Exploratory Data Analysis Project Documentation

TITLE: Coffee Quality Data Analysis

Name: Adhwaitha A

DA/DS: Data Analytics (DA)

Batch Number: B4 (June - Online)(M) - DA & DS

Online/Offline: Online

Table of Contents

SI/no	TITLE
1	INTRODUCTION
2	Aim
3	Business Problem / Problem Statement
4	Project Workflow
5	Data Understanding
6	Data Cleaning - Missing Values Imputation, Outliers, Handling Inconsistent Values
7	Obtaining Derived Metrics
8	Filtering Data for Analysis
9	EDA – Univariate Analysis
10	Segmented Univariate Analysis
11	Bivariate Analysis
12	Multivariate Analysis
13	Overall Insights Obtained from Analysis
14	Conclusion

1.Introduction:

The objective of this project is to perform the exploratory data analysis in **Coffee Quality** for the underlying patterns and insights lying therein. It will contain cleaning of data, deriving metrics, and other analyses for the purpose of deriving meaningful conclusions. This dataset contains the concept that encompasses a variety of factors influencing the taste, aroma, and overall enjoyment of the beverage. Understanding what contributes to high-quality coffee involves considering the entire journey from the coffee plant to the cup.

2.Aim:

The main goal in making top-notch coffee is to create a drink that gives people an amazing taste experience. This means the coffee should have unique and appealing flavors, smells, and feels in the mouth. To do this, you need to look at the whole process, from growing the beans to brewing the final cup…

3. Problem Statement:

The coffee industry struggles to deliver top-notch coffee . It needs to satisfy different consumer tastes while maintaining ethical and sustainable practices across its supply chain. This issue has many sides, including farming methods, processing techniques roasting accuracy, and brewing skills. It also involves environmental and social duties. Tackling these hurdles is key to staying competitive keeping customers happy, and ensuring the industry lasts long-term.

4. Project Workflow

Data Collection --> Data Understanding --> Data Cleaning --> Data Transformation --> Exploratory Data Analysis --> Conclusions

5. Data Understanding

Understanding coffee quality data involves collecting and analyzing information from farming, processing, roasting, and brewing stages. This data is cleaned and integrated to ensure consistency, then analyzed to uncover patterns and significant factors affecting quality. Predictive models are used to forecast outcomes, and continuous monitoring systems are set up

for real-time data analysis. This approach helps businesses make informed decisions, improve quality control, promote sustainability, and enhance the overall coffee experience for consumers.

6. Data Cleaning - Missing Values Imputation , Outliers, Handling Inconsistent Values

1. Missing Values Imputation : Method used, Mean imputation for numerical variables, mode imputation for categorical variables.

2. Outliers : Method used, Z-score method to identify and remove outliers.
3. Inconsistent Values : Method used, Correcting inconsistent entries in categorical variables

7. Derived Metrics

Acquiring derived metrics is a method of figuring out new results from the given data that helps in getting a better understanding. In coffee, this may be the metrics such as cupping scores, defect rates per batch, or the correlation in taste quality between the coffee grown at different altitudes. Through the help of these new metrics, many more trends and patterns can easily be recognized and thus this brings about the ability to make better decisions through enhanced factors. For instance, the quality of green coffee beans is a direct result of the altitude at which it is grown.

8. EDA – Univariate Analysis

Univariate analysis for coffee quality involves examining each variable individually to summarize and identify patterns. This includes analyzing cupping scores altitude processing

methods and roast levels . This approach helps understand central tendencies, variability, and distribution shapes, providing a foundational insight into the data.

9. Segmented univariate analysis

Segmented univariate analysis examines individual variables within specific data segments to uncover patterns.This could involve comparing cupping scores across regions, altitudes, or processing methods. Analyzing average cupping scores by region can reveal distinct flavor profiles, while examining altitudes within each processing method highlights how altitude influences processing. This approach provides deeper insights into the factors affecting coffee quality.

10. Filtering data for analysis

Filtering data for analysis involves selecting specific subsets of data to focus on relevant information and remove noise. For coffee quality, this might mean filtering by harvest year, region, processing method, or roast level to analyze trends and patterns within these specific groups. This process ensures that the analysis is more precise and targeted, leading to more accurate insights and conclusions.

11. Bivariate analysis

Bivariate analysis examines the relationship between two variables to identify patterns and correlations. In the context of coffee quality, this could involve analyzing the relationship between altitude and cupping scores or between processing methods and flavor profiles. This analysis helps uncover how different factors interact and influence each other, providing a deeper understanding of what drives coffee quality.

12. Multivariate analysis

Multivariate analysis explores the relationships among three or more variables simultaneously to understand complex interactions. In coffee quality, this might involve examining how altitude, processing method, and roast level together influence cupping scores. This comprehensive approach helps identify key factors and their combined effects on coffee quality, offering deeper insights for optimizing production processes.

```
                                        #Coffee Quality Database

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

df=pd.read_csv("C:/Users/user/Downloads/coffeeQuality.csv")
df
```

	Unnamed: 0	Species	Owner	Country.of.Origin
0	0	Arabica	metad plc	Ethiopia
1	1	Arabica	metad plc	Ethiopia
2	2	Arabica	grounds for health admin	Guatemala
3	3	Arabica	yidnekachew dabessa	Ethiopia
4	4	Arabica	metad plc	Ethiopia
...	...	...	...	...
1334	1334	Robusta	luis robles	Ecuador
1335	1335	Robusta	luis robles	Ecuador
1336	1336	Robusta	james moore	United States
1337	1337	Robusta	cafe politico	India
1338	1338	Robusta	cafe politico	Vietnam

```
1335                                    robustasa   Lavado 3  own
laboratory
1336                              fazenda cazengo        NaN    cafe
cazengo
1337                                         NaN        NaN
NaN
1338                                         NaN        NaN
NaN

             ICO.Number                                    Company
Altitude  \
0             2014/2015      metad agricultural developmet plc
1950-2200
1             2014/2015      metad agricultural developmet plc
1950-2200
2                   NaN                                     NaN  1600 -
1800 m
3                   NaN  yidnekachew debessa coffee plantation
1800-2200
4             2014/2015      metad agricultural developmet plc
1950-2200
...                 ...                                     ...
...
1334                NaN                              robustasa
NaN
1335                NaN                              robustasa
40
1336                NaN                 global opportunity fund     795
meters
1337  14-1118-2014-0087                          cafe politico
NaN
1338                NaN                          cafe politico
NaN

      ...       Color Category.Two.Defects           Expiration  \
0     ...       Green                    0      April 3rd, 2016
1     ...       Green                    1      April 3rd, 2016
2     ...         NaN                    0       May 31st, 2011
3     ...       Green                    2     March 25th, 2016
4     ...       Green                    2      April 3rd, 2016
...   ...         ...                  ...                  ...
1334  ...  Blue-Green                    1   January 18th, 2017
1335  ...  Blue-Green                    0   January 18th, 2017
1336  ...         NaN                    6  December 23rd, 2015
1337  ...       Green                    1    August 25th, 2015
1338  ...         NaN                    9    August 25th, 2015

                     Certification.Body \
0     METAD Agricultural Development plc
1     METAD Agricultural Development plc
```

```
2            Specialty Coffee Association
3     METAD Agricultural Development plc
4     METAD Agricultural Development plc
...                                  ...
1334         Specialty Coffee Association
1335         Specialty Coffee Association
1336         Specialty Coffee Association
1337         Specialty Coffee Association
1338         Specialty Coffee Association

                         Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720

                         Certification.Contact unit_of_measurement  \
0     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                    m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
...                                       ...                  ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m

     altitude_low_meters altitude_high_meters altitude_mean_meters
0                 1950.0                2200.0                2075.0
1                 1950.0                2200.0                2075.0
2                 1600.0                1800.0                1700.0
3                 1800.0                2200.0                2000.0
4                 1950.0                2200.0                2075.0
...                  ...                   ...                   ...
1334                 NaN                   NaN                   NaN
1335                40.0                  40.0                  40.0
1336               795.0                 795.0                 795.0
1337                 NaN                   NaN                   NaN
1338                 NaN                   NaN                   NaN

[1339 rows x 44 columns]
```

```
df = df.iloc[:, 1:]

df
```

```
      Species                     Owner Country.of.Origin  \
0     Arabica                 metad plc          Ethiopia
1     Arabica                 metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                 metad plc          Ethiopia
...       ...                       ...               ...
1334  Robusta               luis robles           Ecuador
1335  Robusta               luis robles           Ecuador
1336  Robusta               james moore     United States
1337  Robusta             cafe politico             India
1338  Robusta             cafe politico           Vietnam

                                     Farm.Name Lot.Number
Mill  \
0                                    metad plc        NaN       metad
plc
1                                    metad plc        NaN       metad
plc
2     san marcos barrancas "san cristobal cuch        NaN
NaN
3        yidnekachew dabessa coffee plantation        NaN
wolensu
4                                    metad plc        NaN       metad
plc
...                                        ...        ...
...
1334                                 robustasa   Lavado 1     our own
lab
1335                                 robustasa   Lavado 3  own
laboratory
1336                          fazenda cazengo        NaN    cafe
cazengo
1337                                       NaN        NaN
NaN
1338                                       NaN        NaN
NaN

             ICO.Number                                    Company
Altitude  \
0             2014/2015      metad agricultural developmet plc
1950-2200
1             2014/2015      metad agricultural developmet plc
1950-2200
2                   NaN                                    NaN  1600 -
1800 m
```

```
3                    NaN yidnekachew debessa coffee plantation
1800-2200
4             2014/2015      metad agricultural developmet plc
1950-2200
...                   ...                                   ...
...
1334                 NaN                              robustasa
NaN
1335                 NaN                              robustasa
40
1336                 NaN                global opportunity fund     795
meters
1337  14-1118-2014-0087                          cafe politico
NaN
1338                 NaN                          cafe politico
NaN

                            Region  ...        Color
Category.Two.Defects  \
0                      guji-hambela  ...        Green
0
1                      guji-hambela  ...        Green
1
2                               NaN  ...          NaN
0
3                            oromia  ...        Green
2
4                      guji-hambela  ...        Green
2
...                             ...  ...          ...
...
1334               san juan, playas  ...  Blue-Green
1
1335               san juan, playas  ...  Blue-Green
0
1336  kwanza norte province, angola  ...          NaN
6
1337                            NaN  ...        Green
1
1338                            NaN  ...          NaN
9

               Expiration                   Certification.Body  \
0         April 3rd, 2016  METAD Agricultural Development plc
1         April 3rd, 2016  METAD Agricultural Development plc
2          May 31st, 2011         Specialty Coffee Association
3        March 25th, 2016  METAD Agricultural Development plc
4         April 3rd, 2016  METAD Agricultural Development plc
...                   ...                                   ...
```

```
1334   January 18th, 2017        Specialty Coffee Association
1335   January 18th, 2017        Specialty Coffee Association
1336  December 23rd, 2015        Specialty Coffee Association
1337    August 25th, 2015        Specialty Coffee Association
1338    August 25th, 2015        Specialty Coffee Association

                          Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720

                          Certification.Contact unit_of_measurement  \
0     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                    m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
...                                        ...                  ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m

      altitude_low_meters altitude_high_meters  altitude_mean_meters
0                  1950.0                2200.0                2075.0
1                  1950.0                2200.0                2075.0
2                  1600.0                1800.0                1700.0
3                  1800.0                2200.0                2000.0
4                  1950.0                2200.0                2075.0
...                   ...                   ...                   ...
1334                  NaN                   NaN                   NaN
1335                 40.0                  40.0                  40.0
1336                795.0                 795.0                 795.0
1337                  NaN                   NaN                   NaN
1338                  NaN                   NaN                   NaN

[1339 rows x 43 columns]

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1339 entries, 0 to 1338
Data columns (total 43 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Species                 1339 non-null   object
 1   Owner                   1332 non-null   object
 2   Country.of.Origin       1338 non-null   object
 3   Farm.Name               980 non-null    object
 4   Lot.Number              276 non-null    object
 5   Mill                    1021 non-null   object
 6   ICO.Number              1180 non-null   object
 7   Company                 1130 non-null   object
 8   Altitude                1113 non-null   object
 9   Region                  1280 non-null   object
 10  Producer                1107 non-null   object
 11  Number.of.Bags          1338 non-null   float64
 12  Bag.Weight              1339 non-null   object
 13  In.Country.Partner      1339 non-null   object
 14  Harvest.Year            1292 non-null   object
 15  Grading.Date            1339 non-null   object
 16  Owner.1                 1332 non-null   object
 17  Variety                 1113 non-null   object
 18  Processing.Method       1169 non-null   object
 19  Aroma                   1339 non-null   float64
 20  Flavor                  1339 non-null   float64
 21  Aftertaste              1339 non-null   float64
 22  Acidity                 1339 non-null   float64
 23  Body                    1339 non-null   float64
 24  Balance                 1339 non-null   float64
 25  Uniformity              1339 non-null   float64
 26  Clean.Cup               1339 non-null   float64
 27  Sweetness               1339 non-null   float64
 28  Cupper.Points           1339 non-null   float64
 29  Total.Cup.Points        1339 non-null   float64
 30  Moisture                1339 non-null   float64
 31  Category.One.Defects    1339 non-null   int64
 32  Quakers                 1338 non-null   float64
 33  Color                   1069 non-null   object
 34  Category.Two.Defects    1339 non-null   int64
 35  Expiration              1339 non-null   object
 36  Certification.Body      1339 non-null   object
 37  Certification.Address   1339 non-null   object
 38  Certification.Contact   1339 non-null   object
 39  unit_of_measurement     1339 non-null   object
 40  altitude_low_meters     1109 non-null   float64
 41  altitude_high_meters    1109 non-null   float64
 42  altitude_mean_meters    1109 non-null   float64
```

```
dtypes: float64(17), int64(2), object(24)
memory usage: 449.9+ KB

df.describe()

       Number.of.Bags        Aroma       Flavor   Aftertaste
Acidity  \
count     1338.000000  1339.000000  1339.000000  1339.000000
1339.000000
mean       159.085202     7.770187     7.520426     7.401083
7.535706
std        173.698167     5.534440     0.398442     0.404463
0.379827
min          0.000000     0.000000     0.000000     0.000000
0.000000
25%         14.000000     7.420000     7.330000     7.250000
7.330000
50%        175.000000     7.580000     7.580000     7.420000
7.580000
75%        275.000000     7.750000     7.750000     7.580000
7.750000
max       3200.000000   200.000000     8.830000     8.670000
8.750000

              Body      Balance   Uniformity    Clean.Cup    Sweetness
\
count  1339.000000  1339.000000  1339.000000  1339.000000  1339.000000

mean      7.517498     7.518013     9.834877     9.835108     9.856692

std       0.370064     0.408943     0.554591     0.763946     0.616102

min       0.000000     0.000000     0.000000     0.000000     0.000000

25%       7.330000     7.330000    10.000000    10.000000    10.000000

50%       7.500000     7.500000    10.000000    10.000000    10.000000

75%       7.670000     7.750000    10.000000    10.000000    10.000000

max       8.580000     8.750000    10.000000    10.000000    10.000000


       Cupper.Points  Total.Cup.Points     Moisture
Category.One.Defects  \
count    1339.000000       1339.000000  1339.000000
1339.000000
mean        7.503376         82.089851     0.088379
0.479462
std         0.473464          3.500575     0.048287
2.549683
```

```
min          0.000000           0.000000     0.000000
0.000000
25%          7.250000          81.080000     0.090000
0.000000
50%          7.500000          82.500000     0.110000
0.000000
75%          7.750000          83.670000     0.120000
0.000000
max         10.000000          90.580000     0.280000
63.000000
```

```
            Quakers  Category.Two.Defects  altitude_low_meters  \
count  1338.000000           1339.000000          1109.000000
mean      0.173393              3.556385          1750.713315
std       0.832121              5.312541          8669.440545
min       0.000000              0.000000             1.000000
25%       0.000000              0.000000          1100.000000
50%       0.000000              2.000000          1310.640000
75%       0.000000              4.000000          1600.000000
max      11.000000             55.000000        190164.000000
```

```
       altitude_high_meters  altitude_mean_meters
count           1109.000000           1109.000000
mean            1799.347775           1775.030545
std             8668.805771           8668.626080
min                1.000000              1.000000
25%             1100.000000           1100.000000
50%             1350.000000           1310.640000
75%             1650.000000           1600.000000
max           190164.000000         190164.000000
```

```
un = df["Lot.Number"].unique()
a = len(un)
print(a)
```

```
228
```

```
df.drop(columns=["Lot.Number"],inplace=True)
df
```

```
      Species                    Owner Country.of.Origin  \
0     Arabica                metad plc          Ethiopia
1     Arabica                metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                metad plc          Ethiopia
...       ...                      ...                ...
1334  Robusta              luis robles           Ecuador
1335  Robusta              luis robles           Ecuador
1336  Robusta              james moore     United States
```

```
1337  Robusta              cafe politico              India
1338  Robusta              cafe politico            Vietnam

                                     Farm.Name            Mill  \
0                                    metad plc       metad plc
1                                    metad plc       metad plc
2     san marcos barrancas "san cristobal cuch             NaN
3        yidnekachew dabessa coffee plantation         wolensu
4                                    metad plc       metad plc
...                                        ...             ...
1334                                 robustasa     our own lab
1335                                 robustasa  own laboratory
1336                           fazenda cazengo    cafe cazengo
1337                                       NaN             NaN
1338                                       NaN             NaN

             ICO.Number                                   Company
Altitude  \
0             2014/2015      metad agricultural developmet plc
1950-2200
1             2014/2015      metad agricultural developmet plc
1950-2200
2                   NaN                                     NaN  1600 -
1800 m
3                   NaN  yidnekachew debessa coffee plantation
1800-2200
4             2014/2015      metad agricultural developmet plc
1950-2200
...                 ...                                     ...
...
1334                NaN                             robustasa
NaN
1335                NaN                             robustasa
40
1336                NaN                 global opportunity fund     795
meters
1337  14-1118-2014-0087                          cafe politico
NaN
1338                NaN                          cafe politico
NaN

                            Region
Producer  \
0                      guji-hambela                              METAD
PLC
1                      guji-hambela                              METAD
PLC
2                               NaN
NaN
3                           oromia  Yidnekachew Dabessa Coffee
```

```
Plantation
4                        guji-hambela                                  METAD
PLC
...                                   ...
...
1334                 san juan, playas            Café Robusta del Ecuador
S.A.
1335                 san juan, playas            Café Robusta del Ecuador
S.A.
1336  kwanza norte province, angola                                Cafe
Cazengo
1337                               NaN
NaN
1338                               NaN
NaN

      ...        Color Category.Two.Defects           Expiration  \
0     ...        Green                     0      April 3rd, 2016
1     ...        Green                     1      April 3rd, 2016
2     ...          NaN                     0       May 31st, 2011
3     ...        Green                     2     March 25th, 2016
4     ...        Green                     2      April 3rd, 2016
...   ...          ...                   ...                  ...
1334  ...  Blue-Green                      1   January 18th, 2017
1335  ...  Blue-Green                      0   January 18th, 2017
1336  ...          NaN                     6  December 23rd, 2015
1337  ...        Green                     1    August 25th, 2015
1338  ...          NaN                     9    August 25th, 2015

                      Certification.Body  \
0     METAD Agricultural Development plc
1     METAD Agricultural Development plc
2           Specialty Coffee Association
3     METAD Agricultural Development plc
4     METAD Agricultural Development plc
...                                   ...
1334        Specialty Coffee Association
1335        Specialty Coffee Association
1336        Specialty Coffee Association
1337        Specialty Coffee Association
1338        Specialty Coffee Association

                          Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                       ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
```

```
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
```

```
                          Certification.Contact unit_of_measurement  \
0     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                     m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
...                                        ...                   ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m

     altitude_low_meters  altitude_high_meters  altitude_mean_meters
0                 1950.0                2200.0                2075.0
1                 1950.0                2200.0                2075.0
2                 1600.0                1800.0                1700.0
3                 1800.0                2200.0                2000.0
4                 1950.0                2200.0                2075.0
...                  ...                   ...                   ...
1334                 NaN                   NaN                   NaN
1335                40.0                  40.0                  40.0
1336               795.0                 795.0                 795.0
1337                 NaN                   NaN                   NaN
1338                 NaN                   NaN                   NaN

[1339 rows x 42 columns]
```

```
df.isnull().sum()
```

```
Species                     0
Owner                       7
Country.of.Origin           1
Farm.Name                 359
Mill                      318
ICO.Number                159
Company                   209
Altitude                  226
Region                     59
Producer                  232
Number.of.Bags              1
Bag.Weight                  0
In.Country.Partner          0
Harvest.Year               47
Grading.Date                0
```

```
Owner.1                   7
Variety                 226
Processing.Method       170
Aroma                     0
Flavor                    0
Aftertaste                0
Acidity                   0
Body                      0
Balance                   0
Uniformity                0
Clean.Cup                 0
Sweetness                 0
Cupper.Points             0
Total.Cup.Points          0
Moisture                  0
Category.One.Defects      0
Quakers                   1
Color                   270
Category.Two.Defects      0
Expiration                0
Certification.Body        0
Certification.Address     0
Certification.Contact     0
unit_of_measurement       0
altitude_low_meters     230
altitude_high_meters    230
altitude_mean_meters    230
dtype: int64
```

```
c = ['Farm.Name', 'Mill', 'ICO.Number', 'Company', 'Altitude',
'Variety',
                    'Processing.Method', 'Color',
'altitude_low_meters', 'altitude_high_meters', 'altitude_mean_meters']

df.drop(columns=c, inplace=True)

df
```

```
      Species                    Owner Country.of.Origin  \
0     Arabica                metad plc          Ethiopia
1     Arabica                metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                metad plc          Ethiopia
...       ...                      ...               ...
1334  Robusta              luis robles           Ecuador
1335  Robusta              luis robles           Ecuador
1336  Robusta              james moore     United States
1337  Robusta            cafe politico             India
1338  Robusta            cafe politico           Vietnam
```

```
                              Region
Producer  \
0                        guji-hambela                                    METAD
PLC
1                        guji-hambela                                    METAD
PLC
2                                 NaN
NaN
3                             oromia Yidnekachew Dabessa Coffee
Plantation
4                        guji-hambela                                    METAD
PLC
...                               ...
...
1334               san juan, playas          Café Robusta del Ecuador
S.A.
1335               san juan, playas          Café Robusta del Ecuador
S.A.
1336  kwanza norte province, angola                            Cafe
Cazengo
1337                              NaN
NaN
1338                              NaN
NaN

      Number.of.Bags Bag.Weight                   In.Country.Partner  \
0              300.0      60 kg  METAD Agricultural Development plc
1              300.0      60 kg  METAD Agricultural Development plc
2                5.0          1        Specialty Coffee Association
3              320.0      60 kg  METAD Agricultural Development plc
4              300.0      60 kg  METAD Agricultural Development plc
...              ...        ...                                  ...
1334             1.0       2 kg        Specialty Coffee Association
1335             1.0       2 kg        Specialty Coffee Association
1336             1.0       1 kg        Specialty Coffee Association
1337             1.0      5 lbs        Specialty Coffee Association
1338             1.0      5 lbs        Specialty Coffee Association

     Harvest.Year         Grading.Date  ... Total.Cup.Points  Moisture
\
0            2014      April 4th, 2015  ...            90.58      0.12

1            2014      April 4th, 2015  ...            89.92      0.12

2             NaN       May 31st, 2010  ...            89.75      0.00

3            2014     March 26th, 2015  ...            89.00      0.11

4            2014      April 4th, 2015  ...            88.83      0.12
```

```
...           ...                   ...  ...               ...       ...

1334         2016   January 19th, 2016  ...             78.75      0.00

1335         2016   January 19th, 2016  ...             78.08      0.00

1336         2014  December 23rd, 2014  ...             77.17      0.00

1337         2013    August 25th, 2014  ...             75.08      0.10

1338         2013    August 25th, 2014  ...             73.75      0.12


      Category.One.Defects  Quakers  Category.Two.Defects  \
0                        0      0.0                     0
1                        0      0.0                     1
2                        0      0.0                     0
3                        0      0.0                     2
4                        0      0.0                     2
...                    ...      ...                   ...
1334                     0      0.0                     1
1335                     0      0.0                     0
1336                     0      0.0                     6
1337                    20      0.0                     1
1338                    63      0.0                     9

               Expiration                  Certification.Body  \
0         April 3rd, 2016  METAD Agricultural Development plc
1         April 3rd, 2016  METAD Agricultural Development plc
2          May 31st, 2011        Specialty Coffee Association
3        March 25th, 2016  METAD Agricultural Development plc
4         April 3rd, 2016  METAD Agricultural Development plc
...                   ...                                 ...
1334   January 18th, 2017        Specialty Coffee Association
1335   January 18th, 2017        Specialty Coffee Association
1336  December 23rd, 2015        Specialty Coffee Association
1337    August 25th, 2015        Specialty Coffee Association
1338    August 25th, 2015        Specialty Coffee Association

                         Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
```

```
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
```

	Certification.Contact	unit_of_measurement
0	19fef5a731de2db57d16da10287413f5f99bc2dd	m
1	19fef5a731de2db57d16da10287413f5f99bc2dd	m
2	0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660	m
3	19fef5a731de2db57d16da10287413f5f99bc2dd	m
4	19fef5a731de2db57d16da10287413f5f99bc2dd	m
...	...	...
1334	352d0cf7f3e9be14dad7df644ad65efc27605ae2	m
1335	352d0cf7f3e9be14dad7df644ad65efc27605ae2	m
1336	352d0cf7f3e9be14dad7df644ad65efc27605ae2	m
1337	352d0cf7f3e9be14dad7df644ad65efc27605ae2	m
1338	352d0cf7f3e9be14dad7df644ad65efc27605ae2	m

```
[1339 rows x 31 columns]
```

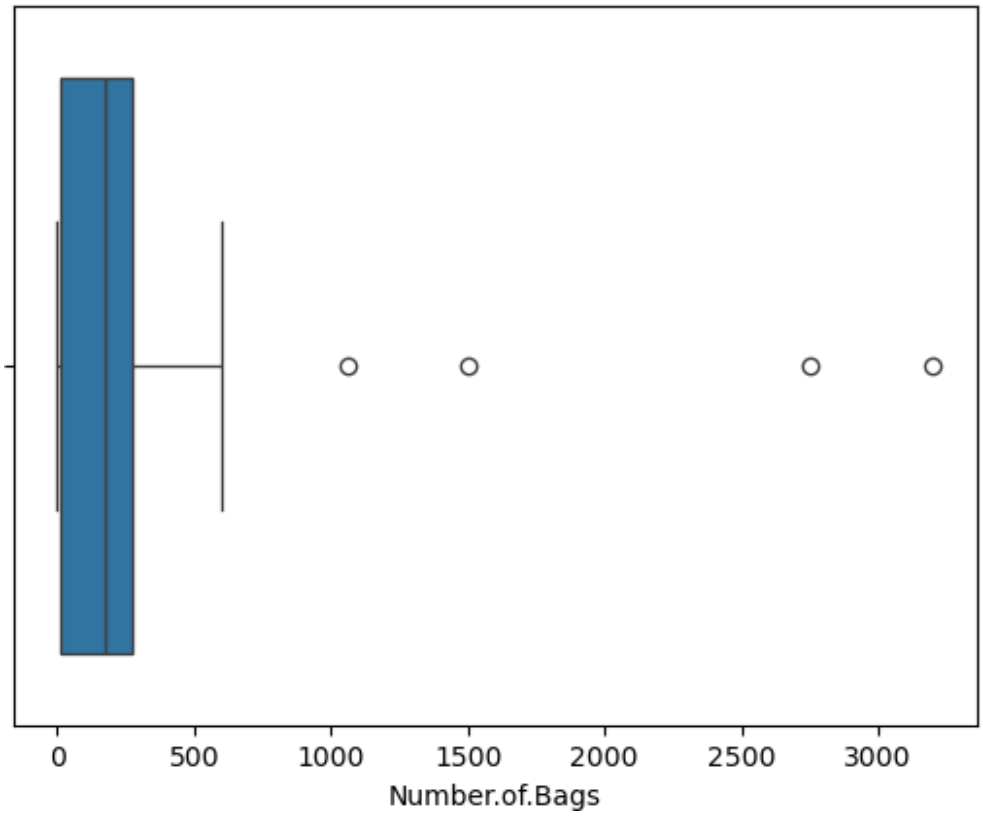
```
df.isnull().sum()
```

```
Species                     0
Owner                       7
Country.of.Origin           1
Region                     59
Producer                  232
Number.of.Bags              1
Bag.Weight                  0
In.Country.Partner          0
Harvest.Year               47
Grading.Date                0
Owner.1                     7
Aroma                       0
Flavor                      0
Aftertaste                  0
Acidity                     0
Body                        0
Balance                     0
Uniformity                  0
Clean.Cup                   0
Sweetness                   0
Cupper.Points               0
Total.Cup.Points            0
Moisture                    0
Category.One.Defects        0
Quakers                     1
Category.Two.Defects        0
Expiration                  0
Certification.Body          0
Certification.Address       0
Certification.Contact       0
```

```
unit_of_measurement         0
dtype: int64
```

```
sns.boxplot(x=df['Number.of.Bags'])
```

```
<Axes: xlabel='Number.of.Bags'>
```




```
df['Number.of.Bags'].sort_values(ascending=True).head()
```

```
704     0.0
1206    1.0
379     1.0
1188    1.0
444     1.0
Name: Number.of.Bags, dtype: float64
```

```
median=df['Number.of.Bags'].median()
median
```

```
175.0
```

```
df["Number.of.Bags"].fillna(median, inplace=True)
df
```

```
      Species                    Owner Country.of.Origin  \
0     Arabica                metad plc          Ethiopia
1     Arabica                metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                metad plc          Ethiopia
...       ...                      ...               ...
1334  Robusta              luis robles           Ecuador
1335  Robusta              luis robles           Ecuador
1336  Robusta              james moore     United States
1337  Robusta            cafe politico             India
1338  Robusta            cafe politico           Vietnam

                             Region
Producer  \
0                       guji-hambela                                  METAD
PLC
1                       guji-hambela                                  METAD
PLC
2                                NaN
NaN
3                           oromia Yidnekachew Dabessa Coffee
Plantation
4                       guji-hambela                                  METAD
PLC
...                              ...
...
1334               san juan, playas          Café Robusta del Ecuador
S.A.
1335               san juan, playas          Café Robusta del Ecuador
S.A.
1336  kwanza norte province, angola                            Cafe
Cazengo
1337                             NaN
NaN
1338                             NaN
NaN

      Number.of.Bags Bag.Weight                    In.Country.Partner  \
0              300.0      60 kg  METAD Agricultural Development plc
1              300.0      60 kg  METAD Agricultural Development plc
2                5.0          1        Specialty Coffee Association
3              320.0      60 kg  METAD Agricultural Development plc
4              300.0      60 kg  METAD Agricultural Development plc
...              ...        ...                                 ...
1334             1.0       2 kg        Specialty Coffee Association
1335             1.0       2 kg        Specialty Coffee Association
1336             1.0       1 kg        Specialty Coffee Association
1337             1.0      5 lbs        Specialty Coffee Association
1338             1.0      5 lbs        Specialty Coffee Association
```

	Harvest.Year	Grading.Date	...	Total.Cup.Points	Moisture
0	2014	April 4th, 2015	...	90.58	0.12
1	2014	April 4th, 2015	...	89.92	0.12
2	NaN	May 31st, 2010	...	89.75	0.00
3	2014	March 26th, 2015	...	89.00	0.11
4	2014	April 4th, 2015	...	88.83	0.12
...	...	...	...	...	...
1334	2016	January 19th, 2016	...	78.75	0.00
1335	2016	January 19th, 2016	...	78.08	0.00
1336	2014	December 23rd, 2014	...	77.17	0.00
1337	2013	August 25th, 2014	...	75.08	0.10
1338	2013	August 25th, 2014	...	73.75	0.12

	Category.One.Defects	Quakers	Category.Two.Defects
0	0	0.0	0
1	0	0.0	1
2	0	0.0	0
3	0	0.0	2
4	0	0.0	2
...	...	...	...
1334	0	0.0	1
1335	0	0.0	0
1336	0	0.0	6
1337	20	0.0	1
1338	63	0.0	9

	Expiration	Certification.Body
0	April 3rd, 2016	METAD Agricultural Development plc
1	April 3rd, 2016	METAD Agricultural Development plc
2	May 31st, 2011	Specialty Coffee Association
3	March 25th, 2016	METAD Agricultural Development plc
4	April 3rd, 2016	METAD Agricultural Development plc
...	...	...
1334	January 18th, 2017	Specialty Coffee Association
1335	January 18th, 2017	Specialty Coffee Association
1336	December 23rd, 2015	Specialty Coffee Association
1337	August 25th, 2015	Specialty Coffee Association
1338	August 25th, 2015	Specialty Coffee Association

```
                          Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720

                          Certification.Contact  unit_of_measurement
0     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                     m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
...                                       ...                   ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m

[1339 rows x 31 columns]
```

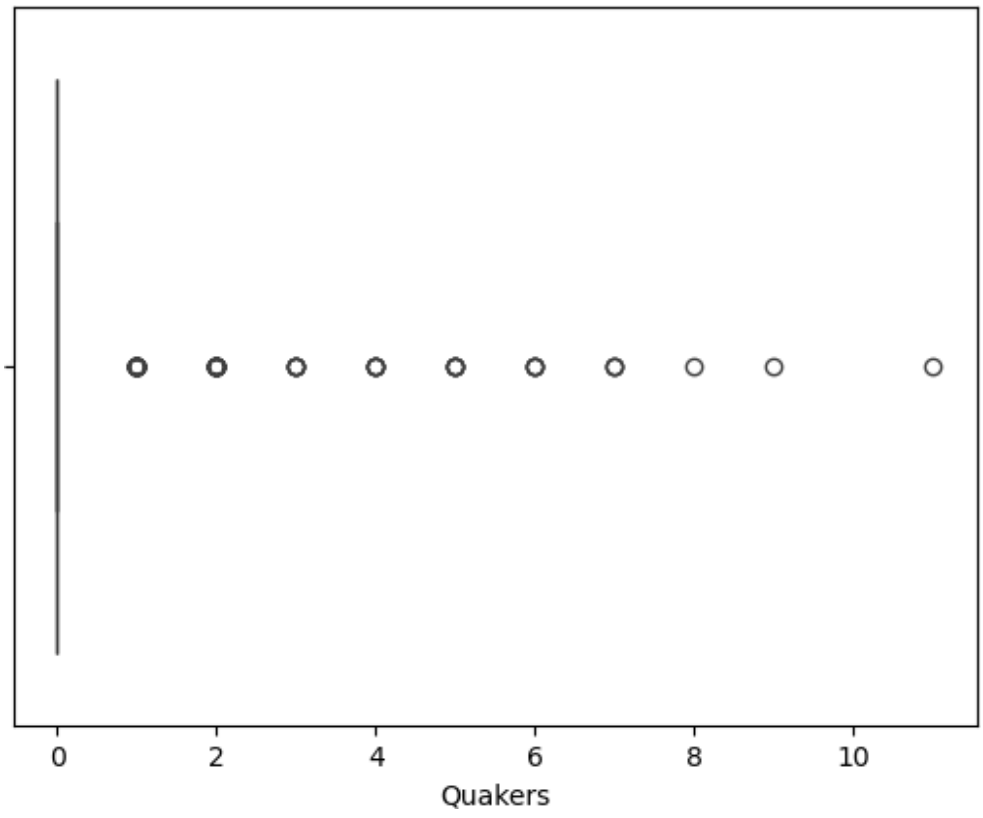
```
df.isnull().sum()
```

```
Species                    0
Owner                      7
Country.of.Origin          1
Region                    59
Producer                 232
Number.of.Bags             0
Bag.Weight                 0
In.Country.Partner         0
Harvest.Year              47
Grading.Date               0
Owner.1                    7
Aroma                      0
Flavor                     0
Aftertaste                 0
Acidity                    0
Body                       0
Balance                    0
Uniformity                 0
Clean.Cup                  0
```

```
Sweetness                    0
Cupper.Points                0
Total.Cup.Points             0
Moisture                     0
Category.One.Defects         0
Quakers                      1
Category.Two.Defects         0
Expiration                   0
Certification.Body           0
Certification.Address        0
Certification.Contact        0
unit_of_measurement          0
dtype: int64
```

```
 sns.boxplot(x=df['Quakers'])
```

```
<Axes: xlabel='Quakers'>
```




```
df['Quakers'].sort_values(ascending=True).head()
```

```
0      0.0
892    0.0
891    0.0
890    0.0
```

```
889    0.0
Name: Quakers, dtype: float64
```

```
mean=df['Quakers'].mean()
mean
```

```
0.17339312406576982
```

```
df["Quakers"].fillna(mean, inplace=True)
df
```

```
      Species                     Owner Country.of.Origin  \
0     Arabica                 metad plc          Ethiopia
1     Arabica                 metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                 metad plc          Ethiopia
...       ...                       ...               ...
1334  Robusta               luis robles           Ecuador
1335  Robusta               luis robles           Ecuador
1336  Robusta               james moore     United States
1337  Robusta             cafe politico             India
1338  Robusta             cafe politico           Vietnam

                              Region
Producer  \
0                       guji-hambela                                  METAD
PLC
1                       guji-hambela                                  METAD
PLC
2                                NaN
NaN
3                             oromia  Yidnekachew Dabessa Coffee
Plantation
4                       guji-hambela                                  METAD
PLC
...                              ...
...
1334               san juan, playas          Café Robusta del Ecuador
S.A.
1335               san juan, playas          Café Robusta del Ecuador
S.A.
1336  kwanza norte province, angola                           Cafe
Cazengo
1337                             NaN
NaN
1338                             NaN
NaN

      Number.of.Bags Bag.Weight                  In.Country.Partner  \
```

```
0              300.0      60 kg  METAD Agricultural Development plc
1              300.0      60 kg  METAD Agricultural Development plc
2                5.0          1        Specialty Coffee Association
3              320.0      60 kg  METAD Agricultural Development plc
4              300.0      60 kg  METAD Agricultural Development plc
...              ...        ...                                  ...
1334             1.0       2 kg        Specialty Coffee Association
1335             1.0       2 kg        Specialty Coffee Association
1336             1.0       1 kg        Specialty Coffee Association
1337             1.0      5 lbs        Specialty Coffee Association
1338             1.0      5 lbs        Specialty Coffee Association

     Harvest.Year          Grading.Date  ... Total.Cup.Points  Moisture
\
0            2014      April 4th, 2015  ...             90.58      0.12

1            2014      April 4th, 2015  ...             89.92      0.12

2             NaN       May 31st, 2010  ...             89.75      0.00

3            2014     March 26th, 2015  ...             89.00      0.11

4            2014      April 4th, 2015  ...             88.83      0.12

...           ...                  ...  ...               ...       ...

1334         2016   January 19th, 2016  ...             78.75      0.00

1335         2016   January 19th, 2016  ...             78.08      0.00

1336         2014  December 23rd, 2014  ...             77.17      0.00

1337         2013    August 25th, 2014  ...             75.08      0.10

1338         2013    August 25th, 2014  ...             73.75      0.12


      Category.One.Defects  Quakers  Category.Two.Defects  \
0                        0      0.0                     0
1                        0      0.0                     1
2                        0      0.0                     0
3                        0      0.0                     2
4                        0      0.0                     2
...                    ...      ...                   ...
1334                     0      0.0                     1
1335                     0      0.0                     0
1336                     0      0.0                     6
1337                    20      0.0                     1
1338                    63      0.0                     9

               Expiration                  Certification.Body  \
```

```
0          April 3rd, 2016  METAD Agricultural Development plc
1          April 3rd, 2016  METAD Agricultural Development plc
2           May 31st, 2011         Specialty Coffee Association
3         March 25th, 2016  METAD Agricultural Development plc
4          April 3rd, 2016  METAD Agricultural Development plc
...                    ...                                  ...
1334    January 18th, 2017         Specialty Coffee Association
1335    January 18th, 2017         Specialty Coffee Association
1336   December 23rd, 2015         Specialty Coffee Association
1337     August 25th, 2015         Specialty Coffee Association
1338     August 25th, 2015         Specialty Coffee Association

                          Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720

                          Certification.Contact  unit_of_measurement
0     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                    m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
...                                        ...                  ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m

[1339 rows x 31 columns]
```

```
df.isnull().sum()
```

```
Species                     0
Owner                       7
Country.of.Origin           1
Region                     59
Producer                  232
Number.of.Bags              0
Bag.Weight                  0
In.Country.Partner          0
```

```
Harvest.Year               47
Grading.Date                0
Owner.1                     7
Aroma                       0
Flavor                      0
Aftertaste                  0
Acidity                     0
Body                        0
Balance                     0
Uniformity                  0
Clean.Cup                   0
Sweetness                   0
Cupper.Points               0
Total.Cup.Points            0
Moisture                    0
Category.One.Defects        0
Quakers                     0
Category.Two.Defects        0
Expiration                  0
Certification.Body          0
Certification.Address       0
Certification.Contact       0
unit_of_measurement         0
dtype: int64
```

```
df['Owner'].sort_values(ascending=True).head()
```

```
899     acacia hills ltd
1212      adam ciruli ye
1082      adam ciruli ye
623           adam kline
673           adam kline
Name: Owner, dtype: object
```

```
mode=df['Owner'].mode()
mode
```

```
0    juan luis alvarado romero
Name: Owner, dtype: object
```

```
df["Owner"].fillna(mode[0], inplace=True)
df
```

```
      Species                     Owner Country.of.Origin  \
0     Arabica                 metad plc          Ethiopia
1     Arabica                 metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                 metad plc          Ethiopia
...       ...                       ...               ...
1334  Robusta              luis robles           Ecuador
```

```
1335  Robusta               luis robles           Ecuador
1336  Robusta               james moore     United States
1337  Robusta             cafe politico             India
1338  Robusta             cafe politico           Vietnam

                              Region
Producer  \
0                       guji-hambela                                  METAD
PLC
1                       guji-hambela                                  METAD
PLC
2                                NaN
NaN
3                            oromia Yidnekachew Dabessa Coffee
Plantation
4                       guji-hambela                                  METAD
PLC
...                              ...
...
1334               san juan, playas          Café Robusta del Ecuador
S.A.
1335               san juan, playas          Café Robusta del Ecuador
S.A.
1336  kwanza norte province, angola                            Cafe
Cazengo
1337                             NaN
NaN
1338                             NaN
NaN

      Number.of.Bags Bag.Weight                  In.Country.Partner  \
0              300.0      60 kg  METAD Agricultural Development plc
1              300.0      60 kg  METAD Agricultural Development plc
2                5.0          1        Specialty Coffee Association
3              320.0      60 kg  METAD Agricultural Development plc
4              300.0      60 kg  METAD Agricultural Development plc
...              ...        ...                                 ...
1334             1.0       2 kg        Specialty Coffee Association
1335             1.0       2 kg        Specialty Coffee Association
1336             1.0       1 kg        Specialty Coffee Association
1337             1.0      5 lbs        Specialty Coffee Association
1338             1.0      5 lbs        Specialty Coffee Association

     Harvest.Year         Grading.Date  ... Total.Cup.Points  Moisture
\
0            2014      April 4th, 2015  ...            90.58      0.12

1            2014      April 4th, 2015  ...            89.92      0.12

2             NaN       May 31st, 2010  ...            89.75      0.00
```

```
3            2014     March 26th, 2015  ...             89.00       0.11
4            2014      April 4th, 2015  ...             88.83       0.12
...           ...                  ...  ...               ...        ...
1334         2016   January 19th, 2016  ...             78.75       0.00
1335         2016   January 19th, 2016  ...             78.08       0.00
1336         2014  December 23rd, 2014  ...             77.17       0.00
1337         2013    August 25th, 2014  ...             75.08       0.10
1338         2013    August 25th, 2014  ...             73.75       0.12

      Category.One.Defects  Quakers  Category.Two.Defects  \
0                        0      0.0                     0
1                        0      0.0                     1
2                        0      0.0                     0
3                        0      0.0                     2
4                        0      0.0                     2
...                    ...      ...                   ...
1334                     0      0.0                     1
1335                     0      0.0                     0
1336                     0      0.0                     6
1337                    20      0.0                     1
1338                    63      0.0                     9

               Expiration                  Certification.Body  \
0         April 3rd, 2016  METAD Agricultural Development plc
1         April 3rd, 2016  METAD Agricultural Development plc
2          May 31st, 2011        Specialty Coffee Association
3        March 25th, 2016  METAD Agricultural Development plc
4         April 3rd, 2016  METAD Agricultural Development plc
...                   ...                                 ...
1334   January 18th, 2017        Specialty Coffee Association
1335   January 18th, 2017        Specialty Coffee Association
1336  December 23rd, 2015        Specialty Coffee Association
1337    August 25th, 2015        Specialty Coffee Association
1338    August 25th, 2015        Specialty Coffee Association

                         Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
```

```
...                                                 ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
```

```
                          Certification.Contact  unit_of_measurement
0     19fef5a731de2db57d16da10287413f5f99bc2dd                      m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                      m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                      m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                      m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                      m
...                                                 ...                    ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                      m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                      m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                      m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                      m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                      m

[1339 rows x 31 columns]
```

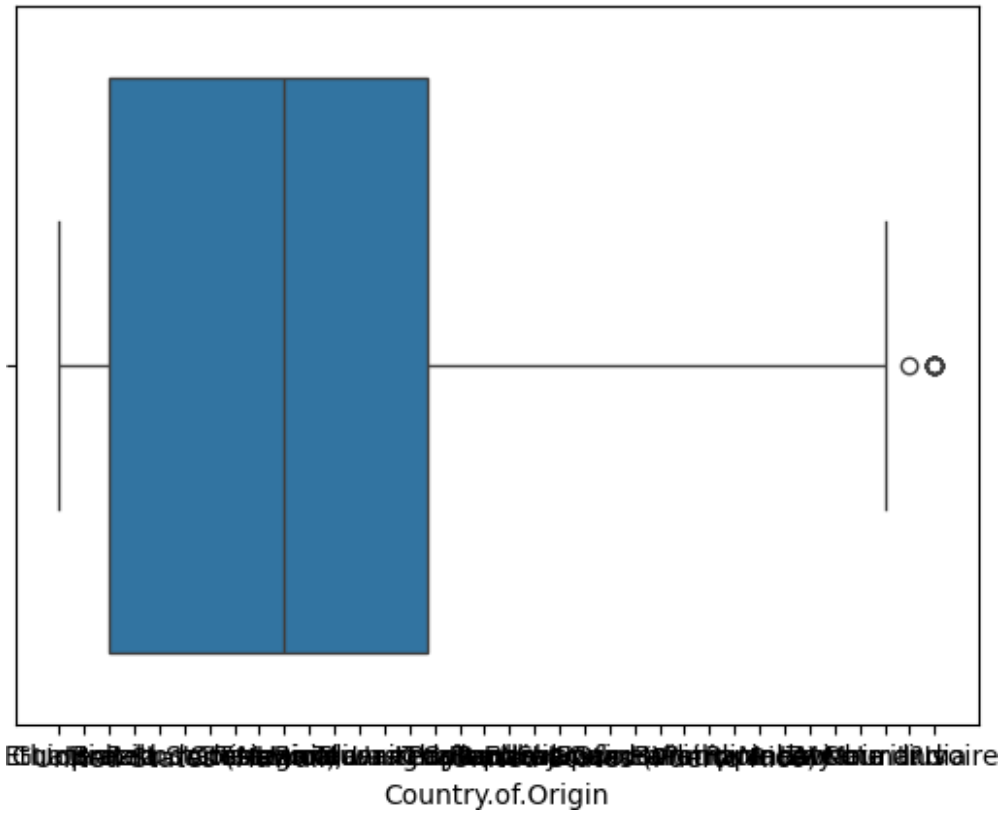
```
df.isnull().sum()
```

```
Species                    0
Owner                      0
Country.of.Origin          1
Region                    59
Producer                 232
Number.of.Bags             0
Bag.Weight                 0
In.Country.Partner         0
Harvest.Year              47
Grading.Date               0
Owner.1                    7
Aroma                      0
Flavor                     0
Aftertaste                 0
Acidity                    0
Body                       0
Balance                    0
Uniformity                 0
Clean.Cup                  0
Sweetness                  0
Cupper.Points              0
Total.Cup.Points           0
Moisture                   0
Category.One.Defects       0
Quakers                    0
Category.Two.Defects       0
```

```
Expiration                  0
Certification.Body          0
Certification.Address       0
Certification.Contact       0
unit_of_measurement         0
dtype: int64
```

```
sns.boxplot(x=df['Country.of.Origin'])
```

```
<Axes: xlabel='Country.of.Origin'>
```




```
df['Country.of.Origin'].sort_values(ascending=True).head()
```

```
1075    Brazil
730     Brazil
483     Brazil
987     Brazil
484     Brazil
Name: Country.of.Origin, dtype: object
```

```
mode=df['Country.of.Origin'].mode()
mode
```

```
0    Mexico
Name: Country.of.Origin, dtype: object
```

```
df["Country.of.Origin"].fillna(mode[0], inplace=True)
df
```

```
      Species                     Owner Country.of.Origin  \
0     Arabica                 metad plc          Ethiopia
1     Arabica                 metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                 metad plc          Ethiopia
...       ...                       ...               ...
1334  Robusta               luis robles           Ecuador
1335  Robusta               luis robles           Ecuador
1336  Robusta               james moore     United States
1337  Robusta             cafe politico             India
1338  Robusta             cafe politico           Vietnam

                             Region
Producer  \
0                       guji-hambela                                  METAD
PLC
1                       guji-hambela                                  METAD
PLC
2                                NaN
NaN
3                            oromia Yidnekachew Dabessa Coffee
Plantation
4                       guji-hambela                                  METAD
PLC
...                              ...
...
1334                san juan, playas          Café Robusta del Ecuador
S.A.
1335                san juan, playas          Café Robusta del Ecuador
S.A.
1336  kwanza norte province, angola                            Cafe
Cazengo
1337                             NaN
NaN
1338                             NaN
NaN

      Number.of.Bags Bag.Weight                  In.Country.Partner  \
0              300.0      60 kg  METAD Agricultural Development plc
1              300.0      60 kg  METAD Agricultural Development plc
2                5.0          1        Specialty Coffee Association
3              320.0      60 kg  METAD Agricultural Development plc
4              300.0      60 kg  METAD Agricultural Development plc
...              ...        ...                                 ...
1334             1.0       2 kg        Specialty Coffee Association
1335             1.0       2 kg        Specialty Coffee Association
```

```
1336             1.0       1 kg        Specialty Coffee Association
1337             1.0      5 lbs        Specialty Coffee Association
1338             1.0      5 lbs        Specialty Coffee Association

     Harvest.Year          Grading.Date  ... Total.Cup.Points  Moisture
\
0            2014      April 4th, 2015  ...            90.58      0.12

1            2014      April 4th, 2015  ...            89.92      0.12

2             NaN       May 31st, 2010  ...            89.75      0.00

3            2014     March 26th, 2015  ...            89.00      0.11

4            2014      April 4th, 2015  ...            88.83      0.12

...           ...                  ...  ...              ...       ...

1334         2016   January 19th, 2016  ...            78.75      0.00

1335         2016   January 19th, 2016  ...            78.08      0.00

1336         2014  December 23rd, 2014  ...            77.17      0.00

1337         2013    August 25th, 2014  ...            75.08      0.10

1338         2013    August 25th, 2014  ...            73.75      0.12


      Category.One.Defects  Quakers  Category.Two.Defects  \
0                        0      0.0                     0
1                        0      0.0                     1
2                        0      0.0                     0
3                        0      0.0                     2
4                        0      0.0                     2
...                    ...      ...                   ...
1334                     0      0.0                     1
1335                     0      0.0                     0
1336                     0      0.0                     6
1337                    20      0.0                     1
1338                    63      0.0                     9

               Expiration                  Certification.Body  \
0         April 3rd, 2016  METAD Agricultural Development plc
1         April 3rd, 2016  METAD Agricultural Development plc
2          May 31st, 2011        Specialty Coffee Association
3        March 25th, 2016  METAD Agricultural Development plc
4         April 3rd, 2016  METAD Agricultural Development plc
...                   ...                                 ...
1334   January 18th, 2017        Specialty Coffee Association
1335   January 18th, 2017        Specialty Coffee Association
```

```
1336   December 23rd, 2015       Specialty Coffee Association
1337      August 25th, 2015      Specialty Coffee Association
1338   August 25th, 2015  Specialty Coffee Association
```

```
                          Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
```

```
                          Certification.Contact  unit_of_measurement
0     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                     m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
...                                        ...                   ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
```

```
[1339 rows x 31 columns]
```

```
df.isnull().sum()
```

```
Species                     0
Owner                       0
Country.of.Origin           0
Region                     59
Producer                  232
Number.of.Bags              0
Bag.Weight                  0
In.Country.Partner          0
Harvest.Year               47
Grading.Date                0
Owner.1                     7
Aroma                       0
Flavor                      0
Aftertaste                  0
Acidity                     0
Body                        0
```

```
Balance                     0
Uniformity                  0
Clean.Cup                   0
Sweetness                   0
Cupper.Points               0
Total.Cup.Points            0
Moisture                    0
Category.One.Defects        0
Quakers                     0
Category.Two.Defects        0
Expiration                  0
Certification.Body          0
Certification.Address       0
Certification.Contact       0
unit_of_measurement         0
dtype: int64
```

```
df['Region'].sort_values(ascending=True).head()
```

```
628    52 narino (exact location: mattituy; municipal...
57                                             acatenango
129                                            acatenango
139                                            acatenango
638                                                  aceh
Name: Region, dtype: object
```

```
mode=df['Region'].mode()
mode
```

```
0    huila
Name: Region, dtype: object
```

```
df["Region"].fillna(mode[0], inplace=True)
df
```

```
      Species                     Owner Country.of.Origin  \
0     Arabica                 metad plc          Ethiopia
1     Arabica                 metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                 metad plc          Ethiopia
...       ...                       ...               ...
1334  Robusta               luis robles           Ecuador
1335  Robusta               luis robles           Ecuador
1336  Robusta               james moore     United States
1337  Robusta             cafe politico             India
1338  Robusta             cafe politico           Vietnam

                            Region
Producer  \
0                      guji-hambela                                  METAD
```

```
PLC
1                        guji-hambela                                    METAD
PLC
2                              huila
NaN
3                             oromia  Yidnekachew Dabessa Coffee
Plantation
4                        guji-hambela                                    METAD
PLC
...                               ...
...
1334               san juan, playas          Café Robusta del Ecuador
S.A.
1335               san juan, playas          Café Robusta del Ecuador
S.A.
1336  kwanza norte province, angola                           Cafe
Cazengo
1337                           huila
NaN
1338                           huila
NaN
```

```
      Number.of.Bags Bag.Weight                   In.Country.Partner  \
0              300.0      60 kg  METAD Agricultural Development plc
1              300.0      60 kg  METAD Agricultural Development plc
2                5.0          1        Specialty Coffee Association
3              320.0      60 kg  METAD Agricultural Development plc
4              300.0      60 kg  METAD Agricultural Development plc
...              ...        ...                                 ...
1334             1.0       2 kg        Specialty Coffee Association
1335             1.0       2 kg        Specialty Coffee Association
1336             1.0       1 kg        Specialty Coffee Association
1337             1.0      5 lbs        Specialty Coffee Association
1338             1.0      5 lbs        Specialty Coffee Association
```

```
     Harvest.Year         Grading.Date  ... Total.Cup.Points  Moisture
\
0            2014      April 4th, 2015  ...            90.58      0.12

1            2014      April 4th, 2015  ...            89.92      0.12

2             NaN       May 31st, 2010  ...            89.75      0.00

3            2014     March 26th, 2015  ...            89.00      0.11

4            2014      April 4th, 2015  ...            88.83      0.12

...           ...                  ...  ...              ...       ...

1334         2016   January 19th, 2016  ...            78.75      0.00
```

```
1335         2016   January 19th, 2016  ...              78.08       0.00

1336         2014  December 23rd, 2014  ...              77.17       0.00

1337         2013    August 25th, 2014  ...              75.08       0.10

1338         2013    August 25th, 2014  ...              73.75       0.12


      Category.One.Defects  Quakers  Category.Two.Defects  \
0                        0      0.0                     0
1                        0      0.0                     1
2                        0      0.0                     0
3                        0      0.0                     2
4                        0      0.0                     2
...                    ...      ...                   ...
1334                     0      0.0                     1
1335                     0      0.0                     0
1336                     0      0.0                     6
1337                    20      0.0                     1
1338                    63      0.0                     9

               Expiration                   Certification.Body  \
0         April 3rd, 2016  METAD Agricultural Development plc
1         April 3rd, 2016  METAD Agricultural Development plc
2          May 31st, 2011         Specialty Coffee Association
3        March 25th, 2016  METAD Agricultural Development plc
4         April 3rd, 2016  METAD Agricultural Development plc
...                   ...                                 ...
1334   January 18th, 2017         Specialty Coffee Association
1335   January 18th, 2017         Specialty Coffee Association
1336  December 23rd, 2015         Specialty Coffee Association
1337    August 25th, 2015         Specialty Coffee Association
1338    August 25th, 2015         Specialty Coffee Association

                         Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720

                         Certification.Contact  unit_of_measurement
```

```
0     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                    m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
...                                                  ...        ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m

[1339 rows x 31 columns]
```

```
df.isnull().sum()
```

```
Species                     0
Owner                       0
Country.of.Origin           0
Region                      0
Producer                  232
Number.of.Bags              0
Bag.Weight                  0
In.Country.Partner          0
Harvest.Year               47
Grading.Date                0
Owner.1                     7
Aroma                       0
Flavor                      0
Aftertaste                  0
Acidity                     0
Body                        0
Balance                     0
Uniformity                  0
Clean.Cup                   0
Sweetness                   0
Cupper.Points               0
Total.Cup.Points            0
Moisture                    0
Category.One.Defects        0
Quakers                     0
Category.Two.Defects        0
Expiration                  0
Certification.Body          0
Certification.Address       0
Certification.Contact       0
unit_of_measurement         0
dtype: int64
```

```
df['Producer'].sort_values(ascending=True).head()
```

```
624                                -
537                                -
255                                -
1269                               1
1061    ADONIS JOSE ORTEZ BELTRAN
Name: Producer, dtype: object

mode=df['Producer'].mode()
mode

0    La Plata
Name: Producer, dtype: object

df["Producer"].fillna(mode[0], inplace=True)
df

      Species                      Owner Country.of.Origin  \
0     Arabica                  metad plc          Ethiopia
1     Arabica                  metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                  metad plc          Ethiopia
...       ...                        ...               ...
1334  Robusta                luis robles           Ecuador
1335  Robusta                luis robles           Ecuador
1336  Robusta                james moore     United States
1337  Robusta              cafe politico             India
1338  Robusta              cafe politico           Vietnam

                              Region
Producer  \
0                       guji-hambela                                    METAD
PLC
1                       guji-hambela                                    METAD
PLC
2                              huila                                     La
Plata
3                             oromia Yidnekachew Dabessa Coffee
Plantation
4                       guji-hambela                                    METAD
PLC
...                                ...
...
1334                san juan, playas          Café Robusta del Ecuador
S.A.
1335                san juan, playas          Café Robusta del Ecuador
S.A.
1336  kwanza norte province, angola                            Cafe
Cazengo
1337                           huila                                     La
```

```
Plata
1338                          huila                                  La
Plata

      Number.of.Bags Bag.Weight                   In.Country.Partner  \
0              300.0      60 kg  METAD Agricultural Development plc
1              300.0      60 kg  METAD Agricultural Development plc
2                5.0          1         Specialty Coffee Association
3              320.0      60 kg  METAD Agricultural Development plc
4              300.0      60 kg  METAD Agricultural Development plc
...              ...        ...                                  ...
1334             1.0       2 kg         Specialty Coffee Association
1335             1.0       2 kg         Specialty Coffee Association
1336             1.0       1 kg         Specialty Coffee Association
1337             1.0      5 lbs         Specialty Coffee Association
1338             1.0      5 lbs         Specialty Coffee Association

     Harvest.Year          Grading.Date  ... Total.Cup.Points  Moisture
\
0            2014      April 4th, 2015  ...             90.58      0.12

1            2014      April 4th, 2015  ...             89.92      0.12

2             NaN       May 31st, 2010  ...             89.75      0.00

3            2014     March 26th, 2015  ...             89.00      0.11

4            2014      April 4th, 2015  ...             88.83      0.12

...           ...                   ...  ...               ...       ...

1334         2016   January 19th, 2016  ...             78.75      0.00

1335         2016   January 19th, 2016  ...             78.08      0.00

1336         2014  December 23rd, 2014  ...             77.17      0.00

1337         2013    August 25th, 2014  ...             75.08      0.10

1338         2013    August 25th, 2014  ...             73.75      0.12


      Category.One.Defects  Quakers  Category.Two.Defects  \
0                        0      0.0                     0
1                        0      0.0                     1
2                        0      0.0                     0
3                        0      0.0                     2
4                        0      0.0                     2
...                    ...      ...                   ...
1334                     0      0.0                     1
1335                     0      0.0                     0
```

```
1336                    0      0.0                     6
1337                   20      0.0                     1
1338                   63      0.0                     9

                Expiration                  Certification.Body  \
0         April 3rd, 2016  METAD Agricultural Development plc
1         April 3rd, 2016  METAD Agricultural Development plc
2          May 31st, 2011        Specialty Coffee Association
3        March 25th, 2016  METAD Agricultural Development plc
4         April 3rd, 2016  METAD Agricultural Development plc
...                   ...                                 ...
1334   January 18th, 2017        Specialty Coffee Association
1335   January 18th, 2017        Specialty Coffee Association
1336  December 23rd, 2015        Specialty Coffee Association
1337    August 25th, 2015        Specialty Coffee Association
1338    August 25th, 2015        Specialty Coffee Association

                         Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720

                         Certification.Contact  unit_of_measurement
0     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                    m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
...                                        ...                  ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m

[1339 rows x 31 columns]
```

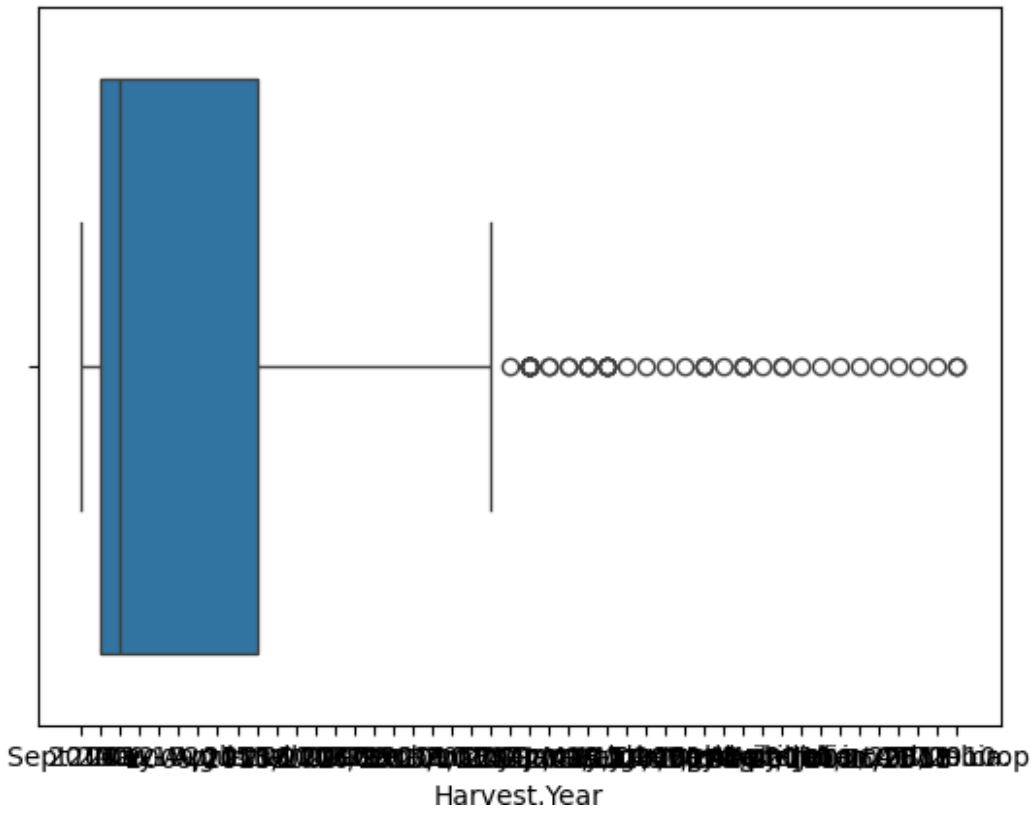
```
df.isnull().sum()
```

```
Species                   0
Owner                     0
Country.of.Origin         0
```

```
Region                       0
Producer                     0
Number.of.Bags               0
Bag.Weight                   0
In.Country.Partner           0
Harvest.Year                47
Grading.Date                 0
Owner.1                      7
Aroma                        0
Flavor                       0
Aftertaste                   0
Acidity                      0
Body                         0
Balance                      0
Uniformity                   0
Clean.Cup                    0
Sweetness                    0
Cupper.Points                0
Total.Cup.Points             0
Moisture                     0
Category.One.Defects         0
Quakers                      0
Category.Two.Defects         0
Expiration                   0
Certification.Body           0
Certification.Address        0
Certification.Contact        0
unit_of_measurement          0
dtype: int64
```

```
sns.boxplot(x=df['Harvest.Year'])
```

```
<Axes: xlabel='Harvest.Year'>
```




```
df['Harvest.Year'].sort_values(ascending=True).head()
```

```
1261      08/09 crop
1191      08/09 crop
437          1T/2011
857          1t/2011
124      2009 - 2010
Name: Harvest.Year, dtype: object
```

```
mode=df['Harvest.Year'].mode()
mode
```

```
0    2012
Name: Harvest.Year, dtype: object
```

```
df["Harvest.Year"].fillna(mode[0], inplace=True)
df
```

```
      Species                    Owner Country.of.Origin  \
0     Arabica                metad plc          Ethiopia
1     Arabica                metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                metad plc          Ethiopia
...       ...                      ...               ...
```

```
1334  Robusta               luis robles           Ecuador
1335  Robusta               luis robles           Ecuador
1336  Robusta               james moore     United States
1337  Robusta             cafe politico             India
1338  Robusta             cafe politico           Vietnam

                              Region
Producer  \
0                       guji-hambela                                    METAD
PLC
1                       guji-hambela                                    METAD
PLC
2                              huila                                     La
Plata
3                             oromia Yidnekachew Dabessa Coffee
Plantation
4                       guji-hambela                                    METAD
PLC
...                               ...
...
1334               san juan, playas          Café Robusta del Ecuador
S.A.
1335               san juan, playas          Café Robusta del Ecuador
S.A.
1336  kwanza norte province, angola                            Cafe
Cazengo
1337                           huila                                     La
Plata
1338                           huila                                     La
Plata

      Number.of.Bags Bag.Weight                   In.Country.Partner  \
0              300.0      60 kg  METAD Agricultural Development plc
1              300.0      60 kg  METAD Agricultural Development plc
2                5.0          1          Specialty Coffee Association
3              320.0      60 kg  METAD Agricultural Development plc
4              300.0      60 kg  METAD Agricultural Development plc
...              ...        ...                                  ...
1334             1.0       2 kg          Specialty Coffee Association
1335             1.0       2 kg          Specialty Coffee Association
1336             1.0       1 kg          Specialty Coffee Association
1337             1.0      5 lbs          Specialty Coffee Association
1338             1.0      5 lbs          Specialty Coffee Association

     Harvest.Year         Grading.Date  ... Total.Cup.Points  Moisture
\
0            2014      April 4th, 2015  ...            90.58      0.12

1            2014      April 4th, 2015  ...            89.92      0.12
```

```
2            2012       May 31st, 2010  ...              89.75       0.00
3            2014     March 26th, 2015  ...              89.00       0.11
4            2014      April 4th, 2015  ...              88.83       0.12
...           ...                  ...  ...                ...        ...
1334         2016   January 19th, 2016  ...              78.75       0.00
1335         2016   January 19th, 2016  ...              78.08       0.00
1336         2014  December 23rd, 2014  ...              77.17       0.00
1337         2013    August 25th, 2014  ...              75.08       0.10
1338         2013    August 25th, 2014  ...              73.75       0.12
```

```
      Category.One.Defects  Quakers  Category.Two.Defects  \
0                        0      0.0                     0
1                        0      0.0                     1
2                        0      0.0                     0
3                        0      0.0                     2
4                        0      0.0                     2
...                    ...      ...                   ...
1334                     0      0.0                     1
1335                     0      0.0                     0
1336                     0      0.0                     6
1337                    20      0.0                     1
1338                    63      0.0                     9
```

```
               Expiration                  Certification.Body  \
0         April 3rd, 2016  METAD Agricultural Development plc
1         April 3rd, 2016  METAD Agricultural Development plc
2          May 31st, 2011        Specialty Coffee Association
3        March 25th, 2016  METAD Agricultural Development plc
4         April 3rd, 2016  METAD Agricultural Development plc
...                   ...                                 ...
1334   January 18th, 2017        Specialty Coffee Association
1335   January 18th, 2017        Specialty Coffee Association
1336  December 23rd, 2015        Specialty Coffee Association
1337    August 25th, 2015        Specialty Coffee Association
1338    August 25th, 2015        Specialty Coffee Association
```

```
                        Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
```

```
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
```

```
                          Certification.Contact  unit_of_measurement
0     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                     m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                     m
...                                        ...                   ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                     m
```

```
[1339 rows x 31 columns]
```

```
df.isnull().sum()
```

```
Species                   0
Owner                     0
Country.of.Origin         0
Region                    0
Producer                  0
Number.of.Bags            0
Bag.Weight                0
In.Country.Partner        0
Harvest.Year              0
Grading.Date              0
Owner.1                   7
Aroma                     0
Flavor                    0
Aftertaste                0
Acidity                   0
Body                      0
Balance                   0
Uniformity                0
Clean.Cup                 0
Sweetness                 0
Cupper.Points             0
Total.Cup.Points          0
Moisture                  0
Category.One.Defects      0
Quakers                   0
```

```
Category.Two.Defects      0
Expiration                0
Certification.Body        0
Certification.Address     0
Certification.Contact     0
unit_of_measurement       0
dtype: int64
```

```
df['Owner.1'].sort_values(ascending=True).head()
```

```
1176    ADRIANA TORRES RICO QUEVEDO
402                           AFCA
49                            AFCA
593                           AFCA
939                           AFCA
Name: Owner.1, dtype: object
```

```
mode=df['Owner.1'].mode()
mode
```

```
0    Juan Luis Alvarado Romero
Name: Owner.1, dtype: object
```

```
df["Owner.1"].fillna(mode[0], inplace=True)
df
```

```
      Species                     Owner Country.of.Origin  \
0     Arabica                 metad plc          Ethiopia
1     Arabica                 metad plc          Ethiopia
2     Arabica  grounds for health admin         Guatemala
3     Arabica       yidnekachew dabessa          Ethiopia
4     Arabica                 metad plc          Ethiopia
...       ...                       ...               ...
1334  Robusta               luis robles           Ecuador
1335  Robusta               luis robles           Ecuador
1336  Robusta               james moore     United States
1337  Robusta             cafe politico             India
1338  Robusta             cafe politico           Vietnam
```

```
...                                ...
...
1334                san juan, playas          Café Robusta del Ecuador
S.A.
1335                san juan, playas          Café Robusta del Ecuador
S.A.
1336  kwanza norte province, angola                            Cafe
Cazengo
1337                            huila                                La
Plata
1338                            huila                                La
Plata

      Number.of.Bags  Bag.Weight                    In.Country.Partner
0              300.0      60 kg   IETAD Agricultural Development plc
1              300.0      60 kg   IETAD Agricultural Development plc
2                5.0          1         Specialty Coffee Association
3              320.0      60 kg   IETAD Agricultural Development plc
4              300.0      60 kg   IETAD Agricultural Development plc
...              ...        ...                                   ...
1334             1.0       2 kg         Specialty Coffee Association
1335             1.0       2 kg         Specialty Coffee Association
1336             1.0       1 kg         Specialty Coffee Association
1337             1.0      5 lbs         Specialty Coffee Association
1338             1.0      5 lbs         Specialty Coffee Association

      Harvest.Year          Grading.Date  ... Total.Cup.Points  Moisture
\
0             2014      April 4th, 2015  ...             90.58      0.12

1             2014      April 4th, 2015  ...             89.92      0.12

2             2012       May 31st, 2010  ...             89.75      0.00

3             2014     March 26th, 2015  ...             89.00      0.11

4             2014      April 4th, 2015  ...             88.83      0.12

...            ...                  ...  ...               ...       ...

1334          2016   January 19th, 2016  ...             78.75      0.00

1335          2016   January 19th, 2016  ...             78.08      0.00

1336          2014  December 23rd, 2014  ...             77.17      0.00

1337          2013    August 25th, 2014  ...             75.08      0.10

1338          2013    August 25th, 2014  ...             73.75      0.12
```

```
      Category.One.Defects  Quakers  Category.Two.Defects  \
0                        0      0.0                     0
1                        0      0.0                     1
2                        0      0.0                     0
3                        0      0.0                     2
4                        0      0.0                     2
...                    ...      ...                   ...
1334                     0      0.0                     1
1335                     0      0.0                     0
1336                     0      0.0                     6
1337                    20      0.0                     1
1338                    63      0.0                     9

               Expiration                   Certification.Body  \
0         April 3rd, 2016  METAD Agricultural Development plc
1         April 3rd, 2016  METAD Agricultural Development plc
2          May 31st, 2011         Specialty Coffee Association
3        March 25th, 2016  METAD Agricultural Development plc
4         April 3rd, 2016  METAD Agricultural Development plc
...                   ...                                 ...
1334   January 18th, 2017         Specialty Coffee Association
1335   January 18th, 2017         Specialty Coffee Association
1336  December 23rd, 2015         Specialty Coffee Association
1337    August 25th, 2015         Specialty Coffee Association
1338    August 25th, 2015         Specialty Coffee Association

                         Certification.Address  \
0     309fcf77415a3661ae83e027f7e5f05dad786e44
1     309fcf77415a3661ae83e027f7e5f05dad786e44
2     36d0d00a3724338ba7937c52a378d085f2172daa
3     309fcf77415a3661ae83e027f7e5f05dad786e44
4     309fcf77415a3661ae83e027f7e5f05dad786e44
...                                        ...
1334  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337  ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1338  ff7c18ad303d4b603ac3f8cff7e611ffc735e720

                         Certification.Contact  unit_of_measurement
0     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
1     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
2     0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                    m
3     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
4     19fef5a731de2db57d16da10287413f5f99bc2dd                    m
...                                        ...                  ...
1334  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1335  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1336  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
1337  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m
```

```
1338  352d0cf7f3e9be14dad7df644ad65efc27605ae2                    m

[1339 rows x 31 columns]
```

```
df.isnull().sum()
```

```
Species                   0
Owner                     0
Country.of.Origin         0
Region                    0
Producer                  0
Number.of.Bags            0
Bag.Weight                0
In.Country.Partner        0
Harvest.Year              0
Grading.Date              0
Owner.1                   0
Aroma                     0
Flavor                    0
Aftertaste                0
Acidity                   0
Body                      0
Balance                   0
Uniformity                0
Clean.Cup                 0
Sweetness                 0
Cupper.Points             0
Total.Cup.Points          0
Moisture                  0
Category.One.Defects      0
Quakers                   0
Category.Two.Defects      0
Expiration                0
Certification.Body        0
Certification.Address     0
Certification.Contact     0
unit_of_measurement       0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1339 entries, 0 to 1338
Data columns (total 31 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Species                 1339 non-null   object
 1   Owner                   1339 non-null   object
 2   Country.of.Origin       1339 non-null   object
 3   Region                  1339 non-null   object
```

```
 4   Producer                1339 non-null   object
 5   Number.of.Bags          1339 non-null   float64
 6   Bag.Weight              1339 non-null   object
 7   In.Country.Partner      1339 non-null   object
 8   Harvest.Year            1339 non-null   object
 9   Grading.Date            1339 non-null   object
 10  Owner.1                 1339 non-null   object
 11  Aroma                   1339 non-null   float64
 12  Flavor                  1339 non-null   float64
 13  Aftertaste              1339 non-null   float64
 14  Acidity                 1339 non-null   float64
 15  Body                    1339 non-null   float64
 16  Balance                 1339 non-null   float64
 17  Uniformity              1339 non-null   float64
 18  Clean.Cup               1339 non-null   float64
 19  Sweetness               1339 non-null   float64
 20  Cupper.Points           1339 non-null   float64
 21  Total.Cup.Points        1339 non-null   float64
 22  Moisture                1339 non-null   float64
 23  Category.One.Defects    1339 non-null   int64
 24  Quakers                 1339 non-null   float64
 25  Category.Two.Defects    1339 non-null   int64
 26  Expiration              1339 non-null   object
 27  Certification.Body      1339 non-null   object
 28  Certification.Address   1339 non-null   object
 29  Certification.Contact   1339 non-null   object
 30  unit_of_measurement     1339 non-null   object
dtypes: float64(14), int64(2), object(15)
memory usage: 324.4+ KB
```

```
df.shape
```

```
(1339, 31)
```

```
                                                #Outlier Dectation
```

```
new_df=df.select_dtypes(exclude=['object'])
```

```
new_df
```

```
      Number.of.Bags  Aroma  Flavor  Aftertaste  Acidity  Body
Balance  \
0              300.0   8.67    8.83        8.67     8.75  8.50
8.42
1              300.0   8.75    8.67        8.50     8.58  8.42
8.42
2                5.0   8.42    8.50        8.42     8.42  8.33
8.42
3              320.0   8.17    8.58        8.42     8.42  8.50
8.25
4              300.0   8.25    8.50        8.25     8.50  8.42
```

```
8.33
...                ...    ...     ...          ...      ...   ...       ..
.
1334               1.0   7.75    7.58         7.33     7.58  5.08
7.83
1335               1.0   7.50    7.67         7.75     7.75  5.17
5.25
1336               1.0   7.33    7.33         7.17     7.42  7.50
7.17
1337               1.0   7.42    6.83         6.75     7.17  7.25
7.00
1338               1.0   6.75    6.67         6.50     6.83  6.92
6.83

      Uniformity  Clean.Cup  Sweetness  Cupper.Points
Total.Cup.Points  \
0           10.00      10.00      10.00           8.75
90.58
1           10.00      10.00      10.00           8.58
89.92
2           10.00      10.00      10.00           9.25
89.75
3           10.00      10.00      10.00           8.67
89.00
4           10.00      10.00      10.00           8.58
88.83
...           ...        ...        ...            ...              ..
.
1334        10.00      10.00       7.75           7.83
78.75
1335        10.00      10.00       8.42           8.58
78.08
1336         9.33       9.33       7.42           7.17
77.17
1337         9.33       9.33       7.08           6.92
75.08
1338         9.33       9.33       6.67           7.92
73.75

      Moisture  Category.One.Defects  Quakers  Category.Two.Defects
0         0.12                     0      0.0                     0
1         0.12                     0      0.0                     1
2         0.00                     0      0.0                     0
3         0.11                     0      0.0                     2
4         0.12                     0      0.0                     2
...        ...                   ...      ...                   ...
1334      0.00                     0      0.0                     1
1335      0.00                     0      0.0                     0
1336      0.00                     0      0.0                     6
```

```
1337      0.10                    20       0.0                     1
1338      0.12                    63       0.0                     9

[1339 rows x 16 columns]
```

```
Q1=new_df.quantile(0.25)
```

```
Q1
```

```
Number.of.Bags          14.00
Aroma                    7.42
Flavor                   7.33
Aftertaste               7.25
Acidity                  7.33
Body                     7.33
Balance                  7.33
Uniformity              10.00
Clean.Cup               10.00
Sweetness               10.00
Cupper.Points            7.25
Total.Cup.Points        81.08
Moisture                 0.09
Category.One.Defects     0.00
Quakers                  0.00
Category.Two.Defects     0.00
Name: 0.25, dtype: float64
```

```
Q3=new_df.quantile(0.75)
```

```
Q3
```

```
Number.of.Bags          275.00
Aroma                     7.75
Flavor                    7.75
Aftertaste                7.58
Acidity                   7.75
Body                      7.67
Balance                   7.75
Uniformity               10.00
Clean.Cup                10.00
Sweetness                10.00
Cupper.Points             7.75
Total.Cup.Points         83.67
Moisture                  0.12
Category.One.Defects      0.00
Quakers                   0.00
Category.Two.Defects      4.00
Name: 0.75, dtype: float64
```

```
IQR=Q3-Q1
```

```
IQR

Number.of.Bags           261.00
Aroma                      0.33
Flavor                     0.42
Aftertaste                 0.33
Acidity                    0.42
Body                       0.34
Balance                    0.42
Uniformity                 0.00
Clean.Cup                  0.00
Sweetness                  0.00
Cupper.Points              0.50
Total.Cup.Points           2.59
Moisture                   0.03
Category.One.Defects       0.00
Quakers                    0.00
Category.Two.Defects       4.00
dtype: float64

a=(new_df<(Q1-1.5*IQR))|(new_df>(Q3+1.5*IQR))
a

      Number.of.Bags  Aroma  Flavor  Aftertaste  Acidity   Body
Balance  \
0              False   True    True        True     True   True
True
1              False   True    True        True     True   True
True
2              False   True    True        True     True   True
True
3              False  False    True        True     True   True
False
4              False   True    True        True     True   True
False
...              ...    ...     ...         ...      ...    ...
..
1334           False  False   False       False    False   True
False
1335           False  False   False       False    False   True
True
1336           False  False   False       False    False False
False
1337           False  False   False        True    False False
False
1338           False   True    True        True    False False
False

      Uniformity Clean.Cup Sweetness Cupper.Points
Total.Cup.Points  \
```

```
0           False       False       False            True
True
1           False       False       False            True
True
2           False       False       False            True
True
3           False       False       False            True
True
4           False       False       False            True
True
...           ...         ...         ...             ...                ..
.
1334        False       False        True           False
False
1335        False       False        True            True
False
1336         True        True        True           False
True
1337         True        True        True           False
True
1338         True        True        True           False
True
```

```
      Moisture   Category.One.Defects   Quakers  Category.Two.Defects
0        False                  False    False                 False
1        False                  False    False                 False
2         True                  False    False                 False
3        False                  False    False                 False
4        False                  False    False                 False
...        ...                    ...      ...                   ...
1334      True                  False    False                 False
1335      True                  False    False                 False
1336      True                  False    False                 False
1337     False                   True    False                 False
1338     False                   True    False                 False

[1339 rows x 16 columns]
```

```
a.sum()
```

```
Number.of.Bags              4
Aroma                      75
Flavor                     44
Aftertaste                 87
Acidity                    25
Body                       34
Balance                    40
Uniformity                187
Clean.Cup                 120
Sweetness                 121
```

```
Cupper.Points             33
Total.Cup.Points          72
Moisture                 305
Category.One.Defects     202
Quakers                   95
Category.Two.Defects      94
dtype: int64
```

```
filter=df[~((new_df<Q1-1.5*IQR)|(new_df>Q1+1.5*IQR)).any(axis=1)]
```

```
filter
```

```
      Species                          Owner Country.of.Origin  \
148   Arabica       juan luis alvarado romero          Guatemala
160   Arabica          yunnan coffee exchange              China
196   Arabica          yunnan coffee exchange              China
203   Arabica       juan luis alvarado romero          Guatemala
206   Arabica           jesus salazar velasco             Mexico
...       ...                             ...                ...
1152  Arabica         calixto guillen vazquez             Mexico
1160  Arabica       juan luis alvarado romero          Guatemala
1166  Arabica   armando luis pohlenz martinez             Mexico
1167  Arabica  exportadora de cafe condor s.a           Colombia
1182  Arabica          lin, che-hao krude 林哲豪      Taiwan

                    Region
Producer \
148                 jalapa                                    ROBERTO
MONTERROSO
160                 yunnan         Menglian Ban’an Coffee Processing
Factory
196                 yunnan  Menglian County Nayun Town Mangzhang Coffee
Farm
203               oriente                                       AMILCAR
LAPOLA
206   san pedro cotzilnam                                  JESUS SALAZAR
VELASCO
...                   ...
...
1152              chiapas                                    VARIOS
PRODUCTORES
1160              oriente                          AGRIPEC DE LA VEGA Y
CIA.
1166         la concordia                       ARMANDO LUIS POHLENZ
MARTINEZ
1167               huila                                          La
Plata
1182         natou county                                    WU SHU
YI 巫叔憶
```

```
      Number.of.Bags Bag.Weight             In.Country.Partner
Harvest.Year  \
148            250.0      69 kg  Asociacion Nacional Del Café
2014
160              3.0      60 kg         Yunnan Coffee Exchange
2015
196              3.0      60 kg         Yunnan Coffee Exchange
2015
203            250.0      69 kg  Asociacion Nacional Del Café
2014
206             10.0       1 kg                       AMECAFE
2012
...              ...        ...                           ...
...
1152           380.0       1 kg                       AMECAFE
2013
1160           250.0      69 kg  Asociacion Nacional Del Café
2013
1166           250.0       1 kg                       AMECAFE
2012
1167           250.0      70 kg                      Almacafé
4T/10
1182            50.0      20 kg  Specialty Coffee Association
2014

              Grading.Date  ... Total.Cup.Points  Moisture  \
148         May 29th, 2014  ...             84.58      0.11
160        April 6th, 2016  ...             84.50      0.10
196        April 6th, 2016  ...             84.25      0.10
203      August 21st, 2014  ...             84.25      0.10
206   September 11th, 2012  ...             84.25      0.12
...                    ...  ...               ...       ...
1152     March 29th, 2013  ...             79.75      0.12
1160       June 27th, 2013  ...             79.67      0.11
1166     August 30th, 2012  ...             79.58      0.13
1167    February 9th, 2011  ...             79.58      0.10
1182    November 7th, 2014  ...             79.25      0.11

      Category.One.Defects  Quakers  Category.Two.Defects  \
148                      0      0.0                     1
160                      0      0.0                     1
196                      0      0.0                     1
203                      0      0.0                     1
206                      0      0.0                     1
...                    ...      ...                   ...
1152                     0      0.0                     5
1160                     0      0.0                     6
1166                     0      0.0                     5
1167                     0      0.0                     4
```

```
1182                     0      0.0                       0

                Expiration            Certification.Body  \
148         May 29th, 2015  Asociacion Nacional Del Café
160        April 6th, 2017        Yunnan Coffee Exchange
196        April 6th, 2017        Yunnan Coffee Exchange
203      August 21st, 2015  Asociacion Nacional Del Café
206   September 11th, 2013                      AMECAFE
...                    ...                          ...
1152     March 29th, 2014                      AMECAFE
1160       June 27th, 2014  Asociacion Nacional Del Café
1166     August 30th, 2013                      AMECAFE
1167    February 9th, 2012                     Almacafé
1182    November 7th, 2015  Specialty Coffee Association

                          Certification.Address  \
148   b1f20fe3a819fd6b2ee0eb8fdc3da256604f1e53
160   683fa6014608fc10ff681b0435b0b2dbe6df988f
196   683fa6014608fc10ff681b0435b0b2dbe6df988f
203   b1f20fe3a819fd6b2ee0eb8fdc3da256604f1e53
206   59e396ad6e22a1c22b248f958e1da2bd8af85272
...                                       ...
1152  59e396ad6e22a1c22b248f958e1da2bd8af85272
1160  b1f20fe3a819fd6b2ee0eb8fdc3da256604f1e53
1166  59e396ad6e22a1c22b248f958e1da2bd8af85272
1167  e493c36c2d076bf273064f7ac23ad562af257a25
1182  36d0d00a3724338ba7937c52a378d085f2172daa

                          Certification.Contact  unit_of_measurement
148   724f04ad10ed31dbb9d260f0dfd221ba48be8a95                    ft
160   f6d87a6c04653c569d4911a66f89d5e30ce83b93                     m
196   f6d87a6c04653c569d4911a66f89d5e30ce83b93                     m
203   724f04ad10ed31dbb9d260f0dfd221ba48be8a95                    ft
206   0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7                     m
...                                        ...                   ...
1152  0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7                     m
1160  724f04ad10ed31dbb9d260f0dfd221ba48be8a95                    ft
1166  0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7                     m
1167  70d3c0c26f89e00fdae6fb39ff54f0d2eb1c38ab                     m
1182  0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                     m

[404 rows x 31 columns]
```

```
filter.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 404 entries, 148 to 1182
Data columns (total 31 columns):
 #   Column                 Non-Null Count  Dtype
```

```
 0   Species                 404 non-null    object
 1   Owner                   404 non-null    object
 2   Country.of.Origin       404 non-null    object
 3   Region                  404 non-null    object
 4   Producer                404 non-null    object
 5   Number.of.Bags          404 non-null    float64
 6   Bag.Weight              404 non-null    object
 7   In.Country.Partner      404 non-null    object
 8   Harvest.Year            404 non-null    object
 9   Grading.Date            404 non-null    object
 10  Owner.1                 404 non-null    object
 11  Aroma                   404 non-null    float64
 12  Flavor                  404 non-null    float64
 13  Aftertaste              404 non-null    float64
 14  Acidity                 404 non-null    float64
 15  Body                    404 non-null    float64
 16  Balance                 404 non-null    float64
 17  Uniformity              404 non-null    float64
 18  Clean.Cup               404 non-null    float64
 19  Sweetness               404 non-null    float64
 20  Cupper.Points           404 non-null    float64
 21  Total.Cup.Points        404 non-null    float64
 22  Moisture                404 non-null    float64
 23  Category.One.Defects    404 non-null    int64
 24  Quakers                 404 non-null    float64
 25  Category.Two.Defects    404 non-null    int64
 26  Expiration              404 non-null    object
 27  Certification.Body      404 non-null    object
 28  Certification.Address   404 non-null    object
 29  Certification.Contact   404 non-null    object
 30  unit_of_measurement     404 non-null    object
dtypes: float64(14), int64(2), object(15)
memory usage: 101.0+ KB
```

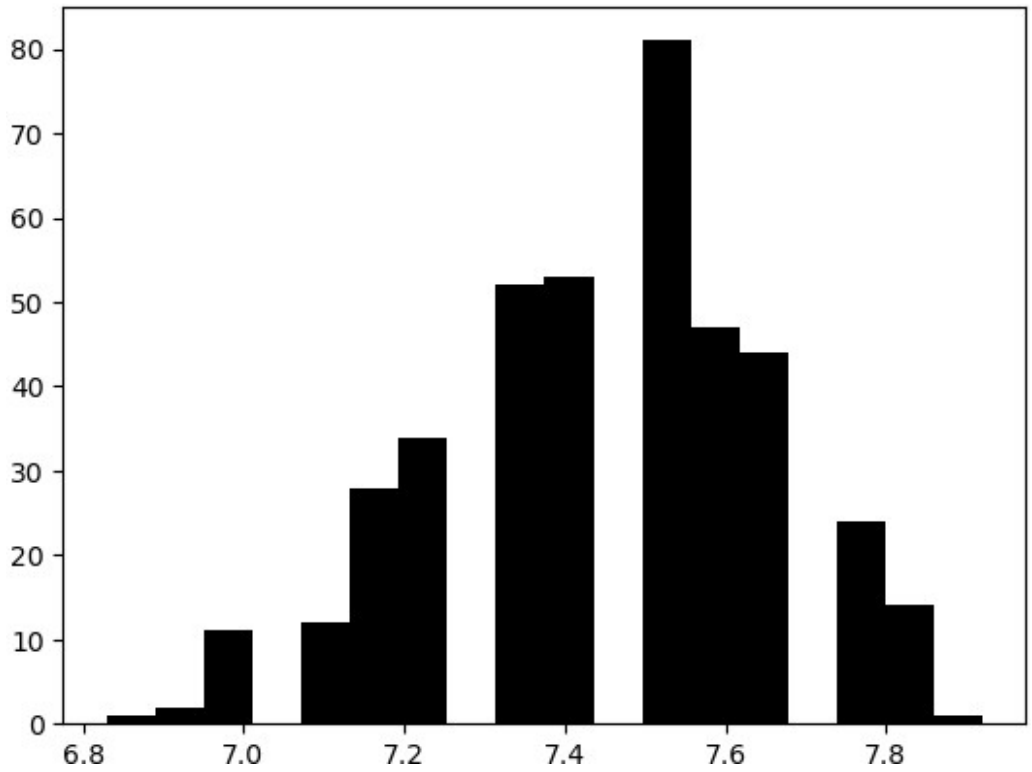
```
filter.shape
```

```
(404, 31)
```

```
                                                    #Univariate
```

```
plt.hist(filter['Balance'],bins=18,color='Black')
```

```
(array([ 1.,  2., 11.,  0., 12., 28., 34.,  0., 52., 53.,  0., 81.,
47.,
        44.,  0., 24., 14.,  1.]),
 array([6.83      , 6.89055556, 6.95111111, 7.01166667, 7.07222222,
        7.13277778, 7.19333333, 7.25388889, 7.31444444, 7.375     ,
        7.43555556, 7.49611111, 7.55666667, 7.61722222, 7.67777778,
        7.73833333, 7.79888889, 7.85944444, 7.92      ]),
 <BarContainer object of 18 artists>)
```




```
df.groupby(['Country.of.Origin']).count()
```

Country.of.Origin	Species	Owner	Region	Producer
Brazil	132	132	132	132
Burundi	2	2	2	2
China	16	16	16	16
Colombia	183	183	183	183
Costa Rica	51	51	51	51
Cote d?Ivoire	1	1	1	1
Ecuador	3	3	3	3
El Salvador	21	21	21	21
Ethiopia	44	44	44	44
Guatemala	181	181	181	181
Haiti	6	6	6	6
Honduras	53	53	53	53
India	14	14	14	14
Indonesia	20	20	20	20
Japan	1	1	1	1
Kenya	25	25	25	25
Laos	3	3	3	3
Malawi	11	11	11	11
Mauritius	1	1	1	1
Mexico	237	237	237	237
Myanmar	8	8	8	8

Nicaragua	26	26	26	26
Panama	4	4	4	4
Papua New Guinea	1	1	1	1
Peru	10	10	10	10
Philippines	5	5	5	5
Rwanda	1	1	1	1
Taiwan	75	75	75	75
Tanzania, United Republic Of	40	40	40	40
Thailand	32	32	32	32
Uganda	36	36	36	36
United States	10	10	10	10
United States (Hawaii)	73	73	73	73
United States (Puerto Rico)	4	4	4	4
Vietnam	8	8	8	8
Zambia	1	1	1	1

In.Country.Partner \ Country.of.Origin	Number.of.Bags	Bag.Weight
Brazil	132	132
Burundi	2	2
China	16	16
Colombia	183	183
Costa Rica	51	51
Cote d?Ivoire	1	1
Ecuador	3	3
El Salvador	21	21
Ethiopia	44	44
Guatemala	181	181
Haiti	6	6
Honduras	53	53
India	14	14
Indonesia	20	20
Japan	1	1

```
1
Kenya                                  25           25
25
Laos                                    3            3
3
Malawi                                 11           11
11
Mauritius                               1            1
1
Mexico                                237          237
237
Myanmar                                 8            8
8
Nicaragua                              26           26
26
Panama                                  4            4
4
Papua New Guinea                        1            1
1
Peru                                   10           10
10
Philippines                             5            5
5
Rwanda                                  1            1
1
Taiwan                                 75           75
75
Tanzania, United Republic Of           40           40
40
Thailand                               32           32
32
Uganda                                 36           36
36
United States                          10           10
10
United States (Hawaii)                 73           73
73
United States (Puerto Rico)             4            4
4
Vietnam                                 8            8
8
Zambia                                  1            1
1

                           Harvest.Year  Grading.Date  Owner.1  ...
\
Country.of.Origin                                             ...

Brazil                              132           132      132  ...
```

Burundi	2	2	2	...
China	16	16	16	...
Colombia	183	183	183	...
Costa Rica	51	51	51	...
Cote d?Ivoire	1	1	1	...
Ecuador	3	3	3	...
El Salvador	21	21	21	...
Ethiopia	44	44	44	...
Guatemala	181	181	181	...
Haiti	6	6	6	...
Honduras	53	53	53	...
India	14	14	14	...
Indonesia	20	20	20	...
Japan	1	1	1	...
Kenya	25	25	25	...
Laos	3	3	3	...
Malawi	11	11	11	...
Mauritius	1	1	1	...
Mexico	237	237	237	...
Myanmar	8	8	8	...
Nicaragua	26	26	26	...
Panama	4	4	4	...
Papua New Guinea	1	1	1	...
Peru	10	10	10	...
Philippines	5	5	5	...
Rwanda	1	1	1	...

Taiwan	75	75	75	...
Tanzania, United Republic Of	40	40	40	...
Thailand	32	32	32	...
Uganda	36	36	36	...
United States	10	10	10	...
United States (Hawaii)	73	73	73	...
United States (Puerto Rico)	4	4	4	...
Vietnam	8	8	8	...
Zambia	1	1	1	...

Country.of.Origin	Total.Cup.Points	Moisture
Brazil	132	132
Burundi	2	2
China	16	16
Colombia	183	183
Costa Rica	51	51
Cote d?Ivoire	1	1
Ecuador	3	3
El Salvador	21	21
Ethiopia	44	44
Guatemala	181	181
Haiti	6	6
Honduras	53	53
India	14	14
Indonesia	20	20
Japan	1	1
Kenya	25	25
Laos	3	3
Malawi	11	11
Mauritius	1	1
Mexico	237	237
Myanmar	8	8
Nicaragua	26	26
Panama	4	4
Papua New Guinea	1	1
Peru	10	10
Philippines	5	5
Rwanda	1	1
Taiwan	75	75
Tanzania, United Republic Of	40	40

Thailand	32	32
Uganda	36	36
United States	10	10
United States (Hawaii)	73	73
United States (Puerto Rico)	4	4
Vietnam	8	8
Zambia	1	1

Country.of.Origin	Category.One.Defects	Quakers
Brazil	132	132
Burundi	2	2
China	16	16
Colombia	183	183
Costa Rica	51	51
Cote d?Ivoire	1	1
Ecuador	3	3
El Salvador	21	21
Ethiopia	44	44
Guatemala	181	181
Haiti	6	6
Honduras	53	53
India	14	14
Indonesia	20	20
Japan	1	1
Kenya	25	25
Laos	3	3
Malawi	11	11
Mauritius	1	1
Mexico	237	237
Myanmar	8	8
Nicaragua	26	26
Panama	4	4
Papua New Guinea	1	1
Peru	10	10
Philippines	5	5
Rwanda	1	1
Taiwan	75	75
Tanzania, United Republic Of	40	40
Thailand	32	32
Uganda	36	36
United States	10	10
United States (Hawaii)	73	73
United States (Puerto Rico)	4	4
Vietnam	8	8
Zambia	1	1

Country.of.Origin	Category.Two.Defects	Expiration

```
Brazil                                          132          132
Burundi                                           2            2
China                                            16           16
Colombia                                        183          183
Costa Rica                                       51           51
Cote d?Ivoire                                     1            1
Ecuador                                           3            3
El Salvador                                      21           21
Ethiopia                                         44           44
Guatemala                                       181          181
Haiti                                             6            6
Honduras                                         53           53
India                                            14           14
Indonesia                                        20           20
Japan                                             1            1
Kenya                                            25           25
Laos                                              3            3
Malawi                                           11           11
Mauritius                                         1            1
Mexico                                          237          237
Myanmar                                           8            8
Nicaragua                                        26           26
Panama                                            4            4
Papua New Guinea                                  1            1
Peru                                             10           10
Philippines                                       5            5
Rwanda                                            1            1
Taiwan                                           75           75
Tanzania, United Republic Of                     40           40
Thailand                                         32           32
Uganda                                           36           36
United States                                    10           10
United States (Hawaii)                           73           73
United States (Puerto Rico)                       4            4
Vietnam                                           8            8
Zambia                                            1            1

                                Certification.Body
Certification.Address \
Country.of.Origin

Brazil                                          132
132
Burundi                                           2
2
China                                            16
16
Colombia                                        183
183
```

```
Costa Rica                                    51
51
Cote d?Ivoire                                  1
1
Ecuador                                        3
3
El Salvador                                   21
21
Ethiopia                                      44
44
Guatemala                                    181
181
Haiti                                          6
6
Honduras                                      53
53
India                                         14
14
Indonesia                                     20
20
Japan                                          1
1
Kenya                                         25
25
Laos                                           3
3
Malawi                                        11
11
Mauritius                                      1
1
Mexico                                       237
237
Myanmar                                        8
8
Nicaragua                                     26
26
Panama                                         4
4
Papua New Guinea                               1
1
Peru                                          10
10
Philippines                                    5
5
Rwanda                                         1
1
Taiwan                                        75
75
Tanzania, United Republic Of                  40
```

```
40
Thailand                                          32
32
Uganda                                            36
36
United States                                     10
10
United States (Hawaii)                            73
73
United States (Puerto Rico)                        4
4
Vietnam                                            8
8
Zambia                                             1
1

                                Certification.Contact
unit_of_measurement
Country.of.Origin

Brazil                                           132
132
Burundi                                            2
2
China                                             16
16
Colombia                                         183
183
Costa Rica                                         51
51
Cote d?Ivoire                                      1
1
Ecuador                                            3
3
El Salvador                                       21
21
Ethiopia                                          44
44
Guatemala                                        181
181
Haiti                                              6
6
Honduras                                          53
53
India                                             14
14
Indonesia                                         20
20
Japan                                              1
```

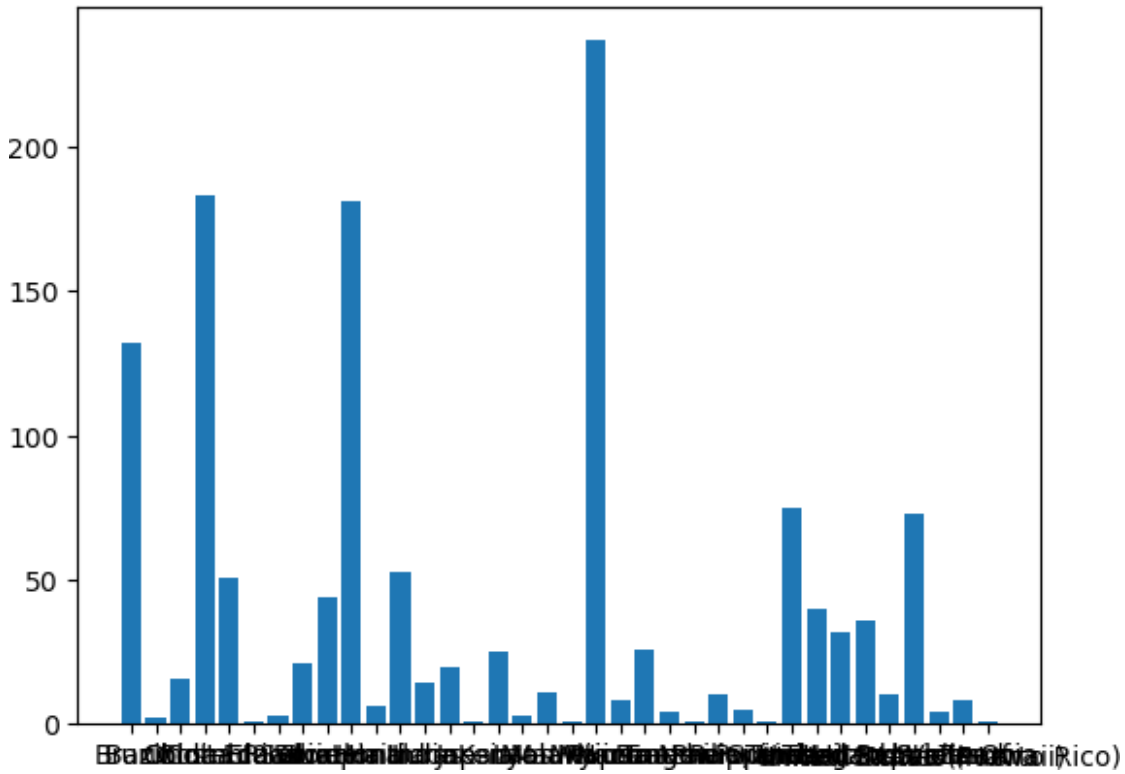
```
1
Kenya                                          25
25
Laos                                            3
3
Malawi                                         11
11
Mauritius                                       1
1
Mexico                                        237
237
Myanmar                                         8
8
Nicaragua                                      26
26
Panama                                          4
4
Papua New Guinea                                1
1
Peru                                           10
10
Philippines                                     5
5
Rwanda                                          1
1
Taiwan                                         75
75
Tanzania, United Republic Of                   40
40
Thailand                                       32
32
Uganda                                         36
36
United States                                  10
10
United States (Hawaii)                         73
73
United States (Puerto Rico)                     4
4
Vietnam                                         8
8
Zambia                                          1
1

[36 rows x 30 columns]
```

```
b=df.groupby(['Country.of.Origin']).size().reset_index(name="count").r
ename(columns={'Country.of.Origin':'CO'})

b
```

	CO	count
0	Brazil	132
1	Burundi	2
2	China	16
3	Colombia	183
4	Costa Rica	51
5	Cote d?Ivoire	1
6	Ecuador	3
7	El Salvador	21
8	Ethiopia	44
9	Guatemala	181
10	Haiti	6
11	Honduras	53
12	India	14
13	Indonesia	20
14	Japan	1
15	Kenya	25
16	Laos	3
17	Malawi	11
18	Mauritius	1
19	Mexico	237
20	Myanmar	8
21	Nicaragua	26
22	Panama	4
23	Papua New Guinea	1
24	Peru	10
25	Philippines	5
26	Rwanda	1
27	Taiwan	75
28	Tanzania, United Republic Of	40
29	Thailand	32
30	Uganda	36
31	United States	10
32	United States (Hawaii)	73
33	United States (Puerto Rico)	4
34	Vietnam	8
35	Zambia	1

```
plt.bar(b['CO'],b['count'])
```

```
<BarContainer object of 36 artists>
```



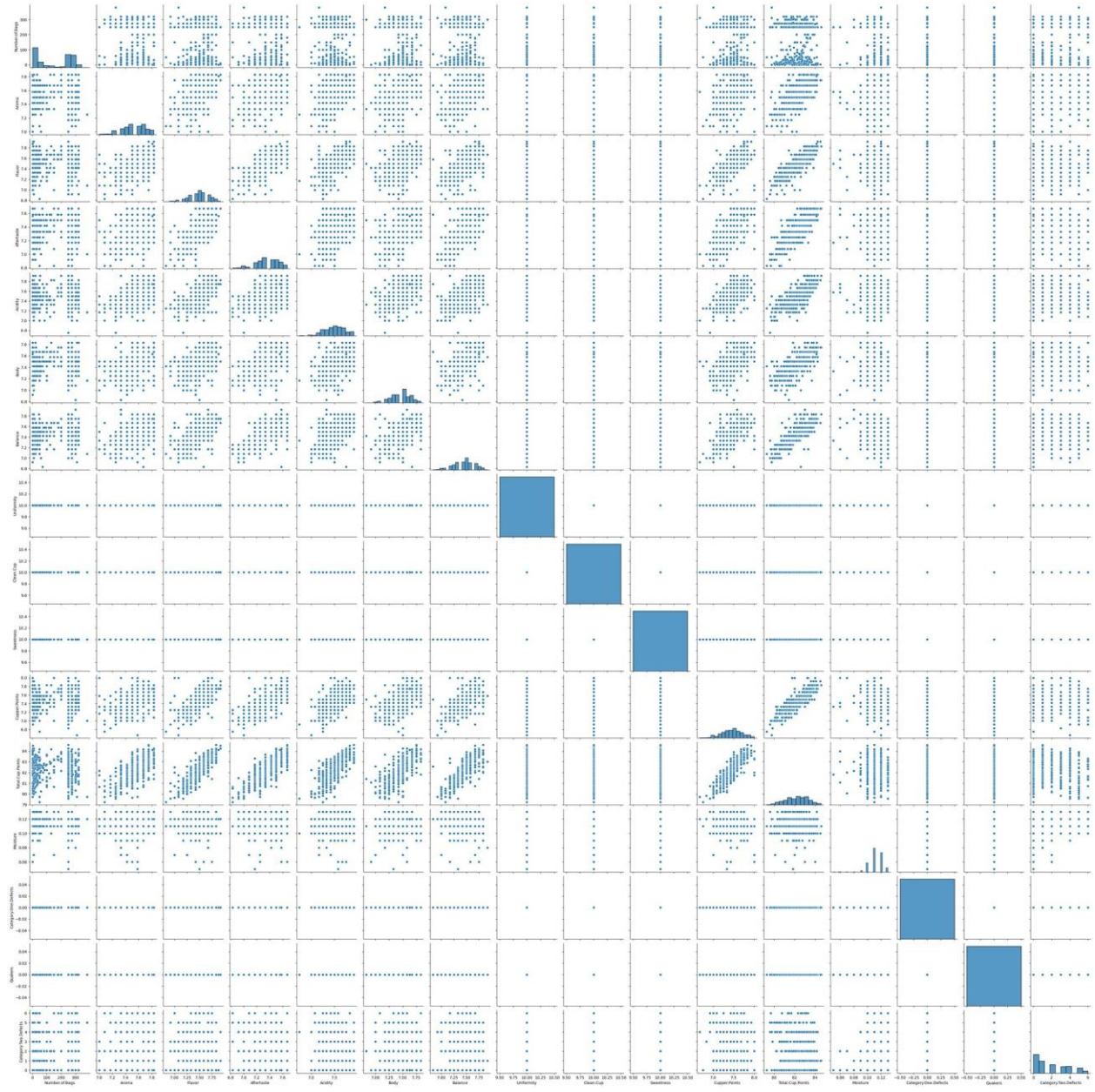

```
b['count%']=b['count']/sum(b['count'])*100
b
```

	CO	count	count%
0	Brazil	132	9.858103
1	Burundi	2	0.149365
2	China	16	1.194922
3	Colombia	183	13.666916
4	Costa Rica	51	3.808813
5	Cote d?Ivoire	1	0.074683
6	Ecuador	3	0.224048
7	El Salvador	21	1.568335
8	Ethiopia	44	3.286034
9	Guatemala	181	13.517550
10	Haiti	6	0.448096
11	Honduras	53	3.958178
12	India	14	1.045556
13	Indonesia	20	1.493652
14	Japan	1	0.074683
15	Kenya	25	1.867065
16	Laos	3	0.224048
17	Malawi	11	0.821509
18	Mauritius	1	0.074683
19	Mexico	237	17.699776
20	Myanmar	8	0.597461

```
21                      Nicaragua     26   1.941748
22                         Panama      4   0.298730
23                Papua New Guinea      1   0.074683
24                           Peru     10   0.746826
25                    Philippines      5   0.373413
26                         Rwanda      1   0.074683
27                         Taiwan     75   5.601195
28  Tanzania, United Republic Of      40   2.987304
29                       Thailand     32   2.389843
30                         Uganda     36   2.688574
31                  United States     10   0.746826
32        United States (Hawaii)     73   5.451830
33   United States (Puerto Rico)      4   0.298730
34                        Vietnam      8   0.597461
35                         Zambia      1   0.074683
```

```
c=sns.pairplot(filter)
c
```

```
<seaborn.axisgrid.PairGrid at 0x2306b7a0d50>
```



```
filter
```

```
      Species                              Owner Country.of.Origin  \
148   Arabica       juan luis alvarado romero          Guatemala
160   Arabica          yunnan coffee exchange              China
196   Arabica          yunnan coffee exchange              China
203   Arabica       juan luis alvarado romero          Guatemala
206   Arabica           jesus salazar velasco             Mexico
...       ...                              ...                ...
1152  Arabica         calixto guillen vazquez             Mexico
1160  Arabica       juan luis alvarado romero          Guatemala
1166  Arabica   armando luis pohlenz martinez             Mexico
```

```
1167  Arabica  exportadora de cafe condor s.a          Colombia
1182  Arabica          lin, che-hao krude 林哲豪     Taiwan

                    Region
Producer \
148                 jalapa                                  ROBERTO
MONTERROSO
160                 yunnan         Menglian Ban'an Coffee Processing
Factory
196                 yunnan  Menglian County Nayun Town Mangzhang Coffee
Farm
203                oriente                                    AMILCAR
LAPOLA
206   san pedro cotzilnam                               JESUS SALAZAR
VELASCO
...                    ...
...
1152               chiapas                                  VARIOS
PRODUCTORES
1160               oriente                          AGRIPEC DE LA VEGA Y
CIA.
1166         la concordia                       ARMANDO LUIS POHLENZ
MARTINEZ
1167                huila                                          La
Plata
1182         natou county                                    WU SHU
YI 巫叔憶

      Number.of.Bags Bag.Weight             In.Country.Partner
Harvest.Year  \
148            250.0      69 kg  Asociacion Nacional Del Café
2014
160              3.0      60 kg        Yunnan Coffee Exchange
2015
196              3.0      60 kg        Yunnan Coffee Exchange
2015
203            250.0      69 kg  Asociacion Nacional Del Café
2014
206             10.0       1 kg                        AMECAFE
2012
...              ...        ...                           ...
...
1152           380.0       1 kg                        AMECAFE
2013
1160           250.0      69 kg  Asociacion Nacional Del Café
2013
1166           250.0       1 kg                        AMECAFE
2012
1167           250.0      70 kg                       Almacafé
```

```
4T/10
1182            50.0      20 kg  Specialty Coffee Association
2014

              Grading.Date  ... Total.Cup.Points  Moisture  \
148          May 29th, 2014  ...             84.58      0.11
160        April 6th, 2016  ...             84.50      0.10
196        April 6th, 2016  ...             84.25      0.10
203      August 21st, 2014  ...             84.25      0.10
206   September 11th, 2012  ...             84.25      0.12
...                    ...  ...               ...       ...
1152     March 29th, 2013  ...             79.75      0.12
1160       June 27th, 2013  ...             79.67      0.11
1166     August 30th, 2012  ...             79.58      0.13
1167    February 9th, 2011  ...             79.58      0.10
1182    November 7th, 2014  ...             79.25      0.11

      Category.One.Defects  Quakers  Category.Two.Defects  \
148                      0      0.0                     1
160                      0      0.0                     1
196                      0      0.0                     1
203                      0      0.0                     1
206                      0      0.0                     1
...                    ...      ...                   ...
1152                     0      0.0                     5
1160                     0      0.0                     6
1166                     0      0.0                     5
1167                     0      0.0                     4
1182                     0      0.0                     0

                Expiration            Certification.Body  \
148          May 29th, 2015  Asociacion Nacional Del Café
160        April 6th, 2017        Yunnan Coffee Exchange
196        April 6th, 2017        Yunnan Coffee Exchange
203      August 21st, 2015  Asociacion Nacional Del Café
206   September 11th, 2013                      AMECAFE
...                    ...                          ...
1152     March 29th, 2014                      AMECAFE
1160       June 27th, 2014  Asociacion Nacional Del Café
1166     August 30th, 2013                      AMECAFE
1167    February 9th, 2012                     Almacafé
1182    November 7th, 2015  Specialty Coffee Association

                         Certification.Address  \
148   b1f20fe3a819fd6b2ee0eb8fdc3da256604f1e53
160   683fa6014608fc10ff681b0435b0b2dbe6df988f
196   683fa6014608fc10ff681b0435b0b2dbe6df988f
203   b1f20fe3a819fd6b2ee0eb8fdc3da256604f1e53
206   59e396ad6e22a1c22b248f958e1da2bd8af85272
...                                       ...
```

```
1152  59e396ad6e22a1c22b248f958e1da2bd8af85272
1160  b1f20fe3a819fd6b2ee0eb8fdc3da256604f1e53
1166  59e396ad6e22a1c22b248f958e1da2bd8af85272
1167  e493c36c2d076bf273064f7ac23ad562af257a25
1182  36d0d00a3724338ba7937c52a378d085f2172daa

                          Certification.Contact  unit_of_measurement
148   724f04ad10ed31dbb9d260f0dfd221ba48be8a95                     ft
160   f6d87a6c04653c569d4911a66f89d5e30ce83b93                      m
196   f6d87a6c04653c569d4911a66f89d5e30ce83b93                      m
203   724f04ad10ed31dbb9d260f0dfd221ba48be8a95                     ft
206   0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7                      m
...                                        ...                    ...
1152  0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7                      m
1160  724f04ad10ed31dbb9d260f0dfd221ba48be8a95                     ft
1166  0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7                      m
1167  70d3c0c26f89e00fdae6fb39ff54f0d2eb1c38ab                      m
1182  0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                      m

[404 rows x 31 columns]
```

```
filter.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 404 entries, 148 to 1182
Data columns (total 31 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Species                 404 non-null    object
 1   Owner                   404 non-null    object
 2   Country.of.Origin       404 non-null    object
 3   Region                  404 non-null    object
 4   Producer                404 non-null    object
 5   Number.of.Bags          404 non-null    float64
 6   Bag.Weight              404 non-null    object
 7   In.Country.Partner      404 non-null    object
 8   Harvest.Year            404 non-null    object
 9   Grading.Date            404 non-null    object
 10  Owner.1                 404 non-null    object
 11  Aroma                   404 non-null    float64
 12  Flavor                  404 non-null    float64
 13  Aftertaste              404 non-null    float64
 14  Acidity                 404 non-null    float64
 15  Body                    404 non-null    float64
 16  Balance                 404 non-null    float64
 17  Uniformity              404 non-null    float64
 18  Clean.Cup               404 non-null    float64
 19  Sweetness               404 non-null    float64
 20  Cupper.Points           404 non-null    float64
 21  Total.Cup.Points        404 non-null    float64
```

```
 22  Moisture                404 non-null    float64
 23  Category.One.Defects    404 non-null    int64
 24  Quakers                 404 non-null    float64
 25  Category.Two.Defects    404 non-null    int64
 26  Expiration              404 non-null    object
 27  Certification.Body      404 non-null    object
 28  Certification.Address   404 non-null    object
 29  Certification.Contact   404 non-null    object
 30  unit_of_measurement     404 non-null    object
dtypes: float64(14), int64(2), object(15)
memory usage: 101.0+ KB
```

```
s=filter.select_dtypes(exclude=['object'])

s
```

```
      Number.of.Bags  Aroma  Flavor  Aftertaste  Acidity  Body
Balance  \
148            250.0   7.75    7.83        7.67     7.92  7.75
7.83
160              3.0   7.75    7.92        7.67     7.83  7.75
7.67
196              3.0   7.83    7.83        7.58     7.75  7.75
7.75
203            250.0   7.83    7.83        7.50     7.92  7.58
7.75
206             10.0   7.83    7.92        7.67     7.83  7.50
7.75
...              ...    ...     ...         ...      ...   ...      ..
.
1152           380.0   7.25    7.08        7.00     7.17  7.17
7.08
1160           250.0   7.50    7.00        6.83     7.08  7.08
7.17
1166           250.0   7.25    6.92        6.92     7.33  7.25
7.00
1167           250.0   7.25    7.17        7.00     6.75  7.17
7.33
1182            50.0   7.08    6.83        6.83     7.25  7.42
7.08

      Uniformity  Clean.Cup  Sweetness  Cupper.Points
Total.Cup.Points  \
148         10.0       10.0       10.0           7.83
84.58
160         10.0       10.0       10.0           7.92
84.50
196         10.0       10.0       10.0           7.75
84.25
203         10.0       10.0       10.0           7.83
```

```
84.25
206         10.0        10.0        10.0             7.75
84.25
...          ...         ...         ...              ...                ..
.
1152        10.0        10.0        10.0             7.00
79.75
1160        10.0        10.0        10.0             7.00
79.67
1166        10.0        10.0        10.0             6.92
79.58
1167        10.0        10.0        10.0             6.92
79.58
1182        10.0        10.0        10.0             6.75
79.25

      Moisture   Category.One.Defects   Quakers  Category.Two.Defects
148       0.11                      0       0.0                     1
160       0.10                      0       0.0                     1
196       0.10                      0       0.0                     1
203       0.10                      0       0.0                     1
206       0.12                      0       0.0                     1
...        ...                    ...       ...                   ...
1152      0.12                      0       0.0                     5
1160      0.11                      0       0.0                     6
1166      0.13                      0       0.0                     5
1167      0.10                      0       0.0                     4
1182      0.11                      0       0.0                     0

[404 rows x 16 columns]

new=s.drop(columns=['Uniformity','Clean.Cup','Sweetness','Category.One
.Defects','Quakers'])

new

      Number.of.Bags  Aroma  Flavor  Aftertaste  Acidity  Body
Balance  \
148            250.0   7.75    7.83        7.67     7.92  7.75
7.83
160              3.0   7.75    7.92        7.67     7.83  7.75
7.67
196              3.0   7.83    7.83        7.58     7.75  7.75
7.75
203            250.0   7.83    7.83        7.50     7.92  7.58
7.75
206             10.0   7.83    7.92        7.67     7.83  7.50
7.75
...              ...    ...     ...         ...      ...   ...      ..
.
```

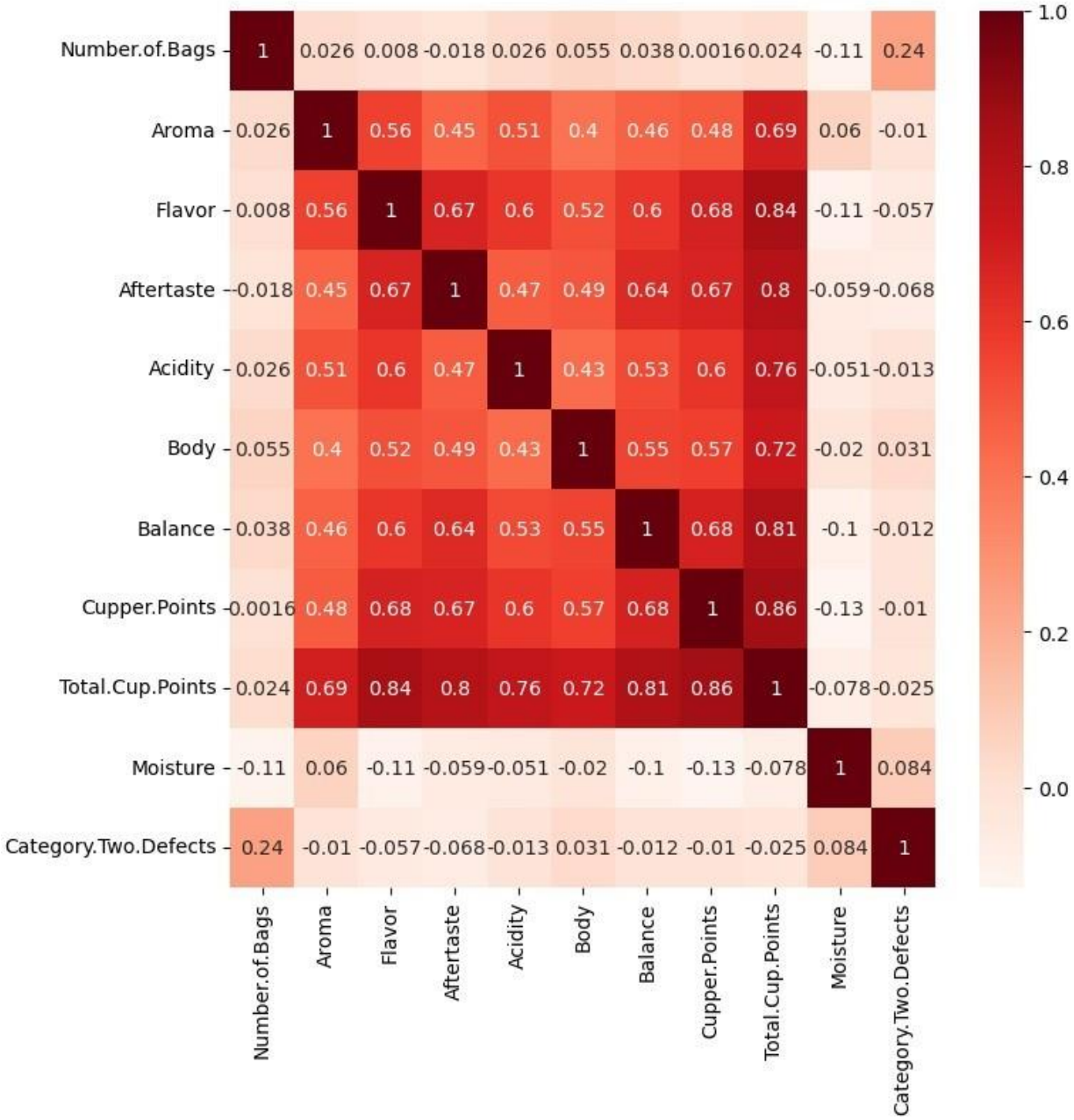
```
1152            380.0   7.25    7.08         7.00     7.17  7.17
7.08
1160            250.0   7.50    7.00         6.83     7.08  7.08
7.17
1166            250.0   7.25    6.92         6.92     7.33  7.25
7.00
1167            250.0   7.25    7.17         7.00     6.75  7.17
7.33
1182             50.0   7.08    6.83         6.83     7.25  7.42
7.08

      Cupper.Points  Total.Cup.Points  Moisture  Category.Two.Defects

148            7.83             84.58      0.11                     1

160            7.92             84.50      0.10                     1

196            7.75             84.25      0.10                     1

203            7.83             84.25      0.10                     1

206            7.75             84.25      0.12                     1

...             ...               ...       ...                   ...

1152           7.00             79.75      0.12                     5

1160           7.00             79.67      0.11                     6

1166           6.92             79.58      0.13                     5

1167           6.92             79.58      0.10                     4

1182           6.75             79.25      0.11                     0


[404 rows x 11 columns]

g=new.corr()

g

                      Number.of.Bags     Aroma    Flavor
Aftertaste  \
Number.of.Bags              1.000000  0.025602  0.007954   -0.018019

Aroma                       0.025602  1.000000  0.559678    0.449410

Flavor                      0.007954  0.559678  1.000000    0.670064

Aftertaste                 -0.018019  0.449410  0.670064    1.000000
```

```
Acidity                    0.026252  0.508095  0.597453    0.473935

Body                       0.054575  0.404418  0.518042    0.494965

Balance                    0.037832  0.457744  0.597673    0.643990

Cupper.Points              0.001627  0.483042  0.678251    0.672963

Total.Cup.Points           0.023737  0.694559  0.841711    0.800253

Moisture                  -0.114628  0.060143 -0.110836   -0.058805

Category.Two.Defects       0.241994 -0.010320 -0.057406   -0.067737
```

```
                        Acidity      Body   Balance  Cupper.Points  \
Number.of.Bags         0.026252  0.054575  0.037832       0.001627
Aroma                  0.508095  0.404418  0.457744       0.483042
Flavor                 0.597453  0.518042  0.597673       0.678251
Aftertaste             0.473935  0.494965  0.643990       0.672963
Acidity                1.000000  0.431081  0.531185       0.604851
Body                   0.431081  1.000000  0.550884       0.565472
Balance                0.531185  0.550884  1.000000       0.677059
Cupper.Points          0.604851  0.565472  0.677059       1.000000
Total.Cup.Points       0.756864  0.717805  0.814135       0.863134
Moisture              -0.050692 -0.019634 -0.099618      -0.128718
Category.Two.Defects  -0.013021  0.030588 -0.012185      -0.010187

                      Total.Cup.Points  Moisture  Category.Two.Defects
Number.of.Bags                0.023737 -0.114628              0.241994

Aroma                         0.694559  0.060143             -0.010320

Flavor                        0.841711 -0.110836             -0.057406

Aftertaste                    0.800253 -0.058805             -0.067737

Acidity                       0.756864 -0.050692             -0.013021

Body                          0.717805 -0.019634              0.030588

Balance                       0.814135 -0.099618             -0.012185

Cupper.Points                 0.863134 -0.128718             -0.010187

Total.Cup.Points              1.000000 -0.078231             -0.025220

Moisture                     -0.078231  1.000000              0.084359
```

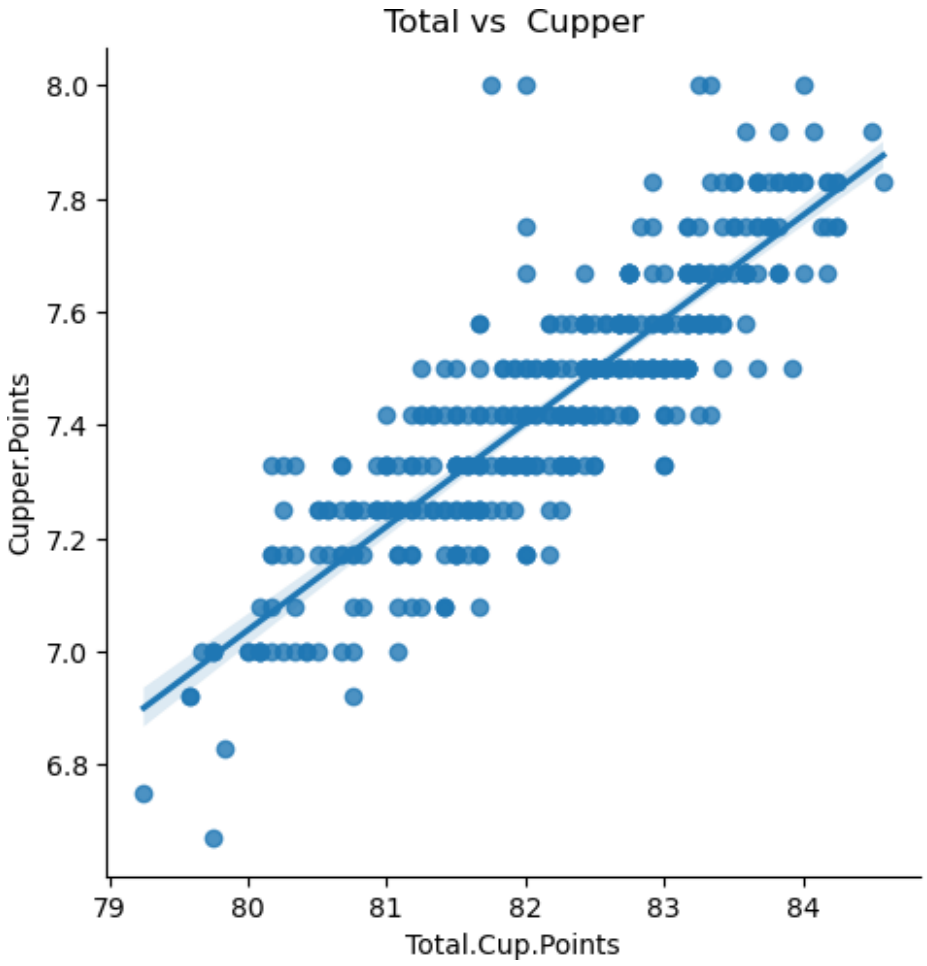
```
Category.Two.Defects          -0.025220   0.084359                1.000000
```

```
plt.figure(figsize=(8,8))
sns.heatmap(g,annot=True,cmap='Reds')
plt.show()
```



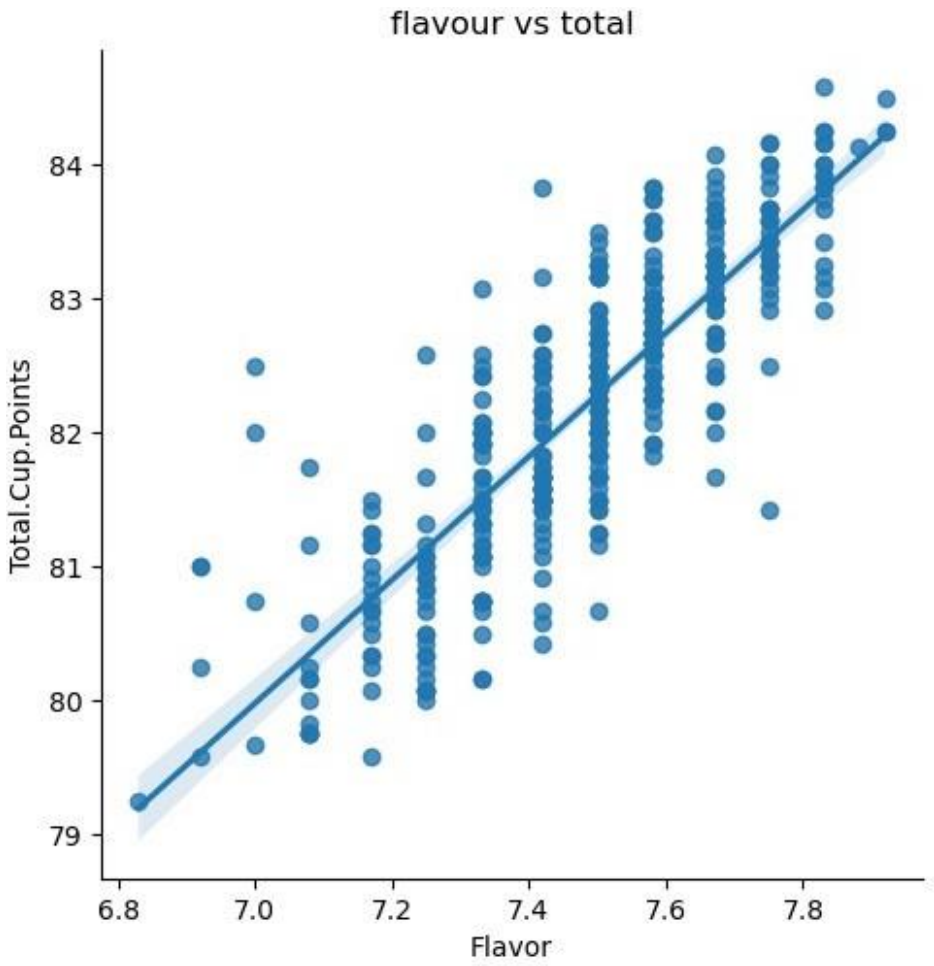

```
sns.lmplot(x='Total.Cup.Points' ,y='Cupper.Points',data=new)
plt.title("Total vs  Cupper")
```

```
Text(0.5, 1.0, 'Total vs  Cupper')
```




```
sns.lmplot(x='Flavor',y='Total.Cup.Points',data=new)
plt.title("flavour vs total")
```

```
Text(0.5, 1.0, 'flavour vs total')
```




```
X = filter[['Flavor', 'Acidity', 'Aftertaste', 'Cupper.Points',
'Aroma', 'Body']]
Y = filter['Total.Cup.Points']

X
```

	Flavor	Acidity	Aftertaste	Cupper.Points	Aroma	Body
148	7.83	7.92	7.67	7.83	7.75	7.75
160	7.92	7.83	7.67	7.92	7.75	7.75
196	7.83	7.75	7.58	7.75	7.83	7.75
203	7.83	7.92	7.50	7.83	7.83	7.58
206	7.92	7.83	7.67	7.75	7.83	7.50
...	...	...	...	...	...	...
1152	7.08	7.17	7.00	7.00	7.25	7.17
1160	7.00	7.08	6.83	7.00	7.50	7.08
1166	6.92	7.33	6.92	6.92	7.25	7.25
1167	7.17	6.75	7.00	6.92	7.25	7.17
1182	6.83	7.25	6.83	6.75	7.08	7.42

```
[404 rows x 6 columns]
```

```
Y
```

```
148     84.58
160     84.50
196     84.25
203     84.25
206     84.25
        ...
1152    79.75
1160    79.67
1166    79.58
1167    79.58
1182    79.25
Name: Total.Cup.Points, Length: 404, dtype: float64
```

```
from sklearn.feature_selection import f_classif
u=f_classif(X,Y)
u
```

```
(array([17.96140216,  9.76119633, 12.8141253 , 21.34390575,
7.47525214,
         8.51363175]),
 array([1.01246262e-74, 1.89407214e-45, 1.15813405e-57, 3.64833219e-
84,
        6.23268531e-35, 7.13981932e-40]))
```

```
filter.groupby(['Flavor','Total.Cup.Points']).count()
```

```
                          Species Owner Country.of.Origin Region
Producer  \
Flavor Total.Cup.Points
```

```
6.83   79.25                    1      1                   1       1
1
6.92   79.58                    1      1                   1       1
1
       80.25                    1      1                   1       1
1
       81.00                    2      2                   2       2
2
7.00   79.67                    1      1                   1       1
1
...                           ...    ...                 ...     ...
...
7.83   84.25                    2      2                   2       2
2
       84.58                    1      1                   1       1
1
7.88   84.13                    1      1                   1       1
1
```

```
7.92   84.25                    2      2                  2        2
2
       84.50                    1      1                  1        1
1

                           Number.of.Bags Bag.Weight
In.Country.Partner  \
Flavor Total.Cup.Points

6.83   79.25                      1          1
1
6.92   79.58                      1          1
1
       80.25                      1          1
1
       81.00                      2          2
2
7.00   79.67                      1          1
1
...                             ...        ...                    ..
.
7.83   84.25                      2          2
2
       84.58                      1          1
1
7.88   84.13                      1          1
1
7.92   84.25                      2          2
2
       84.50                      1          1
1

                           Harvest.Year Grading.Date ...
Cupper.Points  \
Flavor Total.Cup.Points                              ...

6.83   79.25                        1             1  ...
1
6.92   79.58                        1             1  ...
1
       80.25                        1             1  ...
1
       81.00                        2             2  ...
2
7.00   79.67                        1             1  ...
1
...                               ...           ...  ...              ..
.
7.83   84.25                        2             2  ...
2
```

```
       84.58                          1              1  ...
1
7.88   84.13                          1              1  ...
1
7.92   84.25                          2              2  ...
2
       84.50                          1              1  ...
1

                           Moisture Category.One.Defects Quakers \
Flavor Total.Cup.Points
6.83   79.25                      1                    1        1
6.92   79.58                      1                    1        1
       80.25                      1                    1        1
       81.00                      2                    2        2
7.00   79.67                      1                    1        1
...                             ...                  ...      ...
7.83   84.25                      2                    2        2
       84.58                      1                    1        1
7.88   84.13                      1                    1        1
7.92   84.25                      2                    2        2
       84.50                      1                    1        1

                           Category.Two.Defects Expiration
Certification.Body  \
Flavor Total.Cup.Points

6.83   79.25                                 1           1
1
6.92   79.58                                 1           1
1
       80.25                                 1           1
1
       81.00                                 2           2
2
7.00   79.67                                 1           1
1
...                                        ...         ...
...
7.83   84.25                                 2           2
2
       84.58                                 1           1
1
7.88   84.13                                 1           1
1
7.92   84.25                                 2           2
2
       84.50                                 1           1
1
```

```
                          Certification.Address  Certification.Contact
\
Flavor Total.Cup.Points

6.83   79.25                                   1                      1

6.92   79.58                                   1                      1

       80.25                                   1                      1

       81.00                                   2                      2

7.00   79.67                                   1                      1

...                                          ...                    ...

7.83   84.25                                   2                      2

       84.58                                   1                      1

7.88   84.13                                   1                      1

7.92   84.25                                   2                      2

       84.50                                   1                      1


                          unit_of_measurement
Flavor Total.Cup.Points
6.83   79.25                                1
6.92   79.58                                1
       80.25                                1
       81.00                                2
7.00   79.67                                1
...                                       ...
7.83   84.25                                2
       84.58                                1
7.88   84.13                                1
7.92   84.25                                2
       84.50                                1

[198 rows x 29 columns]

filter.groupby(['Flavor','Acidity']).count()

                 Species Owner Country.of.Origin Region Producer \
Flavor Acidity
6.83   7.25           1      1                  1       1         1
6.92   7.25           1      1                  1       1         1
       7.33           3      3                  3       3         3
7.00   7.08           1      1                  1       1         1
```

```
       7.25           1      1                  1       1         1
...                  ...    ...                ...     ...       ...
7.83   7.83           6      6                  6       6         6
       7.92           3      3                  3       3         3
7.88   7.75           1      1                  1       1         1
7.92   7.75           1      1                  1       1         1
       7.83           2      2                  2       2         2

                  Number.of.Bags Bag.Weight In.Country.Partner
Harvest.Year  \
Flavor Acidity

6.83   7.25                   1           1                    1
1
6.92   7.25                   1           1                    1
1
       7.33                   3           3                    3
3
7.00   7.08                   1           1                    1
1
       7.25                   1           1                    1
1
...                         ...         ...                  ...
...
7.83   7.83                   6           6                    6
6
       7.92                   3           3                    3
3
7.88   7.75                   1           1                    1
1
7.92   7.75                   1           1                    1
1
       7.83                   2           2                    2
2

                  Grading.Date ... Total.Cup.Points Moisture \
Flavor Acidity                 ...
6.83   7.25                1  ...                 1         1
6.92   7.25                1  ...                 1         1
       7.33                3  ...                 3         3
7.00   7.08                1  ...                 1         1
       7.25                1  ...                 1         1
...                      ...  ...               ...       ...
7.83   7.83                6  ...                 6         6
       7.92                3  ...                 3         3
7.88   7.75                1  ...                 1         1
7.92   7.75                1  ...                 1         1
       7.83                2  ...                 2         2

                  Category.One.Defects  Quakers  Category.Two.Defects  \
```

Flavor	Acidity			
6.83	7.25	1	1	1
6.92	7.25	1	1	1
	7.33	3	3	3
7.00	7.08	1	1	1
	7.25	1	1	1
...		...	...	...
7.83	7.83	6	6	6
	7.92	3	3	3
7.88	7.75	1	1	1
7.92	7.75	1	1	1
	7.83	2	2	2

		Expiration	Certification.Body	Certification.Address
\				
Flavor	Acidity			
6.83	7.25	1	1	1
6.92	7.25	1	1	1
	7.33	3	3	3
7.00	7.08	1	1	1
	7.25	1	1	1
...		...	...	...
7.83	7.83	6	6	6
	7.92	3	3	3
7.88	7.75	1	1	1
7.92	7.75	1	1	1
	7.83	2	2	2

		Certification.Contact	unit_of_measurement
Flavor	Acidity		
6.83	7.25	1	1
6.92	7.25	1	1
	7.33	3	3
7.00	7.08	1	1
	7.25	1	1
...		...	...
7.83	7.83	6	6
	7.92	3	3
7.88	7.75	1	1

```
7.92   7.75                          1                    1
       7.83                          2                    2

[89 rows x 29 columns]
```

```
filter.groupby(['Flavor','Aftertaste']).count()
```

```
                    Species  Owner  Country.of.Origin  Region  Producer
\
Flavor Aftertaste

6.83   6.83               1      1                  1       1         1

6.92   6.92               1      1                  1       1         1

       7.00               1      1                  1       1         1

       7.17               1      1                  1       1         1

       7.33               1      1                  1       1         1

...                     ...    ...                ...     ...       ...

7.83   7.50               4      4                  4       4         4

       7.58               6      6                  6       6         6

       7.67               3      3                  3       3         3

7.88   7.56               1      1                  1       1         1

7.92   7.67               3      3                  3       3         3


                    Number.of.Bags Bag.Weight In.Country.Partner \
Flavor Aftertaste
6.83   6.83                      1          1                   1
6.92   6.92                      1          1                   1
       7.00                      1          1                   1
       7.17                      1          1                   1
       7.33                      1          1                   1
...                            ...        ...                 ...
7.83   7.50                      4          4                   4
       7.58                      6          6                   6
       7.67                      3          3                   3
7.88   7.56                      1          1                   1
7.92   7.67                      3          3                   3

                    Harvest.Year Grading.Date ...
Total.Cup.Points  \
Flavor Aftertaste                             ...
```

```
6.83   6.83                    1              1  ...                    1
6.92   6.92                    1              1  ...                    1
       7.00                    1              1  ...                    1
       7.17                    1              1  ...                    1
       7.33                    1              1  ...                    1
...                          ...            ...  ...                  ...
7.83   7.50                    4              4  ...                    4
       7.58                    6              6  ...                    6
       7.67                    3              3  ...                    3
7.88   7.56                    1              1  ...                    1
7.92   7.67                    3              3  ...                    3

                     Moisture Category.One.Defects Quakers \
Flavor Aftertaste
6.83   6.83                 1                     1        1
6.92   6.92                 1                     1        1
       7.00                 1                     1        1
       7.17                 1                     1        1
       7.33                 1                     1        1
...                       ...                   ...      ...
7.83   7.50                 4                     4        4
       7.58                 6                     6        6
       7.67                 3                     3        3
7.88   7.56                 1                     1        1
7.92   7.67                 3                     3        3

                     Category.Two.Defects Expiration
Certification.Body  \
Flavor Aftertaste

6.83   6.83                           1           1
1
6.92   6.92                           1           1
1
       7.00                           1           1
1
       7.17                           1           1
1
       7.33                           1           1
```

```
1
...                                    ...          ...                   ..
.
7.83   7.50                             4            4
4
       7.58                             6            6
6
       7.67                             3            3
3
7.88   7.56                             1            1
1
7.92   7.67                             3            3
3

                    Certification.Address Certification.Contact \
Flavor Aftertaste
6.83   6.83                              1                       1
6.92   6.92                              1                       1
       7.00                              1                       1
       7.17                              1                       1
       7.33                              1                       1
...                                    ...                     ...
7.83   7.50                              4                       4
       7.58                              6                       6
       7.67                              3                       3
7.88   7.56                              1                       1
7.92   7.67                              3                       3

                    unit_of_measurement
Flavor Aftertaste
6.83   6.83                           1
6.92   6.92                           1
       7.00                           1
       7.17                           1
       7.33                           1
...                                 ...
7.83   7.50                           4
       7.58                           6
       7.67                           3
7.88   7.56                           1
7.92   7.67                           3

[77 rows x 29 columns]

                                                          #Anova


filter['Flavor'].unique()
```

```
array([12, 14, 11, 13, 10,  7,  9,  8,  6,  5,  2,  3,  4,  1,  0],
      dtype=int64)
```

```
from sklearn.preprocessing import LabelEncoder
label=LabelEncoder()
filter['Flavor']=label.fit_transform(filter['Flavor'])
filter
```

```
C:\Users\user\AppData\Local\Temp\ipykernel_5156\2628298509.py:3:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  filter['Flavor']=label.fit_transform(filter['Flavor'])
```

	Species	Owner	Country.of.Origin	\
148	Arabica	juan luis alvarado romero	Guatemala	
160	Arabica	yunnan coffee exchange	China	
196	Arabica	yunnan coffee exchange	China	
203	Arabica	juan luis alvarado romero	Guatemala	
206	Arabica	jesus salazar velasco	Mexico	
...	...	...	...	
1152	Arabica	calixto guillen vazquez	Mexico	
1160	Arabica	juan luis alvarado romero	Guatemala	
1166	Arabica	armando luis pohlenz martinez	Mexico	
1167	Arabica	exportadora de cafe condor s.a	Colombia	
1182	Arabica	lin, che-hao krude 林哲豪	Taiwan	

```
MARTINEZ
1167                huila                                              La
Plata
1182         natou county                                          WU SHU
YI 巫叔憶

      Number.of.Bags Bag.Weight             In.Country.Partner
Harvest.Year  \
148            250.0      69 kg  Asociacion Nacional Del Café
2014
160              3.0      60 kg         Yunnan Coffee Exchange
2015
196              3.0      60 kg         Yunnan Coffee Exchange
2015
203            250.0      69 kg  Asociacion Nacional Del Café
2014
206             10.0       1 kg                       AMECAFE
2012
...              ...        ...                           ...
...
1152           380.0       1 kg                       AMECAFE
2013
1160           250.0      69 kg  Asociacion Nacional Del Café
2013
1166           250.0       1 kg                       AMECAFE
2012
1167           250.0      70 kg                      Almacafé
4T/10
1182            50.0      20 kg  Specialty Coffee Association
2014

              Grading.Date  ... Total.Cup.Points  Moisture  \
148         May 29th, 2014  ...             84.58      0.11
160        April 6th, 2016  ...             84.50      0.10
196        April 6th, 2016  ...             84.25      0.10
203      August 21st, 2014  ...             84.25      0.10
206   September 11th, 2012  ...             84.25      0.12
...                    ...  ...               ...       ...
1152     March 29th, 2013  ...             79.75      0.12
1160      June 27th, 2013  ...             79.67      0.11
1166    August 30th, 2012  ...             79.58      0.13
1167   February 9th, 2011  ...             79.58      0.10
1182   November 7th, 2014  ...             79.25      0.11

      Category.One.Defects  Quakers  Category.Two.Defects  \
148                      0      0.0                     1
160                      0      0.0                     1
196                      0      0.0                     1
203                      0      0.0                     1
206                      0      0.0                     1
```

```
...                     ...      ...                     ...
1152                      0      0.0                       5
1160                      0      0.0                       6
1166                      0      0.0                       5
1167                      0      0.0                       4
1182                      0      0.0                       0
```

```
                Expiration            Certification.Body  \
148         May 29th, 2015  Asociacion Nacional Del Café
160        April 6th, 2017        Yunnan Coffee Exchange
196        April 6th, 2017        Yunnan Coffee Exchange
203      August 21st, 2015  Asociacion Nacional Del Café
206   September 11th, 2013                       AMECAFE
...                    ...                           ...
1152     March 29th, 2014                       AMECAFE
1160      June 27th, 2014  Asociacion Nacional Del Café
1166     August 30th, 2013                       AMECAFE
1167    February 9th, 2012                      Almacafé
1182    November 7th, 2015  Specialty Coffee Association

                         Certification.Address  \
148   b1f20fe3a819fd6b2ee0eb8fdc3da256604f1e53
160   683fa6014608fc10ff681b0435b0b2dbe6df988f
196   683fa6014608fc10ff681b0435b0b2dbe6df988f
203   b1f20fe3a819fd6b2ee0eb8fdc3da256604f1e53
206   59e396ad6e22a1c22b248f958e1da2bd8af85272
...                                       ...
1152  59e396ad6e22a1c22b248f958e1da2bd8af85272
1160  b1f20fe3a819fd6b2ee0eb8fdc3da256604f1e53
1166  59e396ad6e22a1c22b248f958e1da2bd8af85272
1167  e493c36c2d076bf273064f7ac23ad562af257a25
1182  36d0d00a3724338ba7937c52a378d085f2172daa

                         Certification.Contact  unit_of_measurement
148   724f04ad10ed31dbb9d260f0dfd221ba48be8a95                   ft
160   f6d87a6c04653c569d4911a66f89d5e30ce83b93                    m
196   f6d87a6c04653c569d4911a66f89d5e30ce83b93                    m
203   724f04ad10ed31dbb9d260f0dfd221ba48be8a95                   ft
206   0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7                    m
...                                       ...                  ...
1152  0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7                    m
1160  724f04ad10ed31dbb9d260f0dfd221ba48be8a95                   ft
1166  0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7                    m
1167  70d3c0c26f89e00fdae6fb39ff54f0d2eb1c38ab                    m
1182  0878a7d4b9d35ddbf0fe2ce69a2062cceb45a660                    m

[404 rows x 31 columns]
```

```
df1=filter[['Number.of.Bags','Aroma','Aftertaste','Acidity','Flavor']]
df1
```

```
      Number.of.Bags  Aroma  Aftertaste  Acidity  Flavor
148            250.0   7.75        7.67     7.92      12
160              3.0   7.75        7.67     7.83      14
196              3.0   7.83        7.58     7.75      12
203            250.0   7.83        7.50     7.92      12
206             10.0   7.83        7.67     7.83      14
...              ...    ...         ...      ...     ...
1152           380.0   7.25        7.00     7.17       3
1160           250.0   7.50        6.83     7.08       2
1166           250.0   7.25        6.92     7.33       1
1167           250.0   7.25        7.00     6.75       4
1182            50.0   7.08        6.83     7.25       0

[404 rows x 5 columns]
```

```
u=df1[['Number.of.Bags','Aroma','Aftertaste','Acidity']]
v=df1['Flavor']
u
```

```
      Number.of.Bags  Aroma  Aftertaste  Acidity
148            250.0   7.75        7.67     7.92
160              3.0   7.75        7.67     7.83
196              3.0   7.83        7.58     7.75
203            250.0   7.83        7.50     7.92
206             10.0   7.83        7.67     7.83
...              ...    ...         ...      ...
1152           380.0   7.25        7.00     7.17
1160           250.0   7.50        6.83     7.08
1166           250.0   7.25        6.92     7.33
1167           250.0   7.25        7.00     6.75
1182            50.0   7.08        6.83     7.25

[404 rows x 4 columns]
```

```
v

148     12
160     14
196     12
203     12
206     14
        ..
1152     3
1160     2
1166     1
1167     4
1182     0
Name: Flavor, Length: 404, dtype: int64
```

```
from sklearn.feature_selection import f_classif,SelectKBest
p_values=f_classif(u,v)
p_values
```

```
(array([ 0.85242197, 14.52699572, 26.60334715, 18.27662554]),
 array([6.11543311e-01, 4.34008194e-28, 2.18511309e-48, 6.86968975e-
35]))
```

```
selector=SelectKBest(score_func=f_classif,k=4)
new=selector.fit_transform(u,v)
new
```

```
array([[250.  ,   7.75,   7.67,   7.92],
       [  3.  ,   7.75,   7.67,   7.83],
       [  3.  ,   7.83,   7.58,   7.75],
       ...,
       [250.  ,   7.25,   6.92,   7.33],
       [250.  ,   7.25,   7.  ,   6.75],
       [ 50.  ,   7.08,   6.83,   7.25]])
```

```
p_value=pd.Series(p_values[1])
p_value.index=u.columns
p_value
```

```
Number.of.Bags    6.115433e-01
Aroma             4.340082e-28
Aftertaste        2.185113e-48
Acidity           6.869690e-35
dtype: float64
```

```
#after taste only coffee has more taste
```