Modeling Assigment #4

Brandon Moretz

### Introduction

In this lab we are going the explore a data set consisting of various attributes of approximately twelve thousand commercially available wines in order to predict if they will be sold, if so, how many cases will be sold, and the number of stars an expert would give this wine using a Vivino style rating system. The ultimate purpose of this data analysis and resulting model is to provide a large wine distributor client our recommendations on what kind of wines will be ordered and in what amounts so that we can be operationally prepared to manage the supply chain and resulting required logistics. Additionally, we will attempt to quantity what makes a wine ‘good’ based upon its chemical composition and characteristics.

### Exploratory data analysis

The first step in this analysis project is to take a clean sweep to the data and look for any bad encodings or invalid values. The alcohol content variable has approximately 600 bottles with negative or missing alcohol content, which is not a possible value so these bottles (records) will be excluded from our analysis.

The label appeal variable has the scale -2 to 2, which we will adjust by +3 so that it is on a “normal” 1-5 scale.

### Research

### Conclusion