

Analysis of Red Wine

Harsh Panchal

Department of Mechanical and Industrial Engineering

University of Massachusetts, Amherst

MIE 612: Descriptive Analytics, Fall 2023

Prof. Shannon Roberts

Abstract

Wine quality is very important for wine sellers as a good quality of wine is always in demand in market and due to the limited availability in market the makers always want the quality of wine to be best. This study project does the analysis of the red wine and its different parameters which affects the red wine quality and pH level. The main study is to know that how the parameters like Acidity, Sulfur dioxide affects the quality and Alcohol, Sugar affects the pH level of red wine. Moreover to this the study includes to know whether a correlation exists between the variables. The study is done with the statistical analysis using Regression and Correlation analysis. This study can be helpful for wine makers for making good quality of wine.

Keywords

Red Wine Quality, Wine Quality analysis, Red wine Analysis, pH level of Red wine

Introduction

Red wine is famous all around the world from centuries due to its different types and benefits. The chemical properties of the red wine makes it more interested for all the researchers and wine makers who are having a great knowledge about the chemical properties of the wine.

This study determines the impact of the different chemicals to the quality of the red wine and the pH level of it. Many of the properties like acidity, alcohol; sulfur dioxide plays a role in determining the quality of the red wine.

Alcohol plays one of the major roles is giving a good quality of wine. A perfect proportion of the chemical gives a great tasting experience to the person. There are various different types which have different proportion of the chemicals added to get a desired wine quality.

Another most common chemical used is the sulfur dioxide which is very important chemical for the wine. Together with the alcohol it makes the red wine what it is famous for.

The study shows how the quality of red wine is depended on the multiple factors and chemicals, how different proportions of the chemicals like alcohol, sugar sulfur dioxide gives different quality of red wine.

Methods

In the dataset there are in total 12 variables among them 9 are independent variables and the rest 3 are dependent. All the variables are quantitative variables ranging from a low to high value. For example the variable “quality” ranges from minimum 3 to maximum 8. The variables are mostly related to the chemical properties of the wine.

The methods used are regression and correlation analysis.

Regression is used for knowing the relation between the dependent variable and independent variable like quality and pH. In one graph the independent variables are fixed acidity and total sulfur dioxide and the dependent variable is the quality. In other graph the independent variable is sugar and alcohol and the dependent variable is pH.

The regression is done using the “lm()” function for the variables fixed acidity, total sulfur dioxide and quality.

Here the other method is correlation analysis. It is used for knowing the strength of the linear relationship between the variables. The correlation is also performed between the same pair of independent and dependent variables.

The Correlation is done using the “cor()” function which gives the relationship between the variables.

For plotting the graph ggplot2 library is used and the “ggplot()” function. Here in the graph the result for regression and correlation is added which determines the relation of the independent variables and the dependent variables.

Results

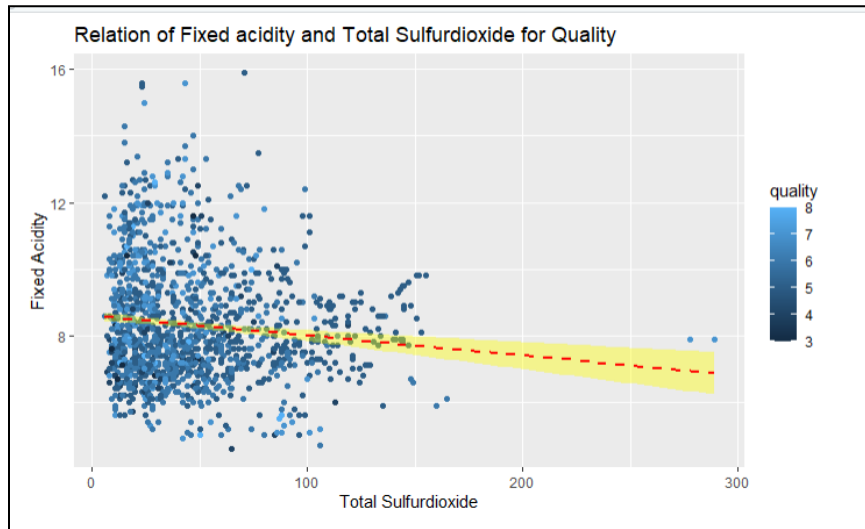


Figure 1: Fixed Acidity vs. Total Sulfur dioxide and the Quality

In the above figure there is a negative line of regression as shown in the dashed red line. The variables are more negatively correlated and also we can see some outliers too. The quality of a specific data point can be known from the color of it.

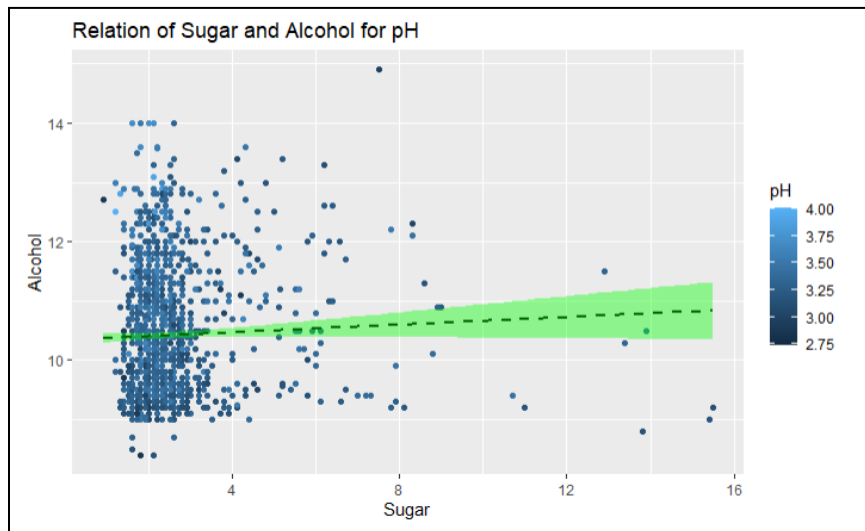


Figure 2: Alcohol vs. Sugar and the pH

In the above figure there is a positive regression line as shown in the dark green dashed line. Here more data points are negatively correlated and there few of the outliers too. So less for better quality less sugar and more alcohol proportion is preferred.

Discussion

In the discussions the main part is to know the what different combinations of the chemical properties influences the red wine quality and how this analysis is helpful to the wine manufactures to make the best quality of red wine possible in market. To produce the desired wine quality and know what proportions of the chemicals gives the best wine.

Moreover, the discussion emphasis to know how each of the chemical whether alcohol, sulfur dioxide or acidity level combined puts a major affects to the quality.

By doing the statistical analysis using the regression and correlation of the variable and there proportion data values I aspire to provide a better understanding of each major chemicals affecting the quality.

In conclusion this section highlights the challenges for making a good quality red wine and using of the analysis method gives a better understanding which becomes helpful for manufactures for making a desired red wine quality.

References

Red Wine dataset

<https://archive.ics.uci.edu/dataset/109/wine>