Name: Borja, Angelo Louis C.

Section: CPE22S3

Performed on: 03/11/2024

Submitted on: 03/11/2024

Submitted to: Engr . Roman M. Richard

```
import numpy as np
import pandas as pd
import statistics as stats
from google.colab import drive
drive.mount('/content/drive')
df_wine = pd.read_csv('/content/drive/MyDrive/CSVS/winequality-red.csv')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.m
```

```
#Column names of df_wine
df_wine.columns
```

```
Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'pH', 'sulphates', 'alcohol', 'quality'],
      dtype='object')
```

```
#Data types of the data
df_wine.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   fixed acidity         1599 non-null   float64
 1   volatile acidity      1599 non-null   float64
 2   citric acid           1599 non-null   float64
 3   residual sugar        1599 non-null   float64
 4   chlorides             1599 non-null   float64
 5   free sulfur dioxide   1599 non-null   float64
 6   total sulfur dioxide  1599 non-null   float64
 7   density               1599 non-null   float64
 8   pH                    1599 non-null   float64
 9   sulphates             1599 non-null   float64
 10  alcohol               1599 non-null   float64
 11  quality               1599 non-null   int64  
dtypes: float64(11), int64(1)
memory usage: 150.0 KB
```

```
#first 10 entries
df_wine.head(10)
```

Next steps: View recommended plots

```
#last 10 entiries
df_wine.tail(10)
```

```
#total number of records
print(df_wine.shape[0])
```

```
      1599
```

```
#Create a new column high quality that display 'Yes' if the value of 'quality' is above 5, otherwise 'Nc
df_wine['high quality'] =  np.where(df_wine['quality'] > 5, 'Yes', 'No')
df_wine.head(10)
```

Next steps: View recommended plots

```
#Create a new Dataframe 'highQuality' that gathers data of high quality red wines
highQuality = df_wine[df_wine['high quality'] == 'Yes'].copy()
highQuality.head()
```

Next steps: View recommended plots

```
#Create a new Dataframe 'lowQuality' that gathers data of low quality red wines
lowQuality = df_wine[df_wine['high quality'] == 'No'].copy()
lowQuality.head()
```

Next steps: View recommended plots

```
#25th and 75th percentile of both dataframes, column total sulfur dioxi
print("highQuality dataframe 25th and 75th [total sulfur dioxide]: "+str(np.percentile((highQuality['tot
print("lowQuality dataframe 25th and 75th [total sulfur dioxide]:  "+str(np.percentile((lowQuality['tota
```

```
highQuality dataframe 25th and 75th [total sulfur dioxide]: 20.0       50.0
lowQuality dataframe 25th and 75th [total sulfur dioxide]:  23.75      78.0
```

```
#median of both dataframes, column total sulfur dioxide
print("Median of the column total sulfur dioxide of the dataframe highQuality: ",np.mean(highQuality['tc
print("Median of the column total sulfur dioxide of the dataframe lowQuality:  ",np.mean(lowQuality['tot
```

```
Median of the column total sulfur dioxide of the dataframe highQuality:  39.352046783
Median of the column total sulfur dioxide of the dataframe lowQuality:   54.645161290
```

```
#standard deviation of both dataframes, column total sulfur dioxide
print("Standard deviation of the column total sulfur dioxide of the dataframe highQuality: ",np.std(high
print("Standard deviation of the column total sulfur dioxide of the dataframe lowQuality:  ",np.std(lowQ
```

```
Standard deviation of the column total sulfur dioxide of the dataframe highQuality:
Standard deviation of the column total sulfur dioxide of the dataframe lowQuality:
```

```
#min and max of both dataframes, column toal sulfur dioxide
print("highQuality dataframe min and max [total sulfur dioxide]: "+str(np.min(highQuality['total sulfur
print("lowQuality dataframe min and max [total sulfur dioxide]: "+str(np.min(lowQuality['total sulfur di
```

```
highQuality dataframe min and max [total sulfur dioxide]: 6.0   289.0
lowQuality dataframe min and max [total sulfur dioxide]: 6.0    155.0
```

```
#Characteristics of a high quality and low quality red wine using their mean(average)
print("\t\t\t\t\tHIGH\t\t\tLOW")
print('Average fixed acidity:        '+str(np.mean(highQuality['fixed acidity']))+"\t\t"+str(np.mean(lowQu
print('Average volatile acidity:     '+str(np.mean(highQuality['volatile acidity']))+"\t"+str(np.mean(lowQ
print('Average citric acid:          '+str(np.mean(highQuality['citric acid']))+"\t"+str(np.mean(lowQualit
print('Average residual sugar:       '+str(np.mean(highQuality['residual sugar']))+"\t"+str(np.mean(lowQua
print('Average chlorides:            '+str(np.mean(highQuality['chlorides']))+"\t"+str(np.mean(lowQuality[
print('Average free sulfur dioxide:  '+str(np.mean(highQuality['free sulfur dioxide']))+"\t\t"+str(np.mear
print('Average total sulfur dioxide: '+str(np.mean(highQuality['total sulfur dioxide']))+"\t\t"+str(np.mea
print('Average density:              '+str(np.mean(highQuality['density']))+"\t"+str(np.mean(lowQuality['c
print('Average pH:                   '+str(np.mean(highQuality['pH']))+"\t"+str(np.mean(lowQuality['pH']))
print('Average sulphates:            '+str(np.mean(highQuality['sulphates']))+"\t"+str(np.mean(lowQuality[
print('Average alcohol:              '+str(np.mean(highQuality['alcohol']))+"\t\t"+str(np.mean(lowQuality[
```

```
                                HIGH                   LOW
```

```
                                        HIGH                    LOW
Average fixed acidity:        8.474035087719297         8.142204301075267
Average volatile acidity:     0.4741461988304093        0.589502688172043
Average citric acid:          0.29988304093567253       0.237755376344086
Average residual sugar:       2.5359649122807015        2.5420698924731187
Average chlorides:            0.08266081871345027       0.09298924731182795
Average free sulfur dioxide:  15.27251461988304         16.567204301075268
Average total sulfur dioxide: 39.35204678362573         54.645161290322584
Average density:              0.9964666432748538        0.997068494623656
Average pH:                   3.3106432748538013        3.3116532258064515
Average sulphates:            0.6926198830409357        0.6185349462365591
Average alcohol:              10.85502923976608         9.926478494623655
```

Data Analysis: all columns except the total sulfur dioxide column results in a big difference between the values of the high quality and low quality wines. Based on this difference, I can conlude that the characteristic that determines a high quality red wine is a low total sulfur dioxide value.