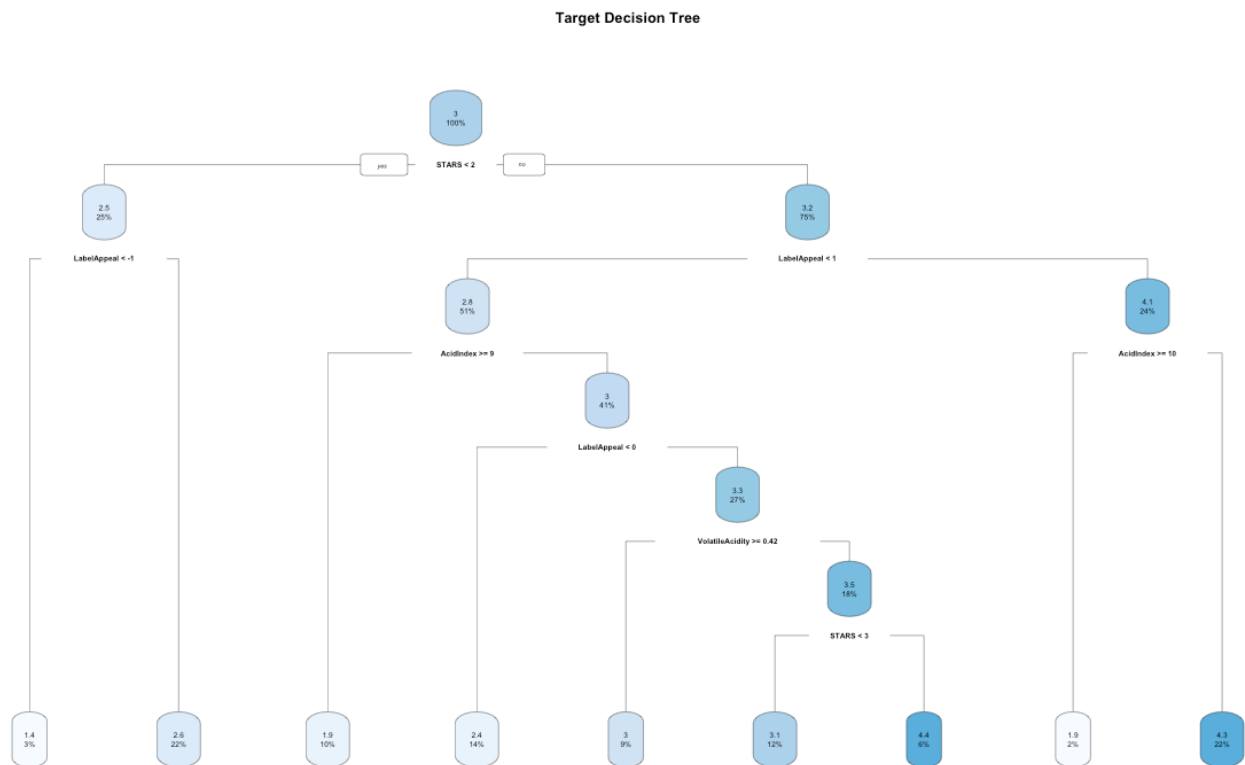


# Wine Project Bingo Bonus

*By Mimi Trinh*

In this bingo bonus, I attempt for 20 points by building a decision tree to predict the target variable. Below is the decision tree output.



Below is the Poisson regression output.

```
Call:
glm(formula = wine$TARGET ~ wine$VolatileAcidity + wine$LabelAppeal +
     wine$AcidIndex + wine$IMP_Chlorides + wine$IMP_FreeSulfurDioxide +
     wine$IMP_TotalSulfurDioxide + wine$IMP_Sulphates + wine$IMP_Alcohol +
     wine$M_STARS + wine$IMP_STARS, family = poisson(link = "log"),
     data = wine)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.1916	-0.6451	0.0135	0.4543	3.7735

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	1.485e+00	4.556e-02	32.600	< 2e-16	***
wine\$VolatileAcidity	-3.194e-02	6.518e-03	-4.900	9.6e-07	***
wine\$LabelAppeal	1.584e-01	6.128e-03	25.851	< 2e-16	***
wine\$AcidIndex	-8.509e-02	5.180e-03	-16.429	< 2e-16	***
wine\$IMP_Chlorides	-3.602e-02	1.646e-02	-2.188	0.028652	*
wine\$IMP_FreeSulfurDioxide	1.017e-04	3.508e-05	2.900	0.003730	**
wine\$IMP_TotalSulfurDioxide	8.290e-05	2.273e-05	3.647	0.000265	***
wine\$IMP_Sulphates	-1.219e-02	5.752e-03	-2.120	0.034042	*
wine\$IMP_Alcohol	3.602e-03	1.407e-03	2.560	0.010472	*
wine\$M_STARS	-1.038e+00	1.696e-02	-61.220	< 2e-16	***
wine\$IMP_STARS	1.887e-01	6.090e-03	30.994	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 22861 on 12794 degrees of freedom  
 Residual deviance: 13831 on 12784 degrees of freedom  
 AIC: 45795

Among nine significant variables in the Poisson model, the decision tree model has four of these variables: STARS, Label Appeal, Acid Index, and Volatile Acidity. The decision tree uses Acid Index and Volatile Index only once whereas it uses STARS and Label Appeal in multiple layers. This is interesting since during the project when we run the correlation plot, STARS and Label Appeal are the only two predictors with a direct, strong, positive correlation with the target variable.

Also, it's interesting that the decision tree model goes from 1.4 to 4.4 in predicted value whereas the actual value in the target variable goes from 0 to 8. Therefore, Poisson, negative binomial, ZIP, and ZNIB are probably better tools to use for this project than decision tree. Perhaps decision tree can be used instead of the zero-inflated model in ZIP and ZINB to determine whether the wine is sold or not, then we can use Poisson and NB to predict the number of cases sold.