In [4]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
```

In [5]:
```python
df = pd.read_csv('Wine Quality Dataset.csv')
df.head()
```

Out[5]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.0	0.27	0.36	20.7	0.045	45.0	170.0	1.0010	3.00	0.45	8.8	6
1	6.3	0.30	0.34	1.6	0.049	14.0	132.0	0.9940	3.30	0.49	9.5	6
2	8.1	0.28	0.40	6.9	0.050	30.0	97.0	0.9951	3.26	0.44	10.1	6
3	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6
4	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6

In [6]:
```python
df.shape
```

Out[6]:
```
(4898, 12)
```

In [7]:
```python
df.index
```

Out[7]:
```
RangeIndex(start=0, stop=4898, step=1)
```

In [8]:
```python
df.columns
```

Out[8]:
```
Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'pH', 'sulphates', 'alcohol', 'quality'],
      dtype='object')
```

In [9]:
```python
df.info()
```

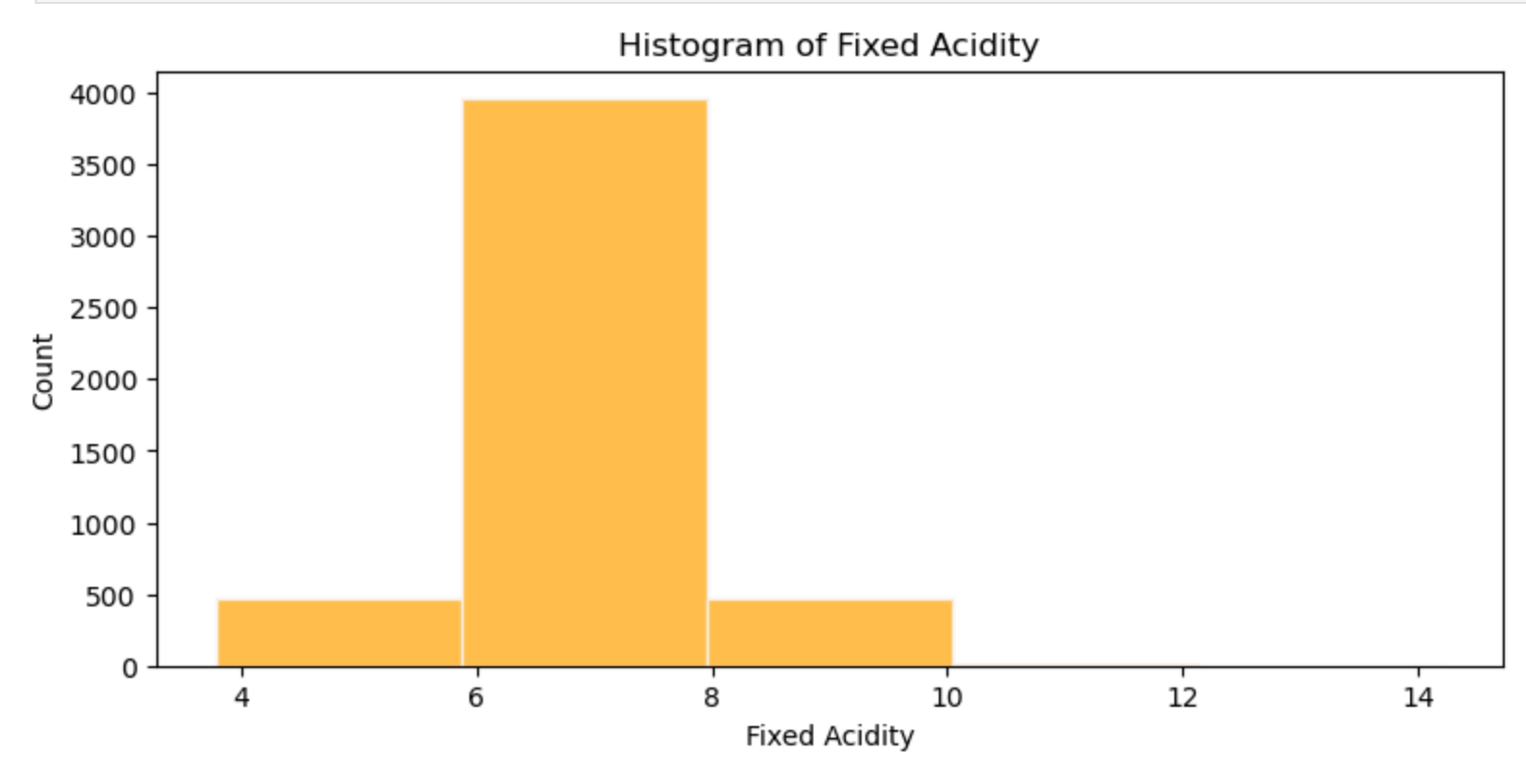
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4898 entries, 0 to 4897
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   fixed acidity         4898 non-null   float64
 1   volatile acidity      4898 non-null   float64
 2   citric acid           4898 non-null   float64
 3   residual sugar        4898 non-null   float64
 4   chlorides             4898 non-null   float64
 5   free sulfur dioxide   4898 non-null   float64
 6   total sulfur dioxide  4898 non-null   float64
 7   density               4898 non-null   float64
 8   pH                    4898 non-null   float64
 9   sulphates             4898 non-null   float64
 10  alcohol               4898 non-null   float64
 11  quality               4898 non-null   int64  
dtypes: float64(11), int64(1)
memory usage: 459.3 KB
```

In [10]:
```python
df.describe()
```

Out[10]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
count	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000
mean	6.854788	0.278241	0.334192	6.391415	0.045772	35.308085	138.360657	0.994027	3.188267	0.489847	10.514267	5.877909
std	0.843868	0.100795	0.121020	5.072058	0.021848	17.007137	42.498065	0.002991	0.151001	0.114126	1.230621	0.885639
min	3.800000	0.080000	0.000000	0.600000	0.009000	2.000000	9.000000	0.987110	2.720000	0.220000	8.000000	3.000000
25%	6.300000	0.210000	0.270000	1.700000	0.036000	23.000000	108.000000	0.991723	3.090000	0.410000	9.500000	5.000000
50%	6.800000	0.260000	0.320000	5.200000	0.043000	34.000000	134.000000	0.993740	3.180000	0.470000	10.400000	6.000000
75%	7.300000	0.320000	0.390000	9.900000	0.050000	46.000000	167.000000	0.996100	3.280000	0.550000	11.400000	6.000000
max	14.200000	1.100000	1.660000	65.800000	0.346000	289.000000	440.000000	1.038980	3.820000	1.080000	14.200000	9.000000

In [11]:
```python
plt.figure(figsize = (9,4))
sns.histplot(data = df, x = 'fixed acidity',color = 'orange',
             edgecolor = 'linen', alpha = 0.7, bins = 5)
plt.title('Histogram of Fixed Acidity')
plt.xlabel('Fixed Acidity')
plt.ylabel('Count')
plt.show()
```



In [12]:
```python
df['fixed acidity'].mean()
```

Out[12]:
```
6.854787668436097
```

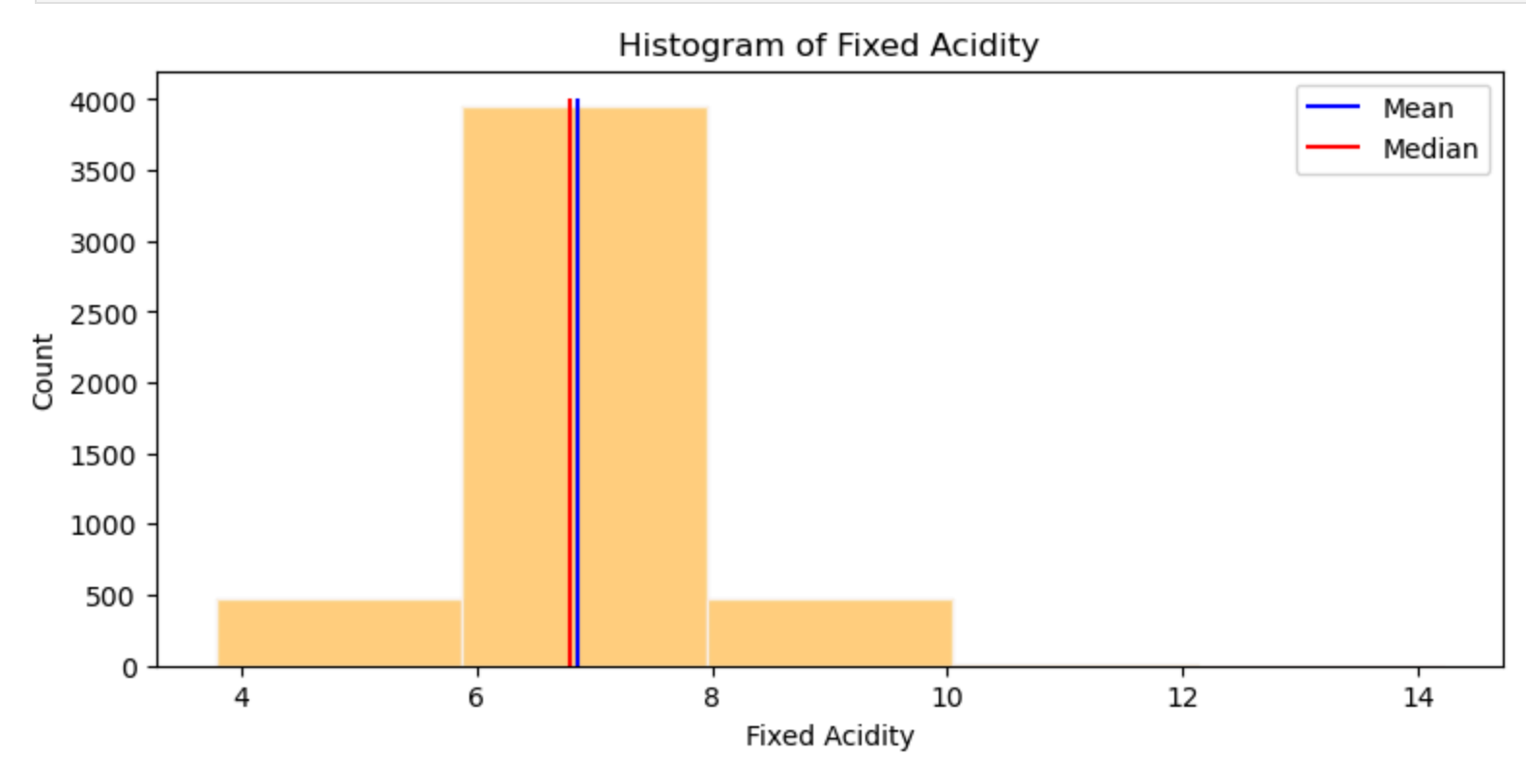
In [13]:
```python
round(df['fixed acidity'].mean(),2)
```

Out[13]:
```
6.85
```

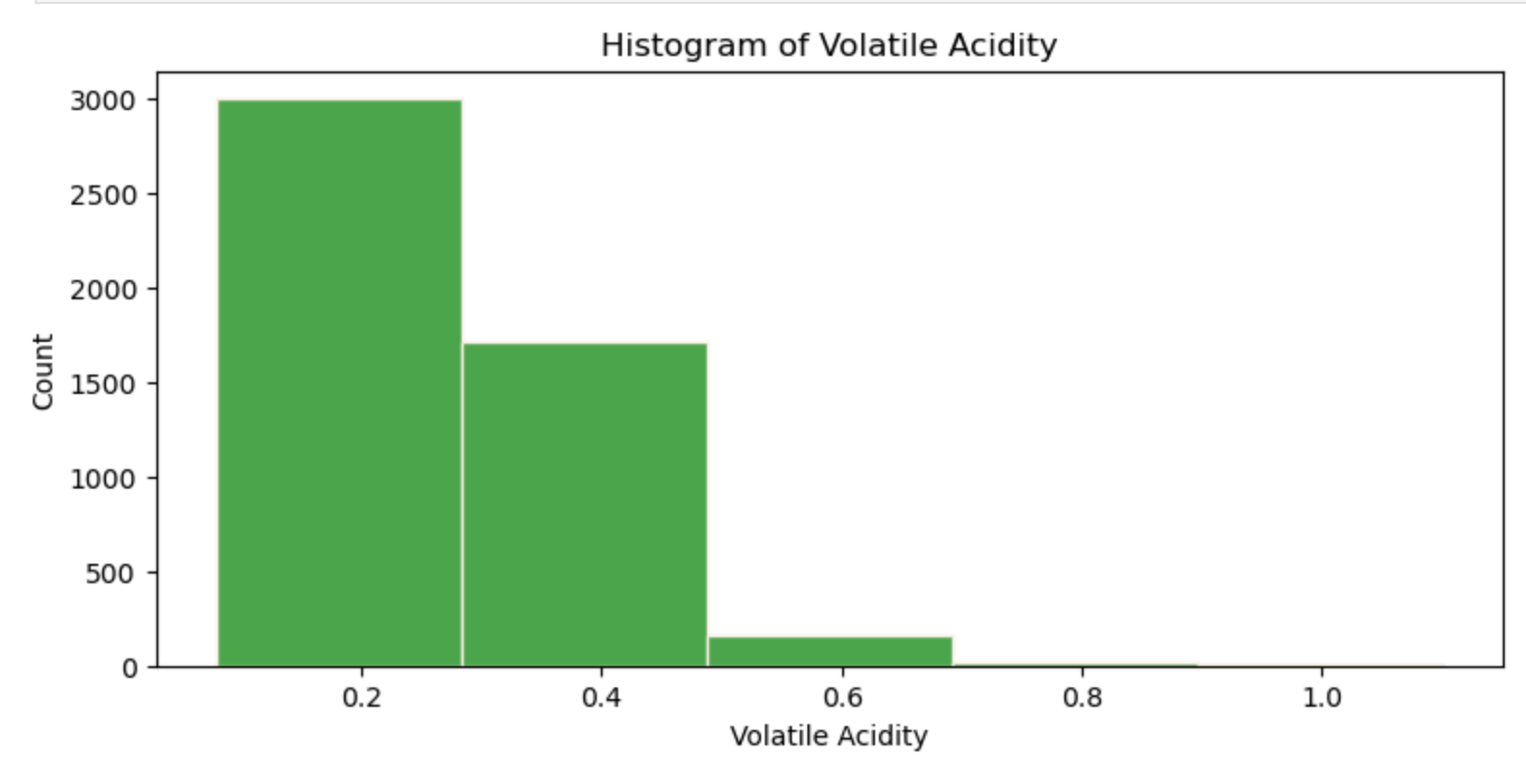
In [14]:
```python
round(df['fixed acidity'].median(),2)
```

Out[14]:
```
6.8
```

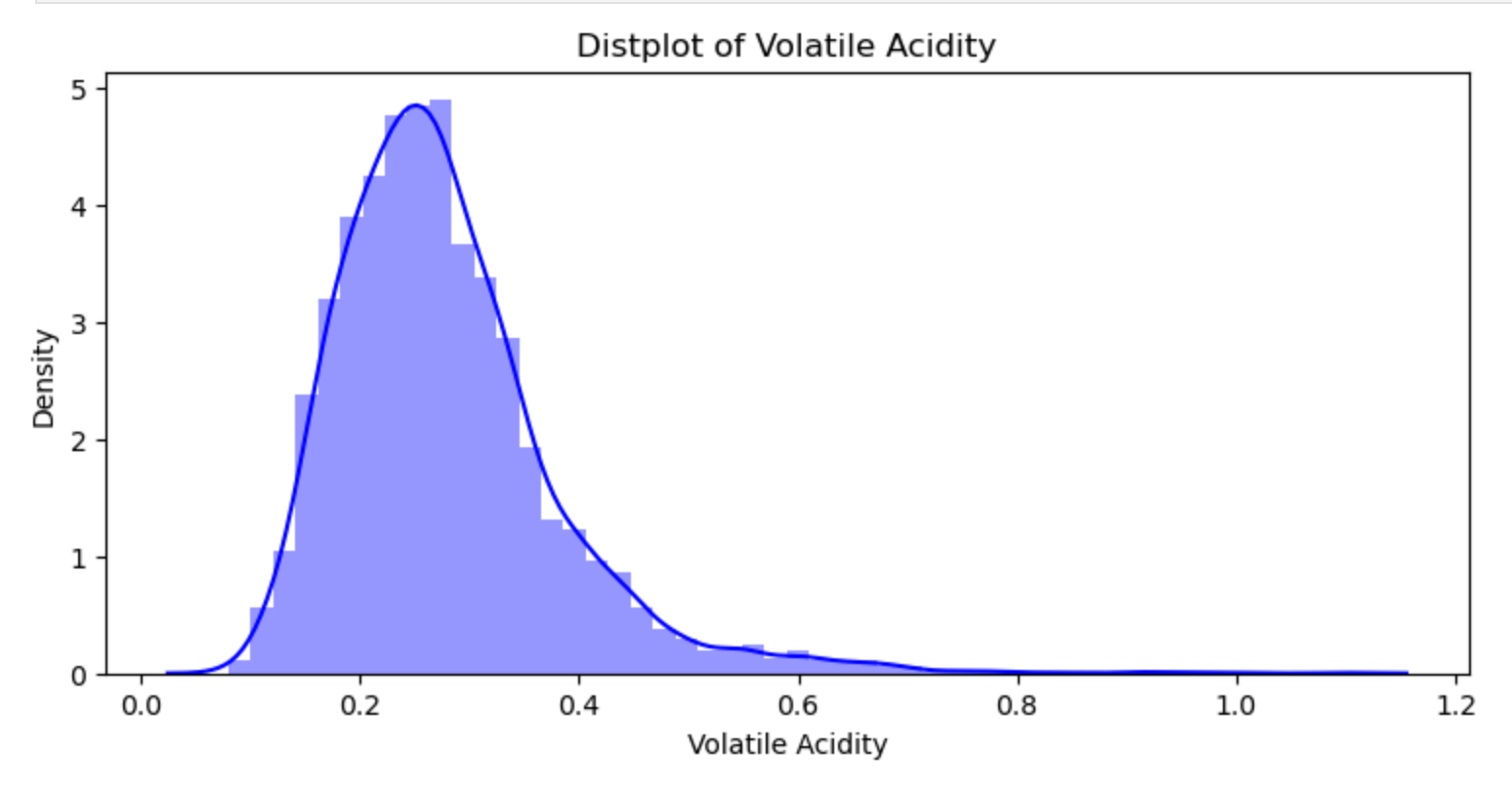
In [15]:
```python
plt.figure(figsize = (9,4))
sns.histplot(data = df ,x = 'fixed acidity', color = 'orange',
edgecolor = 'linen', alpha = 0.5, bins = 5)
plt.title("Histogram of Fixed Acidity")
plt.xlabel('Fixed Acidity')
plt.ylabel('Count')
plt.vlines(df['fixed acidity'].mean(), ymin = 0, ymax = 4000, colors='blue', label='Mean')
plt.vlines(df['fixed acidity'].median(), ymin = 0, ymax = 4000, colors='red', label='Median')
plt.legend()
plt.show()
```



In [16]:
```python
plt.figure(figsize = (9,4))
sns.histplot(data = df ,x = 'volatile acidity', color = 'green',
edgecolor = 'linen', alpha = 0.7, bins = 5)
plt.title("Histogram of Volatile Acidity")
plt.xlabel('Volatile Acidity')
plt.ylabel('Count')
plt.show()
```



In [17]:
```python
plt.figure(figsize = (9,4))
sns.distplot(df['volatile acidity'], color = 'blue')
plt.title("Distplot of Volatile Acidity")
plt.xlabel('Volatile Acidity')
plt.ylabel('Density')
plt.show()
```



In [18]:
```python
df['volatile acidity'].skew()
```

Out[18]:
```
1.5769795029952025
```

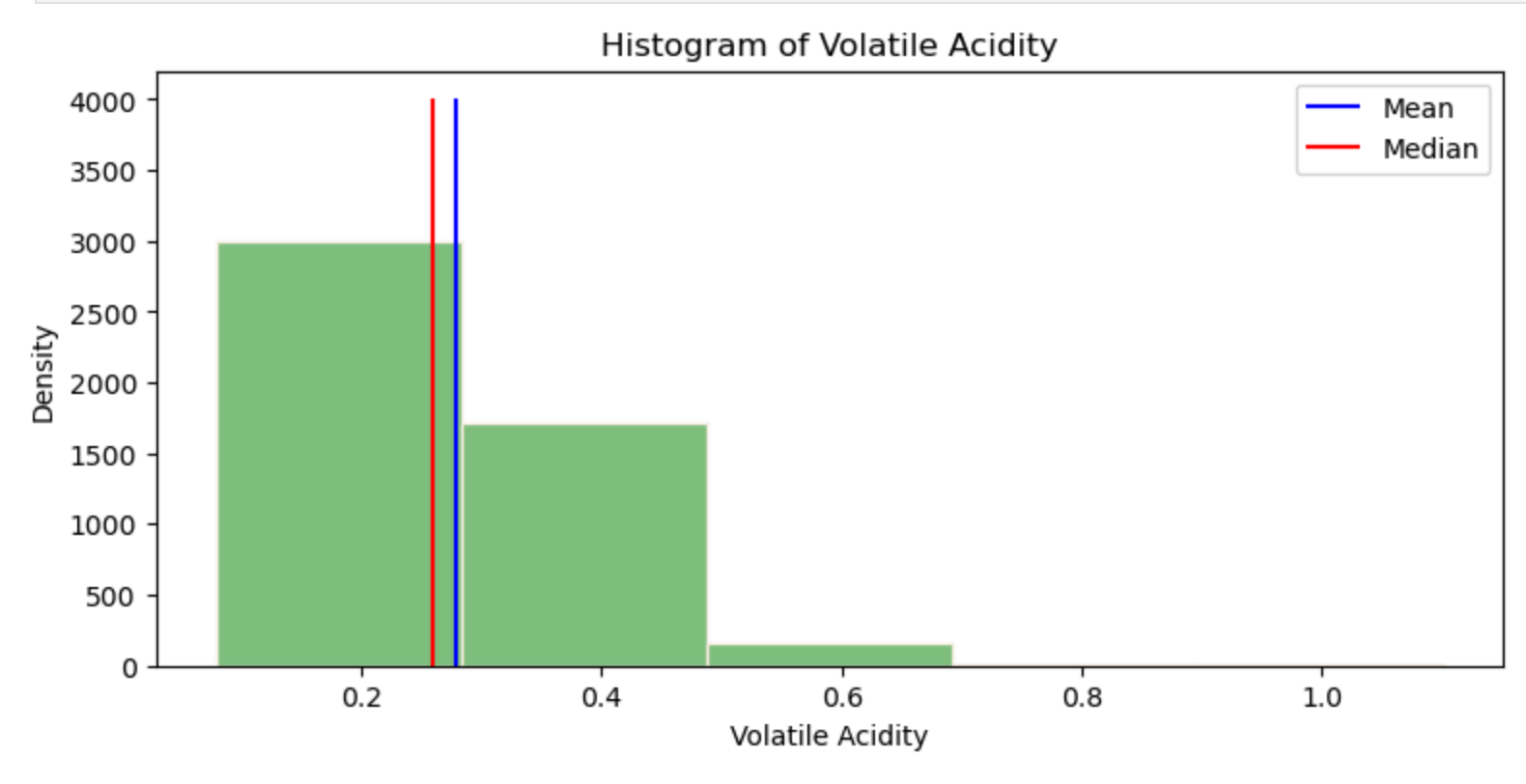
In [19]:
```python
df['volatile acidity'].mean()
```

Out[19]:
```
0.27824111882400976
```

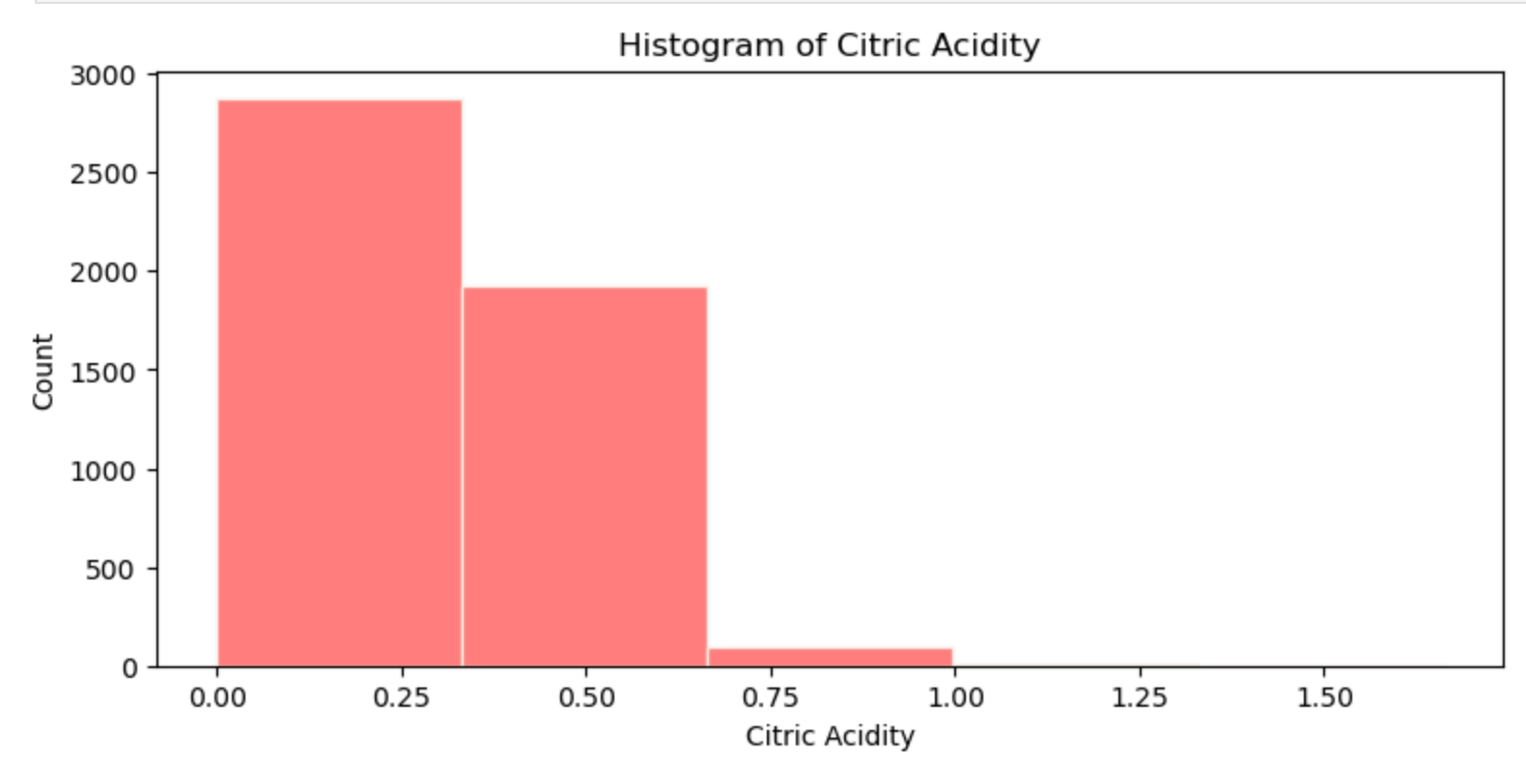
In [20]:
```python
df['volatile acidity'].median()
```

Out[20]:
```
0.26
```

In [21]:
```python
plt.figure(figsize = (9,4))
sns.histplot(data = df ,x = 'volatile acidity', color = 'green',
             edgecolor = 'linen', alpha = 0.5, bins = 5)
plt.title("Histogram of Volatile Acidity")
plt.xlabel('Volatile Acidity')
plt.ylabel('Density')
plt.vlines(df['volatile acidity'].mean(), ymin = 0, ymax = 4000, colors='blue', label='Mean')
plt.vlines(df['volatile acidity'].median(), ymin = 0, ymax = 4000, colors='red', label='Median')
plt.legend()
plt.show()
```



In [22]:
```python
# Create a histogram of the "citric acid" feature
plt.figure(figsize = (9,4))
sns.histplot(data = df ,x = 'citric acid', color = 'red',
edgecolor = 'linen', alpha = 0.5, bins = 5)
plt.title("Histogram of Citric Acidity")
plt.xlabel('Citric Acidity')
plt.ylabel('Count')
plt.show()
```



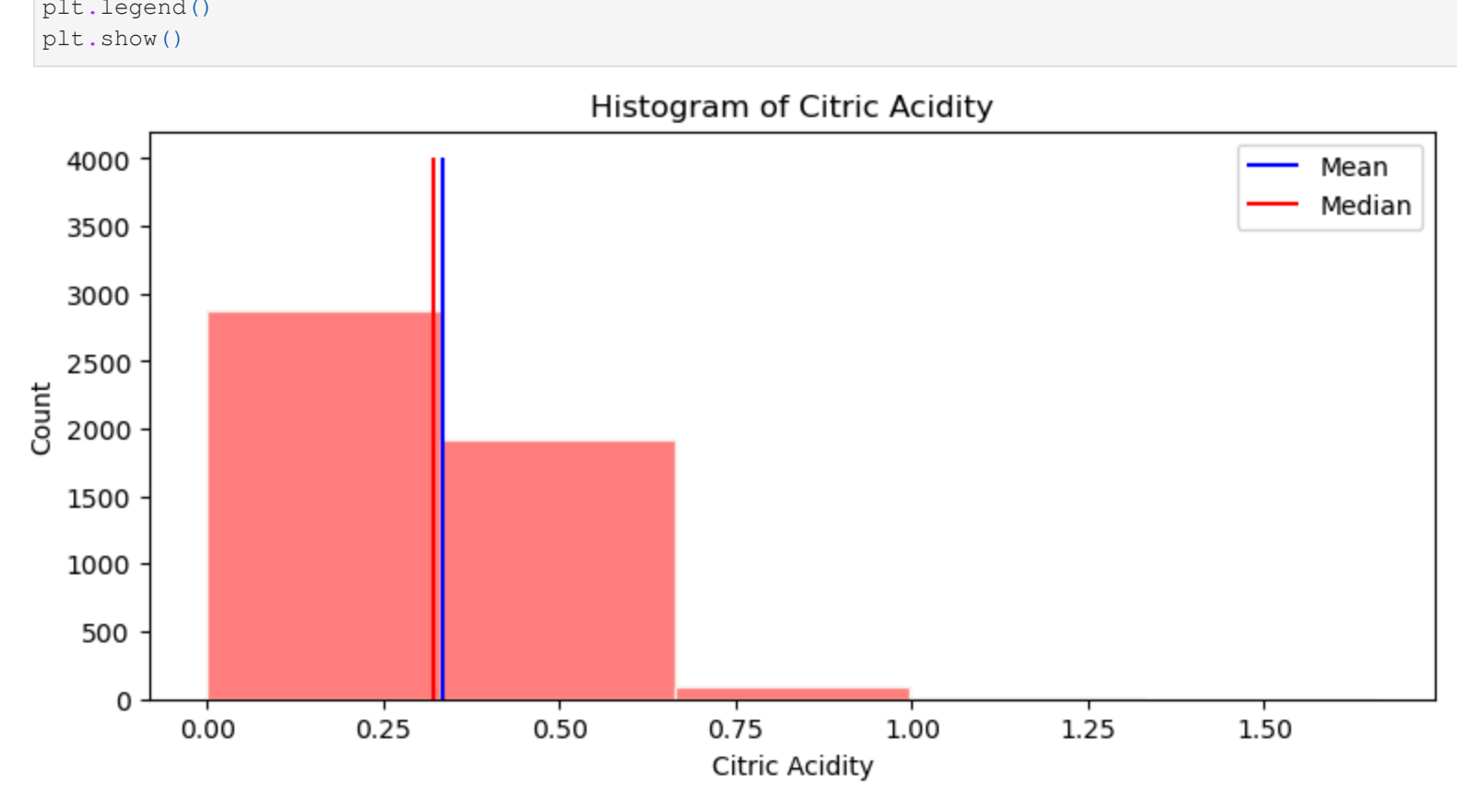
In [23]:
```python
df['citric acid'].mean()
```

Out[23]:
```
0.33419150673744386
```

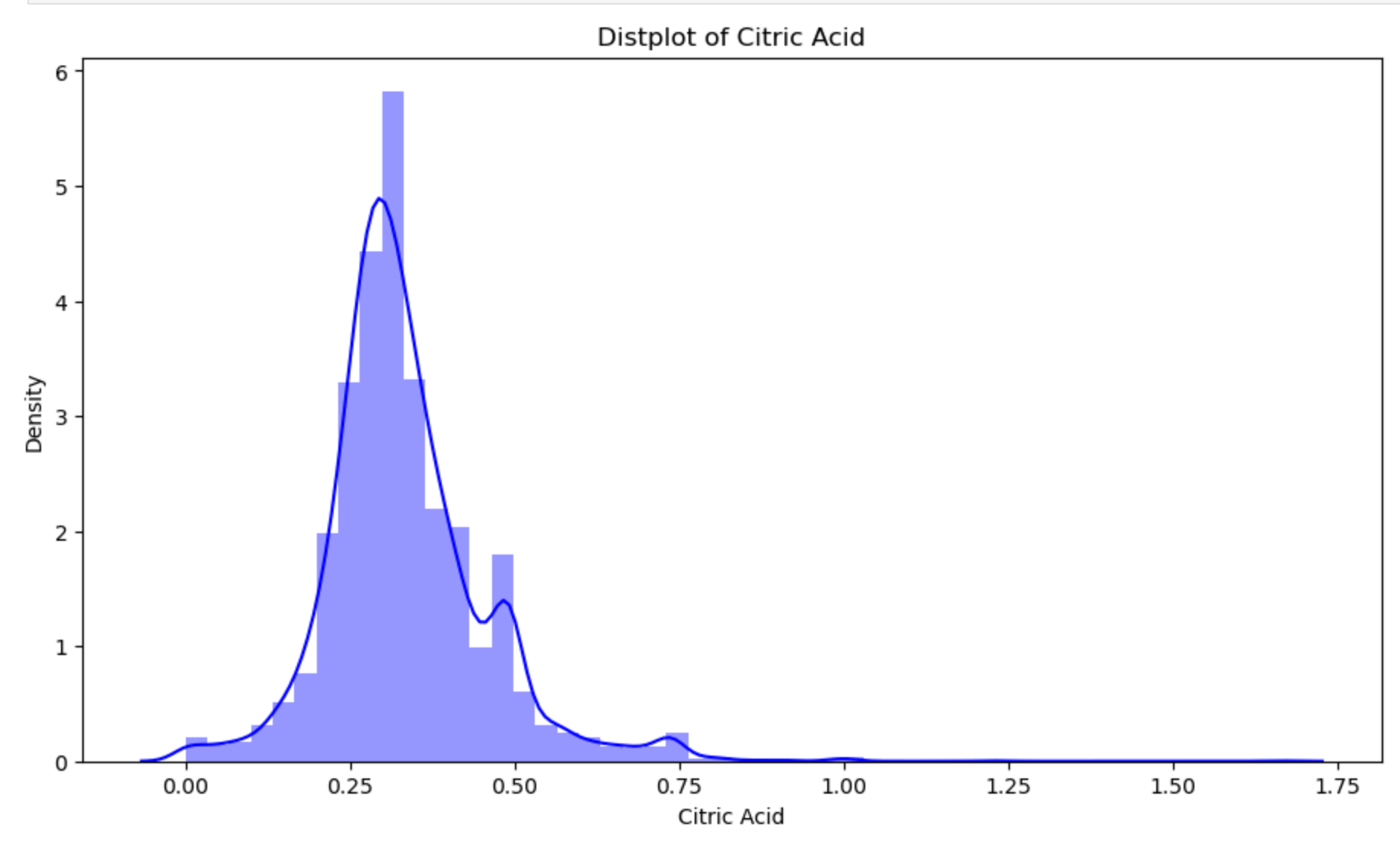
In [24]:
```python
df['citric acid'].median()
```

Out[24]:
```
0.32
```

In [25]:
```python
plt.figure(figsize = (9,4))
sns.histplot(data = df ,x = 'citric acid', color = 'red',
edgecolor = 'linen', alpha = 0.5, bins = 5)
plt.title("Histogram of Citric Acidity")
plt.xlabel('Citric Acidity')
plt.ylabel('Count')
plt.vlines(df['citric acid'].mean(), ymin = 0, ymax = 4000, colors='blue', label='Mean')
plt.vlines(df['citric acid'].median(), ymin = 0, ymax = 4000, colors='red', label='Median')
plt.legend()
plt.show()
```



In [26]:
```python
plt.figure(figsize = (11,6))
sns.distplot(df['citric acid'], color = 'blue')
plt.title("Distplot of Citric Acid")
plt.xlabel('Citric Acid')
plt.ylabel('Density')
plt.show()
```



In [99]:
```python
quality = pd.DataFrame(df['quality'].value_counts())
quality
```

Out[99]:

quality	count
6	2198
5	1457
7	880
8	175
4	163
3	20
9	5

In [100…
```python
quality.index
```

Out[100…
```
Index([6, 5, 7, 8, 4, 3, 9], dtype='int64', name='quality')
```

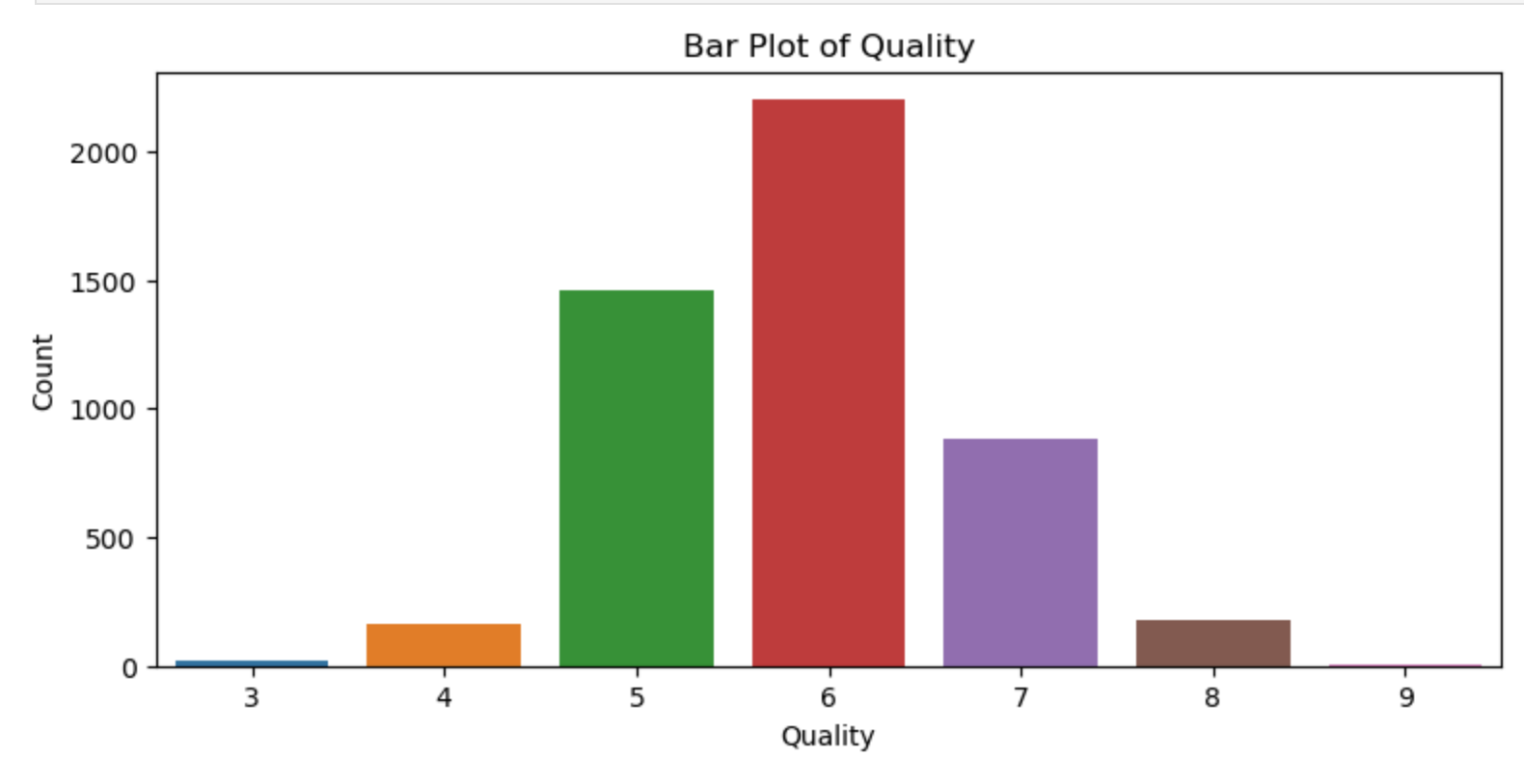
In [102…
```python
df['quality'].value_counts()
```

Out[102…
```
quality
6    2198
5    1457
7     880
8     175
4     163
3      20
9       5
Name: count, dtype: int64
```

In [103…
```python
plt.figure(figsize = (9,4))

sns.barplot(x=quality.index, y=df['quality'].value_counts())

plt.title("Bar Plot of Quality")
plt.xlabel('Quality')
plt.ylabel('Count')
plt.show()
```



In [104…
```python
df['quality'].value_counts().index[0]
```

Out[104…
```
6
```

In [106…
```python
rep_acid = pd.DataFrame(index = ['fixed acidity','volatile acidity','citric acid','quality'],
                        data = [df['fixed acidity'].mean(),df['volatile acidity'].mean(),
                                df['citric acid'].mean(),df['quality'].value_counts().index[0]])
```

In [107…
```python
rep_acid
```

Out[107…

	0
fixed acidity	6.854788
volatile acidity	0.278241
citric acid	0.334192

quality	6.000000

In [ ]: