In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
```

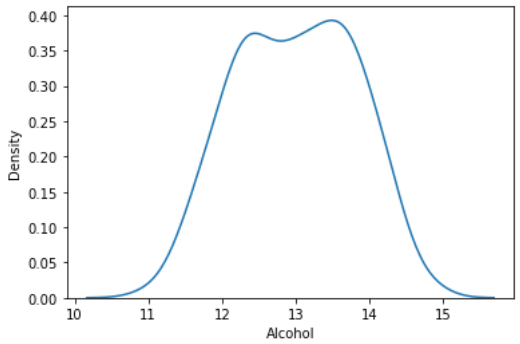
In [2]:
```
df = pd.read_csv("wine_data.csv",header = None , usecols=[0,1,2])
df.columns = ['Class Label','Alcohol','Malic Acid']
df.head()
```

Out[2]:

	Class Label	Alcohol	Malic Acid
0	1	14.23	1.71
1	1	13.20	1.78
2	1	13.16	2.36
3	1	14.37	1.95
4	1	13.24	2.59

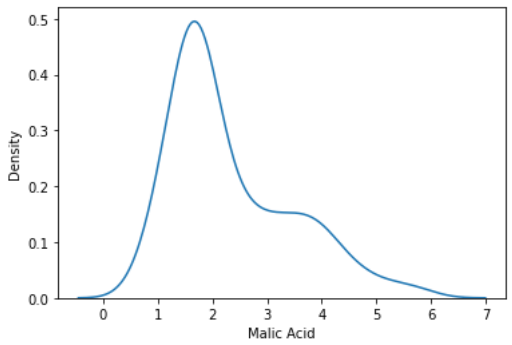
In [3]:
```
sns.kdeplot(df['Alcohol'])
```

Out[3]: <AxesSubplot:xlabel='Alcohol', ylabel='Density'>



In [4]:
```
sns.kdeplot(df['Malic Acid'])
```

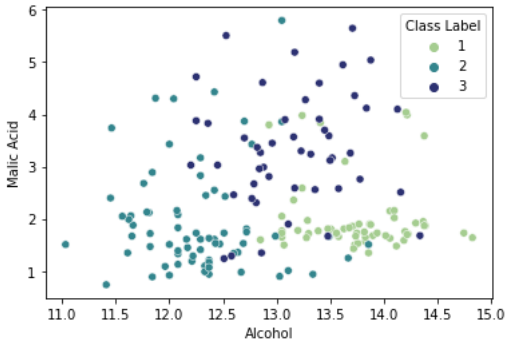
Out[4]: <AxesSubplot:xlabel='Malic Acid', ylabel='Density'>



In [5]:
```
sns.scatterplot(df['Alcohol'],df['Malic Acid'],df['Class Label'],palette = 'crest')
```

```
C:\Users\User27\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword a
rgs: x, y, hue. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an ex
plicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[5]: <AxesSubplot:xlabel='Alcohol', ylabel='Malic Acid'>



In [6]:
```
x = df.drop('Class Label',axis=1)
```

In [7]:
```
x_train,x_test,y_train,y_test = train_test_split(x,df['Class Label'], test_size=0.3)
```

In [8]:
```
x_train.shape, x_test.shape
```

Out[8]: ((124, 2), (54, 2))

In [9]:
```
scaler = MinMaxScaler()
```

In [10]:
```
scaler.fit(x_train)
```

Out[10]: MinMaxScaler()

In [11]:
```
x_train_scaled = pd.DataFrame(scaler.transform(x_train))
x_test_scaled = pd.DataFrame(scaler.transform(x_test))
```

In [12]:
```
x_train_scaled
```

Out[12]:

	0	1
0	0.165789	0.207983
1	0.352632	0.155462
2	0.589474	0.712185
3	0.389474	0.073529
4	0.765789	0.176471
...	...	...
119	0.000000	0.130252
120	0.878947	0.222689
121	0.276316	0.105042
122	0.515789	0.163866
123	0.839474	0.170168

124 rows × 2 columns

In [13]:
```
x_test_scaled
```

Out[13]:

	0	1
0	0.350000	0.617647
1	0.244737	0.042017
2	0.531579	1.031513
3	0.276316	0.050420
4	0.413158	0.329832
5	0.813158	0.123950
6	0.786842	0.165966
7	0.834211	0.182773
8	0.671053	0.355042
9	0.613158	0.350840
10	0.152632	0.096639
11	0.139474	0.243697
12	0.681579	0.852941
13	0.155263	0.231092
14	0.978947	0.176471
15	0.700000	0.497899
16	0.471053	0.521008
17	0.255263	0.006303
18	0.647368	0.161765
19	0.694737	0.075630
20	0.163158	0.163866
21	0.623684	0.634454
22	0.331579	0.109244
23	0.100000	-0.031513
24	0.439474	0.626050
25	0.463158	0.373950
26	0.644737	0.193277
27	0.650000	0.468487
28	0.736842	0.159664
29	0.697368	0.197479
30	0.389474	0.176471
31	0.536842	0.128151
32	0.839474	0.651261
33	0.352632	0.037815
34	0.213158	0.420168
35	0.739474	0.678571
36	0.713158	0.163866
37	0.265789	0.716387
38	0.592105	0.157563
39	0.723684	0.392857
40	0.444737	0.193277
41	0.352632	0.050420
42	0.500000	0.611345
43	0.707895	0.113445
44	0.881579	0.567227
45	0.797368	0.155462
46	0.563158	0.903361
47	0.621053	0.184874
48	0.644737	0.163866
49	0.560526	0.308824
50	0.736842	0.142857
51	0.860526	0.216387
52	0.636842	0.590336
53	0.710526	0.128151

In [14]:
```
x_train_scaled.describe()
```

Out[14]:

	0	1
count	124.000000	124.000000
mean	0.505348	0.300200
std	0.208749	0.227945
min	0.000000	0.000000
25%	0.342105	0.150735
50%	0.493421	0.206933
75%	0.667105	0.449580
max	1.000000	1.000000

In [15]:
```
x_test_scaled.describe()
```

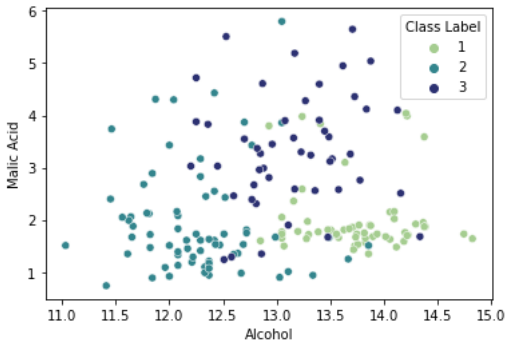
Out[15]:

	0	1
count	54.000000	54.000000
mean	0.548977	0.312247
std	0.223482	0.251525
min	0.100000	-0.031513
25%	0.361842	0.146008
50%	0.617105	0.193277
75%	0.709868	0.490546
max	0.978947	1.031513

In [16]:
```
sns.scatterplot(df['Alcohol'],df['Malic Acid'],df['Class Label'],palette = 'crest')
```

```
C:\Users\User27\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword a
rgs: x, y, hue. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an ex
plicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

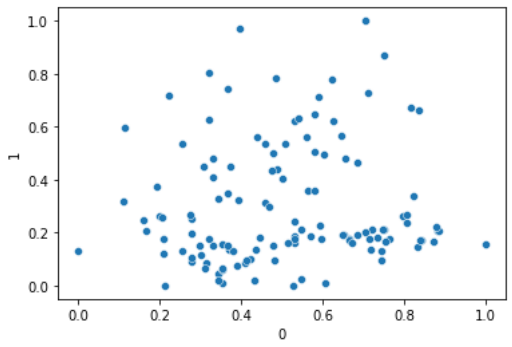
Out[16]: <AxesSubplot:xlabel='Alcohol', ylabel='Malic Acid'>



In [17]:
```
sns.scatterplot(x_train_scaled[0],x_train_scaled[1],palette = 'crest')
```

```
C:\Users\User27\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword a
rgs: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explici
t keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[17]: <AxesSubplot:xlabel='0', ylabel='1'>



Effects of Outliers

In [18]:
```
df.describe()
```

Out[18]:

	Class Label	Alcohol	Malic Acid
count	178.000000	178.000000	178.000000
mean	1.938202	13.000618	2.336348
std	0.775035	0.811827	1.117146
min	1.000000	11.030000	0.740000
25%	1.000000	12.362500	1.602500
50%	2.000000	13.050000	1.865000
75%	3.000000	13.677500	3.082500
max	3.000000	14.830000	5.800000

In [19]:
```
df2 = pd.DataFrame({
    'Alcohol': [30,35,40],
    'Malic Acid' :[15,17,20],
    'Class Label' : [1,2,3]
})
```

In [20]:
```
df3 = df.append(df2)
```

```
C:\Users\User27\AppData\Local\Temp\ipykernel_13812\2057898889.py:1: FutureWarning: The frame.append method is deprecated and wi
ll be removed from pandas in a future version. Use pandas.concat instead.
  df3 = df.append(df2)
```

In [21]:
```
df3.describe()
```

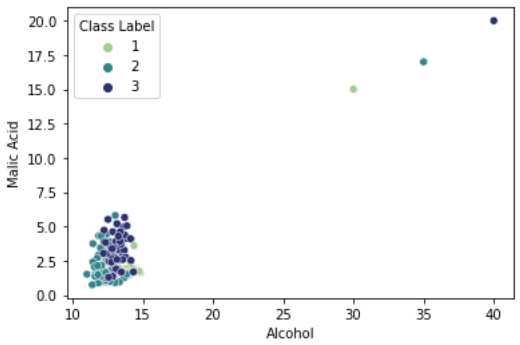
Out[21]:

	Class Label	Alcohol	Malic Acid
count	181.000000	181.000000	181.000000
mean	1.939227	13.365249	2.584917
std	0.775785	2.976302	2.232476
min	1.000000	11.030000	0.740000
25%	1.000000	12.370000	1.610000
50%	2.000000	13.050000	1.870000
75%	3.000000	13.710000	3.170000
max	3.000000	40.000000	20.000000

In [22]:
```
sns.scatterplot(df3['Alcohol'],df3['Malic Acid'],df3['Class Label'],palette = 'crest')
```

```
C:\Users\User27\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword a
rgs: x, y, hue. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an ex
plicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[22]: <AxesSubplot:xlabel='Alcohol', ylabel='Malic Acid'>



In [24]:
```
x1 = df3.drop('Class Label',axis=1)
x1_train,x1_test,y1_train,y1_test = train_test_split(x1,df3['Class Label'], test_size=0.3)
```

In [25]:
```
x1_train_scaled = pd.DataFrame(scaler.fit_transform(x1_train))
x1_test_scaled = pd.DataFrame(scaler.fit_transform(x1_test))
```

In [29]:
```
x1_train_scaled
```

Out[29]:

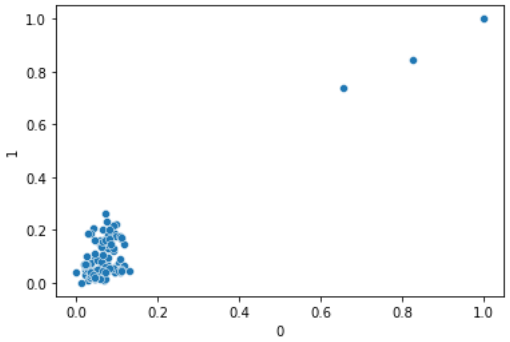
	0	1
0	0.049707	0.040498
1	0.093200	0.187954
2	0.042113	0.206646
3	0.063169	0.031672
4	0.077667	0.046729
...	...	...
121	0.043493	0.108515
122	0.081809	0.164590
123	0.044874	0.018692
124	0.109078	0.044133
125	0.084915	0.147975

126 rows × 2 columns

In [26]:
```
sns.scatterplot(x1_train_scaled[0],x1_train_scaled[1],palette = 'crest')
```

```
C:\Users\User27\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword a
rgs: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explici
t keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[26]: <AxesSubplot:xlabel='0', ylabel='1'>



In [27]:
```
x1_train_scaled.describe()
```

Out[27]:

	0	1
count	126.000000	126.000000
mean	0.084811	0.101930
std	0.121651	0.133779
min	0.000000	0.000000
25%	0.044960	0.044263
50%	0.069382	0.059450
75%	0.091129	0.135644
max	1.000000	1.000000

In [28]:
```
x1_test_scaled.describe()
```

Out[28]:

	0	1
count	55.000000	55.000000
mean	0.495868	0.294715
std	0.246885	0.225564
min	0.000000	0.000000
25%	0.293939	0.154334
50%	0.542424	0.198732
75%	0.686364	0.434461
max	1.000000	1.000000

In [ ]: