chlorides             -0.04       0.02   -1.96     0.05
## FreeSulfurDioxide      0.00       0.00    2.03     0.04
## TotalSulfurDioxide     0.00       0.00    3.27     0.00
## Density               -0.40       0.26   -1.57     0.12
## Sulphates             -0.01       0.01   -1.87     0.06
## LabelAppeal            0.16       0.01   19.69     0.00
## AcidIndex             -0.08       0.01  -12.97     0.00
## STARS1                 0.78       0.03   29.98     0.00
## STARS2                 1.09       0.02   45.04     0.00
## STARS3                 1.21       0.03   47.68     0.00
## STARS4                 1.33       0.03   40.74     0.00
```

```
## TARGET ~ VolatileAcidity + Chlorides + FreeSulfurDioxide + TotalSulfurDioxide +
##     Density + Sulphates + LabelAppeal + AcidIndex + STARS
```

We can notice that *STARS*, *LableAppeal*, *AcidIndex*, *VolatileAcidity* are the significant variables, also *TotalSulfurDioxide* is some what significant here.

For each one-unit increase in VolatileAcidity, the expected log count of the number of sample units sold is decreases by 0.03.

For each one-unit increase in LabelAppeal, the expected log count of the number of sample units sold is increased by 0.16.

The factor variable shown as STARS4 is the expected difference (1.33) in log count between group 4 and the reference group zero (/NA).

Categorical:

```
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)             0.25       0.51    0.49     0.62
## FixedAcidity            0.00       0.00   -1.75     0.08
## VolatileAcidity        -0.03       0.01   -3.31     0.00
## Chlorides              -0.05       0.02   -2.33     0.02
## FreeSulfurDioxide       0.00       0.00    2.74     0.01
## TotalSulfurDioxide      0.00       0.00    3.72     0.00
## Density                -0.58       0.25   -2.28     0.02
## Sulphates              -0.02       0.01   -2.29     0.02
## Alcohol                 0.00       0.00    1.45     0.15
## LabelAppeal-1           0.24       0.05    4.88     0.00
## LabelAppeal0            0.42       0.05    8.84     0.00
## LabelAppeal1            0.55       0.05   11.24     0.00
## LabelAppeal2            0.68       0.06   12.23     0.00
## AcidIndex5              0.11       0.45    0.25     0.80
## AcidIndex13            -0.44       0.47   -0.94     0.35
## AcidIndex14            -0.20       0.47   -0.43     0.67
## AcidIndex15             0.11       0.53    0.21     0.83
## AcidIndex16           -12.73     162.49   -0.08     0.94
## AcidIndex17            -0.28       0.63   -0.44     0.66
## STARS1                  0.79       0.03   30.48     0.00
## STARS2                  1.12       0.02   46.37     0.00
## STARS3                  1.24       0.03   49.06     0.00
## STARS4                  1.37       0.03   42.43     0.00
```

```
## TARGET ~ FixedAcidity + VolatileAcidity + Chlorides + FreeSulfurDioxide +
##     TotalSulfurDioxide + Density + Sulphates + Alcohol + LabelAppeal +
##     AcidIndex + STARS
```

We can notice that *STARS*, *LableAppeal*, *AcidIndex*, *VolatileAcidity* and *TotalSulfurDioxide* are the significant variables.

For example, for each one-unit increase in VolatileAcidity, the expected log count of the number of sample units sold is decreases by 0.03.

The factor variable shown as STARS4 is the expected difference in log count between group 4 and the reference group zero (/NA).

Lets check if there is overdispersion (c-hat, to check if mean exceeding the variance) here, (Residual Deviance)/(Residual df). (If c-hat is 1, then no overdispersion occur)

c-hat for overdispersion check is 1.0730764

Poisson Model - Stepwise Forward

```
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)            0.25       0.51    0.49     0.62
## STARS1                 0.79       0.03   30.48     0.00
## STARS2                 1.12       0.02   46.37     0.00
## STARS3                 1.24       0.03   49.06     0.00
## STARS4                 1.37       0.03   42.43     0.00
## LabelAppeal-1          0.24       0.05    4.88     0.00
## LabelAppeal0           0.42       0.05    8.84     0.00
## LabelAppeal1           0.55       0.05   11.24     0.00
## LabelAppeal2           0.68       0.06   12.23     0.00
## AcidIndex5             0.11       0.45    0.25     0.80
## AcidIndex13           -0.44       0.47   -0.94     0.35
## AcidIndex14           -0.20       0.47   -0.43     0.67
## AcidIndex15            0.11       0.53    0.21     0.83
## AcidIndex16          -12.73     162.49   -0.08     0.94
## AcidIndex17           -0.28       0.63   -0.44     0.66
## TotalSulfurDioxide     0.00       0.00    3.72     0.00
## VolatileAcidity       -0.03       0.01   -3.31     0.00
## FreeSulfurDioxide      0.00       0.00    2.74     0.01
## Chlorides             -0.05       0.02   -2.33     0.02
## Sulphates             -0.02       0.01   -2.29     0.02
## Density               -0.58       0.25   -2.28     0.02
## FixedAcidity           0.00       0.00   -1.75     0.08
## Alcohol                0.00       0.00    1.45     0.15
```

```
## TARGET ~ STARS + LabelAppeal + AcidIndex + TotalSulfurDioxide +
##     VolatileAcidity + FreeSulfurDioxide + Chlorides + Sulphates +
##     Density + FixedAcidity + Alcohol
```

c-hat for overdispersion check is 1.0730764

We notice the very similar results here. (Similar to Stepwise Backward), Hence the same interpretation applies here.

Poisson Model - Manual

Lets include only significant predictors noticed from the data exploration section.

Numerical:

```
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)            0.80       0.05   15.51     0.00
## VolatileAcidity       -0.03       0.01   -3.61     0.00
## Chlorides             -0.04       0.02   -2.01     0.04
## FreeSulfurDioxide      0.00       0.00    1.99     0.05
## TotalSulfurDioxide     0.00       0.00    3.24     0.00
## LabelAppeal            0.16       0.01   19.71     0.00
## AcidIndex             -0.08       0.01  -13.08     0.00
## STARS1                 0.78       0.03   30.00     0.00
## STARS2                 1.09       0.02   45.07     0.00
## STARS3                 1.22       0.03   47.76     0.00
## STARS4                 1.33       0.03   40.78     0.00
```

```
## TARGET ~ VolatileAcidity + Chlorides + FreeSulfurDioxide + TotalSulfurDioxide +
##     LabelAppeal + AcidIndex + STARS
```

We can notice that *STARS*, *LableAppeal*, *AcidIndex*, *VolatileAcidity* are the significant variables, also *TotalSulfurDioxide* is some what significant here.

For each one-unit increase in VolatileAcidity, the expected log count of the number of sample units sold is decreases by 0.03.

For each one-unit increase in LabelAppeal, the expected log count of the number of sample units sold is increased by 0.16.

The factor variable shown as STARS4 is the expected difference (1.33) in log count between group 4 and the reference group zero (/NA).

Categorical:

```
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -0.32       0.45   -0.71     0.48
## STARS1              0.80       0.03   30.71     0.00
## STARS2              1.13       0.02   46.63     0.00
## STARS3              1.25       0.03   49.57     0.00
## STARS4              1.38       0.03   42.82     0.00
## LabelAppeal-1       0.24       0.05    4.93     0.00
## LabelAppeal0        0.43       0.05    8.91     0.00
## LabelAppeal1        0.55       0.05   11.28     0.00
## LabelAppeal2        0.68       0.06   12.22     0.00
## AcidIndex5          0.12       0.45    0.26     0.80
## AcidIndex13        -0.46       0.47   -0.97     0.33
## AcidIndex14        -0.22       0.47   -0.47     0.64
## AcidIndex15         0.12       0.53    0.23     0.82
## AcidIndex16       -12.76     162.29   -0.08     0.94
## AcidIndex17        -0.37       0.63   -0.58     0.56
## VolatileAcidity    -0.03       0.01   -3.37     0.00
```

```
## TARGET ~ STARS + LabelAppeal + AcidIndex + VolatileAcidity
```

We only included the above significant variables we noticed from our correlation here, so this model has got few co-efficients compared with the above.

c-hat for overdispersion check is 1.0779548

Negative Binomial Model - Stepwise Backward

Lets now try with Negative Binomial modeling, which fits greately for over-dispersed count outcome variables.

First, Include all variables and build the model. And then use the stepwise backward.

Categorical

```
##                     Estimate  Std. Error z value Pr(>|z|)
## (Intercept)             0.25        0.51    0.49     0.62
## FixedAcidity            0.00        0.00   -1.75     0.08
## VolatileAcidity        -0.03        0.01   -3.31     0.00
## Chlorides              -0.05        0.02   -2.33     0.02
## FreeSulfurDioxide       0.00        0.00    2.74     0.01
## TotalSulfurDioxide      0.00        0.00    3.72     0.00
## Density                -0.58        0.25   -2.28     0.02
## Sulphates              -0.02        0.01   -2.29     0.02
## Alcohol                 0.00        0.00    1.45     0.15
## LabelAppeal-1           0.24        0.05    4.88     0.00
## LabelAppeal0            0.42        0.05    8.84     0.00
## LabelAppeal1            0.55        0.05   11.24     0.00
## LabelAppeal2            0.68        0.06   12.23     0.00
## AcidIndex5              0.11        0.45    0.25     0.80
## AcidIndex13            -0.44        0.47   -0.94     0.35
## AcidIndex14            -0.20        0.47   -0.43     0.67
## AcidIndex15             0.11        0.53    0.21     0.83
## AcidIndex16           -37.44 38745320.70    0.00     1.00
## AcidIndex17            -0.28        0.63   -0.44     0.66
## STARS1                  0.79        0.03   30.48     0.00
## STARS2                  1.12        0.02   46.37     0.00
## STARS3                  1.24        0.03   49.06     0.00
## STARS4                  1.37        0.03   42.43     0.00
```

```
## TARGET ~ FixedAcidity + VolatileAcidity + Chlorides + FreeSulfurDioxide +
##     TotalSulfurDioxide + Density + Sulphates + Alcohol + LabelAppeal +
##     AcidIndex + STARS
```

We noticed that our dataset do NOT has lot of overdispersion (based on poission model above), so the negative binomial results are very much close to the poission.

For example, for each one-unit increase in VolatileAcidity, the expected log count of the number of sample units sold is decreases by 0.031.

The factor variable shown as STARS1 is the expected difference [0.80] in log count between group 1 and the reference group zero (/NA).

Numerical:

```
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)             1.21       0.26    4.69     0.00
## VolatileAcidity        -0.03       0.01   -3.61     0.00
## Chlorides              -0.04       0.02   -1.96     0.05
## FreeSulfurDioxide       0.00       0.00    2.03     0.04
## TotalSulfurDioxide      0.00       0.00    3.27     0.00
## Density                -0.40       0.26   -1.57     0.12
## Sulphates              -0.01       0.01   -1.87     0.06
## LabelAppeal             0.16       0.01   19.69     0.00
## AcidIndex              -0.08       0.01  -12.97     0.00
## STARS1                  0.78       0.03   29.98     0.00
## STARS2                  1.09       0.02   45.04     0.00
## STARS3                  1.21       0.03   47.67     0.00
## STARS4                  1.33       0.03   40.74     0.00
```

```
## TARGET ~ VolatileAcidity + Chlorides + FreeSulfurDioxide + TotalSulfurDioxide +
##     Density + Sulphates + LabelAppeal + AcidIndex + STARS
```

Results are similar to Poisson as described above, in numerical case as well.

Negative Binomial Model - Stepwise Forward

```
##                     Estimate  Std. Error z value Pr(>|z|)
## (Intercept)             0.25        0.51    0.49     0.62
## STARS1                  0.79        0.03   30.48     0.00
## STARS2                  1.12        0.02   46.37     0.00
## STARS3                  1.24        0.03   49.06     0.00
## STARS4                  1.37        0.03   42.43     0.00
## LabelAppeal-1           0.24        0.05    4.88     0.00
## LabelAppeal0            0.42        0.05    8.84     0.00
## LabelAppeal1            0.55        0.05   11.24     0.00
## LabelAppeal2            0.68        0.06   12.23     0.00
## AcidIndex5              0.11        0.45    0.25     0.80
## AcidIndex13            -0.44        0.47   -0.94     0.35
## AcidIndex14            -0.20        0.47   -0.43     0.67
## AcidIndex15             0.11        0.53    0.21     0.83
## AcidIndex16           -37.44 38745320.70    0.00     1.00
## AcidIndex17            -0.28        0.63   -0.44     0.66
## TotalSulfurDioxide      0.00        0.00    3.72     0.00
## VolatileAcidity        -0.03        0.01   -3.31     0.00
## FreeSulfurDioxide       0.00        0.00    2.74     0.01
## Chlorides              -0.05        0.02   -2.33     0.02
## Sulphates              -0.02        0.01   -2.29     0.02
## Density                -0.58        0.25   -2.28     0.02
## FixedAcidity            0.00        0.00   -1.75     0.08
## Alcohol                 0.00        0.00    1.45     0.15
```

```
## TARGET ~ STARS + LabelAppeal + AcidIndex + TotalSulfurDioxide +
##     VolatileAcidity + FreeSulfurDioxide + Chlorides + Sulphates +
##     Density + FixedAcidity + Alcohol
```

This provides us with the similar results as Stepwise Backward.

Negative Binomial Model - Manual

Lets include only significant predictors noticed from the data exploration section. Since in the dataset with all numeric values Density does not seems significant, so we decide to remove it

Categorical:

```
##                   Estimate  Std. Error z value Pr(>|z|)
## (Intercept)          -0.32        0.45   -0.71     0.48
## STARS1                0.80        0.03   30.71     0.00
## STARS2                1.13        0.02   46.63     0.00
## STARS3                1.25        0.03   49.56     0.00
## STARS4                1.38        0.03   42.82     0.00
## LabelAppeal-1         0.24        0.05    4.93     0.00
## LabelAppeal0          0.43        0.05    8.91     0.00
## LabelAppeal1          0.55        0.05   11.28     0.00
## LabelAppeal2          0.68        0.06   12.22     0.00
## AcidIndex5            0.12        0.45    0.26     0.80
## AcidIndex13          -0.46        0.47   -0.97     0.33
## AcidIndex14          -0.22        0.47   -0.47     0.64
## AcidIndex15           0.12        0.53    0.23     0.82
## AcidIndex16         -37.42 37494215.41    0.00     1.00
## AcidIndex17          -0.37        0.63   -0.58     0.56
## VolatileAcidity      -0.03        0.01   -3.37     0.00
```

```
## TARGET ~ STARS + LabelAppeal + AcidIndex + VolatileAcidity
```

From the above, we can see that, in this model *Acid Index* is not significant, however *STARS*, *LabelAppeal* and *VolatileAcidity* are significant.

Numerical:

```
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)              0.80       0.05   15.51     0.00
## VolatileAcidity         -0.03       0.01   -3.61     0.00
## Chlorides               -0.04       0.02   -2.01     0.04
## FreeSulfurDioxide        0.00       0.00    1.99     0.05
## TotalSulfurDioxide       0.00       0.00    3.24     0.00
## LabelAppeal              0.16       0.01   19.71     0.00
## AcidIndex               -0.08       0.01  -13.08     0.00
## STARS1                   0.78       0.03   30.00     0.00
## STARS2                   1.09       0.02   45.06     0.00
## STARS3                   1.22       0.03   47.76     0.00
## STARS4                   1.33       0.03   40.78     0.00
```

```
## TARGET ~ VolatileAcidity + Chlorides + FreeSulfurDioxide + TotalSulfurDioxide +
##     LabelAppeal + AcidIndex + STARS
```

In numerical case, the signficant variables are pretty much same as in the poisson case (including the coefficient estimates).

We only included the few significant variables in the above manual models (from correlations), hence the manual model has got few co-efficients compared with the non-manual ones.

Linear Model - Stepwise Backward

Lets now just try with multiple linear regression model, and see the outcome.

Numerical

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.93       0.58    6.77     0.00
## VolatileAcidity       -0.09       0.02   -4.86     0.00
## Chlorides             -0.13       0.05   -2.67     0.01
## FreeSulfurDioxide      0.00       0.00    2.68     0.01
## TotalSulfurDioxide     0.00       0.00    4.26     0.00
## Density               -1.13       0.58   -1.95     0.05
## Sulphates             -0.03       0.02   -1.97     0.05
## LabelAppeal            0.47       0.02   25.82     0.00
## AcidIndex             -0.19       0.01  -16.10     0.00
## STARS1                 1.39       0.04   31.84     0.00
## STARS2                 2.39       0.04   56.59     0.00
## STARS3                 2.98       0.05   61.01     0.00
## STARS4                 3.67       0.08   45.95     0.00
```

```
## TARGET ~ VolatileAcidity + Chlorides + FreeSulfurDioxide + TotalSulfurDioxide +
##     Density + Sulphates + LabelAppeal + AcidIndex + STARS
```

In case of linear model, the significant variables are similar to the *poisson* and *negative binomial*, which are *STARS*, *AcidIndex*, *LabelAppeal*, *TotalSulfurDioxide* and *VolatileAcidity*

For example, a unit increase in VolatileAcidity can be result in decrease of 0.09 in TARGET variable, keeping the other variables constant.

Categorical

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)            1.59       1.44    1.10     0.27
## FixedAcidity          -0.01       0.00   -2.17     0.03
## VolatileAcidity       -0.09       0.02   -4.62     0.00
## Chlorides             -0.16       0.05   -3.18     0.00
## FreeSulfurDioxide      0.00       0.00    3.52     0.00
## TotalSulfurDioxide     0.00       0.00    4.91     0.00
## Density               -1.70       0.58   -2.90     0.00
## Sulphates             -0.04       0.02   -2.63     0.01
## Alcohol                0.01       0.00    2.04     0.04
## LabelAppeal-1          0.37       0.08    4.44     0.00
## LabelAppeal0           0.83       0.08   10.18     0.00
## LabelAppeal1           1.26       0.08   14.85     0.00
## LabelAppeal2           1.82       0.11   15.99     0.00
## AcidIndex5             0.53       1.32    0.40     0.69
## AcidIndex13           -0.68       1.34   -0.50     0.62
## AcidIndex14            0.01       1.35    0.01     0.99
## AcidIndex15            0.60       1.45    0.41     0.68
## AcidIndex16           -1.35       1.53   -0.88     0.38
## AcidIndex17           -0.34       1.53   -0.22     0.83
```

```
## STARS1                 1.44       0.04   32.63     0.00
## STARS2                 2.51       0.04   58.68     0.00
## STARS3                 3.06       0.05   62.61     0.00
## STARS4                 3.78       0.08   47.39     0.00
```

```
## TARGET ~ FixedAcidity + VolatileAcidity + Chlorides + FreeSulfurDioxide +
##     TotalSulfurDioxide + Density + Sulphates + Alcohol + LabelAppeal +
##     AcidIndex + STARS
```

For categorical data, the significant variables include *STARS*, *LabelAppeal*, *TotalSulfurDioxide*, *FreeSulfurDioxide*, and *VolatileAcidity*.

Linear Model - Stepwise Forward

```
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.59       1.44    1.10     0.27
## STARS1                  1.44       0.04   32.63     0.00
## STARS2                  2.51       0.04   58.68     0.00
## STARS3                  3.06       0.05   62.61     0.00
## STARS4                  3.78       0.08   47.39     0.00
## LabelAppeal-1           0.37       0.08    4.44     0.00
## LabelAppeal0            0.83       0.08   10.18     0.00
## LabelAppeal1            1.26       0.08   14.85     0.00
## LabelAppeal2            1.82       0.11   15.99     0.00
## AcidIndex5              0.53       1.32    0.40     0.69
## AcidIndex13            -0.68       1.34   -0.50     0.62
## AcidIndex14             0.01       1.35    0.01     0.99
## AcidIndex15             0.60       1.45    0.41     0.68
## AcidIndex16            -1.35       1.53   -0.88     0.38
## AcidIndex17            -0.34       1.53   -0.22     0.83
## TotalSulfurDioxide      0.00       0.00    4.91     0.00
## VolatileAcidity        -0.09       0.02   -4.62     0.00
## FreeSulfurDioxide       0.00       0.00    3.52     0.00
## Chlorides              -0.16       0.05   -3.18     0.00
## Density                -1.70       0.58   -2.90     0.00
## Sulphates              -0.04       0.02   -2.63     0.01
## FixedAcidity           -0.01       0.00   -2.17     0.03
## Alcohol                 0.01       0.00    2.04     0.04
```

```
## TARGET ~ STARS + LabelAppeal + AcidIndex + TotalSulfurDioxide +
##     VolatileAcidity + FreeSulfurDioxide + Chlorides + Density +
##     Sulphates + FixedAcidity + Alcohol
```

Stepwise Forward results are similar to the Stepwise backward linear model.

Linear Model - Manual

Numerical:

```
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)             2.80       0.10   27.91     0.00
```

```
## VolatileAcidity       -0.09       0.02   -4.86     0.00
## Chlorides             -0.13       0.05   -2.72     0.01
## FreeSulfurDioxide      0.00       0.00    2.64     0.01
## TotalSulfurDioxide     0.00       0.00    4.23     0.00
## LabelAppeal            0.47       0.02   25.87     0.00
## AcidIndex             -0.19       0.01  -16.27     0.00
## STARS1                 1.40       0.04   31.87     0.00
## STARS2                 2.40       0.04   56.61     0.00
## STARS3                 2.98       0.05   61.12     0.00
## STARS4                 3.67       0.08   45.97     0.00

## TARGET ~ VolatileAcidity + Chlorides + FreeSulfurDioxide + TotalSulfurDioxide +
##     LabelAppeal + AcidIndex + STARS
```

All the variables we included here are significant. The co-efficients are similar to our poisson, negative binomial models. For example, a unit increase in LabelAppeal would result in 0.46 increase of the TARGET variable.

Categorical:

```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -0.06       1.33   -0.04     0.97
## STARS1               1.45       0.04   32.87     0.00
## STARS2               2.52       0.04   58.90     0.00
## STARS3               3.09       0.05   63.13     0.00
## STARS4               3.81       0.08   47.63     0.00
## LabelAppeal-1        0.37       0.08    4.47     0.00
## LabelAppeal0         0.83       0.08   10.21     0.00
## LabelAppeal1         1.26       0.09   14.81     0.00
## LabelAppeal2         1.81       0.11   15.89     0.00
## AcidIndex5           0.53       1.33    0.40     0.69
## AcidIndex13         -0.74       1.35   -0.55     0.59
## AcidIndex14         -0.06       1.35   -0.05     0.96
## AcidIndex15          0.63       1.46    0.44     0.66
## AcidIndex16         -1.41       1.53   -0.92     0.36
## AcidIndex17         -0.49       1.54   -0.32     0.75
## VolatileAcidity     -0.09       0.02   -4.67     0.00

## TARGET ~ STARS + LabelAppeal + AcidIndex + VolatileAcidity
```

Interpretation of the categorical data is little difficult here, for example, a unit increase in STARS1 in reference to STARS(NA) would result increase of 1.45 in the TARGET variable, keeping the other variables constant.

Model Selection

Lets prepare a validation results data frame by deriving the validation metrics like, RMSE, R^2 (for linear model only) and AIC and number of coefficients etc., for both the dataframes , one that treats the *AcidIndex and LabelAppeal* as categorical, and the other as numerical, to help decide a better model out of the above 15 models.

Validation Results (AcidIndex and LabelAppeal as Categorical)

ModelType	RMSE	Adj_R2	AIC	Coefs
Poisson - Stepwise Backward	2.6125	NA	26200.28	22
Poisson - Stepwise Forward	2.6125	NA	26200.28	22
Poisson - Manual	2.6125	NA	26229.36	15
Negative Binomial - Backward	2.7576	NA	26202.52	22
Negative Binomial - Forward	2.7576	NA	26202.52	22
Negative Binomial - Manual	2.7570	NA	26231.60	15
Linear - Stepwise Backward	1.3710	0.52	24832.12	22
Linear - Stepwise Forward	1.3710	0.52	24832.12	22
Linear - Manual	1.3725	0.52	24890.70	15

Validation Results (AcidIndex and LabelAppeal as numeric)

ModelType	RMSE	Adj_R2	AIC	Coefs
Poisson - Step model	2.5754	NA	26062.22	12
Poisson - Manual	2.5754	NA	26064.13	10
Negative Binomial - Step model	2.5754	NA	26064.46	12
Negative Binomial - Manual	2.5748	NA	26066.37	10
Linear - Step model	1.3298	0.54	24651.77	12
Linear - Manual	1.3306	0.53	24655.39	10

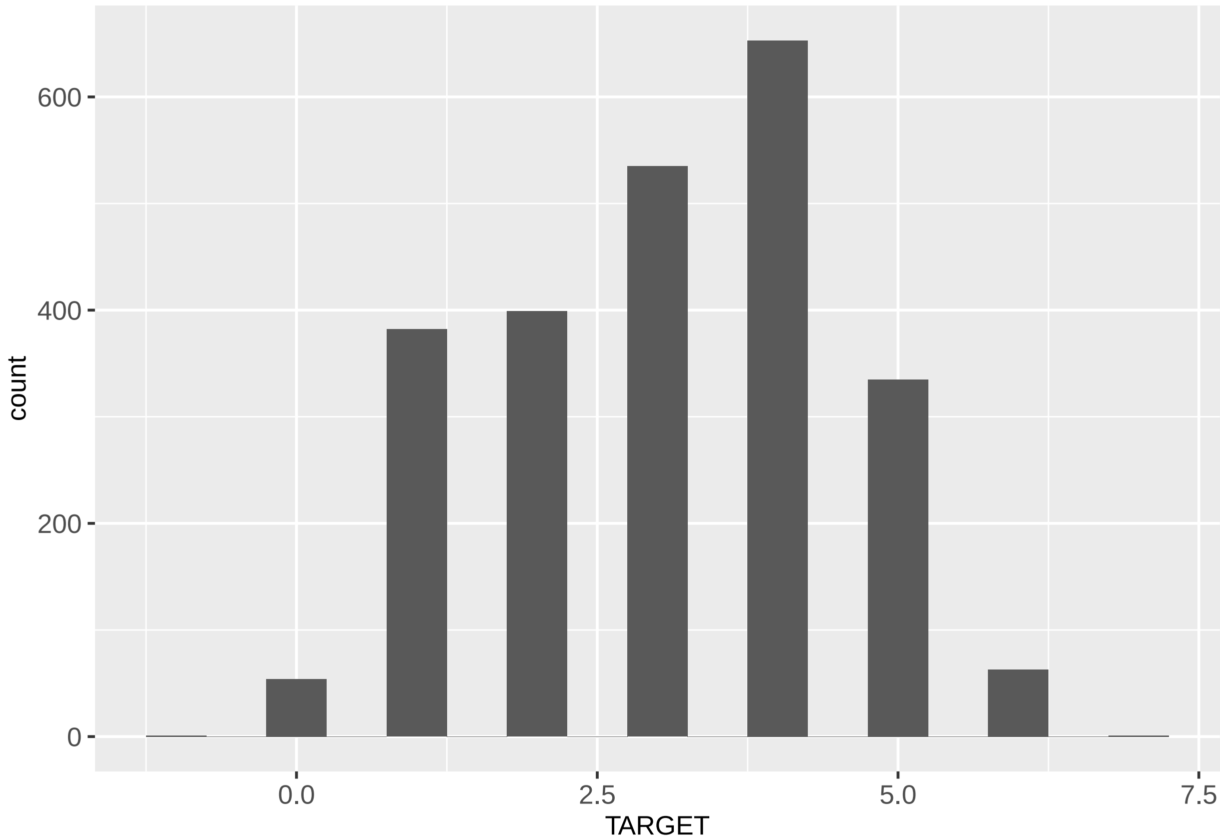
Since we are comparing different types of models, its tricky to select a common metric for these.

For our evaluation, lets consider the model that had least RMSE, AIC and probably minimal number of Coefs - which in our case is the *Linear - Step model* of numerical dataframe.

Evaluation

Lets do the data transformation first for our eval data frame, and then predict.

Lets quickly review the distribution of the *TARGET* variable:



Mean of the below TARGET distribution of eval dataset is 3.08

Var of the TARGET distribution of eval dataset is 2.06

We notice underdispersion here (with both linear as well as poisson models applied and verified on the eval data), our further analysis may include the usage of the *Generalized Poisson Regression* using *VGAM* package. Also, we notice excessive zeros counts, and we might think of using the *Zero-Inflated Poisson Regression* for further analysis.

Table 7: Predictions

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
3	0	-1	1	5428	0	0	2	10766	0	-1	NA
9	2	0	4	5430	4	0	NA	10776	0	-2	0
10	1	0	2	5433	0	0	1	10783	1	0	3
18	1	-1	2	5437	2	-1	3	10789	2	-1	3
21	0	0	1	5440	1	1	NA	10790	2	2	5
30	4	1	5	5442	3	1	5	10797	0	1	1
31	3	0	3	5445	1	0	3	10807	1	0	3
37	0	1	2	5449	2	0	NA	10810	0	0	1
39	0	0	0	5452	2	0	4	10817	1	1	3
47	0	0	2	5460	0	-1	1	10820	1	-1	3
60	1	0	NA	5461	1	-1	2	10822	2	0	4
62	0	1	1	5465	0	-1	1	10828	2	1	4
63	2	0	4	5467	3	0	NA	10829	1	-1	NA
64	0	0	NA	5471	2	1	4	10830	2	-2	3
68	0	-1	1	5474	0	0	2	10831	4	1	6

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
75	2	-1	3	5475	2	-1	NA	10841	3	2	5
76	1	-1	3	5480	0	-1	NA	10847	2	-1	NA
83	0	1	1	5481	2	0	4	10856	0	-1	0
87	2	0	4	5484	0	0	1	10860	0	0	NA
92	3	1	5	5494	2	2	5	10861	3	1	NA
98	2	-2	3	5495	0	-1	1	10863	1	0	3
106	0	0	2	5497	0	0	2	10875	1	0	NA
107	0	2	2	5499	2	-1	3	10884	3	0	4
113	1	-1	2	5507	0	0	1	10895	0	-1	1
120	2	0	4	5510	1	1	3	10897	1	0	3
123	2	2	5	5515	0	0	1	10898	1	0	NA
125	2	-1	3	5516	1	-2	2	10903	0	1	1
126	4	1	6	5517	0	-1	1	10908	0	1	2
128	4	0	5	5524	3	0	4	10924	1	0	2
129	1	-1	2	5530	3	2	5	10926	1	-1	2
131	3	0	NA	5534	1	-1	3	10927	1	0	3
135	0	1	2	5543	0	0	1	10928	1	0	3
141	3	0	4	5545	1	0	3	10933	0	0	1
147	1	0	3	5558	2	0	4	10939	4	1	6
148	0	-1	1	5562	0	0	2	10942	2	1	4
151	2	0	4	5573	4	2	6	10945	2	-1	4
156	1	1	3	5581	3	0	4	10949	2	0	4
157	3	-1	4	5583	3	1	5	10950	1	0	3
174	0	-1	1	5587	3	-1	4	10958	3	2	5
186	0	1	2	5589	1	-1	2	10963	2	0	4
193	1	-1	2	5591	4	1	5	10967	2	-1	4
195	0	0	NA	5596	1	0	3	10971	0	0	NA
212	0	0	1	5606	3	0	4	10972	0	0	2
213	0	0	1	5608	2	1	5	10974	2	0	4
217	1	0	NA	5611	2	-1	3	10976	4	0	5
223	2	0	4	5612	2	1	4	10980	1	0	3
226	1	0	3	5614	3	0	NA	10991	0	0	2
228	2	1	4	5620	2	0	4	10995	3	0	NA
230	3	0	4	5623	3	1	5	11014	3	1	5
241	1	0	2	5624	1	2	3	11017	2	1	NA
243	2	0	4	5626	3	2	5	11019	2	1	4
249	0	0	1	5633	0	-2	0	11022	0	0	NA
281	3	0	4	5635	2	-1	3	11030	3	0	5
288	0	0	1	5640	2	1	NA	11031	2	-1	3
294	1	-2	2	5643	1	0	NA	11041	0	0	2
295	1	-2	2	5644	4	1	6	11042	2	0	4
300	2	2	NA	5653	3	1	5	11044	3	0	NA
302	2	1	4	5663	3	0	NA	11047	2	1	NA
303	0	0	2	5664	3	1	5	11048	1	0	NA
308	0	-1	1	5667	1	-1	2	11049	1	0	3
319	2	1	NA	5671	0	0	1	11052	1	0	3
320	0	0	NA	5673	2	0	4	11058	0	0	1
324	1	0	NA	5676	2	-2	3	11069	2	-1	3
331	1	-1	3	5678	1	0	2	11070	2	-1	NA
343	1	0	3	5698	1	-1	2	11073	3	1	NA
347	1	-1	2	5700	4	1	6	11074	0	-1	1
348	2	0	4	5705	2	0	4	11078	0	0	2

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
350	2	1	4	5706	2	2	NA	11079	0	0	1
357	0	-2	0	5711	0	0	NA	11085	0	0	1
358	2	0	4	5712	3	1	5	11088	2	0	4
360	3	0	NA	5716	2	-1	3	11106	0	0	1
366	1	1	3	5719	1	1	3	11110	4	1	NA
367	1	-1	2	5725	1	-1	NA	11114	3	-1	4
368	3	1	5	5728	4	2	6	11118	1	-1	2
376	1	-1	2	5734	0	-1	0	11129	2	1	5
380	1	0	3	5735	3	0	4	11130	2	0	4
388	0	-2	0	5743	1	-1	3	11131	2	0	4
396	2	1	NA	5754	0	-1	1	11133	1	-1	NA
398	4	0	5	5755	1	0	NA	11138	3	-1	4
403	3	0	NA	5756	2	1	NA	11143	2	0	4
410	1	0	2	5766	1	0	3	11146	4	0	5
412	0	2	2	5770	2	0	4	11153	2	0	4
420	1	-1	2	5774	0	0	1	11162	1	2	4
434	1	-1	2	5775	1	-1	2	11170	4	2	6
440	1	0	3	5776	2	0	4	11171	0	0	1
450	2	0	4	5778	4	1	6	11201	2	-1	3
453	1	0	3	5786	2	-1	3	11216	3	1	5
464	4	0	5	5787	2	0	4	11219	1	0	3
465	2	1	NA	5791	3	1	5	11222	2	1	5
466	4	0	5	5794	2	-1	NA	11234	0	0	1
473	1	-1	2	5803	1	0	3	11238	2	0	4
476	0	0	2	5804	1	1	NA	11244	2	-1	4
478	0	0	2	5808	0	1	1	11246	0	-2	1
479	2	-1	3	5810	2	0	4	11248	0	2	3
493	1	-1	3	5813	2	0	4	11250	0	0	1
497	3	-1	NA	5828	2	-1	4	11256	2	1	4
503	2	0	4	5839	4	1	6	11259	1	-1	2
504	2	0	3	5842	3	0	4	11263	0	0	1
505	2	-2	3	5843	2	-1	2	11264	1	0	NA
507	0	1	1	5844	2	0	4	11270	0	-1	1
513	1	0	3	5847	1	0	3	11274	0	0	1
519	1	-1	2	5851	1	-1	2	11281	2	-1	NA
521	2	0	4	5854	3	-1	NA	11285	0	0	1
522	2	0	4	5857	0	0	1	11300	1	-1	2
545	1	0	NA	5866	0	-2	1	11305	2	-1	3
549	0	0	2	5874	0	0	NA	11317	3	-1	4
551	0	0	1	5886	2	0	NA	11319	1	-1	2
556	3	2	NA	5895	1	0	2	11330	0	1	NA
557	4	1	6	5897	1	-1	3	11334	3	1	5
559	0	-1	1	5898	1	0	3	11335	4	1	NA
560	1	-1	NA	5900	0	1	2	11336	1	0	3
566	2	0	4	5902	0	1	2	11356	3	0	4
569	2	1	4	5908	0	0	1	11358	2	0	3
573	1	-1	NA	5909	1	0	3	11360	0	0	1
578	0	0	1	5912	4	1	6	11364	0	0	1
579	3	1	NA	5913	0	0	1	11373	4	1	6
582	3	1	5	5917	4	0	5	11379	3	-1	4
596	3	0	NA	5918	1	0	3	11382	3	0	NA
598	1	-1	NA	5921	3	-1	NA	11383	0	0	2

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
599	0	-1	1	5931	0	-2	1	11385	3	0	4
602	1	0	3	5942	2	0	4	11387	2	-1	NA
605	0	-1	1	5943	2	-1	NA	11391	1	0	3
617	1	-1	NA	5950	2	0	NA	11397	1	-1	2
619	4	1	5	5954	0	0	1	11404	1	-2	2
630	1	1	3	5983	2	-1	NA	11405	0	1	1
634	2	0	4	5995	1	0	NA	11409	2	1	4
643	1	-1	NA	6002	1	0	2	11419	2	-1	3
645	0	-1	0	6005	2	0	4	11430	3	0	5
647	2	1	4	6009	4	2	6	11434	4	1	6
649	1	1	3	6011	1	-1	NA	11436	1	-1	2
656	2	0	4	6012	4	0	5	11440	1	1	NA
657	3	1	5	6019	0	-1	NA	11443	1	-1	NA
658	1	0	NA	6021	1	1	3	11449	2	-2	3
667	3	-1	4	6029	3	1	NA	11452	2	1	NA
692	1	0	NA	6036	2	1	NA	11453	0	-1	NA
693	3	0	4	6037	1	0	3	11456	3	1	4
698	0	1	1	6038	0	-1	0	11457	0	0	1
699	2	0	4	6043	1	1	2	11459	2	1	4
700	4	2	NA	6045	3	0	4	11471	0	0	1
704	2	-1	3	6047	0	-1	1	11476	2	-1	3
707	1	1	NA	6048	1	0	3	11479	0	-1	1
708	2	1	4	6061	2	0	4	11481	2	0	4
709	1	0	3	6063	1	1	3	11485	1	1	3
713	0	-1	0	6064	3	1	5	11486	1	0	NA
714	2	-1	3	6068	3	0	4	11487	1	0	3
716	1	-1	2	6069	0	0	1	11488	1	-1	2
718	2	0	4	6070	2	0	4	11498	0	1	NA
722	3	0	4	6071	3	1	5	11506	1	-1	3
729	3	2	NA	6074	3	0	NA	11511	3	1	5
731	1	0	2	6079	2	0	4	11515	0	-1	NA
733	2	0	3	6082	1	0	NA	11518	0	-1	NA
746	1	0	3	6088	3	0	5	11521	1	-1	NA
747	2	0	4	6094	1	0	3	11523	1	0	NA
748	0	-1	NA	6095	0	0	1	11524	0	1	NA
753	0	-1	NA	6098	1	0	3	11525	2	0	4
757	0	-1	1	6102	2	0	NA	11528	0	-1	1
763	2	-1	4	6105	0	0	2	11530	1	0	3
767	3	1	5	6113	4	1	NA	11531	2	-1	3
774	1	0	3	6116	1	0	3	11533	1	0	3
776	0	1	2	6120	0	0	2	11535	2	0	4
788	0	0	1	6121	1	-1	2	11537	2	0	4
794	2	0	4	6126	2	-2	3	11538	2	-2	3
799	0	0	2	6144	2	1	3	11541	0	-2	NA
803	3	0	4	6145	2	1	4	11548	3	0	4
806	3	0	4	6153	1	-1	2	11552	1	-1	3
807	2	0	4	6156	2	0	NA	11558	0	1	2
811	3	0	5	6159	3	0	4	11560	0	0	2
816	3	2	NA	6162	0	1	NA	11566	0	0	1
818	2	0	3	6184	1	-1	3	11572	0	0	1
819	0	-1	1	6188	1	0	3	11573	3	0	5
831	3	1	5	6189	2	-1	3	11582	3	0	4

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
835	3	1	5	6191	1	-1	NA	11586	1	-1	2
837	0	-1	1	6211	0	1	2	11590	3	0	4
841	0	1	NA	6216	2	0	4	11591	1	-1	2
846	0	-1	1	6218	0	-1	1	11601	2	0	4
856	4	1	NA	6222	0	0	1	11611	2	0	NA
861	2	1	4	6235	2	-1	3	11617	0	-1	NA
862	0	2	NA	6245	0	-1	1	11619	2	1	4
863	2	-1	3	6248	2	0	4	11624	2	0	4
865	2	0	NA	6253	2	0	4	11626	4	1	6
871	1	-1	3	6256	0	1	1	11644	1	0	2
879	0	0	1	6257	2	1	4	11652	0	-2	0
880	1	-1	2	6259	0	-2	NA	11656	0	-1	1
881	2	0	3	6266	2	1	4	11658	3	-2	3
885	2	0	4	6268	2	0	NA	11659	3	1	5
887	1	0	NA	6275	0	-2	0	11663	2	-1	NA
892	0	1	NA	6280	1	0	3	11665	2	0	4
898	2	0	4	6283	1	-1	NA	11683	3	0	4
900	0	-2	0	6288	4	0	NA	11685	0	-1	1
904	0	0	NA	6289	1	-1	2	11691	1	0	3
906	3	1	5	6301	2	2	5	11694	2	0	3
910	2	0	4	6308	2	1	4	11698	0	0	1
912	3	0	4	6314	2	0	4	11700	2	0	4
913	1	-1	3	6315	0	-1	0	11703	1	0	2
919	4	0	5	6316	2	0	4	11705	1	-1	2
924	0	-1	NA	6317	1	0	3	11710	3	-1	NA
925	1	0	3	6318	3	-2	3	11711	2	-1	3
930	2	-1	3	6323	3	-1	4	11714	0	-1	1
940	1	-2	NA	6329	3	-1	5	11731	2	1	NA
941	1	1	4	6336	0	-1	1	11732	0	0	NA
946	0	0	1	6341	2	1	5	11742	1	0	3
949	3	1	5	6348	3	-1	4	11744	2	0	4
951	0	-1	1	6349	2	1	NA	11745	1	0	3
962	2	0	4	6365	0	0	1	11749	0	1	NA
966	1	-1	2	6372	2	0	3	11756	1	1	3
967	4	0	6	6376	1	0	NA	11761	0	-1	1
971	0	1	2	6378	0	0	1	11762	3	0	4
981	3	-1	4	6379	2	1	NA	11766	3	0	5
982	2	-1	NA	6382	0	-1	1	11767	4	2	7
983	0	0	1	6383	2	-1	3	11769	1	0	3
984	0	1	2	6389	3	1	5	11770	2	0	4
989	2	-2	2	6390	1	-2	NA	11771	3	0	4
990	2	1	4	6392	3	0	5	11777	2	-1	3
992	1	0	3	6394	0	1	2	11778	3	1	5
995	3	1	5	6402	0	1	2	11779	0	0	1
996	0	-1	1	6404	1	2	4	11788	0	0	2
998	0	-1	0	6405	1	-1	3	11790	2	1	4
1001	4	-1	5	6406	0	0	1	11794	2	0	4
1007	0	0	1	6409	4	0	NA	11801	2	0	NA
1008	1	-1	NA	6410	3	1	5	11807	0	0	1
1016	1	-1	3	6411	2	0	4	11812	3	0	4
1022	1	-1	NA	6421	3	1	4	11817	0	1	2
1027	3	1	5	6428	4	2	6	11818	1	0	NA

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
1032	0	1	NA	6429	1	-1	2	11825	0	1	2
1033	2	0	4	6432	3	0	4	11828	3	0	NA
1041	3	1	4	6436	0	-1	NA	11833	3	0	4
1065	0	-1	NA	6437	2	0	4	11837	3	1	NA
1074	0	0	2	6438	0	2	2	11838	0	0	1
1075	0	0	1	6445	1	0	3	11842	0	0	2
1081	0	-1	1	6447	3	0	NA	11853	3	0	4
1094	3	0	4	6450	2	1	4	11857	3	0	NA
1099	2	-1	3	6462	2	1	4	11858	1	0	2
1105	1	0	3	6467	2	0	4	11860	3	1	5
1123	0	1	2	6478	2	0	4	11867	2	0	NA
1135	0	0	NA	6484	2	1	4	11868	4	0	5
1142	1	-1	2	6492	2	1	4	11871	3	0	NA
1155	1	-1	2	6497	3	2	NA	11875	1	2	NA
1169	1	0	NA	6504	2	-1	3	11881	3	2	5
1176	1	1	NA	6505	0	-1	1	11890	0	1	2
1178	2	0	4	6513	2	-1	3	11892	3	0	NA
1180	2	0	4	6525	4	0	5	11894	2	-1	3
1184	0	0	1	6526	2	1	NA	11896	2	1	NA
1185	0	-1	1	6528	1	-1	3	11903	2	1	4
1193	0	0	1	6540	0	1	NA	11905	3	-1	4
1196	0	-1	1	6542	0	1	NA	11907	1	0	3
1199	0	-2	1	6544	3	1	5	11909	4	2	6
1203	2	-1	2	6548	2	0	4	11911	1	0	3
1205	1	-1	2	6552	2	-1	NA	11915	0	-1	1
1207	1	-1	2	6558	2	0	4	11918	2	0	NA
1208	0	0	1	6567	2	-2	3	11920	4	0	5
1212	0	-2	NA	6569	3	0	4	11923	3	0	4
1213	0	0	NA	6572	4	1	5	11924	1	0	3
1222	0	-1	1	6577	3	2	5	11926	1	-2	2
1223	0	0	2	6581	2	-1	4	11931	2	0	NA
1226	3	0	4	6588	2	-1	3	11933	3	0	4
1227	3	2	5	6591	0	1	2	11940	2	1	5
1229	0	-1	1	6594	1	-1	2	11951	1	1	3
1230	3	0	5	6600	2	2	5	11953	1	0	3
1231	1	0	3	6602	1	1	NA	11973	0	0	NA
1241	0	-1	1	6604	1	0	3	11984	0	-1	NA
1243	3	0	4	6605	2	0	4	11985	0	-1	1
1244	4	1	6	6614	1	-1	NA	11991	1	-1	3
1246	3	0	5	6616	0	-2	NA	12002	1	0	3
1248	1	-1	3	6621	0	0	2	12006	2	-1	NA
1249	2	0	4	6640	3	0	5	12008	3	0	5
1252	2	0	4	6641	2	-1	NA	12013	0	-1	1
1261	1	1	3	6643	1	0	2	12015	3	-1	4
1275	3	-1	4	6644	0	-1	NA	12016	3	0	5
1281	0	0	2	6649	1	-2	2	12023	0	0	NA
1285	3	0	NA	6650	2	0	NA	12029	0	0	2
1288	0	1	2	6655	4	1	6	12036	0	0	1
1290	2	0	4	6661	1	-1	2	12038	1	1	3
1291	0	0	1	6672	4	0	5	12041	0	1	2
1304	2	1	4	6677	0	-1	NA	12049	2	1	NA
1305	3	-2	3	6688	1	0	2	12050	0	1	NA

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
1323	2	1	4	6689	3	1	NA	12054	1	1	3
1342	0	0	NA	6691	1	-1	NA	12060	2	-1	3
1348	1	-1	NA	6692	2	1	5	12062	3	1	5
1353	2	0	4	6694	3	-1	4	12065	1	-1	2
1363	2	-1	3	6702	1	-1	NA	12079	0	-1	1
1371	2	0	4	6714	0	0	NA	12083	4	1	6
1372	0	-1	1	6716	3	1	5	12090	2	1	5
1378	0	0	1	6724	2	0	4	12091	2	1	4
1381	3	0	4	6725	0	-1	1	12094	3	0	4
1382	3	0	4	6730	2	0	4	12099	2	0	4
1393	2	1	NA	6735	1	0	3	12101	1	1	3
1394	3	1	5	6738	1	0	3	12110	1	0	3
1398	3	2	5	6739	1	0	3	12116	3	2	NA
1404	1	1	NA	6743	2	0	4	12122	3	-1	4
1405	2	0	4	6747	1	0	3	12127	4	2	6
1419	0	0	1	6750	4	2	6	12133	2	0	4
1421	2	-2	3	6751	2	-1	NA	12142	0	1	2
1426	0	-1	1	6753	3	0	4	12147	3	0	NA
1431	0	0	NA	6754	1	1	3	12156	1	-1	2
1435	2	-1	NA	6755	0	-2	1	12157	3	1	5
1437	0	0	2	6762	2	0	4	12158	4	1	NA
1438	0	1	NA	6764	2	0	4	12161	3	0	4
1442	0	-2	NA	6772	0	1	2	12163	3	0	4
1464	0	-1	NA	6774	0	-1	1	12166	1	0	2
1471	1	1	NA	6787	2	-2	3	12170	0	0	NA
1473	2	1	4	6789	2	0	3	12174	3	1	NA
1476	2	-1	4	6793	3	0	4	12183	0	0	NA
1478	1	-1	2	6798	0	1	2	12188	0	0	NA
1479	2	0	4	6799	0	1	1	12189	3	0	5
1487	3	1	5	6800	2	0	4	12192	2	0	NA
1492	2	1	4	6802	0	-1	NA	12201	2	-1	4
1496	2	-1	4	6808	0	-1	1	12204	0	-1	0
1497	0	1	2	6809	2	0	4	12207	0	-1	0
1515	1	0	NA	6812	2	1	NA	12208	1	-1	2
1519	0	0	1	6814	1	0	NA	12209	2	0	NA
1522	2	-1	3	6816	1	-1	2	12210	4	2	6
1526	1	1	4	6822	0	0	1	12217	2	-1	3
1537	1	0	3	6829	2	0	NA	12227	0	0	1
1538	3	0	5	6834	1	-1	3	12231	3	-1	NA
1540	1	-1	2	6836	3	0	5	12232	2	0	4
1543	2	0	NA	6839	3	-1	NA	12239	3	1	5
1548	0	-1	1	6840	0	-1	NA	12240	1	-1	2
1549	0	0	1	6843	0	1	2	12251	2	-1	3
1556	2	-1	3	6846	1	-1	2	12256	2	1	4
1564	0	0	1	6848	0	-1	1	12261	0	0	NA
1570	3	-1	4	6852	1	-1	NA	12263	2	-2	3
1577	1	-1	2	6856	4	1	5	12266	0	-1	0
1585	3	0	5	6860	2	1	4	12267	1	0	3
1590	2	1	4	6866	3	0	5	12268	2	0	4
1592	0	-1	1	6870	3	0	4	12279	3	1	NA
1594	0	-1	1	6878	0	1	NA	12280	2	1	NA
1596	4	2	6	6880	3	1	5	12283	1	0	3

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
1598	4	0	5	6885	2	2	NA	12284	0	1	1
1603	0	0	2	6897	3	0	4	12285	4	1	NA
1607	0	-1	1	6902	3	-1	4	12286	3	1	5
1612	4	1	5	6904	3	0	4	12292	2	1	5
1627	2	1	4	6907	3	0	4	12295	0	0	2
1629	1	1	3	6909	0	1	1	12301	2	0	NA
1630	2	-1	3	6914	1	-2	2	12314	1	-1	2
1640	4	1	5	6915	2	2	5	12315	0	0	1
1641	3	0	4	6922	0	-1	1	12318	0	0	NA
1646	2	0	4	6924	1	-1	2	12332	3	1	5
1662	0	1	2	6933	2	0	3	12334	0	0	1
1668	0	0	NA	6934	2	1	4	12337	2	0	4
1671	0	1	1	6941	2	0	4	12338	3	1	5
1672	3	0	5	6957	1	0	NA	12349	3	0	5
1673	4	1	NA	6960	1	2	3	12350	3	-1	4
1686	2	0	4	6969	0	0	NA	12359	4	1	5
1688	3	0	4	6975	1	0	3	12360	3	1	5
1696	3	-1	4	6980	0	-1	1	12373	2	1	NA
1701	4	0	5	6983	0	-1	1	12374	0	-1	1
1707	2	0	NA	6987	2	-2	3	12380	2	1	4
1708	1	1	3	6994	1	1	3	12382	2	-1	3
1713	2	0	4	6997	2	1	5	12383	2	-1	3
1715	3	-1	NA	7002	1	1	3	12390	2	1	4
1717	1	0	3	7010	1	-1	2	12398	2	1	4
1721	2	-1	3	7015	1	1	NA	12405	2	0	4
1724	1	0	NA	7019	1	0	3	12407	1	0	3
1725	2	-1	3	7022	2	1	NA	12410	4	1	6
1730	2	0	4	7025	1	-1	2	12418	3	2	5
1731	3	0	4	7029	1	2	4	12421	3	1	NA
1734	1	2	3	7031	1	-1	3	12422	1	1	3
1740	1	1	3	7037	1	-1	NA	12439	0	0	NA
1748	1	0	3	7038	2	0	4	12444	2	0	4
1749	2	1	4	7043	0	-2	1	12463	4	0	5
1750	4	1	5	7049	1	1	3	12465	0	-1	NA
1763	1	0	3	7052	1	1	3	12470	3	0	NA
1768	3	0	5	7053	3	0	5	12471	2	0	4
1773	1	-1	NA	7056	2	-1	3	12480	3	0	4
1777	0	1	2	7057	3	0	5	12482	3	0	4
1778	0	0	1	7080	0	0	NA	12484	2	-1	3
1780	1	0	3	7086	0	-1	1	12487	3	0	4
1782	0	0	1	7087	2	0	3	12491	2	0	NA
1784	2	1	4	7105	2	0	NA	12503	4	0	NA
1786	2	0	4	7108	2	-2	NA	12507	0	0	1
1787	2	1	NA	7121	0	-2	0	12526	1	-1	2
1792	0	0	1	7122	1	-1	2	12533	1	0	3
1800	1	-1	NA	7125	0	0	2	12540	0	0	1
1801	2	-1	3	7132	1	1	3	12543	1	-1	2
1803	1	0	2	7134	2	0	3	12552	0	0	NA
1804	3	-1	4	7151	1	0	NA	12555	3	2	5
1807	1	-1	2	7152	3	-1	4	12556	3	0	4
1818	3	2	5	7157	1	0	3	12570	3	0	4
1821	1	1	3	7159	0	0	1	12579	2	0	3

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
1822	3	0	5	7166	0	1	2	12588	1	-1	2
1828	1	0	3	7167	0	0	1	12600	0	1	NA
1833	2	0	4	7177	1	0	3	12615	3	1	5
1844	3	0	4	7179	0	-1	1	12624	2	0	4
1847	1	0	3	7181	0	2	2	12629	0	-1	NA
1850	1	-1	3	7183	2	2	5	12634	2	-1	3
1854	2	0	4	7186	2	0	4	12638	0	0	1
1858	3	1	5	7193	3	1	5	12646	1	1	3
1864	3	0	4	7205	2	-1	NA	12650	0	0	NA
1867	1	0	NA	7207	0	0	2	12665	1	0	NA
1876	1	0	NA	7209	3	0	4	12674	2	-1	NA
1880	0	0	NA	7216	0	1	NA	12676	0	0	NA
1881	1	-1	NA	7232	1	0	NA	12678	1	0	NA
1891	1	0	3	7235	0	-1	NA	12685	1	1	3
1894	2	0	NA	7238	1	0	2	12690	1	0	3
1895	2	1	4	7240	1	-1	2	12698	2	0	4
1901	0	1	NA	7243	0	-1	1	12702	2	1	4
1905	2	1	4	7252	1	-1	2	12704	1	0	3
1912	3	2	5	7269	1	-1	2	12705	1	0	2
1918	2	-1	3	7275	2	0	NA	12710	3	-1	4
1921	2	0	4	7281	1	0	NA	12715	3	1	5
1923	2	-1	3	7283	0	-2	1	12720	2	0	4
1924	0	1	2	7287	3	2	5	12734	1	0	2
1931	1	-1	NA	7289	3	2	5	12744	2	0	4
1941	3	0	5	7291	3	0	NA	12747	0	0	1
1950	0	-2	1	7294	0	2	NA	12757	3	1	5
1951	3	0	5	7304	3	1	5	12758	1	-2	NA
1954	2	1	5	7308	2	1	4	12766	0	-1	1
1961	2	0	4	7313	2	-1	3	12782	2	-1	3
1966	2	0	NA	7319	0	0	1	12787	0	0	NA
1979	2	0	4	7325	0	-1	NA	12799	1	-1	2
1982	0	0	NA	7326	1	0	3	12804	2	1	5
1987	1	0	3	7330	3	1	5	12809	2	-1	3
1997	1	0	NA	7332	2	1	4	12813	2	-1	3
2004	3	0	4	7337	0	-1	1	12816	3	-1	3
2011	4	0	5	7341	0	0	1	12821	2	0	NA
2015	1	1	3	7346	3	-1	4	12826	2	-2	3
2025	4	2	6	7353	0	-2	1	12831	2	0	4
2033	0	0	NA	7354	3	0	4	12832	1	1	4
2034	2	1	4	7361	0	0	NA	12833	2	-1	3
2035	0	0	NA	7366	0	1	2	12835	2	0	4
2036	0	-2	1	7368	3	1	5	12842	0	-1	1
2053	1	0	3	7372	3	0	4	12844	2	0	4
2059	3	-1	NA	7375	2	-1	3	12847	0	-1	1
2060	0	-1	1	7377	3	2	5	12852	1	0	3
2073	1	0	NA	7380	0	1	2	12856	2	0	4
2084	1	0	3	7382	3	0	NA	12857	2	-1	3
2089	3	-1	4	7385	1	0	3	12858	4	1	NA
2092	0	-1	1	7392	1	0	2	12861	1	0	3
2109	4	0	5	7395	2	1	4	12869	2	0	4
2129	3	1	4	7397	3	-1	4	12876	1	2	4
2134	3	2	5	7403	3	-1	4	12877	0	0	NA

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
2135	2	1	5	7406	3	-1	NA	12879	1	1	3
2148	1	1	NA	7409	0	0	0	12882	2	1	4
2149	0	1	NA	7410	2	0	4	12883	4	-1	NA
2150	1	-1	2	7412	2	1	4	12887	1	-1	2
2165	1	0	3	7419	2	-1	3	12889	2	0	NA
2166	1	0	NA	7425	2	-1	4	12891	3	1	5
2168	4	0	NA	7435	4	2	6	12894	3	-1	4
2170	0	-2	0	7438	2	0	4	12895	0	0	NA
2171	1	-1	2	7440	2	1	4	12899	0	1	2
2172	1	1	3	7447	2	-1	NA	12905	4	2	6
2176	3	1	5	7449	2	-2	NA	12913	0	0	NA
2182	2	-1	3	7456	3	-1	4	12916	0	0	1
2189	1	-1	2	7464	2	0	4	12917	1	0	3
2191	2	1	NA	7478	1	-1	NA	12925	3	-1	4
2197	1	0	3	7480	2	-1	NA	12934	3	1	5
2202	0	1	2	7481	3	1	5	12939	3	1	4
2203	2	0	4	7483	2	0	3	12943	0	1	2
2204	0	-1	0	7484	0	-1	1	12950	3	1	5
2206	2	1	4	7491	3	0	4	12961	0	-1	0
2218	1	0	2	7494	1	-1	2	12963	1	0	3
2219	2	0	4	7501	1	-1	NA	12973	0	1	2
2221	0	0	1	7503	2	0	NA	12979	2	0	4
2226	0	0	1	7509	1	-1	2	12980	0	-1	1
2228	2	-1	NA	7517	0	1	1	12981	0	-1	NA
2232	2	0	NA	7518	4	1	6	12982	1	1	3
2236	0	0	0	7519	1	-2	2	12992	1	1	3
2241	1	-1	3	7521	3	1	NA	12994	0	0	1
2245	3	1	5	7522	1	-1	NA	12999	2	-1	3
2251	3	-1	NA	7536	2	0	4	13002	2	0	4
2255	3	1	5	7539	0	-1	1	13004	0	0	1
2256	2	0	4	7547	3	2	5	13010	1	1	NA
2259	0	0	1	7549	1	0	3	13013	2	0	3
2263	2	0	4	7552	1	0	NA	13015	2	1	4
2264	2	0	NA	7554	0	-1	NA	13019	2	2	5
2267	0	0	1	7556	2	0	3	13030	1	-2	NA
2273	1	-1	2	7564	2	0	NA	13031	3	1	NA
2277	2	1	4	7566	0	0	1	13036	1	1	3
2287	3	0	4	7570	3	0	5	13037	3	1	5
2289	2	-1	4	7571	1	1	3	13042	0	2	2
2291	0	-1	1	7572	2	0	4	13054	1	-1	3
2296	1	-1	2	7575	1	-1	2	13060	0	1	2
2299	0	-1	NA	7586	2	1	3	13072	2	0	NA
2306	2	-1	3	7589	3	1	5	13073	1	0	NA
2314	0	1	2	7590	1	0	3	13079	3	1	5
2317	1	-1	3	7597	2	0	4	13081	0	-1	1
2318	3	0	4	7602	2	1	4	13086	1	0	NA
2321	3	0	4	7604	3	0	4	13087	2	0	4
2324	2	1	NA	7605	1	0	3	13090	0	0	2
2340	2	-1	3	7612	1	0	2	13098	1	0	3
2343	2	-2	NA	7615	3	0	4	13100	1	0	3
2349	0	-1	1	7617	1	0	NA	13105	0	0	1
2352	3	1	5	7624	2	0	4	13106	1	0	3

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
2353	0	-1	1	7632	3	0	5	13107	2	0	4
2365	1	0	2	7639	2	1	4	13113	4	2	6
2370	0	0	NA	7642	1	1	3	13115	3	0	5
2378	2	-1	NA	7643	2	0	3	13117	0	-1	1
2390	0	-1	0	7649	2	0	4	13118	2	-1	3
2399	0	1	2	7650	0	-1	NA	13121	0	-1	NA
2402	0	0	1	7653	2	0	4	13137	1	-2	2
2403	0	1	2	7654	1	1	NA	13146	0	0	NA
2404	0	0	2	7657	3	1	5	13150	3	1	NA
2414	3	1	5	7662	0	-1	1	13151	2	1	4
2422	2	0	4	7669	2	0	4	13152	1	1	NA
2424	0	1	1	7671	1	0	3	13156	3	-1	NA
2430	2	0	4	7675	0	0	1	13165	4	1	5
2435	2	-1	3	7678	3	0	5	13169	2	1	4
2439	0	0	NA	7682	3	0	NA	13178	2	0	4
2442	2	1	4	7688	0	-1	1	13180	1	1	4
2445	2	0	4	7689	0	-1	1	13183	3	0	4
2449	0	1	2	7690	3	1	5	13184	0	1	1
2451	0	0	1	7692	1	0	2	13188	2	-1	4
2461	1	1	3	7699	0	1	2	13191	2	0	4
2464	2	0	NA	7705	2	-1	4	13196	0	-1	1
2465	2	0	4	7712	3	-1	NA	13203	2	0	NA
2472	2	-1	3	7726	2	1	4	13206	1	-1	3
2476	1	-1	2	7728	1	-2	NA	13211	2	1	NA
2482	1	0	3	7735	2	1	3	13219	0	1	2
2487	3	1	5	7737	1	-1	2	13223	3	2	5
2498	2	1	4	7739	2	0	4	13226	2	0	4
2501	2	0	NA	7743	0	1	2	13228	1	-1	2
2504	1	0	3	7744	2	2	NA	13230	2	1	5
2511	0	0	1	7746	1	-1	2	13240	2	1	4
2518	3	1	5	7749	0	-1	1	13249	3	1	5
2521	3	1	NA	7750	3	0	4	13250	0	-1	1
2530	2	-1	3	7752	2	0	4	13256	2	0	4
2543	2	1	4	7755	3	1	5	13261	0	0	2
2545	2	1	4	7756	2	0	NA	13263	1	0	2
2561	3	-1	NA	7762	3	0	NA	13268	2	1	4
2566	2	0	NA	7764	0	1	2	13275	0	1	2
2572	1	0	3	7769	1	-1	2	13277	3	1	NA
2577	1	1	3	7770	1	-1	2	13283	2	0	4
2578	0	0	2	7776	3	1	5	13284	0	0	1
2580	1	1	3	7778	3	1	5	13285	1	0	3
2581	3	0	5	7784	1	-1	2	13286	3	0	NA
2582	2	1	4	7786	2	-1	NA	13287	0	-1	1
2584	1	0	NA	7789	2	-1	4	13290	1	0	3
2590	3	0	5	7793	2	1	4	13291	0	1	NA
2598	2	-1	3	7794	0	-1	1	13294	3	1	5
2602	0	0	NA	7804	2	-1	NA	13295	2	-1	3
2605	3	1	5	7811	2	1	NA	13303	2	-1	3
2616	2	0	4	7813	1	0	3	13306	0	-2	0
2618	2	0	3	7815	1	1	3	13311	2	1	5
2619	0	-1	1	7817	0	0	1	13322	0	0	1
2624	3	1	5	7818	2	0	NA	13331	0	0	1

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
2632	0	0	2	7821	4	1	6	13337	2	1	5
2640	2	0	4	7825	3	-1	4	13344	0	0	2
2646	3	0	4	7830	3	1	5	13362	2	1	5
2651	0	1	NA	7832	2	0	NA	13364	1	-1	2
2660	0	0	1	7835	0	0	NA	13366	0	0	NA
2661	3	-1	4	7839	3	2	5	13368	2	-1	3
2668	0	-2	0	7842	3	1	5	13370	1	-1	2
2670	0	1	2	7849	3	0	4	13377	3	1	5
2680	1	1	3	7856	1	-1	NA	13378	1	-1	2
2681	2	-1	4	7857	1	-1	3	13388	0	0	1
2689	0	0	NA	7863	2	1	4	13392	2	1	4
2694	2	2	5	7866	3	0	5	13398	2	1	4
2695	0	0	NA	7871	2	-1	4	13403	2	1	NA
2696	1	0	NA	7875	1	-1	NA	13404	2	1	4
2702	1	0	3	7882	1	-1	2	13409	2	0	4
2704	2	0	3	7887	1	0	2	13416	0	2	2
2708	4	1	5	7888	3	-1	4	13422	0	0	1
2709	1	1	3	7891	1	-2	2	13427	2	0	4
2714	3	0	NA	7895	2	1	NA	13433	3	1	5
2716	2	1	NA	7901	3	0	5	13438	2	-1	3
2723	0	0	1	7906	1	-1	2	13441	4	1	NA
2725	2	-1	NA	7908	0	0	1	13449	3	1	5
2738	0	2	NA	7917	1	-2	2	13450	1	1	3
2750	2	0	4	7924	1	0	3	13453	1	-1	2
2756	0	0	2	7948	2	0	4	13460	1	1	3
2758	3	0	4	7950	4	1	6	13461	2	1	5
2766	1	0	NA	7955	1	0	2	13465	2	0	4
2767	0	1	2	7957	0	1	2	13468	0	1	2
2771	2	-2	3	7959	1	-1	NA	13481	2	0	4
2775	2	0	4	7967	4	0	5	13485	1	-1	2
2776	2	-1	NA	7969	1	0	NA	13487	3	1	NA
2779	0	0	2	7971	2	1	4	13490	2	0	NA
2780	1	0	3	7974	2	-1	3	13493	1	0	3
2781	0	0	2	7976	1	0	NA	13497	1	-1	2
2782	3	0	5	7986	4	1	5	13508	0	1	2
2783	2	0	4	7987	0	-1	1	13516	0	0	NA
2796	2	0	4	7993	3	0	4	13525	2	0	3
2798	2	2	4	7996	2	0	4	13533	4	0	NA
2800	1	1	NA	7998	1	1	3	13535	0	0	NA
2803	4	2	6	8018	3	-1	3	13538	2	0	4
2806	3	1	NA	8019	1	-1	NA	13545	1	1	NA
2813	2	-1	3	8027	0	-1	NA	13566	3	-1	4
2818	0	-1	1	8036	0	0	1	13581	1	0	3
2821	3	1	5	8040	2	1	5	13584	0	-1	1
2825	3	0	5	8044	1	2	4	13588	1	-1	3
2829	1	-1	2	8050	2	-1	3	13596	1	0	3
2830	1	0	3	8052	0	0	1	13600	3	2	6
2833	3	0	4	8054	0	1	1	13604	2	0	4
2839	0	-1	NA	8057	2	0	4	13608	1	1	3
2843	4	0	5	8058	3	0	5	13611	1	0	3
2846	2	1	4	8059	1	0	NA	13612	0	0	1
2847	1	1	3	8066	3	-2	4	13615	2	0	4

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
2848	4	1	NA	8070	3	2	5	13616	3	0	4
2856	1	1	NA	8072	3	2	5	13618	2	0	4
2863	0	1	1	8078	1	-1	3	13625	0	-1	NA
2867	3	1	5	8079	2	-1	3	13628	3	-1	4
2869	2	0	4	8080	3	0	4	13629	0	0	1
2873	3	0	5	8081	3	0	4	13630	3	1	5
2874	2	0	4	8088	0	0	1	13633	2	0	4
2875	2	-1	3	8091	1	0	3	13637	2	0	4
2880	0	-1	1	8094	1	0	3	13640	0	-1	1
2886	3	-1	4	8095	1	-1	NA	13641	1	0	2
2887	2	0	NA	8099	3	0	5	13651	2	1	NA
2888	2	0	3	8101	3	1	5	13674	2	0	4
2889	1	-1	2	8102	3	1	5	13684	2	0	4
2890	0	1	2	8116	3	0	5	13690	1	-2	NA
2892	0	-1	1	8125	3	0	5	13707	2	0	4
2901	1	1	NA	8134	0	2	2	13709	3	1	5
2902	1	0	2	8139	0	1	2	13710	3	-1	NA
2905	1	-1	2	8141	2	1	4	13713	2	2	5
2917	1	0	NA	8147	1	0	3	13724	0	2	NA
2922	1	0	3	8158	4	1	NA	13725	0	2	2
2924	1	2	4	8160	1	-1	3	13731	1	-1	2
2930	0	2	2	8165	2	-1	NA	13736	1	-1	2
2931	3	1	5	8187	1	-1	2	13740	0	1	2
2946	0	-1	1	8205	0	0	NA	13745	3	2	NA
2955	2	0	4	8209	1	1	3	13748	0	-1	1
2962	2	-1	NA	8211	2	0	NA	13751	0	1	2
2964	0	0	1	8232	2	0	4	13758	0	-1	1
2965	0	0	1	8236	2	1	4	13762	0	0	1
2967	2	2	5	8237	3	1	5	13764	3	0	5
2970	4	1	5	8238	4	1	6	13765	0	0	2
2973	3	2	6	8245	3	0	4	13769	1	-1	NA
2974	0	-1	1	8256	1	0	NA	13770	1	0	2
2976	0	0	2	8268	3	-1	NA	13774	3	1	5
2977	0	0	2	8269	1	-1	3	13787	1	0	3
2978	0	0	NA	8270	2	1	NA	13791	1	1	2
2986	3	0	5	8286	1	0	3	13802	1	-1	2
2988	3	1	5	8289	2	0	NA	13807	1	0	3
2989	0	1	NA	8301	3	-1	4	13808	1	-2	2
2995	2	0	4	8305	0	-1	1	13809	0	1	2
3005	3	1	5	8310	1	0	3	13810	2	0	NA
3011	3	1	5	8312	0	-2	NA	13822	3	1	NA
3013	1	0	3	8318	3	1	5	13823	2	-1	3
3019	2	0	NA	8321	3	0	4	13825	2	0	4
3021	1	-2	2	8328	0	1	NA	13826	0	1	2
3022	3	1	5	8331	2	1	NA	13833	2	0	3
3029	0	-1	1	8334	1	0	NA	13837	2	-1	3
3037	2	0	4	8344	2	0	4	13842	2	0	3
3042	1	-1	2	8345	1	-1	2	13846	2	0	3
3043	2	0	4	8352	3	0	5	13852	3	1	4
3049	2	0	4	8358	2	0	NA	13853	0	-2	0
3050	4	1	6	8359	1	-1	NA	13858	0	-1	1
3053	0	-1	1	8360	1	-1	3	13860	0	0	1

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
3058	0	0	1	8365	2	-1	3	13866	0	0	2
3062	0	1	2	8366	2	-1	3	13886	0	0	NA
3063	3	0	NA	8369	3	1	5	13887	0	1	1
3065	0	1	NA	8373	3	1	5	13890	1	2	4
3080	1	-1	3	8378	2	0	4	13891	2	-1	4
3088	0	1	NA	8392	2	-1	3	13893	1	-1	2
3093	0	-1	1	8397	2	-1	3	13902	3	2	5
3096	3	1	NA	8399	1	-1	2	13903	0	0	NA
3101	4	1	6	8400	1	1	3	13908	2	1	4
3103	0	-2	1	8405	1	0	3	13912	1	-2	2
3107	3	2	5	8406	2	1	4	13924	0	1	2
3109	3	0	5	8410	0	1	2	13928	1	0	3
3111	4	1	6	8413	3	0	4	13929	1	1	3
3113	2	1	4	8414	0	-1	0	13938	2	1	4
3116	3	1	5	8416	4	1	NA	13939	1	-1	NA
3132	1	0	NA	8426	1	-1	2	13941	3	1	5
3141	3	1	5	8434	2	1	4	13951	1	1	3
3153	0	0	NA	8439	0	2	NA	13962	0	-2	0
3154	0	0	1	8440	0	-2	0	13964	1	-1	2
3160	0	0	1	8475	3	0	5	13967	2	-1	3
3167	0	1	2	8480	3	1	5	13971	3	-1	4
3170	2	-1	4	8497	1	-1	2	13972	3	0	4
3173	1	-1	2	8499	0	1	1	13975	0	2	1
3174	2	0	4	8500	1	1	3	13977	1	0	3
3177	3	1	5	8501	2	-1	3	13979	3	0	5
3179	3	0	NA	8502	2	-1	NA	13983	0	-1	1
3184	0	-1	1	8518	2	2	5	13984	2	0	3
3190	1	0	NA	8520	2	0	4	13987	0	0	1
3193	3	0	NA	8523	2	1	4	13994	3	0	NA
3199	3	0	4	8525	0	-2	0	13999	1	0	3
3201	0	0	NA	8532	1	0	NA	14003	2	2	5
3202	1	-1	3	8535	1	0	2	14008	2	1	5
3203	1	1	3	8543	0	0	2	14011	2	2	NA
3206	2	-1	NA	8554	0	1	2	14012	1	0	3
3209	2	-1	3	8560	4	0	5	14016	2	-1	3
3210	1	0	2	8561	4	-1	5	14017	1	0	NA
3217	3	0	5	8563	0	-1	1	14020	0	0	NA
3220	3	0	NA	8566	0	0	1	14027	0	1	2
3228	1	-1	NA	8570	2	0	4	14038	2	0	3
3232	1	0	3	8572	3	-1	4	14040	3	1	5
3239	0	0	NA	8582	1	-1	1	14042	1	-1	3
3243	2	0	4	8583	0	-1	1	14055	0	1	NA
3245	1	0	3	8587	0	0	1	14057	2	1	4
3246	2	1	NA	8592	1	-1	2	14060	0	1	2
3251	4	0	5	8593	0	1	2	14081	0	-1	NA
3253	4	0	5	8607	0	-1	1	14091	2	0	4
3257	3	0	NA	8609	2	1	4	14111	2	-1	4
3260	0	0	1	8610	2	0	NA	14117	3	1	5
3261	1	1	3	8614	2	0	NA	14121	2	0	NA
3263	3	1	5	8616	4	0	5	14122	3	1	NA
3278	2	0	4	8622	4	-1	5	14125	0	0	2
3281	1	0	3	8623	2	0	4	14129	0	0	1

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
3283	3	1	5	8624	0	-1	1	14135	0	-1	0
3290	1	-1	2	8633	4	1	6	14148	0	2	NA
3297	2	0	4	8641	3	2	NA	14157	2	0	4
3304	3	0	4	8644	4	1	5	14161	0	0	1
3305	2	-1	4	8649	2	1	4	14163	0	0	NA
3307	2	2	4	8653	1	0	3	14172	3	-1	NA
3308	1	1	3	8657	3	2	5	14180	0	1	2
3313	1	1	3	8658	1	0	3	14182	2	1	NA
3314	1	0	3	8663	2	0	3	14188	0	1	2
3317	0	2	2	8672	1	-1	2	14191	1	0	3
3348	0	1	2	8680	0	-1	0	14201	4	2	6
3350	1	-1	2	8684	1	0	3	14202	2	0	4
3359	2	0	4	8687	0	0	1	14213	2	0	4
3367	3	0	4	8688	2	0	4	14220	0	0	1
3376	1	0	3	8690	2	0	4	14224	1	-1	2
3378	3	-1	4	8712	0	1	1	14231	0	-1	NA
3384	0	0	1	8717	1	1	3	14241	3	1	5
3386	1	0	3	8730	3	1	5	14243	3	0	NA
3387	2	0	NA	8739	2	0	4	14245	1	-1	3
3388	2	-1	3	8744	2	1	4	14247	1	-2	NA
3390	2	1	4	8747	3	1	5	14248	1	0	3
3391	0	0	1	8748	3	0	4	14252	1	0	3
3396	0	-1	-1	8751	2	-1	3	14254	3	0	4
3398	2	0	4	8758	2	0	3	14260	3	-1	4
3404	4	0	5	8761	0	-1	NA	14269	2	1	4
3406	0	1	2	8763	0	1	2	14272	0	1	2
3407	1	1	4	8764	2	0	4	14274	2	0	4
3414	3	1	4	8765	3	2	5	14279	1	1	3
3419	2	-1	4	8773	4	0	NA	14280	3	1	5
3423	1	-1	NA	8780	1	0	3	14290	2	-1	4
3427	1	0	3	8781	2	-1	3	14298	4	2	NA
3432	0	0	0	8782	2	0	NA	14308	3	0	4
3434	1	0	3	8785	0	0	0	14313	2	0	4
3438	0	0	1	8786	1	1	NA	14316	2	0	4
3442	0	0	1	8797	0	-1	1	14319	4	1	NA
3443	0	0	1	8799	3	0	4	14322	0	1	NA
3448	0	0	1	8807	0	-1	NA	14323	2	-1	NA
3456	2	-1	3	8816	2	0	NA	14325	3	2	5
3464	4	1	5	8817	3	-1	4	14337	3	0	4
3470	0	0	1	8826	3	1	NA	14339	3	0	4
3475	2	0	4	8833	1	0	3	14341	1	0	3
3477	3	-1	NA	8834	1	0	NA	14342	3	1	5
3490	1	0	NA	8835	1	-1	NA	14346	4	1	6
3493	2	0	4	8840	3	1	5	14351	2	0	4
3502	2	-1	NA	8843	1	1	3	14354	0	0	1
3508	3	0	NA	8849	2	-1	3	14355	0	0	1
3516	0	-1	NA	8855	1	1	3	14358	3	2	5
3517	3	1	NA	8861	1	0	2	14359	0	-1	NA
3525	1	0	NA	8862	3	0	4	14364	2	0	4
3532	2	0	3	8865	3	-2	NA	14374	0	-2	0
3535	2	0	NA	8868	4	0	5	14376	1	-1	NA
3536	2	1	5	8870	2	1	4	14382	1	0	3

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
3540	2	-1	3	8880	2	1	NA	14384	1	0	3
3547	1	0	3	8885	0	0	2	14393	0	0	NA
3550	3	-1	4	8894	0	0	2	14398	1	0	3
3557	1	-1	NA	8895	3	0	5	14403	3	0	4
3562	0	-1	1	8899	3	0	5	14406	0	0	NA
3563	2	-1	3	8912	0	0	2	14408	2	0	NA
3564	0	-1	1	8922	1	-1	3	14411	0	1	2
3570	1	1	3	8924	2	-1	NA	14414	2	0	4
3573	1	1	NA	8928	2	0	4	14418	1	-1	NA
3577	1	1	3	8932	0	1	2	14423	2	0	NA
3579	3	-1	NA	8943	3	1	5	14442	3	0	4
3581	0	-1	1	8945	4	1	NA	14443	2	0	4
3587	2	0	3	8946	2	1	4	14444	2	-1	3
3602	3	-1	4	8954	1	0	2	14446	3	-1	NA
3609	2	0	3	8958	2	0	NA	14455	1	0	3
3612	2	1	4	8960	0	-1	1	14456	3	0	5
3621	1	1	3	8965	1	0	2	14458	2	0	4
3642	0	1	NA	8966	2	1	4	14464	2	1	NA
3647	0	0	1	8967	1	-1	2	14466	3	1	4
3649	1	0	NA	8969	2	-1	3	14467	1	0	3
3654	1	0	NA	8980	3	0	NA	14469	3	-1	4
3660	2	1	4	8984	3	0	NA	14483	0	1	2
3665	3	0	NA	8985	1	0	NA	14484	1	0	3
3669	2	0	4	8988	1	0	NA	14490	1	0	3
3673	2	0	4	8989	2	1	4	14491	2	-2	3
3675	0	0	2	8995	0	0	1	14494	1	0	3
3678	2	0	4	9004	1	1	3	14496	2	0	4
3680	1	1	3	9010	1	0	NA	14503	0	-1	1
3686	3	2	5	9012	0	0	1	14504	1	0	3
3693	2	1	4	9018	2	0	4	14505	2	-1	3
3710	1	0	3	9036	0	-2	0	14506	0	0	1
3713	3	0	4	9037	1	-1	2	14507	0	0	1
3718	3	1	NA	9040	0	1	NA	14512	0	-1	NA
3725	2	-1	4	9041	2	2	5	14520	2	1	4
3726	1	1	3	9044	4	1	5	14527	1	-1	2
3747	1	0	2	9045	1	0	NA	14531	3	1	5
3753	0	0	1	9047	1	0	2	14532	0	0	NA
3754	4	0	5	9049	0	0	NA	14535	1	-1	2
3760	4	0	5	9061	0	0	2	14543	0	0	1
3763	1	-1	2	9062	1	0	3	14554	2	0	4
3765	3	0	4	9076	2	0	4	14556	3	2	5
3769	4	0	5	9079	1	-1	2	14557	2	0	4
3771	2	0	4	9081	2	0	4	14561	1	-1	2
3784	1	1	3	9082	2	1	4	14562	3	-1	NA
3787	2	1	NA	9089	0	-1	0	14567	0	-2	0
3794	0	0	1	9092	2	-1	3	14568	0	-2	1
3796	2	0	4	9094	0	1	NA	14574	3	-1	4
3798	2	1	4	9115	0	0	1	14575	1	0	3
3809	2	0	4	9117	3	0	4	14579	3	-1	4
3812	3	1	5	9118	1	0	3	14581	2	0	4
3819	0	-2	0	9120	0	-1	1	14582	1	-2	NA
3828	3	0	NA	9124	1	0	NA	14586	0	0	1

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
3831	3	1	5	9128	0	-2	NA	14591	1	0	3
3833	1	1	3	9135	0	0	1	14598	1	0	3
3837	3	1	5	9136	1	0	3	14599	1	0	NA
3839	0	-1	1	9138	3	0	4	14600	0	-1	1
3843	0	0	2	9157	3	0	NA	14612	4	0	5
3846	2	0	NA	9176	0	0	1	14613	0	-1	1
3854	3	1	5	9183	0	0	1	14624	3	1	4
3861	0	0	1	9187	1	0	3	14626	1	0	3
3864	2	0	4	9188	0	-1	1	14630	3	0	5
3868	1	0	3	9190	3	0	4	14633	1	1	3
3869	3	1	5	9197	2	0	NA	14639	1	0	NA
3870	1	0	3	9200	3	-1	4	14642	2	0	NA
3883	1	-1	3	9201	0	0	2	14643	4	0	5
3886	1	0	3	9203	0	0	1	14649	1	1	3
3889	2	0	NA	9212	2	0	4	14650	0	0	2
3894	0	-1	0	9213	0	-1	1	14653	1	0	2
3907	1	0	3	9214	3	0	NA	14655	1	-1	2
3910	2	0	4	9217	1	1	3	14656	1	-1	2
3913	0	0	1	9219	1	1	4	14662	2	0	4
3914	1	0	3	9220	3	1	5	14663	0	1	2
3921	2	2	5	9221	4	1	5	14673	1	0	3
3923	0	1	2	9237	0	0	2	14674	1	-1	2
3929	2	-2	3	9240	2	1	4	14676	2	0	4
3931	0	-2	1	9241	1	-1	2	14682	0	0	2
3932	3	1	5	9248	2	-1	4	14685	4	0	5
3937	0	0	NA	9253	4	2	6	14689	1	0	3
3943	1	1	3	9259	2	1	4	14693	1	0	3
3956	2	1	4	9267	0	1	2	14697	1	0	3
3957	1	0	3	9271	1	0	3	14700	1	0	3
3961	4	1	6	9273	0	-1	1	14704	0	1	NA
3971	1	1	3	9285	4	1	6	14710	2	-1	4
4004	0	-1	1	9290	2	0	NA	14719	2	0	3
4005	1	1	3	9291	2	0	3	14724	2	1	5
4006	3	0	4	9293	0	0	1	14728	3	0	4
4011	1	-1	3	9294	2	0	4	14735	2	0	3
4013	3	1	5	9301	2	-1	4	14736	1	-1	NA
4014	3	2	5	9302	3	0	NA	14741	0	0	NA
4016	0	-1	0	9312	1	0	2	14744	1	1	2
4017	3	1	5	9316	2	0	4	14753	0	0	1
4020	1	0	NA	9319	0	1	2	14756	3	0	4
4022	2	0	4	9328	3	1	5	14762	3	1	5
4026	0	0	1	9331	2	1	4	14765	4	0	5
4032	0	-1	1	9338	0	-1	0	14783	4	0	5
4043	1	-1	2	9350	1	0	3	14784	1	-1	2
4045	1	1	3	9356	2	0	4	14786	0	0	1
4048	3	1	5	9359	1	-1	2	14790	1	-1	1
4051	3	1	NA	9362	1	1	3	14793	2	-1	3
4052	2	0	4	9364	1	0	3	14796	4	1	6
4056	2	-1	3	9370	1	1	3	14801	0	1	2
4059	1	0	2	9380	0	-1	1	14807	0	-1	1
4069	3	1	NA	9386	2	1	NA	14812	2	0	4
4074	2	0	4	9394	0	0	1	14815	3	1	5

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
4076	1	0	3	9407	2	0	4	14831	3	0	4
4077	0	-1	NA	9411	1	0	3	14833	4	1	6
4079	1	0	NA	9422	2	0	4	14836	4	1	6
4081	2	0	NA	9423	0	0	1	14856	3	0	5
4088	0	0	1	9429	0	0	1	14859	1	0	NA
4105	1	1	NA	9433	0	-1	1	14861	1	1	3
4125	2	0	4	9439	0	0	1	14863	0	-1	0
4134	2	-1	3	9451	3	0	4	14865	0	0	2
4139	1	0	2	9452	1	1	3	14880	0	0	1
4146	1	-1	2	9453	0	-1	NA	14881	2	0	4
4149	3	0	NA	9460	3	1	NA	14883	0	-1	1
4151	0	1	1	9465	2	2	5	14884	3	1	NA
4155	1	0	NA	9470	3	0	4	14894	2	0	4
4157	1	1	3	9476	2	0	NA	14896	3	0	4
4168	3	1	NA	9485	2	0	4	14899	2	-1	4
4170	1	-1	2	9486	0	-1	1	14900	2	-1	4
4174	1	-1	2	9488	1	0	3	14901	0	0	NA
4179	3	1	5	9507	4	1	5	14906	2	0	4
4185	3	0	4	9508	0	-1	1	14907	0	0	1
4199	0	-1	1	9517	4	2	6	14915	3	1	5
4205	0	0	1	9521	2	2	5	14919	1	-1	NA
4208	1	1	NA	9528	2	-1	3	14926	4	1	5
4211	2	-2	3	9532	1	0	NA	14927	1	-1	2
4212	0	-1	0	9536	2	-1	3	14933	2	-1	3
4215	1	0	3	9540	3	1	5	14937	2	-1	3
4217	2	-1	NA	9542	2	1	4	14939	2	-1	NA
4219	0	0	2	9546	3	-1	4	14940	0	0	1
4226	3	0	4	9548	3	0	4	14943	0	-1	1
4227	2	-1	NA	9549	4	1	NA	14953	2	0	4
4229	0	-1	1	9554	4	0	5	14954	1	-1	2
4231	1	0	2	9555	2	1	4	14969	2	0	4
4233	0	-1	1	9558	0	-1	1	14999	4	2	NA
4237	1	0	NA	9573	0	0	1	15008	2	0	4
4243	2	1	4	9575	4	0	5	15009	2	0	NA
4248	3	0	5	9584	2	-1	3	15018	2	1	4
4255	3	1	5	9586	2	0	4	15023	2	0	4
4262	1	-1	3	9588	3	1	5	15025	0	0	1
4266	0	-1	1	9591	0	0	1	15034	2	0	NA
4268	0	1	2	9592	3	0	NA	15036	3	0	4
4270	1	-1	NA	9597	3	2	NA	15051	0	1	2
4273	0	1	NA	9600	3	0	4	15052	2	-1	3
4276	2	1	4	9603	2	0	NA	15064	2	0	4
4277	2	0	4	9605	1	-1	2	15070	2	0	3
4279	1	-1	NA	9614	0	0	NA	15074	2	1	4
4299	2	0	4	9616	3	1	NA	15077	2	1	4
4313	0	1	2	9622	2	0	4	15081	2	0	4
4322	3	0	NA	9624	3	0	4	15086	4	1	6
4324	0	0	1	9629	0	0	1	15093	0	-1	0
4328	2	-1	3	9633	2	0	NA	15094	0	2	2
4331	1	0	3	9640	3	0	4	15103	1	-1	2
4335	1	-1	2	9644	1	-1	2	15104	0	-2	NA
4337	2	0	4	9645	0	0	1	15110	0	0	2

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
4338	0	-1	1	9646	1	0	3	15112	1	-2	2
4343	1	-1	2	9648	1	0	3	15115	2	2	NA
4347	1	-2	2	9649	0	-1	1	15131	0	0	2
4355	2	0	4	9660	2	0	4	15139	3	-2	3
4357	1	0	NA	9664	2	1	NA	15141	2	0	3
4359	4	1	6	9675	0	1	NA	15148	0	-1	1
4362	1	-2	NA	9679	1	0	3	15154	4	1	NA
4368	1	-1	3	9680	2	-1	3	15156	2	-1	3
4374	3	1	NA	9682	0	0	NA	15161	1	0	2
4375	3	1	5	9697	1	0	3	15167	2	-1	3
4378	2	0	4	9701	3	-1	4	15178	3	0	NA
4381	0	0	2	9704	2	-1	3	15205	3	0	5
4387	2	1	4	9705	0	1	2	15207	0	1	NA
4400	1	-1	2	9707	3	0	4	15222	2	0	4
4423	2	1	4	9714	0	1	2	15223	3	0	5
4424	1	0	NA	9718	1	0	3	15225	3	0	4
4428	3	0	4	9722	3	1	5	15228	2	1	NA
4433	3	2	6	9739	1	0	3	15239	1	-1	NA
4436	0	-1	NA	9747	4	1	5	15241	0	0	1
4437	0	0	1	9751	0	-1	1	15246	0	0	1
4439	4	1	6	9757	1	-1	2	15247	0	-1	1
4449	2	0	4	9759	3	-1	4	15249	1	0	3
4456	2	0	4	9760	3	0	NA	15255	3	-1	NA
4463	4	1	5	9764	2	-1	3	15257	0	0	1
4467	1	-1	NA	9776	0	0	NA	15267	0	0	1
4468	0	0	2	9778	1	0	3	15277	2	-1	3
4469	1	0	3	9786	0	-1	0	15280	3	1	5
4472	2	0	4	9803	2	0	4	15289	2	1	NA
4473	3	0	4	9804	2	1	NA	15297	0	0	1
4476	1	0	3	9815	3	1	5	15302	0	0	1
4500	1	-1	3	9824	1	-2	NA	15304	0	-1	1
4509	2	0	4	9825	0	0	1	15312	0	-1	1
4513	1	-1	NA	9826	1	0	3	15321	0	-1	NA
4521	0	0	1	9827	2	0	4	15325	0	0	1
4527	1	0	3	9833	2	0	4	15326	2	1	4
4530	1	-2	2	9835	0	-1	0	15333	3	1	NA
4532	1	1	3	9860	2	0	4	15337	0	0	1
4533	2	-1	3	9865	1	1	3	15338	1	2	4
4535	1	-1	3	9871	3	-1	4	15340	2	2	5
4536	3	1	4	9874	0	-1	1	15342	1	-1	NA
4542	2	0	4	9880	1	1	3	15344	2	-1	4
4551	1	0	2	9882	0	-1	NA	15347	0	0	1
4554	2	0	4	9885	1	0	3	15349	3	1	NA
4555	1	-2	2	9888	2	1	4	15355	1	1	3
4564	0	0	1	9892	1	-2	2	15359	0	0	1
4572	1	-1	NA	9893	3	1	5	15366	0	-1	1
4573	4	2	6	9896	0	-1	1	15367	0	-1	NA
4577	0	-1	1	9902	1	1	3	15368	0	-1	1
4579	3	1	NA	9906	2	0	4	15369	2	1	NA
4583	2	0	NA	9910	3	1	5	15380	2	0	4
4584	4	1	6	9914	0	1	2	15381	2	-1	3
4596	2	-1	3	9918	0	1	2	15387	0	1	1

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
4599	2	0	4	9920	0	-1	1	15388	1	0	3
4607	2	0	4	9926	2	2	4	15389	0	2	2
4609	0	-1	0	9931	3	2	5	15392	0	0	2
4610	0	0	1	9935	2	0	4	15400	2	-1	3
4616	1	-2	2	9945	2	1	4	15405	1	1	3
4617	1	0	3	9953	0	1	2	15407	2	0	4
4633	3	0	5	9957	1	0	NA	15408	4	2	6
4638	2	0	NA	9963	2	1	NA	15411	1	0	3
4641	1	0	NA	9972	2	1	4	15413	3	0	4
4653	4	2	6	9976	3	0	NA	15418	4	0	5
4655	3	0	4	9979	1	0	NA	15419	3	1	5
4659	0	1	2	9980	0	0	1	15421	1	0	3
4669	0	-1	1	9982	0	1	NA	15425	1	0	3
4678	0	0	1	9991	2	0	NA	15436	2	0	4
4685	2	1	4	10000	2	2	5	15438	3	0	4
4686	2	0	NA	10003	2	0	4	15440	3	0	4
4691	0	-1	1	10005	1	-1	2	15443	2	0	4
4695	2	1	4	10014	2	0	4	15460	1	0	3
4698	2	0	4	10032	1	1	3	15464	0	-1	1
4700	4	0	5	10034	1	0	NA	15465	2	-1	3
4711	2	-1	4	10041	1	-1	NA	15473	0	0	1
4722	3	0	NA	10042	2	0	4	15475	2	-2	3
4727	3	-1	4	10044	3	1	5	15483	0	0	1
4756	4	1	6	10045	0	-2	0	15494	4	0	5
4762	0	0	1	10054	2	0	4	15495	4	2	6
4763	2	-1	NA	10061	4	1	NA	15498	3	-1	4
4766	4	0	5	10062	0	1	2	15499	1	0	3
4770	0	0	1	10073	1	1	NA	15500	0	1	2
4784	3	-1	4	10081	0	1	1	15501	0	-1	0
4791	1	1	3	10084	0	0	NA	15510	1	-1	2
4795	3	1	5	10086	0	1	1	15512	1	-1	2
4799	0	-1	1	10093	0	-1	NA	15516	1	-1	3
4802	3	0	4	10101	3	1	NA	15518	3	1	NA
4805	2	0	4	10105	3	0	4	15519	2	0	4
4814	2	1	4	10110	2	0	NA	15524	2	-1	3
4816	0	0	1	10113	1	0	3	15527	0	0	1
4817	2	0	4	10115	1	0	3	15529	0	0	2
4822	1	0	3	10119	2	0	4	15530	0	0	2
4827	1	-1	2	10121	3	-1	4	15538	0	0	1
4833	4	0	5	10124	0	1	2	15539	1	0	3
4836	0	1	2	10126	4	1	5	15541	0	0	1
4842	1	0	NA	10127	1	1	3	15546	0	-1	1
4844	1	0	NA	10145	1	0	3	15547	1	0	NA
4845	1	-1	3	10147	0	-1	1	15548	0	1	NA
4849	3	0	4	10148	1	-1	2	15552	1	-1	2
4850	0	0	1	10162	1	-1	NA	15556	3	1	5
4860	3	0	4	10163	1	0	3	15567	2	-1	NA
4863	1	0	2	10166	1	-2	2	15572	1	-1	2
4871	2	0	4	10172	2	0	4	15573	3	1	4
4878	2	0	4	10173	1	0	2	15574	2	0	3
4881	1	0	3	10175	1	-1	2	15577	0	-1	1
4888	1	2	NA	10180	2	-1	NA	15579	2	0	4

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
4900	4	2	6	10186	0	1	2	15581	1	-1	2
4906	1	1	4	10192	3	-1	4	15589	0	-1	1
4909	0	1	NA	10199	2	-1	3	15596	1	0	2
4916	2	1	4	10209	0	0	1	15598	3	1	5
4918	4	1	5	10210	4	2	6	15599	3	0	4
4926	1	0	NA	10214	2	0	NA	15605	3	2	5
4928	1	-1	NA	10215	0	1	NA	15606	3	0	5
4941	1	0	2	10216	0	0	NA	15608	2	0	4
4946	3	0	5	10232	1	1	3	15616	2	1	4
4949	0	0	1	10239	3	1	5	15618	0	0	1
4956	1	0	3	10249	2	1	5	15621	0	0	1
4966	3	0	4	10253	3	1	5	15626	0	0	1
4969	1	0	2	10255	0	0	1	15638	1	0	3
4973	2	0	4	10262	2	-1	NA	15639	1	-2	NA
4978	3	1	5	10264	1	-1	2	15642	1	-2	2
4982	2	0	3	10266	0	1	2	15644	1	0	NA
4985	2	1	4	10268	0	0	2	15646	4	1	NA
4991	1	1	3	10271	2	-1	3	15649	2	1	4
4998	2	0	3	10272	3	1	5	15656	0	-1	1
5000	1	-1	2	10276	0	0	1	15659	0	1	NA
5004	2	0	NA	10277	1	0	2	15680	0	-1	1
5005	1	-1	2	10279	2	-1	3	15686	3	0	4
5011	2	1	4	10281	1	-1	2	15693	0	1	2
5016	0	0	0	10285	1	-1	NA	15697	3	1	5
5018	3	-1	4	10294	0	-1	1	15699	4	0	5
5034	2	0	4	10300	2	-1	3	15701	2	1	4
5038	0	0	1	10304	1	0	3	15705	0	0	2
5042	3	2	NA	10307	1	1	NA	15714	1	0	3
5046	1	0	3	10309	3	1	5	15722	3	0	4
5051	0	0	NA	10310	0	0	2	15728	2	0	NA
5054	0	0	NA	10312	0	0	NA	15734	2	-1	4
5057	3	0	4	10321	2	-1	4	15752	1	-1	2
5062	2	0	4	10332	0	1	1	15756	2	0	4
5063	2	1	NA	10336	2	0	4	15760	2	0	4
5065	1	0	NA	10368	1	0	3	15762	3	0	4
5066	0	-1	1	10369	2	2	5	15767	3	-1	4
5076	1	-1	3	10375	0	-2	0	15768	0	0	NA
5089	2	-1	3	10376	0	-1	1	15773	3	1	NA
5092	2	-2	3	10379	2	0	NA	15774	2	0	NA
5093	3	0	4	10380	1	0	3	15781	1	1	NA
5094	3	0	4	10383	1	-1	NA	15782	1	0	3
5098	2	0	4	10385	3	1	5	15784	4	2	6
5102	3	-2	3	10387	2	0	NA	15791	1	1	NA
5112	3	1	5	10397	1	0	3	15796	2	1	4
5117	1	0	3	10412	1	1	NA	15798	3	1	5
5127	2	0	NA	10413	0	1	NA	15806	0	-1	1
5130	1	-1	3	10418	1	-1	NA	15814	0	0	1
5131	1	-1	2	10420	2	-1	4	15819	1	1	3
5132	2	0	NA	10426	2	0	4	15825	2	0	4
5135	0	0	2	10427	1	0	3	15826	2	1	4
5136	1	0	NA	10428	2	-1	3	15831	3	0	NA
5147	2	1	5	10430	0	0	2	15835	4	1	6

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
5157	2	0	4	10435	0	0	NA	15836	0	-1	1
5160	1	-1	2	10436	0	-2	NA	15839	4	0	5
5165	0	-1	1	10446	4	1	6	15845	1	0	3
5166	0	0	1	10448	3	-1	4	15858	0	0	1
5172	1	-1	2	10449	2	1	4	15859	2	0	4
5173	1	-1	3	10463	1	-1	2	15876	2	1	4
5179	0	0	NA	10469	1	-1	2	15878	3	0	NA
5184	3	1	5	10470	3	0	NA	15880	0	1	2
5187	1	-1	NA	10471	2	0	NA	15886	3	0	NA
5191	1	0	3	10473	3	1	NA	15888	1	1	NA
5193	0	-1	1	10476	2	-1	3	15891	1	1	3
5194	0	0	1	10482	1	0	3	15900	1	-2	NA
5199	1	-1	NA	10500	2	-1	3	15902	3	0	4
5212	0	0	1	10511	3	1	5	15904	1	0	3
5213	1	0	3	10512	2	1	4	15908	1	-1	2
5224	2	0	4	10514	2	0	4	15910	0	-1	1
5226	3	0	4	10515	3	0	5	15917	1	1	4
5239	3	1	5	10526	0	0	NA	15919	3	1	5
5252	2	1	NA	10546	0	-1	NA	15924	1	0	3
5264	0	0	1	10549	2	0	4	15927	1	0	3
5266	1	-2	NA	10553	0	-1	NA	15937	0	0	NA
5271	3	1	5	10558	1	-1	2	15946	2	-1	4
5273	2	0	4	10575	0	0	1	15949	2	-1	3
5276	2	-2	3	10581	1	0	3	15957	2	1	4
5278	3	0	5	10583	1	1	NA	15961	3	0	4
5281	1	-1	2	10584	0	-1	1	15964	1	-1	3
5283	3	0	4	10585	0	1	2	15965	2	-1	4
5291	0	2	2	10610	0	1	2	15966	0	1	2
5294	3	2	6	10611	0	-1	0	15978	0	0	1
5296	2	0	4	10616	2	2	4	15983	0	1	2
5297	0	0	1	10618	1	0	3	15987	3	-2	NA
5313	2	0	4	10628	0	-1	0	15988	1	-1	3
5314	1	0	3	10632	0	-1	1	15998	1	-1	3
5321	2	-1	NA	10642	2	0	4	16004	1	1	3
5325	2	-1	3	10648	1	1	3	16008	3	-1	4
5326	0	0	1	10649	2	0	4	16011	0	1	NA
5328	3	-1	NA	10650	2	-1	NA	16023	1	0	3
5334	1	0	3	10654	1	-2	2	16024	1	-1	NA
5338	2	1	4	10656	3	1	5	16025	1	0	3
5344	1	-2	1	10661	3	1	5	16048	2	2	NA
5348	0	-1	1	10663	0	2	2	16050	1	1	3
5352	0	0	2	10672	1	0	3	16051	0	0	1
5353	2	0	4	10678	3	1	5	16057	0	1	2
5354	0	1	2	10685	3	0	NA	16059	3	1	5
5361	0	-1	NA	10690	3	1	5	16060	2	0	3
5364	1	1	3	10702	2	0	4	16075	3	0	5
5365	2	0	4	10706	1	1	3	16094	3	2	5
5367	0	-2	1	10708	1	-1	2	16096	3	1	5
5379	3	1	4	10716	2	-1	3	16116	0	-1	NA
5382	2	-1	3	10717	4	0	NA	16118	0	0	2
5386	3	1	NA	10720	3	1	5	16121	2	-2	3
5395	2	0	4	10729	0	0	1	16122	2	0	4

IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT	IDX	STRS	LBL	TGT
5410	2	1	NA	10730	3	1	5	16124	4	1	5
5411	1	1	3	10745	0	1	2	16125	2	0	4
5416	3	1	4	10753	1	-1	NA	16126	1	0	NA
5424	1	2	4	10754	1	-1	2	16130	3	0	4
5426	1	0	3	10762	1	-1	2	NA	NA	NA	NA

Appendix

```
library(dplyr)
library(psych)
library(ggplot2)
library(gridExtra)
library(reshape2)
library(car)
library(recommenderlab)
library(PerformanceAnalytics)
library(knitr)
library(faraway)
library(MASS)

wine.trn <- read.csv("https://raw.githubusercontent.com/Nguyver/DATA621-HW/master/HW5/wine-training-data
    header = TRUE, sep = ",", stringsAsFactors = FALSE, na.strings = c("NA", ""))

wine.evl <- read.csv("https://raw.githubusercontent.com/Nguyver/DATA621-HW/master/HW5/wine-evaluation-da
    header = TRUE, sep = ",", stringsAsFactors = FALSE, na.strings = c("NA", ""))
summary(wine.trn)

colnames(wine.trn)[1] <- "INDEX"
glimpse(wine.trn)

na_count <- sapply(wine.trn, function(y) sum(length(which(is.na(y)))))

na_countPrc <- round(sapply(wine.trn, function(y) sum(length(which(is.na(y))))/length(y) *
    100), 2)

na.df <- filter(data.frame(ColName = colnames(wine.trn), NA_Count = na_count, NA_Percent = na_countPrc)
    NA_Count > 0)

knitr::kable(filter(na.df, NA_Count > 0))

ggplot_missing <- function(x) {

    x %>% is.na %>% melt %>% ggplot(data = ., aes(x = Var2, y = Var1)) + geom_raster(aes(fill = value))
        scale_fill_grey(name = "", labels = c("Present", "Missing")) + theme_minimal() +
        theme(axis.text.x = element_text(angle = 45, vjust = 0.5)) + labs(x = "Variables in Dataset",
        y = "Rows / observations")
}

ggplot_missing(wine.trn)
```

```
g1 <- ggplot(wine.trn, aes(x = factor(STARS), y = TARGET)) + geom_violin(aes(fill = factor(STARS))) +
    geom_boxplot(width = 0.2)

g2 <- ggplot(wine.trn, aes(x = TARGET, fill = factor(STARS))) + geom_density(alpha = 0.5)

blank <- rectGrob(gp = gpar(col = "white"))  # make a white spacer grob
grid.arrange(g1, blank, g2, heights = c(0.6, 0.05, 0.4), nrow = 3)

ggplot(wine.trn, aes(x = TARGET)) + geom_histogram(binwidth = 0.5) + theme(axis.text = element_text(siz
    axis.title = element_text(size = 10))

wine.trn1 <- wine.trn[, -1]

layout(matrix(1:15, 3, 5, byrow = TRUE))
par(mar = c(2, 1, 2, 1))
for (i in 1:ncol(wine.trn1)) hist(wine.trn1[, i], main = names(wine.trn1)[i])

layout(matrix(1:15, 3, 5, byrow = TRUE))
par(mar = c(2, 1, 2, 1))
for (i in 1:ncol(wine.trn1)) boxplot(wine.trn1[, i], main = names(wine.trn1)[i])

g1 <- ggplot(wine.trn, aes(x = factor(AcidIndex), y = TARGET)) + geom_violin(aes(fill = factor(AcidInde

g2 <- ggplot(wine.trn, aes(x = factor(LabelAppeal), y = TARGET)) + geom_violin(aes(fill = factor(LabelA

blank <- rectGrob(gp = gpar(col = "white"))  # make a white spacer grob
grid.arrange(g1, blank, g2, heights = c(0.7, 0.05, 0.25), nrow = 3)

cor.matrix <- cor(wine.trn1[, 1:ncol(wine.trn1)], use = "complete.obs")
chart.Correlation(cor.matrix, histogram = TRUE, pch = 25)


wine.trn1$Alcohol[is.na(wine.trn1$Alcohol)] <- 0
wine.trn1$STARS[is.na(wine.trn1$STARS)] <- 0

wine.trn1$STARS <- as.factor(wine.trn1$STARS)
wine.trn1.numeric$STARS <- as.factor(wine.trn1.numeric$STARS)

wine.trn1$AcidIndex <- as.factor(wine.trn1$AcidIndex)
wine.trn1$LabelAppeal <- as.factor(wine.trn1$LabelAppeal)

wine.trn1.numeric.omit.na <- na.omit(wine.trn1.numeric)
wine.trn.omit.na <- na.omit(wine.trn1)

full.pois.numeric <- glm(TARGET ~ ., data = wine.trn1.numeric.omit.na, family = poisson())

full.pois <- glm(TARGET ~ ., data = wine.trn.omit.na, family = poisson())
# Lets check for Multi-Collinearity - lets find vif value and drop those that has

vifFit1.numeric <- faraway::vif(full.pois.numeric)
vifFit1 <- faraway::vif(full.pois)
# sort by descending
```

```r
vif.df.numeric <- as.data.frame(sort(vifFit1.numeric, decreasing = T))
vif.df <- as.data.frame(sort(vifFit1, decreasing = T))
names(vif.df) <- c("Multicolinearity score")
knitr::kable(vif.df.numeric)

knitr::kable(vif.df)

wine.trn.omit.na$AcidIndex[wine.trn.omit.na$AcidIndex %in% c(6, 7, 8, 9, 10, 11,
    12)] <- 5

full.pois <- glm(TARGET ~ ., data = wine.trn.omit.na, family = poisson())
# Lets check for Multi-Collinearity - lets find vif value and drop those that has
vifFit1 <- faraway::vif(full.pois)
# sort by descending
vif.df <- as.data.frame(sort(vifFit1, decreasing = T))
names(vif.df) <- c("Multicolinearity score")
knitr::kable(vif.df)

set.seed(3)

s0 = sample(1:nrow(wine.trn1.numeric.omit.na), 0.8 * nrow(wine.trn1.numeric.omit.na))
wine.training0 = wine.trn1.numeric.omit.na[s0, ]
wine.test0 = wine.trn1.numeric.omit.na[-s0, ]

s = sample(1:nrow(wine.trn.omit.na), 0.8 * nrow(wine.trn.omit.na))
wine.training = wine.trn.omit.na[s, ]
wine.test = wine.trn.omit.na[-s, ]

# http://www.ats.ucla.edu/stat/r/dae/poissonreg.htm
# http://www.ats.ucla.edu/stat/r/dae/nbreg.htm

full.pois0 <- step(glm(TARGET ~ ., data = wine.training0, family = poisson()), trace = FALSE)
pois.backward.step <- step(glm(TARGET ~ ., data = wine.training, family = poisson()),
    trace = FALSE)

round(summary(pois.backward.step)$coef, 2)
formula(pois.backward.step)

# full.pois <- glm(TARGET ~ ., data=wine.training, family=poisson())
# pois.backward.step = step(full.pois , trace = FALSE)

round(summary(full.pois0)$coef, 2)
formula(full.pois0)

round(summary(pois.backward.step)$coef, 2)
formula(pois.backward.step)

# reference: http://theses.ulaval.ca/archimede/fichiers/21842/apa.html

null.model.pois <- glm(TARGET ~ 1, data = wine.training, family = poisson())
pois.forward.step = step(null.model.pois, scope = list(lower = formula(null.model.pois),
    upper = formula(full.pois)), direction = "forward", trace = FALSE)
```

```r
round(coef(summary(pois.forward.step)), 2)
formula(pois.forward.step)

pois.manual0 <- step(glm(TARGET ~ VolatileAcidity + Chlorides + FreeSulfurDioxide +
    TotalSulfurDioxide + LabelAppeal + AcidIndex + STARS, data = wine.training0,
    family = poisson()), trace = FALSE)

full.pois.manual <- glm(TARGET ~ STARS + LabelAppeal + AcidIndex + VolatileAcidity,
    data = wine.training, family = poisson())
pois.manual = step(full.pois.manual, trace = FALSE)

round(summary(pois.manual0)$coef, 2)
formula(pois.manual0)

round(summary(pois.manual)$coef, 2)
formula(pois.manual)

full.nbm0 <- step(glm.nb(TARGET ~ ., data = wine.training0), trace = FALSE)

full.nbm <- glm.nb(TARGET ~ ., data = wine.training)
nbm.backward.step = step(full.nbm, trace = FALSE)

round(summary(nbm.backward.step)$coef, 2)
formula(nbm.backward.step)

round(summary(full.nbm0)$coef, 2)
formula(full.nbm0)

null.model.nbm <- glm.nb(TARGET ~ 1, data = wine.training)
nbm.forward.step = step(null.model.nbm, scope = list(lower = formula(null.model.nbm),
    upper = formula(full.nbm)), direction = "forward", trace = FALSE)

round(summary(nbm.forward.step)$coef, 2)
formula(nbm.forward.step)

nbm.manual0 <- step(glm.nb(TARGET ~ VolatileAcidity + Chlorides + FreeSulfurDioxide +
    TotalSulfurDioxide + LabelAppeal + AcidIndex + STARS, data = wine.training0),
    trace = FALSE)

full.nbm.manual <- glm.nb(TARGET ~ STARS + LabelAppeal + AcidIndex + VolatileAcidity,
    data = wine.training)
nbm.manual = step(full.nbm.manual, trace = FALSE)

round(summary(nbm.manual)$coef, 2)
formula(nbm.manual)

round(summary(nbm.manual0)$coef, 2)
formula(nbm.manual0)

full.lm0 <- step(lm(TARGET ~ ., data = wine.training0), trace = FALSE)
round(summary(full.lm0)$coef, 2)
formula(full.lm0)
```

```r
full.lm <- lm(TARGET ~ ., data = wine.training)
lm.backward.step = step(full.lm, trace = FALSE)

round(summary(lm.backward.step)$coef, 2)
formula(lm.backward.step)

nothing.mod.lnr <- lm(TARGET ~ 1, data = wine.training)
lm.forward.step <- step(nothing.mod.lnr, scope = list(lower = formula(nothing.mod.lnr),
    upper = formula(full.lm)), direction = "forward", trace = FALSE)

round(summary(lm.forward.step)$coef, 2)
formula(lm.forward.step)

lm.manual0 <- step(lm(TARGET ~ VolatileAcidity + Chlorides + FreeSulfurDioxide +
    TotalSulfurDioxide + LabelAppeal + AcidIndex + STARS, data = wine.training0),
    trace = FALSE)
round(summary(lm.manual0)$coef, 2)
formula(lm.manual0)

full.lm.manual <- lm(TARGET ~ STARS + LabelAppeal + AcidIndex + VolatileAcidity,
    data = wine.training)
lm.manual = step(full.lm.manual, trace = FALSE)

round(summary(lm.manual)$coef, 2)
formula(lm.manual)

# RMSE - Root Mean Square Error ( / CrossValidation)
rmse <- function(testDataset, model) {
    return(round(sqrt(mean((predict(model, testDataset) - testDataset$TARGET)^2)),
        4))
}

validationResults <- data.frame(ModelType = c("Poisson - Stepwise Backward", "Poisson - Stepwise Forwar
    "Poisson - Manual", "Negative Binomial - Backward", "Negative Binomial - Forward",
    "Negative Binomial - Manual", "Linear - Stepwise Backward", "Linear - Stepwise Forward",
    "Linear - Manual"), RMSE = c(rmse(wine.test, pois.backward.step), rmse(wine.test,
    pois.forward.step), rmse(wine.test, pois.manual), rmse(wine.test, nbm.backward.step),
    rmse(wine.test, nbm.forward.step), rmse(wine.test, nbm.manual), rmse(wine.test,
        lm.backward.step), rmse(wine.test, lm.forward.step), rmse(wine.test, lm.manual)),
    Adj_R2 = c(NA, NA, NA, NA, NA, NA, round(summary(lm.backward.step)$adj.r.squared,
        2), round(summary(lm.forward.step)$adj.r.squared, 2), round(summary(lm.manual)$adj.r.squared,
        2)), AIC = c(AIC(pois.backward.step), AIC(pois.forward.step), AIC(pois.manual),
        AIC(nbm.backward.step), AIC(nbm.forward.step), AIC(nbm.manual), AIC(lm.backward.step),
        AIC(lm.forward.step), AIC(lm.manual)), Coefs = c(length(pois.backward.step$coefficients) -
        1, length(pois.forward.step$coefficients) - 1, length(pois.manual$coefficients) -
        1, length(nbm.backward.step$coefficients) - 1, length(nbm.forward.step$coefficients) -
        1, length(nbm.manual$coefficients) - 1, length(lm.backward.step$coefficients) -
        1, length(lm.forward.step$coefficients) - 1, length(lm.manual$coefficients) -
        1))

kable(validationResults)
```

```r
validationResults0 <- data.frame(ModelType = c("Poisson - Step model", "Poisson - Manual",
    "Negative Binomial - Step model", "Negative Binomial - Manual", "Linear - Step model",
    "Linear - Manual"), RMSE = c(rmse(wine.test0, full.pois0), rmse(wine.test0, full.pois0),
    rmse(wine.test0, full.nbm0), rmse(wine.test0, nbm.manual0), rmse(wine.test0,
        full.lm0), rmse(wine.test0, lm.manual0)), Adj_R2 = c(NA, NA, NA, NA, round(summary(full.lm0)$adj
    2), round(summary(lm.manual0)$adj.r.squared, 2)), AIC = c(AIC(full.pois0), AIC(pois.manual0),
    AIC(full.nbm0), AIC(nbm.manual0), AIC(full.lm0), AIC(lm.manual0)), Coefs = c(length(full.pois0$coef
    1, length(pois.manual0$coefficients) - 1, length(full.nbm0$coefficients) - 1,
    length(nbm.manual0$coefficients) - 1, length(full.lm0$coefficients) - 1, length(lm.manual0$coeffici
        1))


kable(validationResults0)

colnames(wine.evl)[1] <- "INDEX"
wine.evl$Alcohol[is.na(wine.evl$Alcohol)] <- 0
wine.evl$STARS[is.na(wine.evl$STARS)] <- 0

wine.evl$STARS <- as.factor(wine.evl$STARS)
# wine.evl$AcidIndex <- as.factor(wine.evl$AcidIndex) wine.evl$LabelAppeal <-
# as.factor(wine.evl$LabelAppeal) wine.evl$AcidIndex[wine.evl$AcidIndex %in%
# c(6,7,8,9,10,11,12) ]<- 5

wine.evl$TARGET <- round(predict(full.lm0, newdata = wine.evl, type = "response"))


wine.evl.omit.na <- na.omit(wine.evl)
ggplot(wine.evl.omit.na, aes(x = TARGET)) + geom_histogram(binwidth = 0.5) + theme(axis.text = element_t
    axis.title = element_text(size = 10))
mn <- mean(wine.evl.omit.na$TARGET)
vr <- var(wine.evl.omit.na$TARGET)

wine.evl.short <- wine.evl[, c("INDEX", "STARS", "LabelAppeal", "TARGET")]
names(wine.evl.short) <- c("IDX", "STRS", "LBL", "TGT")

wine.evl.short.1 <- wine.evl.short[1:ceiling(nrow(wine.evl.short)/3), ]
wine.evl.short.2 <- wine.evl.short[(ceiling(nrow(wine.evl.short)/3) + 1):(ceiling(nrow(wine.evl.short)/3
    2), ]
wine.evl.short.3 <- wine.evl.short[((ceiling(nrow(wine.evl.short)/3) * 2) + 1):nrow(wine.evl.short),
    ]

wine.evl.short.3 <- rbind(wine.evl.short.3, c(NA, NA, NA, NA))

wine.evl.split <- cbind(wine.evl.short.1, wine.evl.short.2, wine.evl.short.3)

kable(wine.evl.split, caption = "Predictions")
```