Semester	T.E. Semester VI – Computer Engineering
Subject	Data Warehousing and Mining
Subject Professor In-charge	Prof. Kavita Shirsat
Assisting Teachers	Prof. Kavita Shirsat
Laboratory	Lab 312 A

Student Name	Deep Salunkhe	
Roll Number	21102A0014	
Grade and Subject Teacher's Signature		

Experiment Number	02	
Experiment Title	To Visualize and interpret the Data	
Resources / Apparatus Required	Hardware: Computer system	Software: Python, Jupyter noteBook
Description	1. Data visualization is the graphical representation of data and information using visual elements such as charts, graphs, and maps. 2. It is a powerful tool for gaining insights, identifying patterns, and understanding complex relationships in data, making it easier to interpret and communicate information. 3. Data visualization enables the presentation of large datasets in a concise and intuitive manner, allowing users to quickly grasp key trends and outliers. 4. It is widely used across various domains, including business, science, finance, and academia, to make data-driven decisions and communicate findings effectively. 5. Effective data visualization requires thoughtful selection of visual elements, color schemes, and design principles to ensure clarity, accuracy, and meaningful representation of the underlying data.	

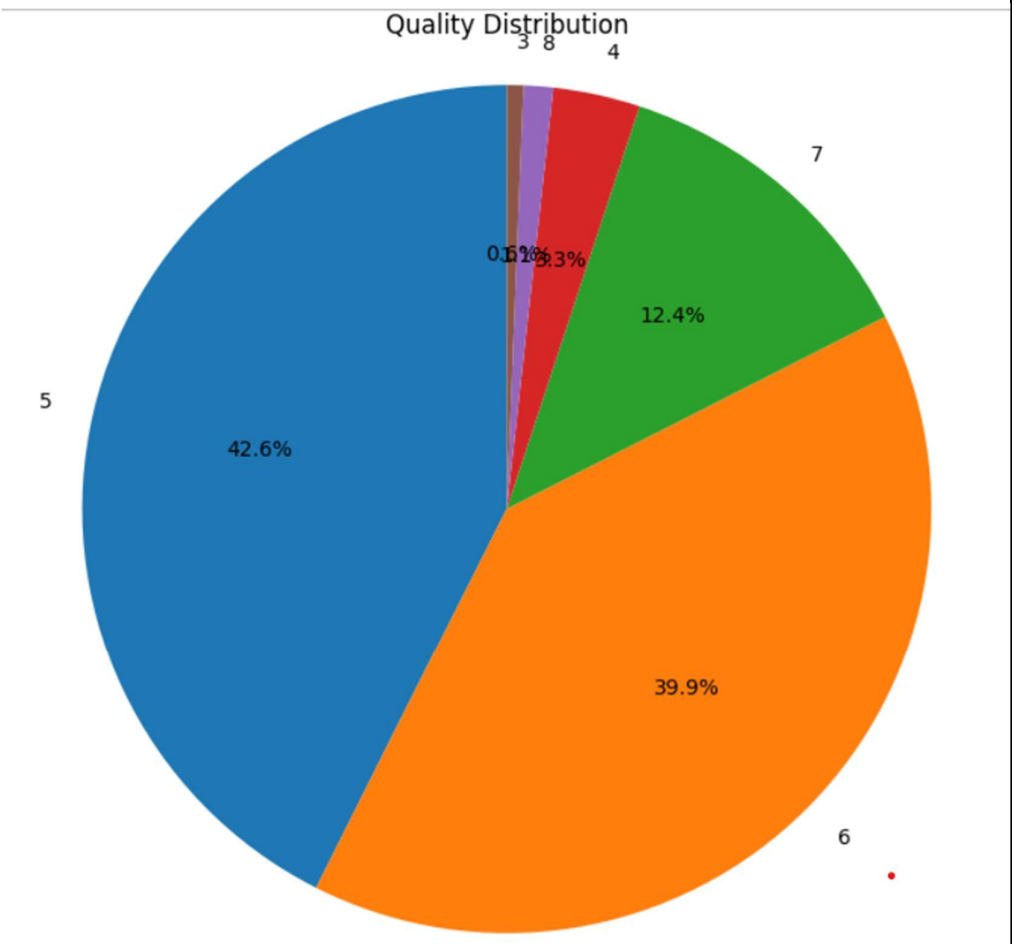
Implementation

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
df=pd.read_csv('winequality-red.csv')
```

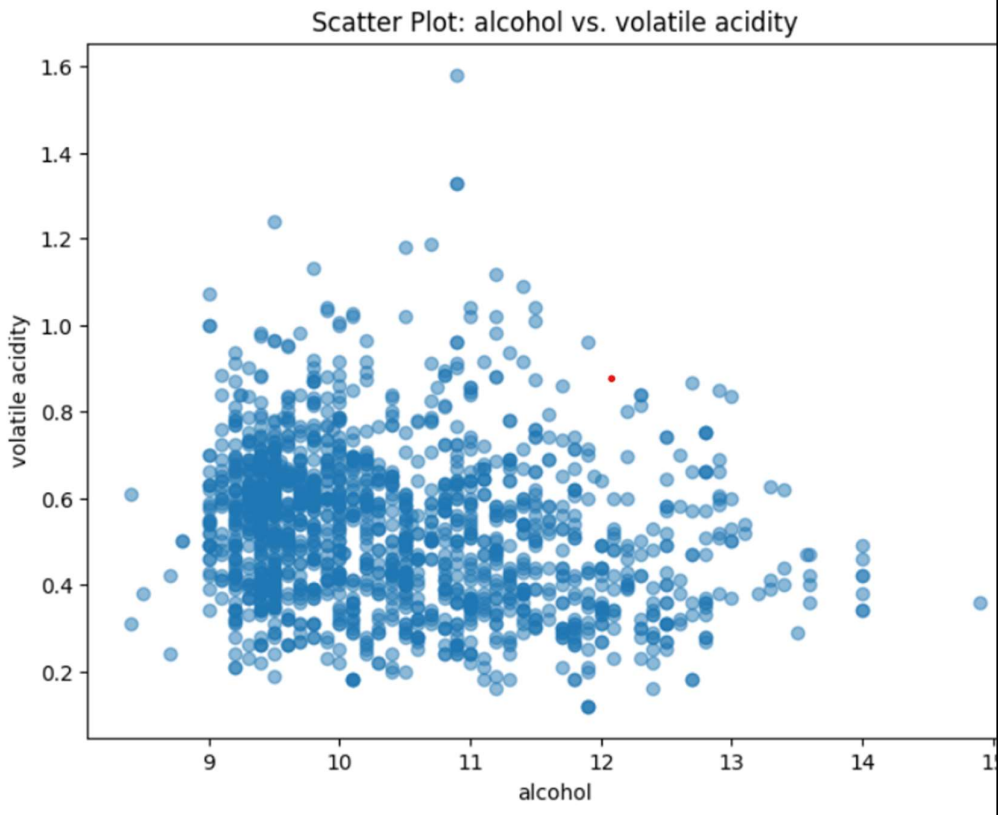
Lab 2=>ploting the visiualization of the given data for interpritation

```
# Pie Chart of quality
quality_counts = df['quality'].value_counts()

# Create the pie chart
plt.figure(figsize=(8, 8))
plt.pie(quality_counts, labels=quality_counts.index, autopct='%1.1f%%', sta
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circ
plt.title('Quality Distribution')
plt.show()
```



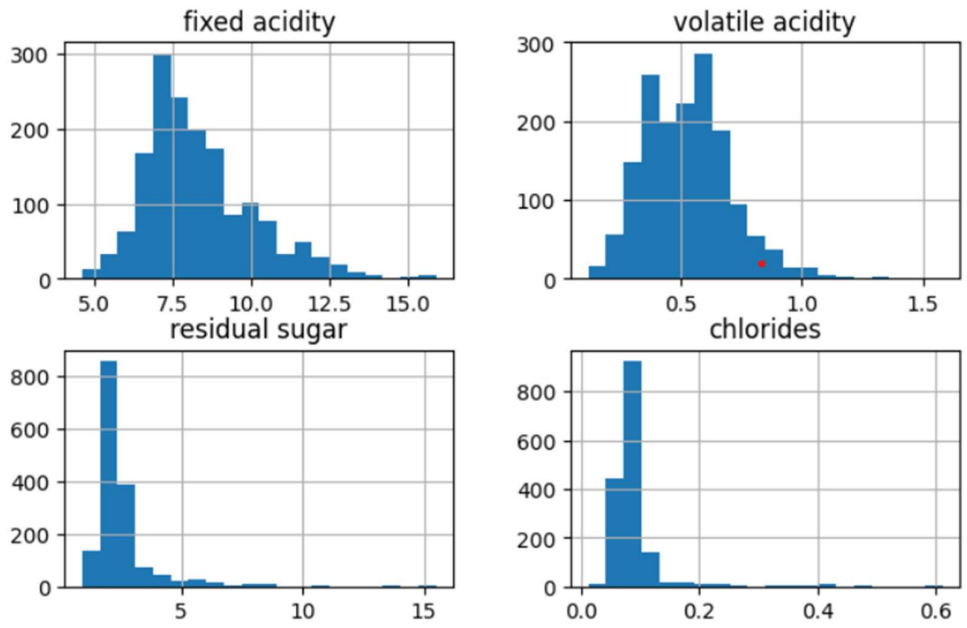
The inference from the above chart=> 1)the max wine is of quality 6 and 5

```
#Scatter graph
# Extract the "quality" and "pH" columns from the DataFrame
x = df['alcohol']
y = df['volatile acidity']

# Create the scatter plot
plt.figure(figsize=(8, 6))
plt.scatter(x, y, marker='o', alpha=0.5)
plt.xlabel('alcohol')
plt.ylabel('volatile acidity')
plt.title('Scatter Plot: alcohol vs. volatile acidity')
plt.show()
```



From this we can see that even if the amount of alcohol increases the increase in acidity is insignificant

```
#Histograms
df.hist(figsize=(12, 10), bins=20)
plt.suptitle('Histograms for Numeric Features', fontsize=16)
plt.show()
```



This graph helps to understand the frequencies of the ranges of each features

Conclusion:	In conclusion, data visualization plays a vital role in the data domain as it enhances understanding, identifies patterns, supports data exploration, facilitates effective communication, and serves as a powerful decision-making tool. By transforming complex data into visually appealing and intuitive representations, data visualization empowers individuals and organizations to harness the full potential of data, make informed decisions, and drive success in various domains.