

Exploratory Data Analysis

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas_profiling as pp
from sklearn.preprocessing import PowerTransformer
from sklearn.preprocessing import StandardScaler
```

In [2]:

```
df = pd.read_excel("Dry_Bean_Dataset.xlsx")
df.head()
```

Out[2]:

	Area	Perimeter	MajorAxisLength	MinorAxisLength	AspectRatio	Eccentricity	ConvexArea	EquivDiameter	Extent	Solidity	roundness
0	28395	610.291	208.178117	173.888747	1.197191	0.549812	28715	190.141097	0.763923	0.988856	0.958027
1	28734	638.018	200.524796	182.734419	1.097356	0.411785	29172	191.272750	0.783968	0.984986	0.887034
2	29380	624.110	212.826130	175.931143	1.209713	0.562727	29690	193.410904	0.778113	0.989559	0.947849
3	30008	645.884	210.557999	182.516516	1.153638	0.498616	30724	195.467062	0.782681	0.976696	0.903936
4	30140	620.134	201.847882	190.279279	1.060798	0.333680	30417	195.896503	0.773098	0.990893	0.984877

In [3]:

```
eda = pp.ProfileReport(df)
eda.to_file("Bean_Report.html")
```

In [4]:

```
df = df.drop_duplicates()
X = df.iloc[:, :16]
y = df.iloc[:, 16:]
y = y.reset_index().drop(columns = "index")
scaler = StandardScaler()
X = pd.DataFrame(scaler.fit_transform(X), columns = X.columns)
pt = PowerTransformer(method = "yeo-johnson")
X = pd.DataFrame(pt.fit_transform(X), columns = X.columns)
```

In [5]:

```
eda = pp.ProfileReport(X)
eda.to_file("Bean_Transformed_Report.html")
```

In [9]:

```
pairplot = sns.pairplot(df, hue='Class')
plt.savefig("pairplot.jpg", dpi = 600)
```



