```python
import pandas as pd
import numpy as np
file_name="/content/winequalityN.csv"
df=pd.read_csv(file_name)
df.head(7)
```

| | type | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density |
|---|---|---|---|---|---|---|---|---|---|
| 0 | white | 7.0 | 0.27 | 0.36 | 20.7 | 0.045 | 45.0 | 170.0 | 1.0010 |
| 1 | white | 6.3 | 0.30 | 0.34 | 1.6 | 0.049 | 14.0 | 132.0 | 0.9940 |
| 2 | white | 8.1 | 0.28 | 0.40 | 6.9 | 0.050 | 30.0 | 97.0 | 0.9951 |
| 3 | white | 7.2 | 0.23 | 0.32 | 8.5 | 0.058 | 47.0 | 186.0 | 0.9956 |
| 4 | white | 7.2 | 0.23 | 0.32 | 8.5 | 0.058 | 47.0 | 186.0 | 0.9956 |
| 5 | white | 8.1 | 0.28 | 0.40 | 6.9 | 0.050 | 30.0 | 97.0 | 0.9951 |
| 6 | white | 6.2 | 0.32 | 0.16 | 7.0 | 0.045 | 30.0 | 136.0 | 0.9949 |

```python
df.tail(10)
```

| | type | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | dens: |
|---|---|---|---|---|---|---|---|---|---|
| 6487 | red | 6.6 | 0.725 | 0.20 | 7.8 | 0.073 | 29.0 | 79.0 | 0.997 |
| 6488 | red | 6.3 | 0.550 | 0.15 | 1.8 | 0.077 | 26.0 | 35.0 | 0.993 |
| 6489 | red | 5.4 | 0.740 | 0.09 | 1.7 | 0.089 | 16.0 | 26.0 | 0.994 |
| 6490 | red | 6.3 | 0.510 | 0.13 | 2.3 | 0.076 | 29.0 | 40.0 | 0.995 |
| 6491 | red | 6.8 | 0.620 | 0.08 | 1.9 | 0.068 | 28.0 | 38.0 | 0.996 |
| 6492 | red | 6.2 | 0.600 | 0.08 | 2.0 | 0.090 | 32.0 | 44.0 | 0.994 |
| 6493 | red | 5.9 | 0.550 | 0.10 | 2.2 | 0.062 | 39.0 | 51.0 | 0.995 |
| 6494 | red | 6.3 | 0.510 | 0.13 | 2.3 | 0.076 | 29.0 | 40.0 | 0.995 |
| 6495 | red | 5.9 | 0.645 | 0.12 | 2.0 | 0.075 | 32.0 | 44.0 | 0.995 |
| 6496 | red | 6.0 | 0.310 | 0.47 | 3.6 | 0.067 | 18.0 | 42.0 | 0.995 |

```python
df.describe()
```

|       | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide |
|-------|---------------|------------------|-------------|----------------|-----------|---------------------|
| count | 6487.000000   | 6489.000000      | 6494.000000 | 6495.000000    | 6495.000000 | 6497.000000       |
| mean  | 7.216579      | 0.339691         | 0.318722    | 5.444326       | 0.056042  | 30.525319           |
| std   | 1.296750      | 0.164649         | 0.145265    | 4.758125       | 0.035036  | 17.749400           |

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6497 entries, 0 to 6496
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   type                  6497 non-null   object
 1   fixed acidity         6487 non-null   float64
 2   volatile acidity      6489 non-null   float64
 3   citric acid           6494 non-null   float64
 4   residual sugar        6495 non-null   float64
 5   chlorides             6495 non-null   float64
 6   free sulfur dioxide   6497 non-null   float64
 7   total sulfur dioxide  6497 non-null   float64
 8   density               6497 non-null   float64
 9   pH                    6488 non-null   float64
 10  sulphates             6493 non-null   float64
 11  alcohol               6497 non-null   float64
 12  quality               6497 non-null   int64
dtypes: float64(11), int64(1), object(1)
memory usage: 660.0+ KB
```

```
df["alcohol"].mean()
```
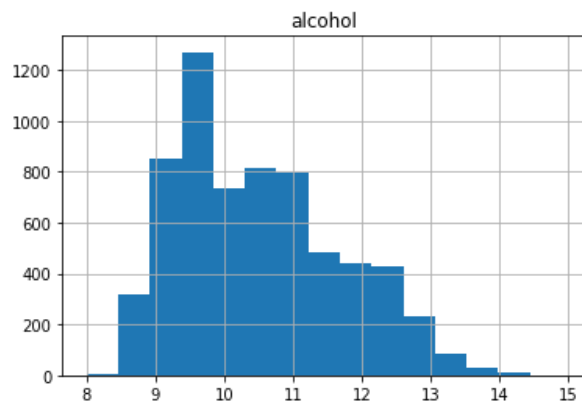
```
10.491800831149455
```

```
df["sulphates"].mean()
```

```
0.531215154782073
```

```
df.hist(column='alcohol',bins=15)
```
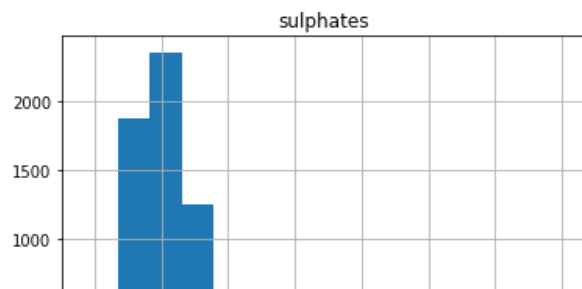
```
array([[<AxesSubplot:title={'center':'alcohol'}>]], dtype=object)
```
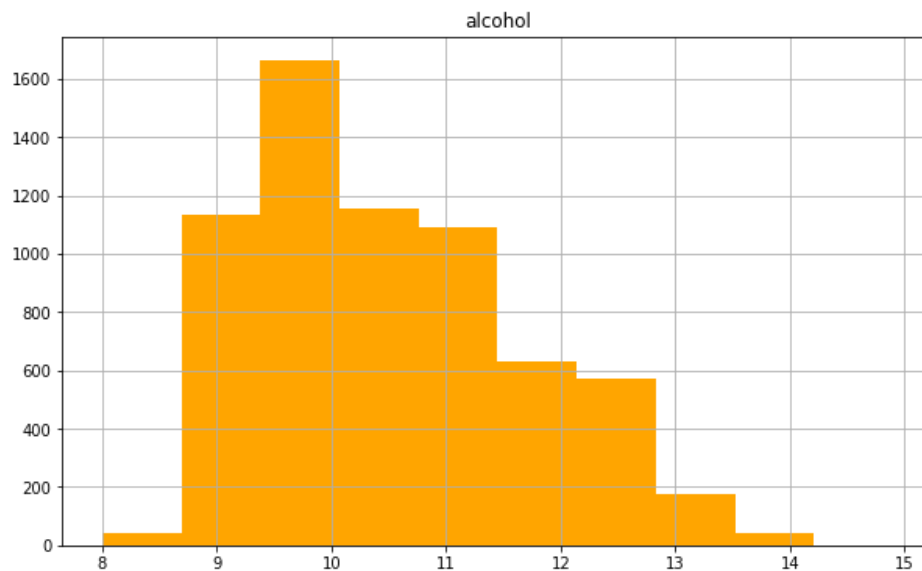


```
df.hist(column='sulphates',bins=15)
```

```
array([[<AxesSubplot:title={'center':'sulphates'}>]], dtype=object)
```



```
from pickle import FALSE
df.hist(column='alcohol',bins=10,grid=FALSE,figsize=(10,6),color='orange')
```
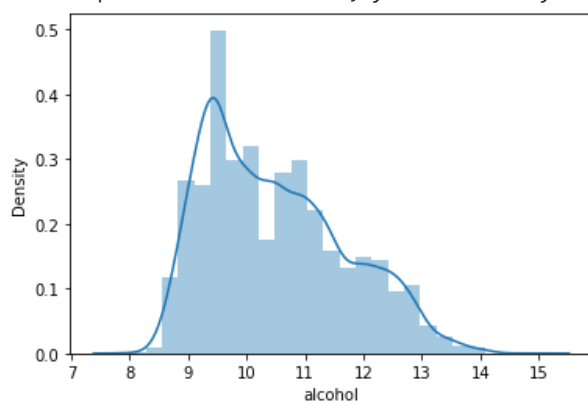
```
array([[<AxesSubplot:title={'center':'alcohol'}>]], dtype=object)
```
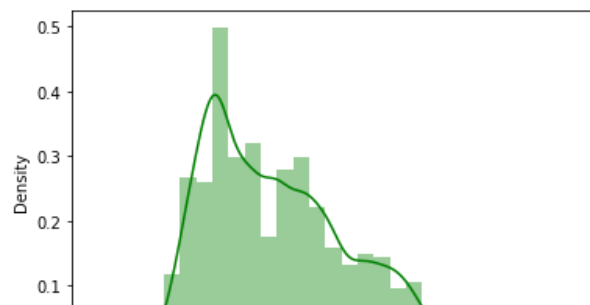


```
import seaborn as sns
sns.distplot(df['alcohol'],bins=25,kde=FALSE)
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/distributions.py:2619: FutureWarnin
  warnings.warn(msg, FutureWarning)
<AxesSubplot:xlabel='alcohol', ylabel='Density'>
```



```
sns.distplot(df['alcohol'],bins=25,kde=True,color="green")
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/distributions.py:2619: FutureWarnin
  warnings.warn(msg, FutureWarning)
<AxesSubplot:xlabel='alcohol', ylabel='Density'>
```
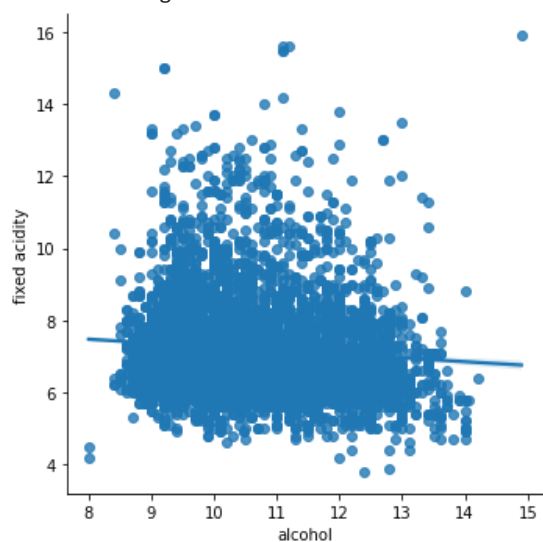


```
df["sulphates"].value_counts().head()
```

```
0.50    275
0.46    243
0.54    234
0.44    232
0.38    214
Name: sulphates, dtype: int64
```

```
sns.lmplot(x="alcohol",y="fixed acidity",data=df)
```
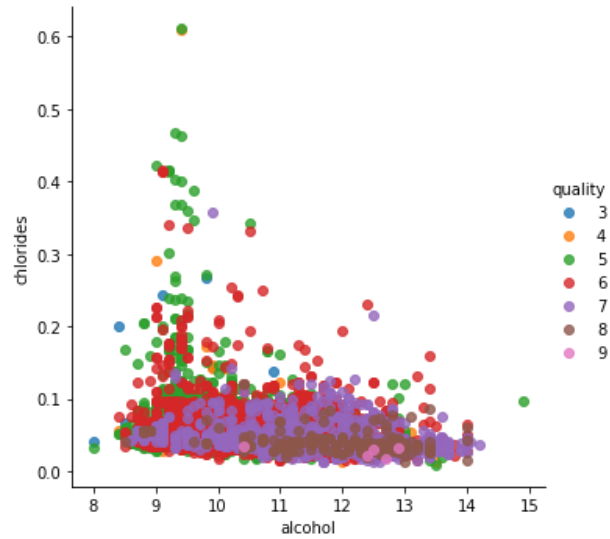
```
<seaborn.axisgrid.FacetGrid at 0x7faef9a60fa0>
```



```
sns.lmplot(x="alcohol",y="density",data=df)
```
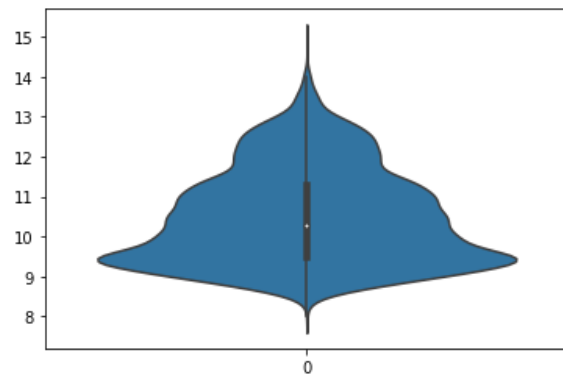
```
<seaborn.axisgrid.FacetGrid at 0x7faef9902a60>
```

```
1.04
```

```python
sns.lmplot(x="alcohol",y="chlorides",data=df,fit_reg=False,hue='quality')
```

```
<seaborn.axisgrid.FacetGrid at 0x7faef98080a0>
```



```python
sns.violinplot(data=df["alcohol"])
```

```
<AxesSubplot:>
```