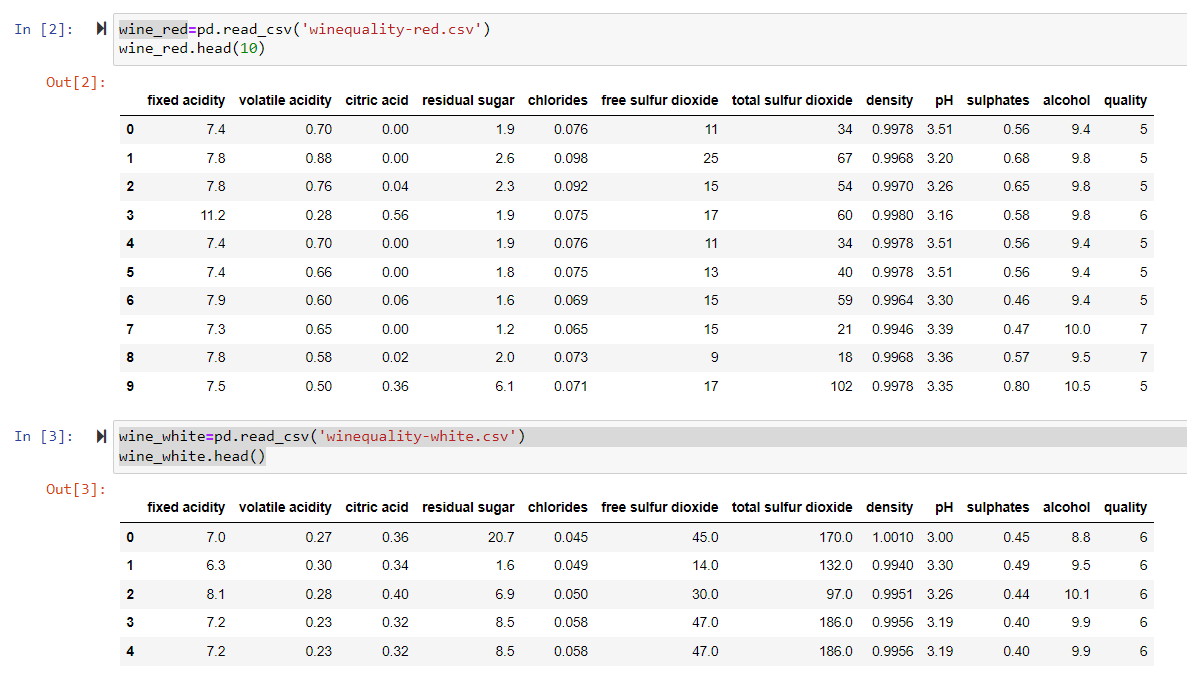
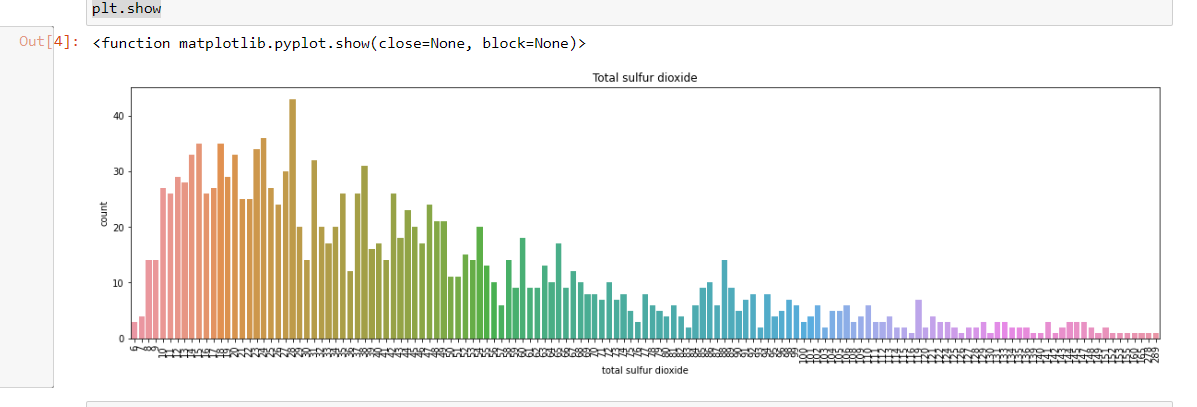
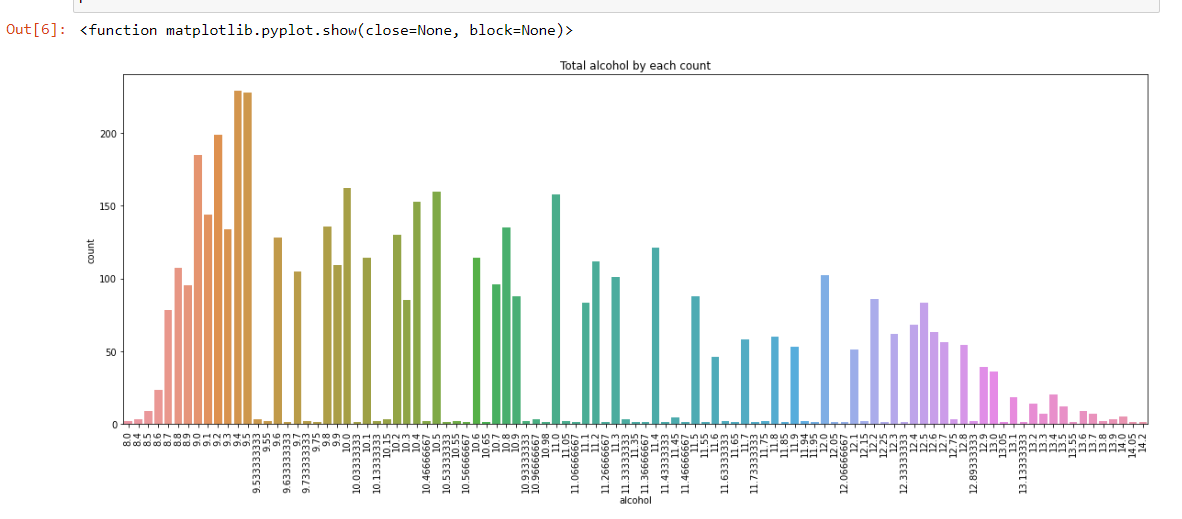
**1.Display the data of white wine and red wine**

I am sending a bar chart for representing data in graphical way







**2.Get the size of both white wine and red wine.**

------------- WHITE WINE -------------

Number of rows: 4898

Number of columns: 12

-------------- RED WINE --------------

Number of rows: 1599

Number of columns: 12

--------------------------------------

**3.Get the number of null in both type of wines**

------------- WHITE WINE -------------

fixed acidity 0

volatile acidity 0

citric acid 0

residual sugar 0

chlorides 0

free sulfur dioxide 0

total sulfur dioxide 0

density 0

pH 0

sulphates 0

alcohol 0

quality 0

dtype: int64

-------------- RED WINE --------------

fixed acidity 0

volatile acidity 0

citric acid 0

residual sugar 0

chlorides 0

free sulfur dioxide 0

total sulfur dioxide 0

density 0

pH 0

sulphates 0

alcohol 0

quality 0

dtype: int64

-----------------------------------------

**4.Check for duplicates and remove them in both type of wine data**

------------- WHITE WINE -------------

No of rows duplicated: 937

Total no of rows: 4898

Total no of rows after removing duplicate rows: 3961

-------------- RED WINE --------------

No of rows duplicated: 240

Total no of rows: 1599

Total no of rows after removing duplicate rows: 1359

--------------------------------------

**5.Get the number of unique values in all Feature and print them**

------------- WHITE WINE -------------

No of Unique values in fixed acidity: 68

No of Unique values in volatile acidity: 125

No of Unique values in citric acid: 87

No of Unique values in residual sugar: 310

No of Unique values in chlorides: 160

No of Unique values in free sulfur dioxide: 132

No of Unique values in total sulfur dioxide: 251

No of Unique values in density: 890

No of Unique values in pH: 103

No of Unique values in sulphates: 79

No of Unique values in alcohol: 103

No of Unique values in quality: 7

-------------- RED WINE --------------

No of Unique values in fixed acidity: 96

No of Unique values in volatile acidity: 143

No of Unique values in citric acid: 80

No of Unique values in residual sugar: 91

No of Unique values in chlorides: 153

No of Unique values in free sulfur dioxide: 60

No of Unique values in total sulfur dioxide: 144

No of Unique values in density: 436

No of Unique values in pH: 89

No of Unique values in sulphates: 96

No of Unique values in alcohol: 65

No of Unique values in quality: 6

--------------------------------------

**6.Print the mean density of both the sets**

Mean density of white wine: 0.9940273764801896

Mean density of red wine: 0.9967466791744831

**7.Combine both the datasets with color of the wine in column**



**8.Remove the null values if any**

fixed acidity 0

volatile acidity 0

citric acid 0

residual sugar 0

chlorides 0

free sulfur dioxide 0

total sulfur dioxide 0

density 0

pH 0

sulphates 0

alcohol 0

quality 0

color 0

dtype: int64

**9.Remove the outliers if any**

no outliers

**10. Do wines with higher alcohol content receive better ratings?**

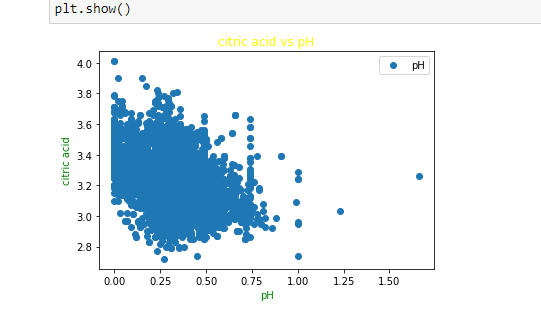
By looking into the details, we can see that good quality wines have higher levels of alcohol on average, have a lower volatile acidity on average, higher levels of sulphates on average, and higher levels of residual sugar on average

Based on our data YES.

**11.Find correlation between citric acid content and PH.**

If the correlation coefficient between citric acid content and pH is negative, that indicates a negative correlation between the two variables. In other words, as the citric acid content increases, the pH tends to decrease.

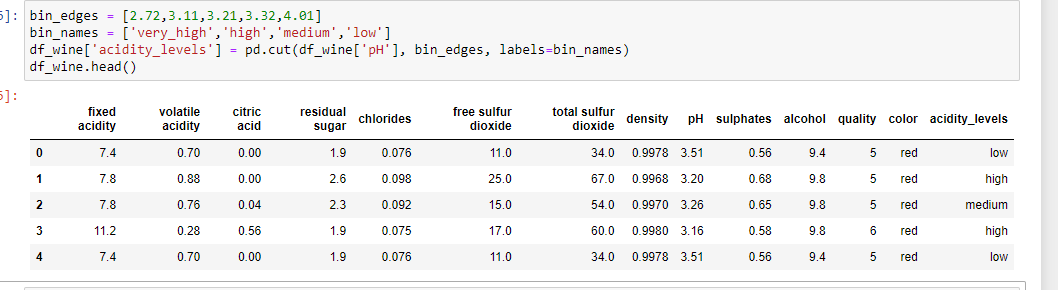
**12.Find regression of PH on citric acid**



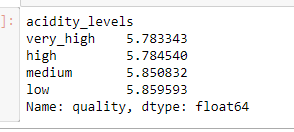
**13.Find how many variables are independent.**

So, in this data set, the number of independent variables would be 11 (fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol) as these variables can potentially influence or predict the outcome of the dependent variable "quality".

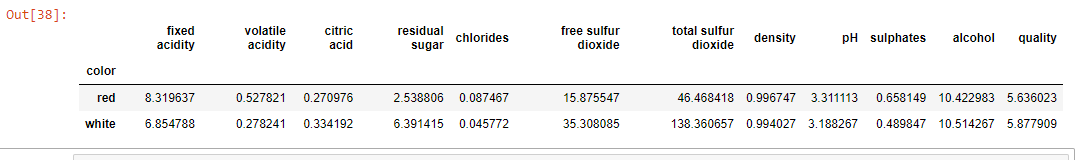
**14.Create a new column as the acidity level. Divide the existing data into 5 groups based on the acidity level. The levels are to be named ['High', 'Moderately\_High', 'Medium', 'Low']. Acidity may be obtained from the PH.**



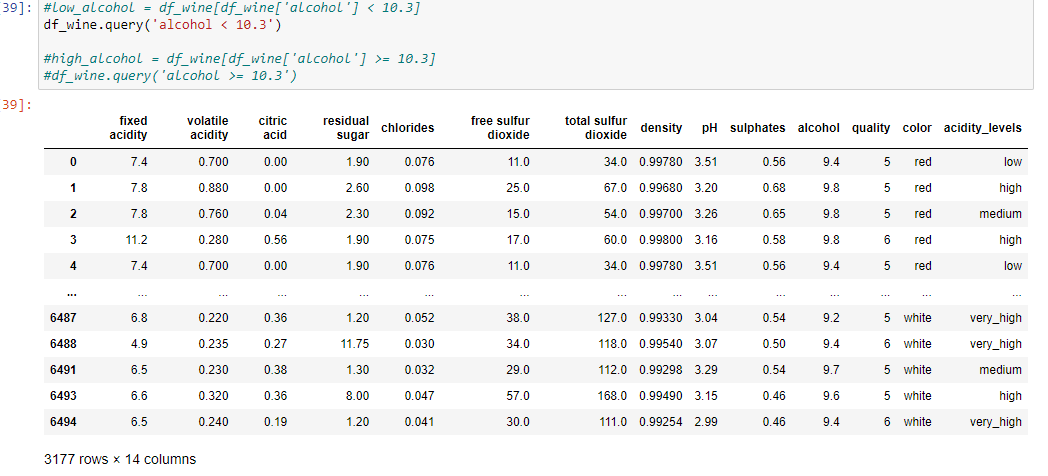
**15.Find the mean quality of each acidity level with groupby. Use groupby**



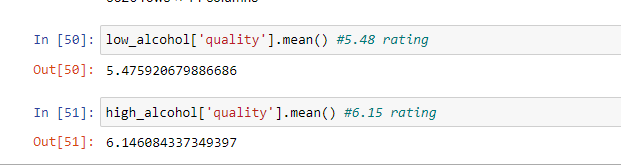
**16.Is a certain type of wine (red or white) associated with higher quality?**



**17.Select samples with alcohol content less than the median**



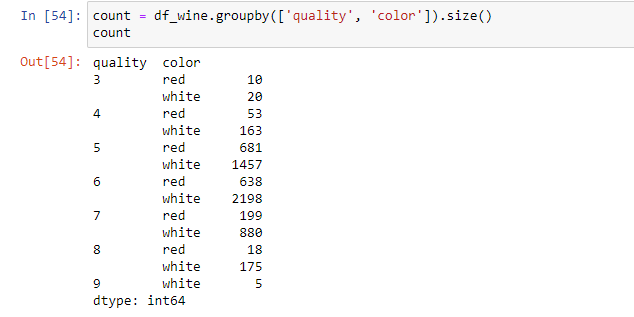
**18.Get the quality rating for the low alcohol and high alcohol groups.**



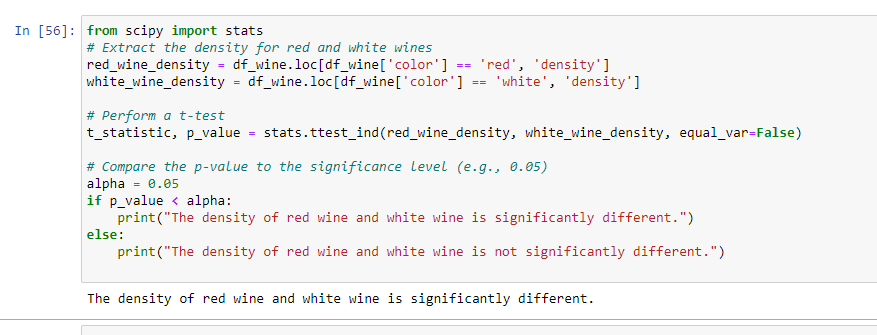
**19.Do sweeter wines get more ratings?**

The relationship between the residual sugar content and the quality rating of wine is not necessarily straightforward, as there are many factors that can affect the overall quality and taste of a wine. Some people may prefer sweeter wines and rate them higher, while others may prefer drier wines and rate them higher. The residual sugar content is just one of many factors that can contribute to the taste and quality of a wine.

**20.Get the number of counts based on ratings and color of wine.**



**21.Is the density of red wine and white wine similar?**



**22.Is the confidence interval for the differences of the proportion of white wine with rating 3 and the proportion of red wine with rating 3.**