

# Data 621 - HW5

Devin Teran, Atina Karim, Tom Hill, Amit Kapoor

5/23/2021

## Contents

Overview	1
Data Exploration	1
Data Preparation	2
Build Models	2
Select Models	2
Code Appendix	2

## Overview

In this assignment, we 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.

The 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.

## Data Exploration

Below is the description of the variables of interest in the data set.

VARIABLE NAME	DEFINITION	THEORETICAL EFFECT
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	

VARIABLE NAME	DEFINITION	THEORETICAL EFFECT
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	

## Data Preparation

## Build Models

## Select Models

## Code Appendix

```
knitr::opts_chunk$set(echo=FALSE, error=FALSE, warning=FALSE, message=FALSE, fig.align = "center")
# Libraries

set.seed(621)
# training data
wine_train <- read.csv('https://raw.githubusercontent.com/hillt5/DATA_621/master/HW5/wine-training-data')
# test data
wine_test <- read.csv('https://raw.githubusercontent.com/hillt5/DATA_621/master/HW5/wine-evaluation-data')
```