

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
import pandas as pd
import io
data = pd.read_csv('https://raw.githubusercontent.com/JanviAjudiya/DS_Pra-2_Data-pr
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1339 entries, 0 to 1338
Data columns (total 44 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Unnamed: 0                            1339 non-null   int64
1   Species                               1339 non-null   object
2   Owner                                 1332 non-null   object
3   Country.of.Origin                     1338 non-null   object
4   Farm.Name                             980 non-null    object
5   Lot.Number                            276 non-null    object
6   Mill                                  1021 non-null   object
7   ICO.Number                            1182 non-null   object
8   Company                               1130 non-null   object
9   Altitude                              1113 non-null   object
10  Region                                1280 non-null   object
11  Producer                              1107 non-null   object
12  Number.of.Bags                        1339 non-null   int64
13  Bag.Weight                            1339 non-null   object
14  In.Country.Partner                    1339 non-null   object
15  Harvest.Year                          1292 non-null   object
16  Grading.Date                          1339 non-null   object
17  Owner.1                               1332 non-null   object
18  Variety                               1113 non-null   object
19  Processing.Method                     1169 non-null   object
20  Aroma                                 1339 non-null   float64
21  Flavor                                1339 non-null   float64
22  Aftertaste                            1339 non-null   float64
23  Acidity                               1339 non-null   float64
24  Body                                  1339 non-null   float64
25  Balance                               1339 non-null   float64
26  Uniformity                            1339 non-null   float64
27  Clean.Cup                             1339 non-null   float64
28  Sweetness                             1339 non-null   float64
29  Cupper.Points                         1339 non-null   float64
30  Total.Cup.Points                      1339 non-null   float64
31  Moisture                              1339 non-null   float64
32  Category.One.Defects                  1339 non-null   int64
33  Quakers                               1338 non-null   float64
34  Color                                 1121 non-null   object
35  Category.Two.Defects                  1339 non-null   int64
36  Expiration                            1339 non-null   object
37  Certification.Body                    1339 non-null   object
```

✓ 0s completed at 1:53 PM



```
data.head()
```

	Unnamed: 0	Species	Owner	Country.of.Origin	Farm.Name	Lot.Number	M
0	0	Arabica	metad plc	Ethiopia	metad plc	NaN	m
1	1	Arabica	metad plc	Ethiopia	metad plc	NaN	m
2	2	Arabica	grounds for health admin	Guatemala	san marcos barrancas "san cristobal cuch	NaN	
3	3	Arabica	yidnekachew dabessa	Ethiopia	yidnekachew dabessa coffee plantation	NaN	wolc
4	4	Arabica	metad plc	Ethiopia	metad plc	NaN	m

```
from sklearn.preprocessing import LabelEncoder , OneHotEncoder
data['Species'].value_counts()
```

```
Arabica    1311
Robusta    28
Name: Species, dtype: int64
```

### (i) Label Encoder

```
le=LabelEncoder()
```

```
le.fit(data['Species'])
```

```
Name: Number.of.Bags, Length: 131, dtype: int64
```

```
le.classes_
```

```
array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10,
       11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,
       22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32,
       33, 35, 36, 37, 38, 39, 40, 42, 43, 44, 45,
       48, 49, 50, 51, 53, 54, 56, 58, 60, 62, 65,
       66, 69, 70, 74, 75, 77, 80, 82, 84, 85, 90,
       93, 94, 100, 114, 120, 121, 123, 125, 127, 129, 130,
      134, 135, 138, 140, 149, 150, 160, 165, 166, 167, 170,
      175, 180, 198, 200, 202, 209, 220, 223, 226, 230, 232,
      235, 240, 243, 245, 248, 250, 252, 253, 256, 270, 274,
      275, 280, 285, 288, 300, 302, 304, 305, 310, 320, 325,
      360, 377, 380, 400, 440, 450, 500, 550, 600, 1062])
```

## (ii) Onehot Encoder

```
data['In.Country.Partner'].value_counts()
```

```
Specialty Coffee Association
AMECAFE
Almacafé
Asociacion Nacional Del Café
Brazil Specialty Coffee Association
Instituto Hondureño del Café
Blossom Valley International
Africa Fine Coffee Association
Specialty Coffee Association of Costa Rica
NUCOFFEE
Uganda Coffee Development Authority
Kenya Coffee Traders Association
Ethiopia Commodity Exchange
Specialty Coffee Institute of Asia
METAD Agricultural Development plc
Yunnan Coffee Exchange
Salvadoran Coffee Council
```

```
'Blossom Valley International', 'Blossom Valley International\n',  
'Brazil Specialty Coffee Association',  
'Central De Organizaciones Productoras De Café y Cacao Del Perú - Cent  
'Centro Agroecológico del Café A.C.', 'Coffee Quality Institute',  
'Ethiopia Commodity Exchange', 'Instituto Hondureño del Café',  
'Kenya Coffee Traders Association',  
'METAD Agricultural Development plc', 'NUCOFFEE',  
'Salvadoran Coffee Council', 'Specialty Coffee Ass',  
'Specialty Coffee Association',  
'Specialty Coffee Association of Costa Rica',  
'Specialty Coffee Association of Indonesia',  
'Specialty Coffee Institute of Asia', 'Tanzanian Coffee Board',  
'Torch Coffee Lab Yunnan', 'Uganda Coffee Development Authority',  
'Yunnan Coffee Exchange'], dtype=object)]
```

```
transformed_data = pd.DataFrame(transformed_data ,  
                                columns = ['AMECAFE', 'Africa Fine Coffee Associati  
'Asociacion Nacional Del Café',  
'Asociación Mexicana De Cafés y Cafeterías De Especialidad A.C.',  
'Asociación de Cafés Especiales de Nicaragua',  
'Blossom Valley International', 'Blossom Valley International\n',  
'Brazil Specialty Coffee Association',  
'Central De Organizaciones Productoras De Café y Cacao Del Perú - Central C  
'Centro Agroecológico del Café A.C.', 'Coffee Quality Institute',  
'Ethiopia Commodity Exchange', 'Instituto Hondureño del Café',  
'Kenya Coffee Traders Association',  
'METAD Agricultural Development plc', 'NUCOFFEE',  
'Salvadoran Coffee Council', 'Specialty Coffee Ass',  
'Specialty Coffee Association',  
'Specialty Coffee Association of Costa Rica',  
'Specialty Coffee Association of Indonesia',  
'Specialty Coffee Institute of Asia', 'Tanzanian Coffee Board',
```

Africa Fine Coffee Association  
Almacafé  
Asociacion Nacional Del Café  
Asociación Mexicana De Cafés y Cafeterías De Especialidad A.C.  
Asociación de Cafés Especiales de Nicaragua  
Blossom Valley International  
Blossom Valley International\n  
Brazil Specialty Coffee Association  
Central De Organizaciones Productoras De Café y Cacao Del Perú - Central Café  
Centro Agroecológico del Café A.C.  
Coffee Quality Institute  
Ethiopia Commodity Exchange  
Instituto Hondureño del Café  
Kenya Coffee Traders Association  
METAD Agricultural Development plc  
NUCOFFEE  
Salvadoran Coffee Council  
Specialty Coffee Ass  
Specialty Coffee Association  
Specialty Coffee Association of Costa Rica  
Specialty Coffee Association of Indonesia  
Specialty Coffee Institute of Asia  
Tanzanian Coffee Board  
Torch Coffee Lab Yunnan  
Uganda Coffee Development Authority  
Yunnan Coffee Exchange  
Name: 90, dtype: float64

```
len(numeric_columns) , len(data.columns)
```

```
(21, 44)
```

```
numeric_columns.remove('Aroma')  
numeric_columns.remove('Flavor')
```

```
temp_data = data[numeric_columns]  
temp_data
```

	Unnamed: 0	Species	Number.of.Bags	Aftertaste	Acidity	Body	Balance	Uni
0	0	0	114	8.67	8.75	8.50	8.42	
1	1	0	114	8.50	8.58	8.42	8.42	
2	2	0	5	8.42	8.42	8.33	8.42	

<b>4</b>	0.002990	0.0	0.876923	0.951557	0.971429	0.981352	0.952000
...	...	...	...	...	...	...	...
<b>1334</b>	0.997010	1.0	0.007692	0.845444	0.866286	0.592075	0.894857
<b>1335</b>	0.997758	1.0	0.007692	0.893887	0.885714	0.602564	0.600000
<b>1336</b>	0.998505	1.0	0.007692	0.826990	0.848000	0.874126	0.819429
<b>1337</b>	0.999253	1.0	0.007692	0.778547	0.819429	0.844988	0.800000
<b>1338</b>	1.000000	1.0	0.007692	0.749712	0.780571	0.806527	0.780571

1339 rows × 14 columns

```
Altitude      226
Region         59
Producer      232
Number.of.Bags    0
Bag.Weight      0
In.Country.Partner  0
Harvest.Year    47
```









