```python
In [1]:  import numpy as np
         import pandas as pd
         import warnings
         warnings.filterwarnings('ignore')
```

```python
In [2]:  red_wine_data = pd.read_csv('https://raw.githubusercontent.com/btkhimsar/DataSets
```

```python
In [3]:  red_wine_data.head()
```

Out[3]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4

```python
In [4]:  red_wine_data.shape
```

Out[4]: (1599, 12)

```python
In [5]:  red_wine_data.describe()
```

Out[5]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.467792
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.895324
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.000000
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.000000
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.000000
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.000000

In [6]:
```python
red_wine_data.columns
```

Out[6]: Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'pH', 'sulphates', 'alcohol', 'quality'],
      dtype='object')

In [7]:
```python
red_wine_data['quality'].unique()
```

Out[7]: array([5, 6, 7, 4, 8, 3], dtype=int64)

In [8]:
```python
red_wine_data['quality'].nunique()
```

Out[8]: 6

In [9]:
```python
red_wine_data['quality'].value_counts()
```

Out[9]: 5   681
6   638
7   199
4   53
8   18
3   10
Name: quality, dtype: int64

In [10]:
```python
red_wine_data.rename(columns={'fixed acidity':'fixed_acidity','volatile acidity'
```

Out[10]:

	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	free_sulfur_dioxide	total_
0	7.4	0.700	0.00	1.9	0.076	11.0	
1	7.8	0.880	0.00	2.6	0.098	25.0	
2	7.8	0.760	0.04	2.3	0.092	15.0	
3	11.2	0.280	0.56	1.9	0.075	17.0	
4	7.4	0.700	0.00	1.9	0.076	11.0	
...	...	...	...	...	...	...	
1594	6.2	0.600	0.08	2.0	0.090	32.0	
1595	5.9	0.550	0.10	2.2	0.062	39.0	
1596	6.3	0.510	0.13	2.3	0.076	29.0	
1597	5.9	0.645	0.12	2.0	0.075	32.0	
1598	6.0	0.310	0.47	3.6	0.067	18.0	

1599 rows × 12 columns

In [11]: `red_wine_data.isna()`

Out[11]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcol
0	False	False	False	False	False	False	False	False	False	False	Fa
1	False	False	False	False	False	False	False	False	False	False	Fa
2	False	False	False	False	False	False	False	False	False	False	Fa
3	False	False	False	False	False	False	False	False	False	False	Fa
4	False	False	False	False	False	False	False	False	False	False	Fa
...	...	...	...	...	...	...	...	...	...	...	
1594	False	False	False	False	False	False	False	False	False	False	Fa
1595	False	False	False	False	False	False	False	False	False	False	Fa
1596	False	False	False	False	False	False	False	False	False	False	Fa
1597	False	False	False	False	False	False	False	False	False	False	Fa
1598	False	False	False	False	False	False	False	False	False	False	Fa

1599 rows × 12 columns

In [12]: `red_wine_data.isna().sum()`

Out[12]:
```
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
```

In [13]:
```python
red_wine_data.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
```
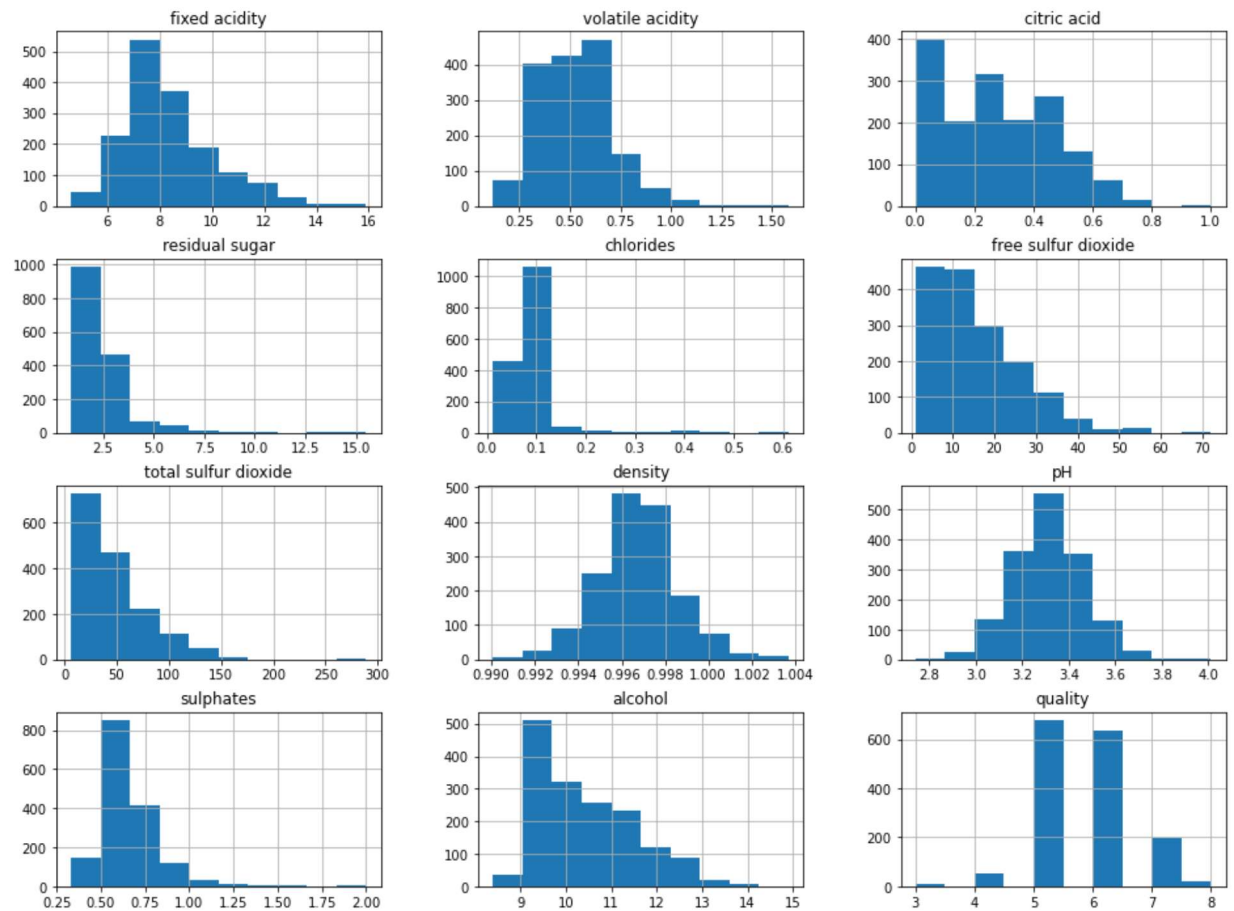
In [14]:
```python
duplicate=red_wine_data[red_wine_data.duplicated()]
duplicate.shape
```

Out[14]: (240, 12)

In [15]:
```python
import matplotlib.pyplot as plt
import seaborn as sns
```

In [16]:
```python
y=red_wine_data['quality']
X=red_wine_data.drop(red_wine_data['quality'],inplace=True)
```

```
In [17]: red_wine_data.hist(bins=10,figsize=(16,12))
         plt.show()
```
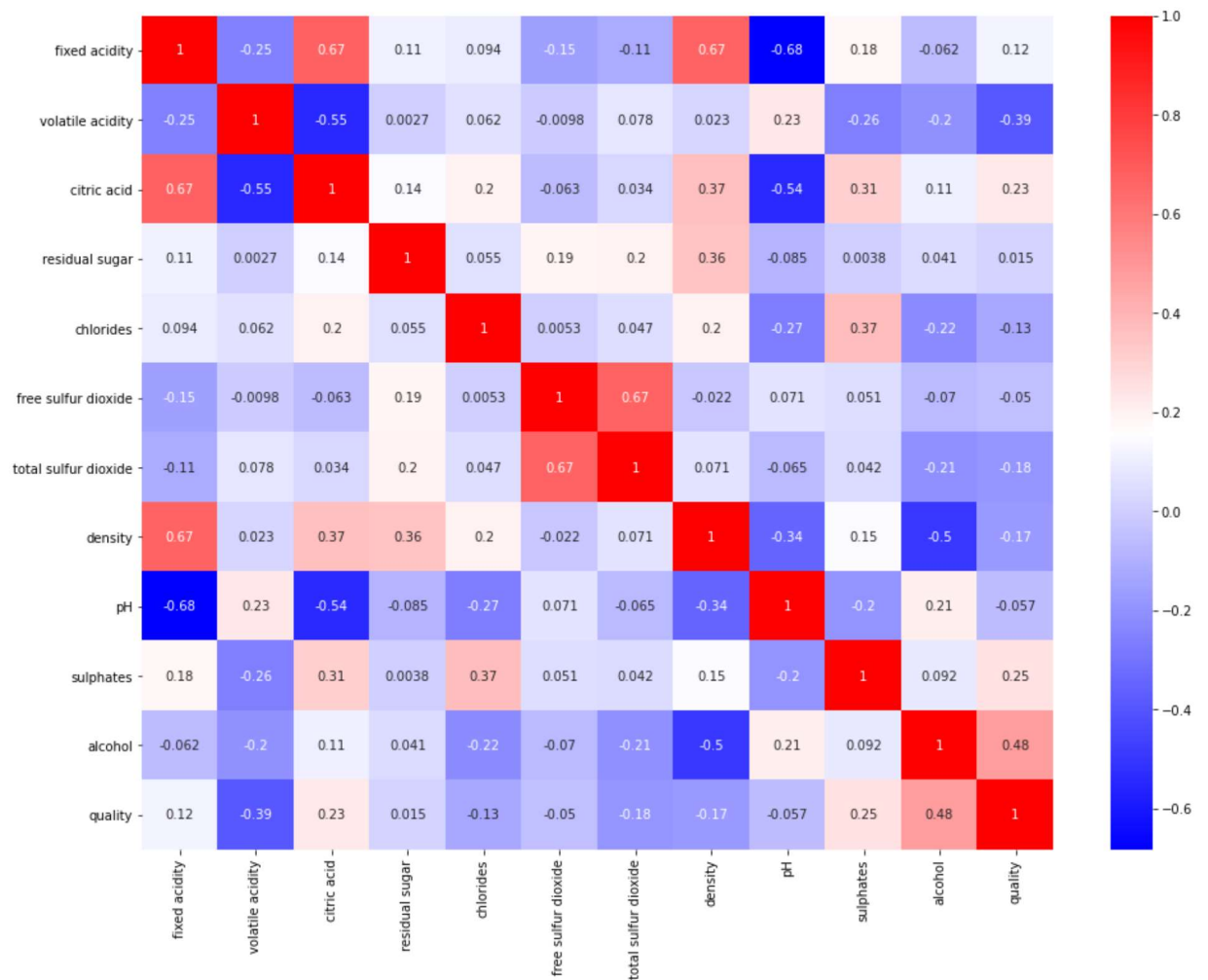
In [18]: `red_wine_data.corr()`

Out[18]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	
fixed acidity	1.000000	-0.254398	0.671410	0.114616	0.093650	-0.154536	-0.114387	0.668582	-C
volatile acidity	-0.254398	1.000000	-0.551206	0.002691	0.061790	-0.009841	0.077683	0.022580	C
citric acid	0.671410	-0.551206	1.000000	0.142135	0.203428	-0.062613	0.033668	0.365729	-C
residual sugar	0.114616	0.002691	0.142135	1.000000	0.054904	0.186785	0.202718	0.355577	-C
chlorides	0.093650	0.061790	0.203428	0.054904	1.000000	0.005293	0.047110	0.200734	-C
free sulfur dioxide	-0.154536	-0.009841	-0.062613	0.186785	0.005293	1.000000	0.667660	-0.021810	C
total sulfur dioxide	-0.114387	0.077683	0.033668	0.202718	0.047110	0.667660	1.000000	0.070902	-C
density	0.668582	0.022580	0.365729	0.355577	0.200734	-0.021810	0.070902	1.000000	-C
pH	-0.682678	0.233098	-0.540720	-0.084971	-0.265023	0.071316	-0.065474	-0.342598	1
sulphates	0.182851	-0.260545	0.311429	0.003772	0.370793	0.051157	0.042388	0.148474	-C
alcohol	-0.062303	-0.201714	0.107572	0.040552	-0.222395	-0.070349	-0.206664	-0.496641	C
quality	0.124330	-0.391616	0.228214	0.014504	-0.128786	-0.050443	-0.184133	-0.174029	-C

In [19]:
```python
plt.figure(figsize=(16,12))
sns.heatmap(red_wine_data.corr(),cmap='bwr',annot=True)
```

Out[19]: <AxesSubplot:>

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
fixed acidity	1	-0.25	0.67	0.11	0.094	-0.15	-0.11	0.67	-0.68	0.18	-0.062	0.12
volatile acidity	-0.25	1	-0.55	0.0027	0.062	-0.0098	0.078	0.023	0.23	-0.26	-0.2	-0.39
citric acid	0.67	-0.55	1	0.14	0.2	-0.063	0.034	0.37	-0.54	0.31	0.11	0.23
residual sugar	0.11	0.0027	0.14	1	0.055	0.19	0.2	0.36	-0.085	0.0038	0.041	0.015
chlorides	0.094	0.062	0.2	0.055	1	0.0053	0.047	0.2	-0.27	0.37	-0.22	-0.13
free sulfur dioxide	-0.15	-0.0098	-0.063	0.19	0.0053	1	0.67	-0.022	0.071	0.051	-0.07	-0.05
total sulfur dioxide	-0.11	0.078	0.034	0.2	0.047	0.67	1	0.071	-0.065	0.042	-0.21	-0.18
density	0.67	0.023	0.37	0.36	0.2	-0.022	0.071	1	-0.34	0.15	-0.5	-0.17
pH	-0.68	0.23	-0.54	-0.085	-0.27	0.071	-0.065	-0.34	1	-0.2	0.21	-0.057
sulphates	0.18	-0.26	0.31	0.0038	0.37	0.051	0.042	0.15	-0.2	1	0.092	0.25
alcohol	-0.062	-0.2	0.11	0.041	-0.22	-0.07	-0.21	-0.5	0.21	0.092	1	0.48
quality	0.12	-0.39	0.23	0.015	-0.13	-0.05	-0.18	-0.17	-0.057	0.25	0.48	1

In [20]: `sns.pairplot(red_wine_data)`

Out[20]: `<seaborn.axisgrid.PairGrid at 0x22cb8753f70>`
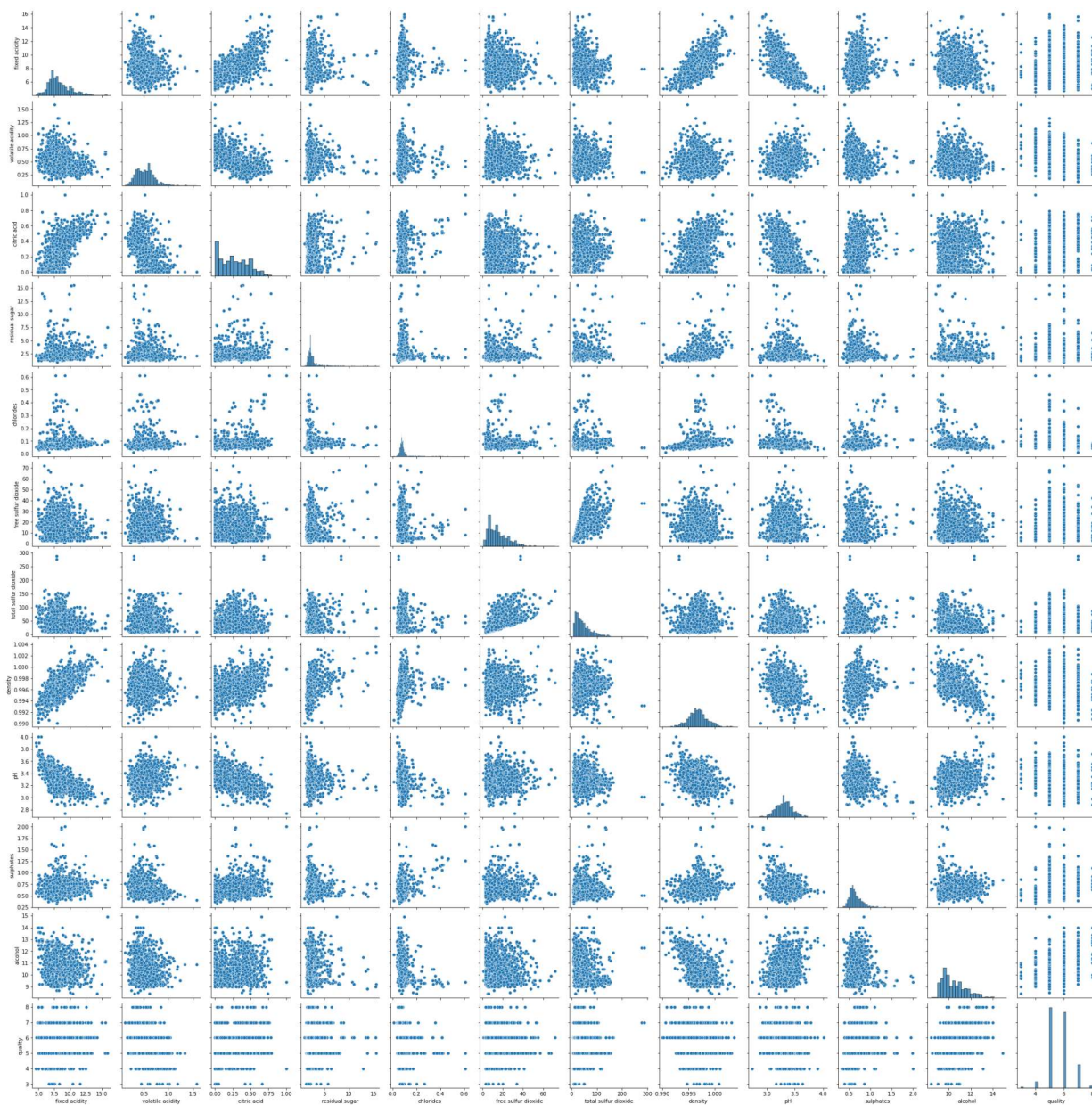
In [1]: `#CUSTOMER CREDIT CARD FRAUD DETECTION`

In [ ]: