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

```
df = pd.read_csv('/content/Coffee_Qlty.csv')
```

```
def dfinfo(df):
  datainf = {
      'dtp': df.dtypes,
      'cnt':df.count(),
      'unq': df.nunique(),
      'nul': df.isna().sum(),
      'dup': df.duplicated().sum()
      }
  res = pd.DataFrame(datainf)
  return res
```

```
dfinfo(df)
```

	dtp	cnt	unq	nul	dup
REC_ID	int64	1339	1339	0	0
Species	object	1339	2	0	0
Continent.of.Origin	object	1338	5	1	0
Country.of.Origin	object	1338	34	1	0
Harvest.Year	float64	1279	10	60	0
Expiration	object	1339	565	0	0
Variety	object	1113	29	226	0
Color	object	1069	3	270	0
Processing.Method	object	1169	5	170	0
Aroma	float64	1339	33	0	0
Flavor	float64	1339	35	0	0
Aftertaste	float64	1339	35	0	0
Acidity	float64	1339	31	0	0
Body	float64	1339	33	0	0
Balance	float64	1339	33	0	0
Uniformity	float64	1339	10	0	0
Clean.Cup	float64	1339	11	0	0
Sweetness	float64	1339	17	0	0
Moisture	float64	1339	23	0	0
Quakers	int64	1339	11	0	0
Category.One.Defects	int64	1339	18	0	0
Category.Two.Defects	int64	1339	38	0	0

```
df2 = df.dropna(subset = ['Continent.of.Origin'])
```

```
df2['Harvest.Year'].unique()
```

```
array([2014.,   nan, 2013., 2012., 2010., 2009., 2015., 2011., 2016.,
       2017., 2018.])
```

```
hymed = df2['Harvest.Year'].median()
```

```
df2['Harvest.Year'] = df2['Harvest.Year'].fillna(hymed)
df2['Harvest.Year'] = df2['Harvest.Year'].astype(int)
```

Show hidden output

```
df2['Variety'].fillna('Not Known', inplace  = True)
df2['Color'].fillna('Not Known', inplace  = True)
df2['Processing.Method'].fillna('Not Known', inplace  = True)
```

```
df3 = df2.rename(columns = {
    'REC_ID':'Record Id',
'Species':'Species',
'Continent.of.Origin':'Continent Of Origin',
'Country.of.Origin':'Country Of Origin',
'Harvest.Year':'Harvest Year',
'Expiration':'Expiration',
'Variety':'Variety',
'Color':'Color',
'Processing.Method':'Processing Method',
'Aroma':'Aroma',
'Flavor':'Flavor',
'Aftertaste':'Aftertaste',
'Acidity':'Acidity',
'Body':'Body',
'Balance':'Balance',
'Uniformity':'Uniformity',
'Clean.Cup':'Clean Cup',
'Sweetness':'Sweetness',
'Moisture':'Moisture',
'Quakers':'Quakers',
'Category.One.Defects':'Category One Defects',
'Category.Two.Defects':'Category Two Defects',
})
```

```
dfinfo(df3)
```

	dtp	cnt	unq	nul	dup
Record Id	int64	1338	1338	0	0
Species	object	1338	2	0	0
Continent Of Origin	object	1338	5	0	0
Country Of Origin	object	1338	34	0	0
Harvest Year	int64	1338	10	0	0
Expiration	object	1338	565	0	0
Variety	object	1338	30	0	0
Color	object	1338	4	0	0
Processing Method	object	1338	6	0	0
Aroma	float64	1338	33	0	0
Flavor	float64	1338	35	0	0
Aftertaste	float64	1338	35	0	0
Acidity	float64	1338	31	0	0
Body	float64	1338	33	0	0
Balance	float64	1338	33	0	0
Uniformity	float64	1338	10	0	0
Clean Cup	float64	1338	11	0	0
Sweetness	float64	1338	17	0	0
Moisture	float64	1338	23	0	0