```python
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
```

```python
wine_dataset = pd.read_csv('/content/winequality-red (1).csv')
```

```python
wine_dataset.shape
```

> (1599, 12)

```python
wine_dataset.head()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulph
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	

Next steps: [Generate code with **wine_dataset**]   [🔘 View recommended plots]   [New interactive sheet]

```python
wine_dataset.describe()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	s di
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.0
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.4
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.8
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.0
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.0
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.0
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.0
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.0

```
wine_dataset.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
```
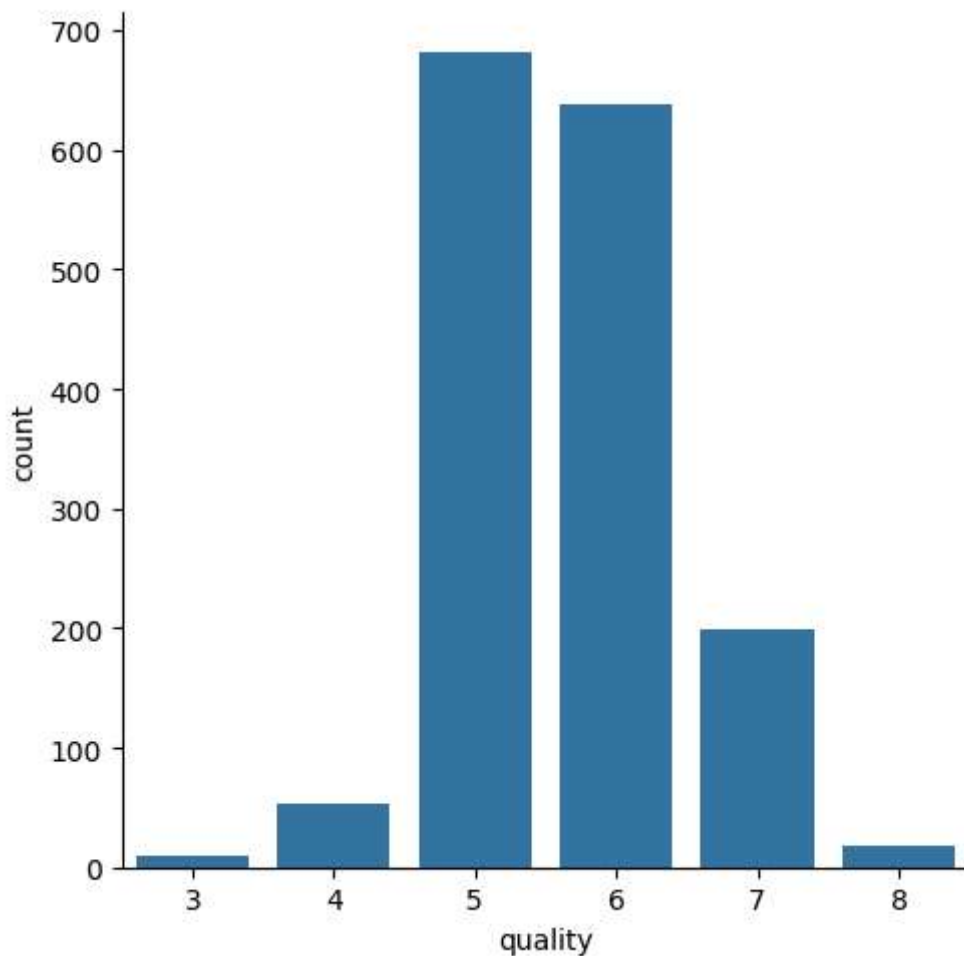
```
wine_dataset.isnull().sum()
```

	0
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

```
sns.catplot(x='quality', data = wine_dataset, kind = 'count')
```

⤵ `<seaborn.axisgrid.FacetGrid at 0x7839bab03be0>`


```python
fig = plt.figure(figsize=(15,10))

plt.subplot(3,4,1)
sns.barplot(x='quality',y='fixed acidity',data=wine_dataset)

plt.subplot(3,4,2)
sns.barplot(x='quality',y='volatile acidity',data=wine_dataset)

plt.subplot(3,4,3)
sns.barplot(x='quality',y='citric acid',data=wine_dataset)

plt.subplot(3,4,4)
sns.barplot(x='quality',y='residual sugar',data=wine_dataset)

plt.subplot(3,4,5)
sns.barplot(x='quality',y='chlorides',data=wine_dataset)

plt.subplot(3,4,6)
sns.barplot(x='quality',y='free sulfur dioxide',data=wine_dataset)

plt.subplot(3,4,7)
sns.barplot(x='quality',y='total sulfur dioxide',data=wine_dataset)
```
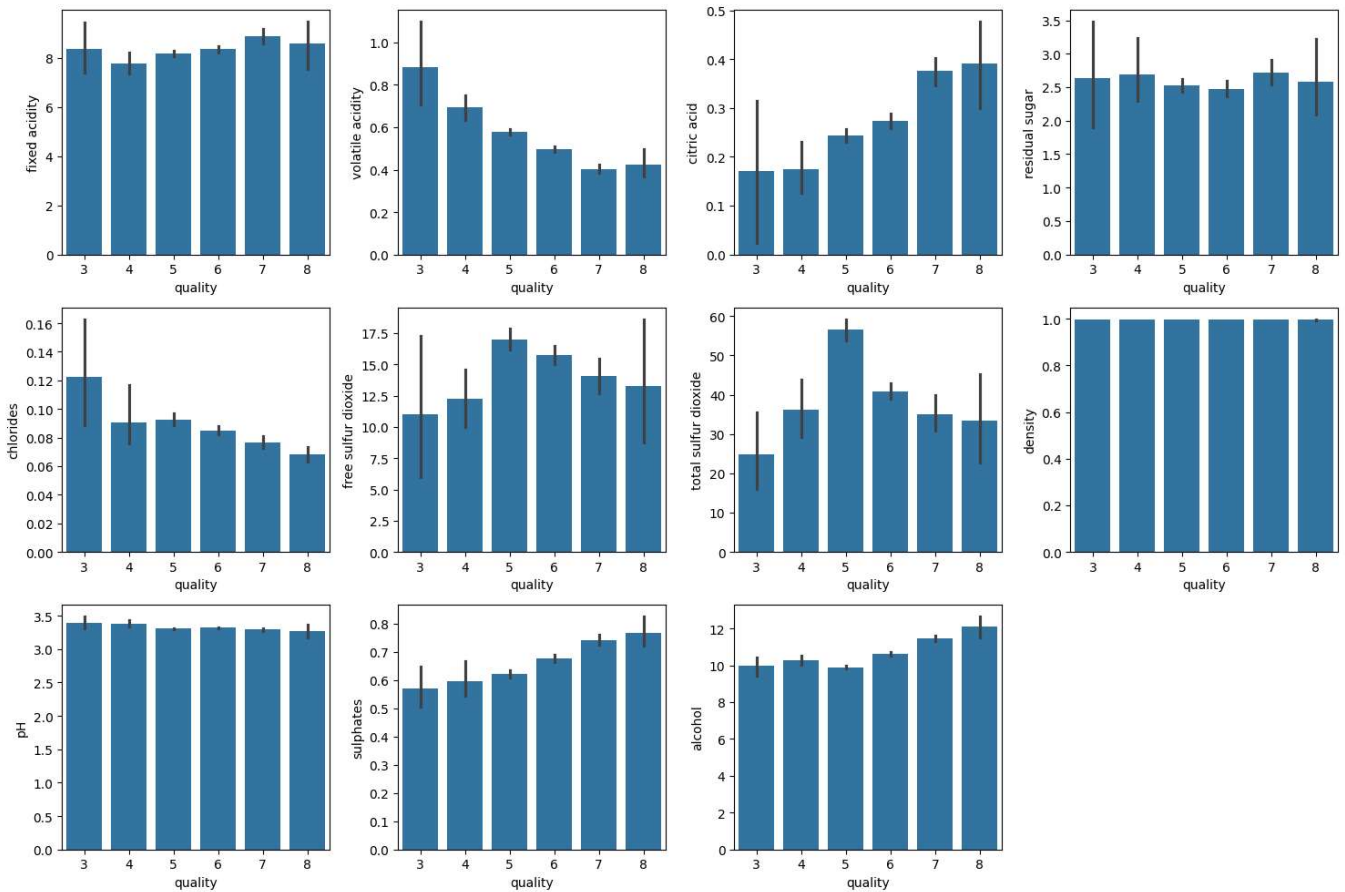
```python
plt.subplot(3,4,8)
sns.barplot(x='quality',y='density',data=wine_dataset)

plt.subplot(3,4,9)
sns.barplot(x='quality',y='pH',data=wine_dataset)

plt.subplot(3,4,10)
sns.barplot(x='quality',y='sulphates',data=wine_dataset)

plt.subplot(3,4,11)
sns.barplot(x='quality',y='alcohol',data=wine_dataset)

plt.tight_layout()
```
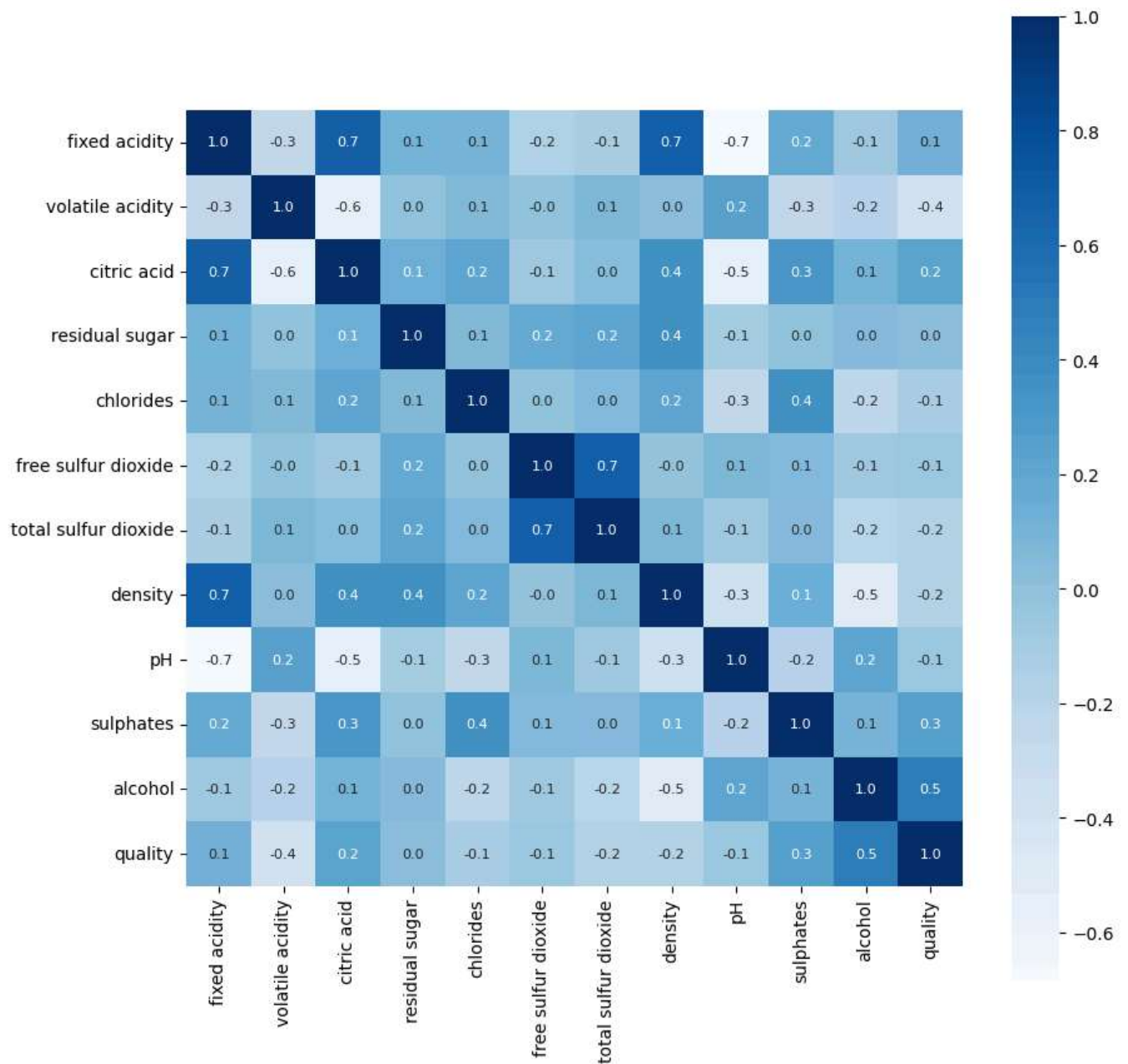
```
correlation = wine_dataset.corr()
```

```
# constructing a heatmap to understand the correlation between the columns
plt.figure(figsize=(10,10))
sns.heatmap(correlation, cbar=True, square=True, fmt = '.1f', annot = True, annot_kws={'size
```

<Axes: >

```
X = wine_dataset.drop('quality',axis=1)
X
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	su
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	
2	7.8	0.760	0.04	2.3	0.092	15.0	54.0	0.99700	3.26	
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.99800	3.16	
4	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	
...	...	...	...	...	...	...	...	...	...	
1594	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	
1595	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	
1596	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	
1597	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	
1598	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.99549	3.39	

Next steps: Generate code with X   ◉ View recommended plots   New interactive sheet

```
Y = wine_dataset['quality'].apply(lambda y_value: 1 if y_value>=7 else 0 )
Y
```

quality