In [2]: import pandas as pd
 import numpy as np
 import matplotlib.pyplot as plt
 import seaborn as sns
 import sklearn
 from sklearn.model\_selection import train\_test\_split
 from sklearn.linear\_model import LinearRegression

In [3]: df=pd.read\_csv("WineQT.csv")

In [4]: df.head()

## Out[4]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	рН	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

In [5]: df.describe()

## Out[5]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfui dioxide
count	1143.000000	1143.000000	1143.000000	1143.000000	1143.000000	1143.000000	1143.000000
mean	8.311111	0.531339	0.268364	2.532152	0.086933	15.615486	45.914698
std	1.747595	0.179633	0.196686	1.355917	0.047267	10.250486	32.782130
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000
25%	7.100000	0.392500	0.090000	1.900000	0.070000	7.000000	21.000000
50%	7.900000	0.520000	0.250000	2.200000	0.079000	13.000000	37.000000
75%	9.100000	0.640000	0.420000	2.600000	0.090000	21.000000	61.000000
max	15.900000	1.580000	1.000000	15.500000	0.611000	68.000000	289.000000
4							

## 

#	Column	Non-Null Count	Dtype
0	fixed acidity	1143 non-null	float64
1	volatile acidity	1143 non-null	float64
2	citric acid	1143 non-null	float64
3	residual sugar	1143 non-null	float64
4	chlorides	1143 non-null	float64
5	free sulfur dioxide	1143 non-null	float64
6	total sulfur dioxide	1143 non-null	float64
7	density	1143 non-null	float64
8	рН	1143 non-null	float64
9	sulphates	1143 non-null	float64
10	alcohol	1143 non-null	float64
11	quality	1143 non-null	int64
12	Id	1143 non-null	int64

dtypes: float64(11), int64(2)

Data columns (total 13 columns):

memory usage: 116.2 KB

## In [7]: df.isnull().sum()

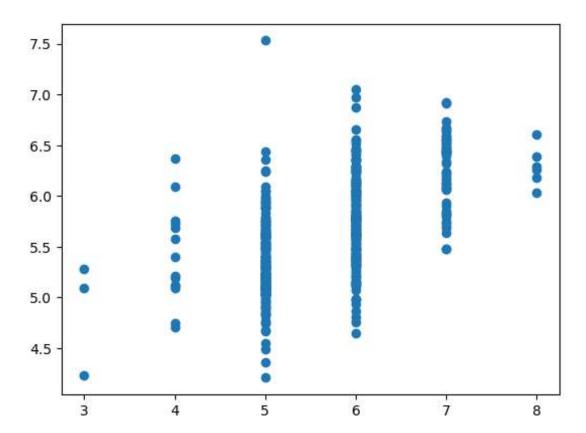
```
Out[7]: fixed acidity
                                 0
        volatile acidity
                                 0
        citric acid
                                 0
        residual sugar
                                 0
        chlorides
                                 0
        free sulfur dioxide
                                 0
        total sulfur dioxide
                                 0
        density
                                 0
        рΗ
                                 0
        sulphates
                                 0
        alcohol
                                 0
        quality
                                 0
        Ιd
                                 0
```

dtype: int64

```
In [15]: | sns.heatmap(df.corr(),annot=True)
Out[15]: <Axes: >
                                                                                                - 1.0
                   fixed acidity - 1 0.250.670.170.11-0.160.110.680.690.170.075.12-0.28
                volatile acidity -0.25 1 -0.50.005805-6.002.078.0170.22-0.28-0.2-0.40.007
                                                                                                - 0.8
                      citric acid -0.67-0.54 1 0.180.250.050.0370.38-0.550.330.110.24-0.14
                                                                                                - 0.6
                 residual sugar -0.1-0.0058.18 1 0.0710.170.190.38-0.120.010.058.02-0.04
                      chlorides -0.110.0560.250.071 1 0.016.0480.21-0.280.37-0.230.120.08
                                                                                                - 0.4
             free sulfur dioxide -0.160.000.058.170.015 1 0.660.054.078.03-0.040.068.095
                                                                                                - 0.2
            total sulfur dioxide -0.1D.078.0370.190.0480.66 1 0.050.059.0270.190.180.11
                        density -0.680.0170.380.380.210.0540.05 1 -0.350.14-0.490.180.36
                                                                                                - 0.0
                             pH -0.690.22-0.550.120.280.07-0.0590.35 1 -0.190.230.0520.13
                                                                                                -0.2
                      sulphates -0.17-0.280.330.0170.370.039.0270.14-0.19 1 0.0940.26 -0.1
                        alcohol -0.0750.2 0.110.0580.230.0470.190.490.230.094 1 0.48 0.24
                                                                                                 -0.4
                        quality -0.12-0.410.240.0220.120.0630.180.180.0520.260.48 1
                             ld -0.26.0079.140.046.088.0950.110.360.13-0.1 0.240.07
                                                       ree sulfur dioxide
                                                            total sulfur dioxide
                                                                                  quality
                                          citric acid
                                                                              alcohol
                                  fixed acidity
                                                   chlorides
                                                                     Hd
                                                                          sulphates
                                                                                       D
                                      volatile acidity
                                               esidual sugar
                                                                 density
In [17]: df.columns
Out[17]: Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
                    'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
                    'pH', 'sulphates', 'alcohol', 'quality', 'Id'],
                  dtype='object')
In [26]: x=df[['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
                    'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
                    'pH', 'sulphates', 'alcohol']]
           y=df['quality']
In [38]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.4,random_state=4
In [42]: | lr=LinearRegression()
```

Out[51]: <matplotlib.collections.PathCollection at 0x23f469ffca0>

In [51]: plt.scatter(y\_test,predictions)



In [49]: sns.distplot((y\_test-predictions),bins=50)

C:\Users\computer\AppData\Local\Temp\ipykernel\_8372\1061164399.py:1: UserWarn
ing:

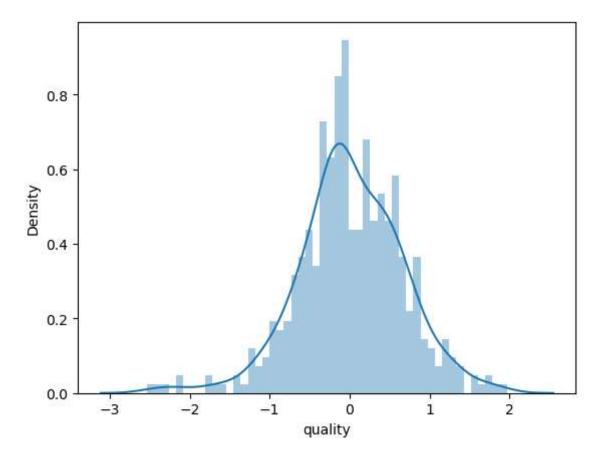
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <a href="https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751">https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751</a> (https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

sns.distplot((y\_test-predictions),bins=50)

Out[49]: <Axes: xlabel='quality', ylabel='Density'>



In [ ]: