# Text Description automatically generated

# Project Report by Team 14:

# Topic: Red Wine Quality

# Team Members:

|  |  |
| --- | --- |
| Sai Teja Alladi | UTA ID: 1001980895 |
| Pruthviraj Ganji | UTA ID: 1001988085 |
| Sanjeeth Reddy Kandi | UTA ID: 1002005821 |
| Sai Mrinal Reddy Aleti | UTA ID: 1001965817 |

**Red Wine Quality Prediction**

**Executive Summary –**

We are a well reputed Wine dealer, we intend to sell **only high Quality** wine in the market at a much better price and with great taste compared to the present competitors.

We assess the quality of wine based on the several factore of wine such as Alcohol, Ph-value, Acidity and Residual Sugars and others. For that we are taking a very low underrated wine which Is not very popular too.

**Problem Statement:**

The Prediction of wine quality from the present high range value wines and all other wines in the market. So that for example we’ve taken the present Golconda wine which is very underrated and not very popular in the market, then to present the quality of this wine after comparing with all others wines.

**Problem Solution:**

Using this data set containing 1599 observations containing all wine factor details and values of some specific wines. We "Team 14" predict on the data set and provide data insights and recommendations and get the best wine quality and the best in the market.

**Data Description:**

The data set has 1599 observations with 12 different variables as explained below,

**Independent Variables:**

* Fixed acidity
* Volatile acidity
* Citric acid
* Residual sugar
* Chlorides
* Free sulfur
* Total sulfur dioxide
* Density
* Ph
* Sulphates
* Alcohol

**Dependent (Target) Variable:**

* Quality – 5—9

**Data Visualization:**

First process is to check which are the factors have the least impact on making the wine, so that we can consider them as least important factor to decide the wine quality based upon that factor graph comparing with the other factor deciding as the quality.

The below bar chart depicts the analysis between alcohol and quality by that we can declare that this factor has the best analysis with the factor.

Chart, bar chart

Description automatically generated

The following bar chart depicts, how the quality of wine is varied with respect to the Ph-value.

Chart

Description automatically generated

The following scatterplot shows the relationship between the citric acid and the Ph-value.

Chart, scatter chart

Description automatically generated

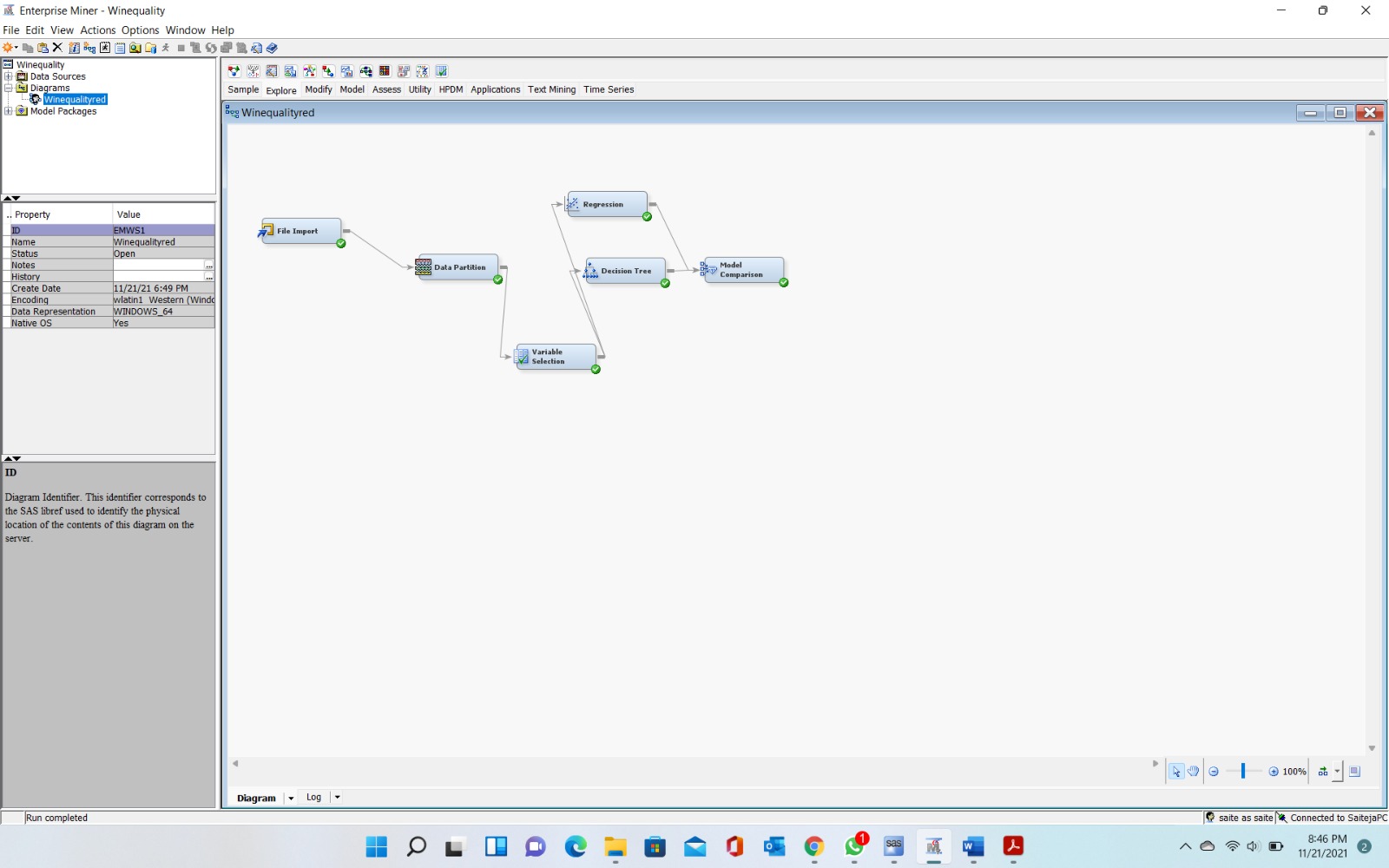
* Data prediction techniques we have used are decision trees and logistic regression.

Some of the input variables are

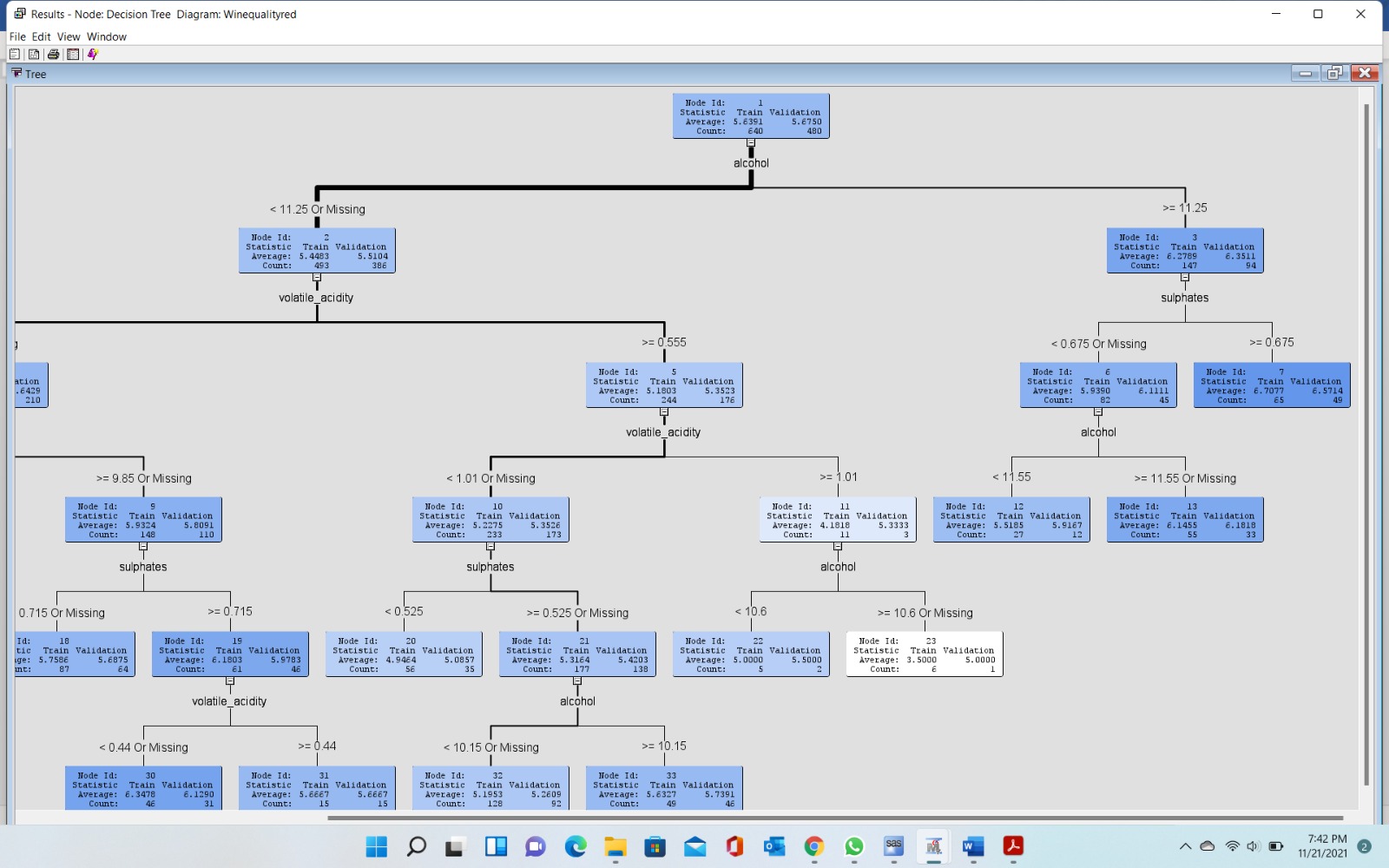
* Alcohol
* Volatile acidity
* Sulphates

We will be partitioning our dataset into 3 parts:

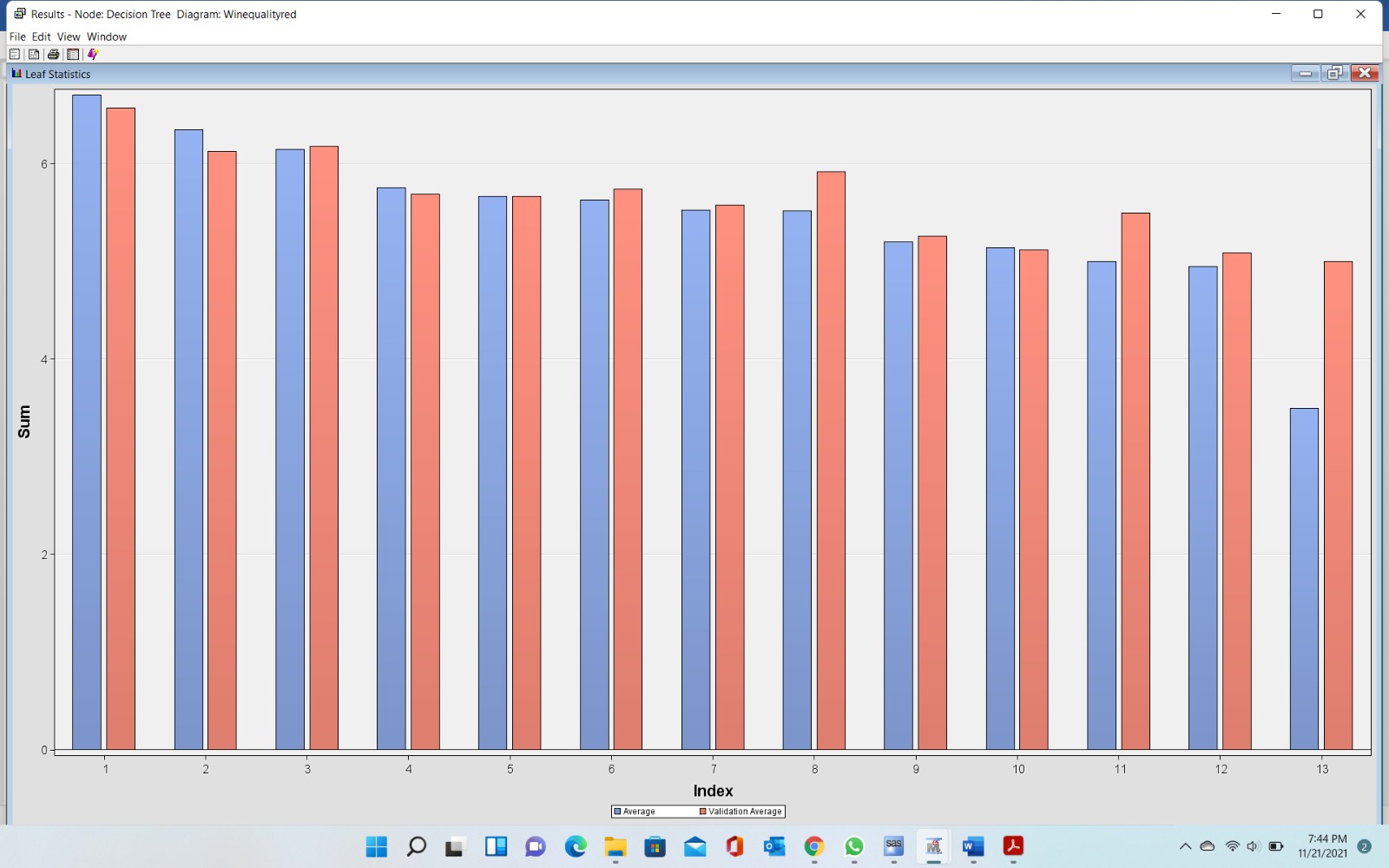
* 40% as the Training data
* 30% as the Validation data
* 30% as the Test data



Decision tree:



Input variables:



Logistic regression:

Chart, line chart

Description automatically generated

Graphical user interface, chart

Description automatically generated

Methodology

REGRESSION:

* Based on the analysis; we can come up with the following hypothesis:

H0: quality of the red wine is not determined by the mentioned factors

H1: quality of the red wine is determined by the mentioned factors

* Based on the ANOVA, we can say that since the F calculated value is 81.35 which is greater than the F tabulated value at 0.05 significance level, we can conclude that we should reject the null hypothesis and give a conclusion that quality of the red wine is determined by the mentioned factors
* Moreover, the table below shows that the R square is 0.3606 which shows that the model can explain 36.06 of the quality of wine. This percentage can be adjusted by conducting backward elimination on the model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| **Model** | 11 | 375.75440 | 34.15949 | 81.35 | <.0001 |
| **Error** | 1587 | 666.41070 | 0.41992 |  |  |
| **Corrected Total** | 1598 | 1042.16510 |  |  |  |

PARAMETER ESTIMATES:

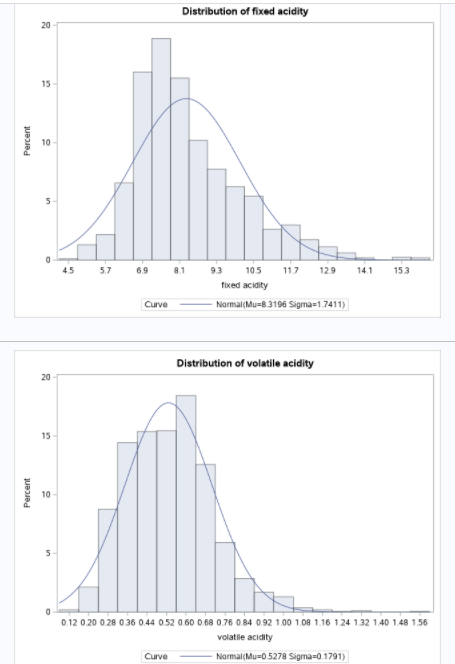
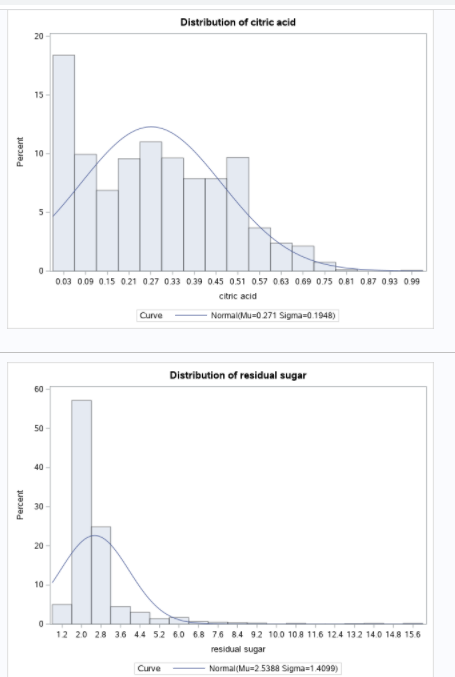
* The regression of the data shows that by treating Quality of the wine as my dependent variable, the parameters estimates that are significant in this analysis include volatile acidity, chlorides, total sulfur dioxide, sulphates and alcohol.

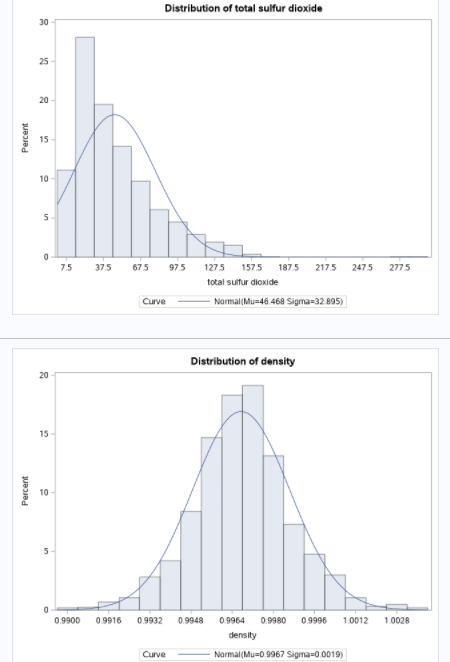
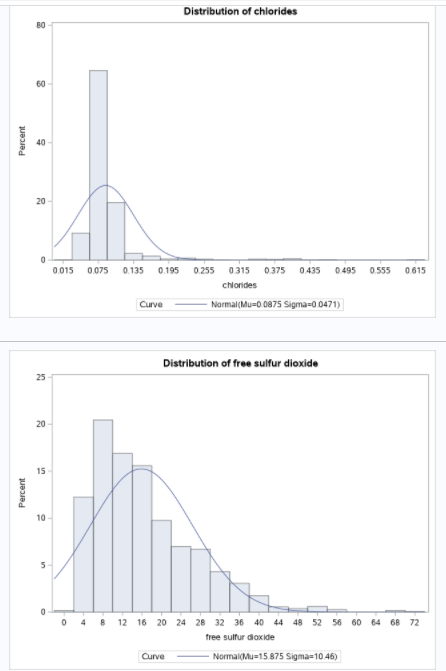
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | | |
| **Variable** | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| |
| **Intercept** | Intercept | 1 | 21.96521 | 21.19457 | 1.04 | 0.3002 |
| **fixed acidity** | fixed acidity | 1 | 0.02499 | 0.02595 | 0.96 | 0.3357 |
| **volatile acidity** | volatile acidity | 1 | -1.08359 | 0.12110 | -8.95 | <.0001 |
| **citric acid** | citric acid | 1 | -0.18256 | 0.14718 | -1.24 | 0.2150 |
| **residual sugar** | residual sugar | 1 | 0.01633 | 0.01500 | 1.09 | 0.2765 |
| **chlorides** | chlorides | 1 | -1.87423 | 0.41928 | -4.47 | <.0001 |
| **free sulfur dioxide** | free sulfur dioxide | 1 | 0.00436 | 0.00217 | 2.01 | 0.0447 |
| **total sulfur dioxide** | total sulfur dioxide | 1 | -0.00326 | 0.00072873 | -4.48 | <.0001 |
| **density** | density | 1 | -17.88116 | 21.63310 | -0.83 | 0.4086 |
| **pH** | pH | 1 | -0.41365 | 0.19160 | -2.16 | 0.0310 |
| **sulphates** | sulphates | 1 | 0.91633 | 0.11434 | 8.01 | <.0001 |
| **alcohol** | alcohol | 1 | 0.27620 | 0.02648 | 10.43 | <.0001 |

* The Regression equation is :

***Quality= 21.96521+0.02499fixed acidity-1.08359Volatile acidity-1.08359citric acid+0.01633residual sugar-1.87423chlorides+0.00436 free sulfur dioxide-0.00326 total sulfur dioxide-17.88116 density-0.41365 pH+0.91633 sulphates+0.27620 alcohol***

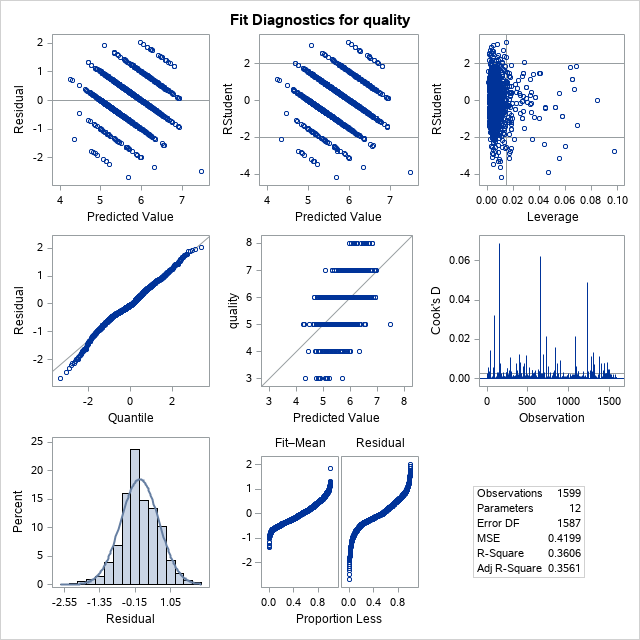
**Visualizations:**



**Diagnostics for quality**:

All residual values compared with all observations and graphs.



**Results:**

**ANALAYSIS FOR THE GOLCONDA WINE**

These are the predicted values based upon the analysis from the dataset and from the quality upon 18 values which have the good quality which is greater than and equal to 8.

Finalized Values which has the quality 8.

Graphical user interface, application, table, Excel

Description automatically generated

Output for the Golconda wine analysis:

| **Variable** | **Label** | **Mean** | **Std Dev** | **Minimum** | **Maximum** | **N** |
| --- | --- | --- | --- | --- | --- | --- |
| fixed acidity  volatile acidity  citric acid  residual sugar  chlorides  free sulfur dioxide  total sulfur dioxide  density  pH  sulphates  alcohol  quality | fixed acidity  volatile acidity  citric acid  residual sugar  chlorides  free sulfur dioxide  total sulfur dioxide  density  pH  sulphates  alcohol  quality | 8.5666667  0.4233333  0.3911111  2.5777778  0.0684444  13.2777778  33.4444444  0.9952122  3.2672222  0.7677778  12.0944444  8.0000000 | 2.1196559  0.1449138  0.1995256  1.2950383  0.0116781  11.1556135  25.4332395  0.0023783  0.2006403  0.1153795  1.2240109  0 | 5.0000000  0.2600000  0.0300000  1.4000000  0.0440000  3.0000000  12.0000000  0.9908000  2.8800000  0.6300000  9.8000000  8.0000000 | 12.6000000  0.8500000  0.7200000  6.4000000  0.0860000  42.0000000  88.0000000  0.9988000  3.7200000  1.1000000  14.0000000  8.0000000 | 18  18  18  18  18  18  18  18  18  18  18  18 |

**Conclusion:**

We can conclude that from the above predicted values that the most high quality and higher cost wine is basically which is present in the low quality wine at very cheaper price.

The quality depends on the factors of the acid and the ph value so that it can be made in that known directions and quantities.

So, that we hereby said that Golconda wine is the best wine in all aspects and prospects when comparing with any other wine and the world’s top most wine.