EDA on Coffee Quality Dataset

Identification of attributes and implementing the DataSet

In [6]:
```
#Importing necessary libraries to perform EDA
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [7]:
```
#To read/open csv file
df=pd.read_csv("df_Coffee.csv")
print("First/Top 2 records>>>")
#Top 2 records serially
df.head(2)
```

```
First/Top 2 records>>>
```

Out[7]:

	Unnamed: 0	ID	Country of Origin	Farm Name	Lot Number	Mill	ICO Number	Company	Altitude	Region	...	Total Cup Points	Moisture Percentage	Category One Defects	C
0	0	0	Colombia	Finca El Paraiso	CQU2022015	Finca El Paraiso	NaN	Coffee Quality Union	1700-1930	Piendamo,Cauca	...	89.33	11.8	0	
1	1	1	Taiwan	Royal Bean Geisha Estate	The 2022 Pacific Rim Coffee Summit,T037	Royal Bean Geisha Estate	NaN	Taiwan Coffee Laboratory	1200	Chiayi	...	87.58	10.5	0	

2 rows × 41 columns

In [9]:
```
#Breif information of all attributes
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 207 entries, 0 to 206
Data columns (total 41 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Unnamed: 0             207 non-null    int64
 1   ID                     207 non-null    int64
 2   Country of Origin      207 non-null    object
 3   Farm Name              205 non-null    object
 4   Lot Number             206 non-null    object
 5   Mill                   204 non-null    object
 6   ICO Number             75 non-null     object
 7   Company                207 non-null    object
 8   Altitude               206 non-null    object
 9   Region                 205 non-null    object
 10  Producer               206 non-null    object
 11  Number of Bags         207 non-null    int64
 12  Bag Weight             207 non-null    object
 13  In-Country Partner     207 non-null    object
 14  Harvest Year           207 non-null    object
 15  Grading Date           207 non-null    object
 16  Owner                  207 non-null    object
 17  Variety                201 non-null    object
 18  Status                 207 non-null    object
 19  Processing Method      202 non-null    object
 20  Aroma                  207 non-null    float64
 21  Flavor                 207 non-null    float64
 22  Aftertaste             207 non-null    float64
 23  Acidity                207 non-null    float64
 24  Body                   207 non-null    float64
 25  Balance                207 non-null    float64
 26  Uniformity             207 non-null    float64
 27  Clean Cup              207 non-null    float64
 28  Sweetness              207 non-null    float64
 29  Overall                207 non-null    float64
 30  Defects                207 non-null    float64
 31  Total Cup Points       207 non-null    float64
 32  Moisture Percentage    207 non-null    float64
 33  Category One Defects   207 non-null    int64
 34  Quakers                207 non-null    int64
 35  Color                  207 non-null    object
 36  Category Two Defects   207 non-null    int64
 37  Expiration             207 non-null    object
 38  Certification Body     207 non-null    object
 39  Certification Address  207 non-null    object
 40  Certification Contact  207 non-null    object
dtypes: float64(13), int64(6), object(22)
memory usage: 66.4+ KB
```

In [10]:
```
#Statistical infromation of the given coffee data set attribute
df.describe()
```

Out[10]:

	Unnamed: 0	ID	Number of Bags	Aroma	Flavor	Aftertaste	Acidity	Body	Balance	Uniformity	Clean Cup	Swee
count	207.000000	207.000000	207.000000	207.000000	207.000000	207.000000	207.00000	207.000000	207.000000	207.000000	207.0	
mean	103.000000	103.000000	155.449275	7.721063	7.744734	7.599758	7.69029	7.640918	7.644058	9.990338	10.0	
std	59.899917	59.899917	244.484868	0.287626	0.279613	0.275911	0.25951	0.233499	0.256299	0.103306	0.0	
min	0.000000	0.000000	1.000000	6.500000	6.750000	6.670000	6.83000	6.830000	6.670000	8.670000	10.0	
25%	51.500000	51.500000	1.000000	7.580000	7.580000	7.420000	7.50000	7.500000	7.500000	10.000000	10.0	
50%	103.000000	103.000000	14.000000	7.670000	7.750000	7.580000	7.67000	7.670000	7.670000	10.000000	10.0	
75%	154.500000	154.500000	275.000000	7.920000	7.920000	7.750000	7.87500	7.750000	7.790000	10.000000	10.0	
max	206.000000	206.000000	2240.000000	8.580000	8.500000	8.420000	8.58000	8.250000	8.420000	10.000000	10.0	

In [11]:
```
# To count the total rows and columns present in the data set
df.shape
```

Out[11]:
```
(207, 41)
```

In [12]:
```
# Column names
df.columns
```

Out[12]:
```
Index(['Unnamed: 0', 'ID', 'Country of Origin', 'Farm Name', 'Lot Number',
       'Mill', 'ICO Number', 'Company', 'Altitude', 'Region', 'Producer',
       'Number of Bags', 'Bag Weight', 'In-Country Partner', 'Harvest Year',
       'Grading Date', 'Owner', 'Variety', 'Status', 'Processing Method',
       'Aroma', 'Flavor', 'Aftertaste', 'Acidity', 'Body', 'Balance',
       'Uniformity', 'Clean Cup', 'Sweetness', 'Overall', 'Defects',
       'Total Cup Points', 'Moisture Percentage', 'Category One Defects',
       'Quakers', 'Color', 'Category Two Defects', 'Expiration',
       'Certification Body', 'Certification Address', 'Certification Contact'],
      dtype='object')
```

In [13]:
```
# Determining the start and the stop index
df.index
```

Out[13]:
```
RangeIndex(start=0, stop=207, step=1)
```

Observation

- In this block,there were basic syntax to get a brief overview of the dataset.Here are few conclusions drawn: -There are 207 rows and 40 columns. -Also drawn the statistical inference. -The index starts from 0 and goes till 207.

DataCleaning

In [42]:
```
#Removing Unwanted column
df.drop('Unnamed: 0',axis=1)
```

Out[42]:

	ID	Country of Origin	Farm Name	Lot Number	Mill	ICO Number	Company	Altitude	Region	Producer	...	Total Cup Points
0	0	Colombia	Finca El Paraiso	CQU2022015	Finca El Paraiso	NaN	Coffee Quality Union	1700-1930	Piendamo,Cauca	Diego Samuel Bermudez	...	89.33
1	1	Taiwan	Royal Bean Geisha Estate	The 2022 Pacific Rim Coffee Summit,T037	Royal Bean Geisha Estate	NaN	Taiwan Coffee Laboratory	1200	Chiayi	曾福森	...	87.58
2	2	Laos	OKLAO coffee farms	The 2022 Pacific Rim Coffee Summit,LA01	oklao coffee processing plant	NaN	Taiwan Coffee Laboratory	1300	Laos Borofen Plateau	WU TAO CHI	...	87.42
3	3	Costa Rica	La Cumbre	CQU2022017	La Montana Tarrazu MIll	NaN	Coffee Quality Union	1900	Los Santos,Tarrazu	Santa Maria de Dota	...	87.17
4	4	Colombia	Finca Santuario	CQU2023002	Finca Santuario	NaN	Coffee Quality Union	1850-2100	Popayan,Cauca	Camilo Merizalde	...	87.08
...	...	...	...	...	...	...	...	...	...	...	...	...
202	202	Brazil	Fazenda Conquista	019/22	Dry Mill	NaN	Ipanema Coffees	950	Sul de Minas	Ipanema Coffees	...	80.08
203	203	Nicaragua	Finca San Felipe	017-053-0155	Beneficio Atlantic Sébaco	017-053-0155	Exportadora Atlantic S.A	1200	Matagalpa	Exportadora Atlantic S.A.	...	80.00
204	204	Laos	-	105/3/VL7285-005	DRY MILL	105/3/VL7285-005	Marubeni Corporation	1300	Bolaven Plateau	LAO MINH TIEN COFFEE SOLE CO.,LTD	...	79.67
205	205	El Salvador	Rosario de Maria II, Area de La Pila	0423A01	Optimum Coffee, San Salvador, El Salvador	NaN	Aprentium Enterprises LLC	1200	Volcan de San Vicente, La Paz, El Salvador	Roselia Yglesias	...	78.08
206	206	Brazil	Walter Matter	1058 y 1059	Beneficio humedo/seco	002/1208/1016	Descafeinadores Mexicano SA. de CV	850-1100	Minas Gerais	Walter Matter	...	78.00

207 rows × 40 columns

In [46]:
```
#To rename a column from bag_weight to bag_weight\kg
df1=df.rename(
    columns={
        'bag_weight': 'bag_weight\kg'
    })
```

In [46]:
```
# Top 5 rows from df1 dataSet
df1.head()
```

Out[46]:

	Unnamed: 0	ID	Country of Origin	Farm Name	Lot Number	Mill	ICO Number	Company	Altitude	Region	...	Total Cup Points	Moisture Percentage	Catego Oı Defec
0	0	0	Colombia	Finca El Paraiso	CQU2022015	Finca El Paraiso	NaN	Coffee Quality Union	1700-1930	Piendamo,Cauca	...	89.33	11.8	
1	1	1	Taiwan	Royal Bean Geisha Estate	The 2022 Pacific Rim Coffee Summit,T037	Royal Bean Geisha Estate	NaN	Taiwan Coffee Laboratory	1200	Chiayi	...	87.58	10.5	
2	2	2	Laos	OKLAO coffee farms	The 2022 Pacific Rim Coffee Summit,LA01	oklao coffee processing plant	NaN	Taiwan Coffee Laboratory	1300	Laos Borofen Plateau	...	87.42	10.4	
3	3	3	Costa Rica	La Cumbre	CQU2022017	La Montana Tarrazu MIll	NaN	Coffee Quality Union	1900	Los Santos,Tarrazu	...	87.17	11.8	
4	4	4	Colombia	Finca Santuario	CQU2023002	Finca Santuario	NaN	Coffee Quality Union	1850-2100	Popayan,Cauca	...	87.08	11.6	

5 rows × 41 columns

In [15]:
```
dfn=df1.select_dtypes(exclude='object')
```

In [16]:
```
# To check whether the given rows have duplicate values
dfn.duplicated()
```

Out[16]:
```
0      False
1      False
2      False
3      False
4      False
       ...
202    False
203    False
204    False
205    False
206    False
Length: 207, dtype: bool
```

In [17]:
```
dfn.duplicated().sum()
```

Out[17]:
```
0
```

In [18]:
```
# To check whether the given rows have null values
dfn.isnull()
```

Out[18]:

	Unnamed: 0	ID	Number of Bags	Aroma	Flavor	Aftertaste	Acidity	Body	Balance	Uniformity	Clean Cup	Sweetness	Overall	Defects	Total Cup Points	F
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
202	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
203	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
204	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
205	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	
206	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	

207 rows × 19 columns

In [19]:
```
dfn.isnull().sum()
```

Out[19]:
```
Unnamed: 0              0
ID                      0
Number of Bags          0
Aroma                   0
Flavor                  0
Aftertaste              0
Acidity                 0
Body                    0
Balance                 0
Uniformity              0
Clean Cup               0
Sweetness               0
Overall                 0
Defects                 0
Total Cup Points        0
Moisture Percentage     0
Category One Defects    0
Quakers                 0
Category Two Defects    0
dtype: int64
```

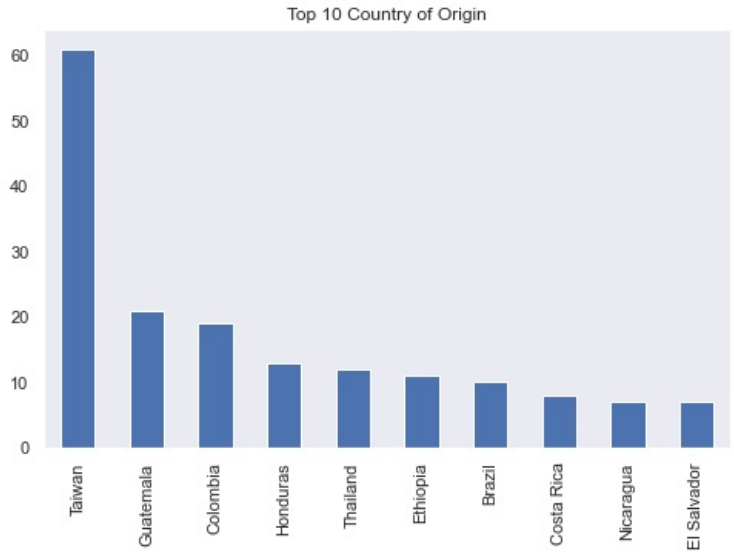
Observations

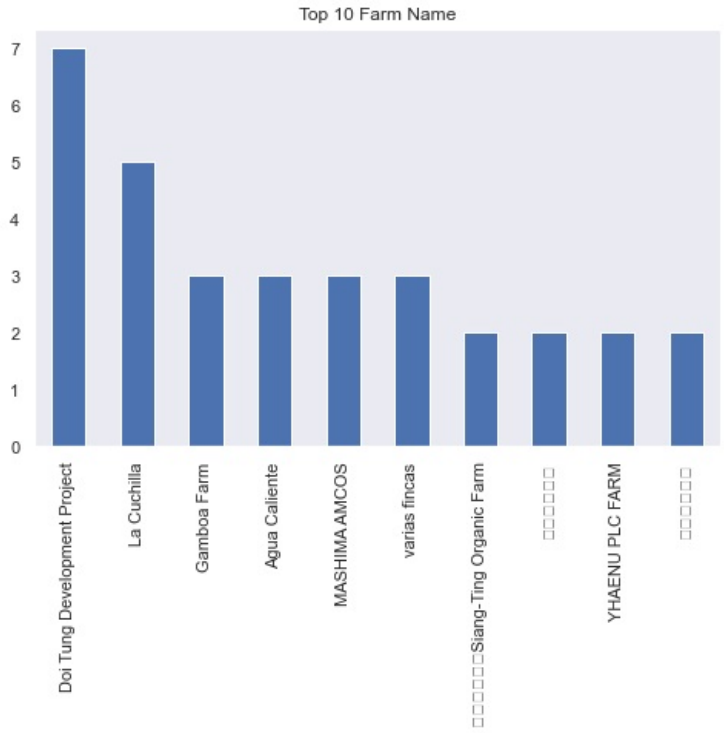
-There existed a column with no significance 'Unnamed: 0',that has been dropped.
-All the values are checked whether there are null present in the dataset,but none of them has null values.
-All the values are checked whether there are duplicated present in the dataset,but none of them has duplicated values.

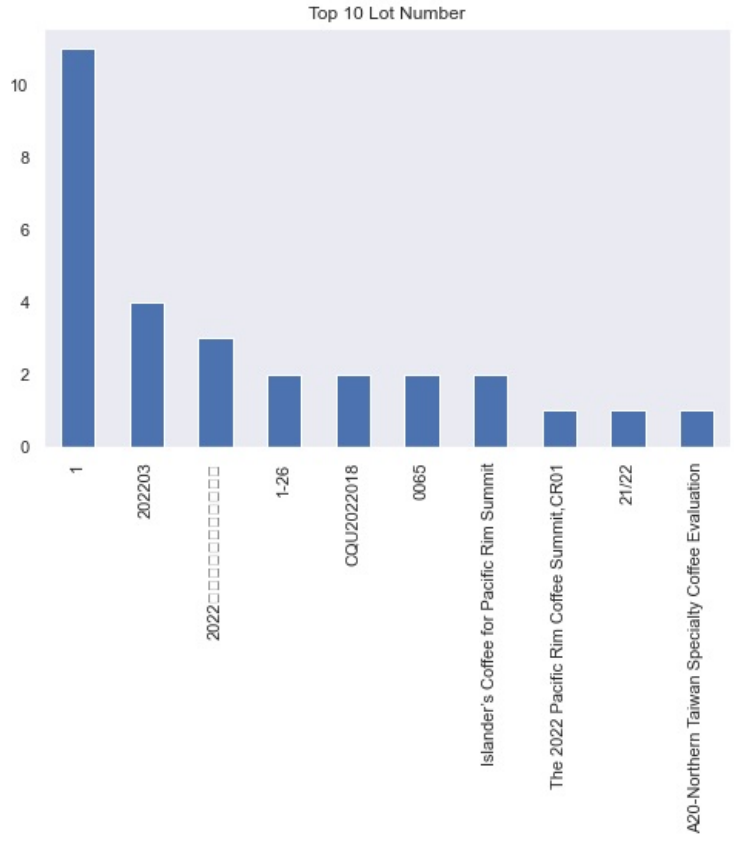
EDA on coffee data set

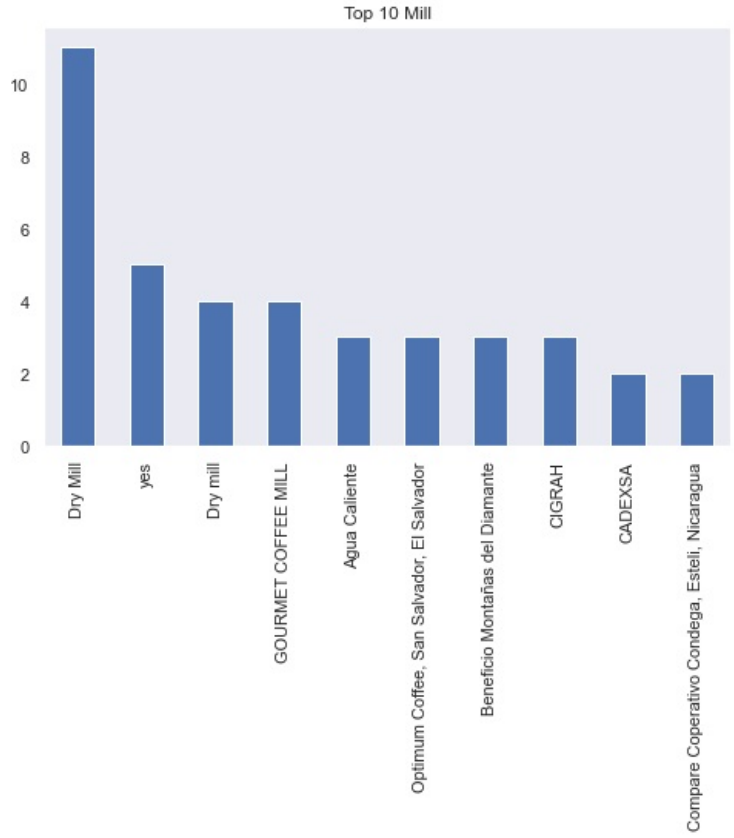
In [20]:
```
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
sns.set()
```

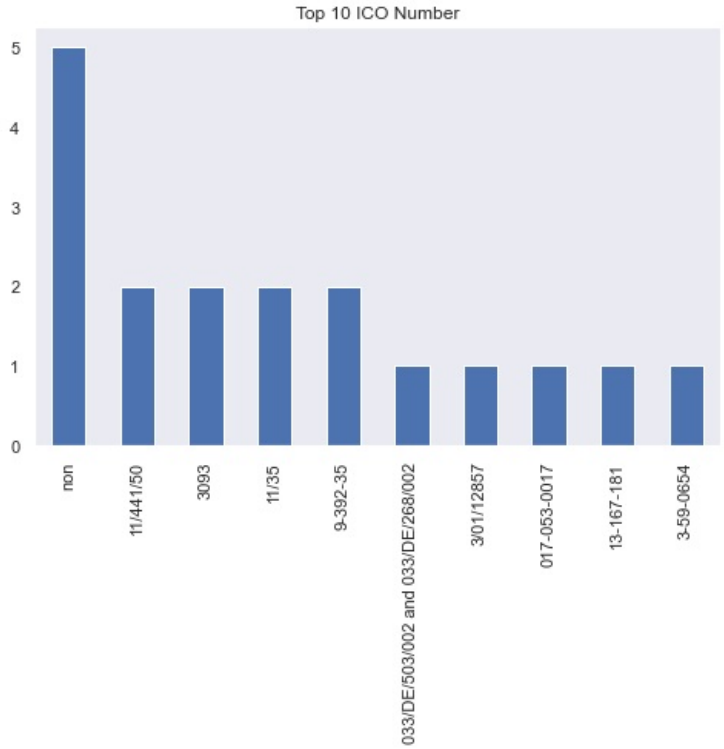
In [48]:
```
cat_list = df1.select_dtypes(include=["object"]).columns.tolist()
for col in cat_list:
    plt.figure(figsize=(8,5))
    top10 = df1[col].value_counts()[:10]
    top10.plot(kind='bar')
    plt.title("Top 10 " + col)
    plt.grid(visible=False)
    plt.show()
```

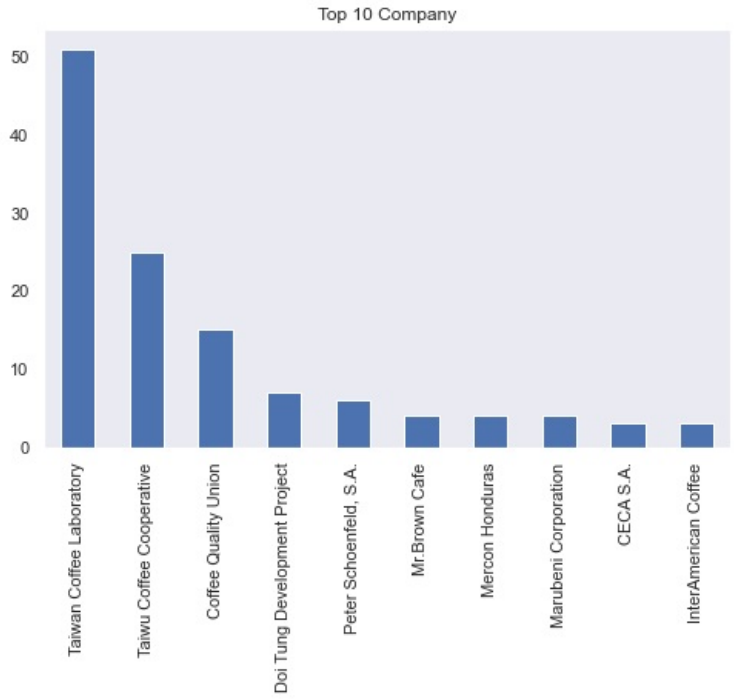


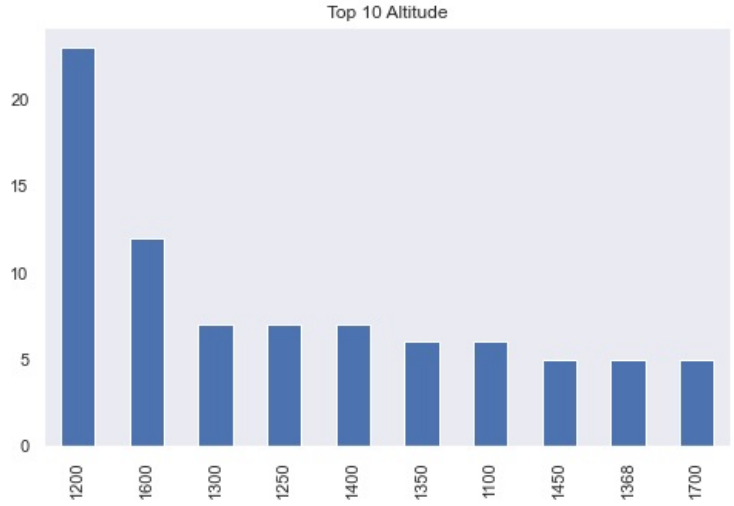

Top 10 Farm Name
7
6
5
4
3
2
1
0
Doi Tung Development Project
La Cuchilla
Gamboa Farm
Agua Caliente
MASHIMA AMCOS
varias fincas
□□□□□□Siang-Ting Organic Farm
□□□□□□
YHAENU PLC FARM
□□□□□

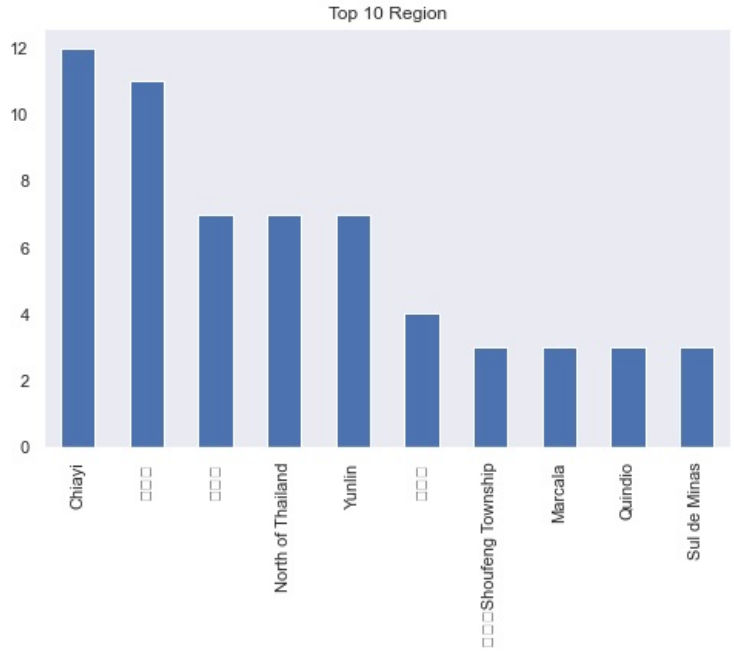

Top 10 Lot Number
10
8
6
4
2
0
1
202203
2022□□□□□□□□□□
1-26
COU2022018
0065
Islander's Coffee for Pacific Rim Summit
The 2022 Pacific Rim Coffee Summit,CR01
21/22
A20-Northern Taiwan Specialty Coffee Evaluation

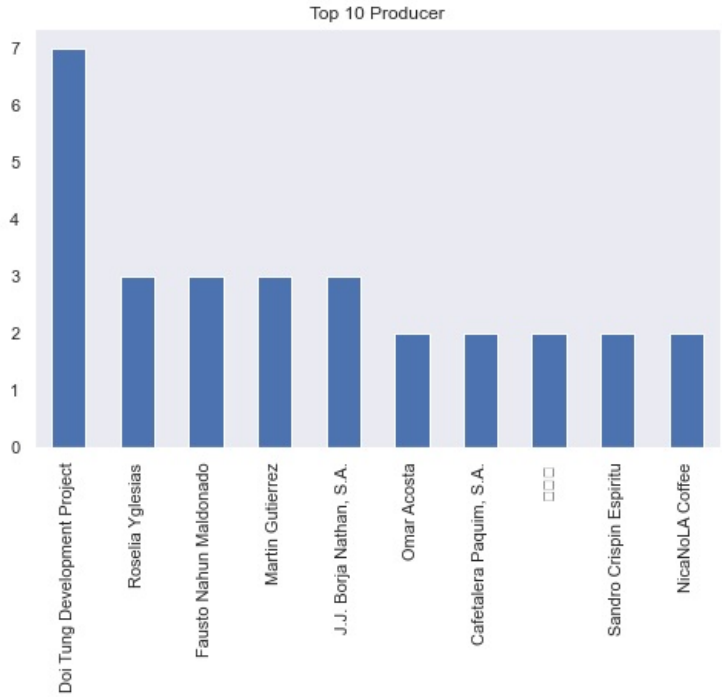

Top 10 Mill
10
8
6
4
2
0
Dry Mill
yes
Dry mill
GOURMET COFFEE MILL
Agua Caliente
Optimum Coffee, San Salvador, El Salvador
Beneficio Montañas del Diamante
CIGRAH
CADEXSA
Compare Coperativo Condega, Esteli, Nicaragua

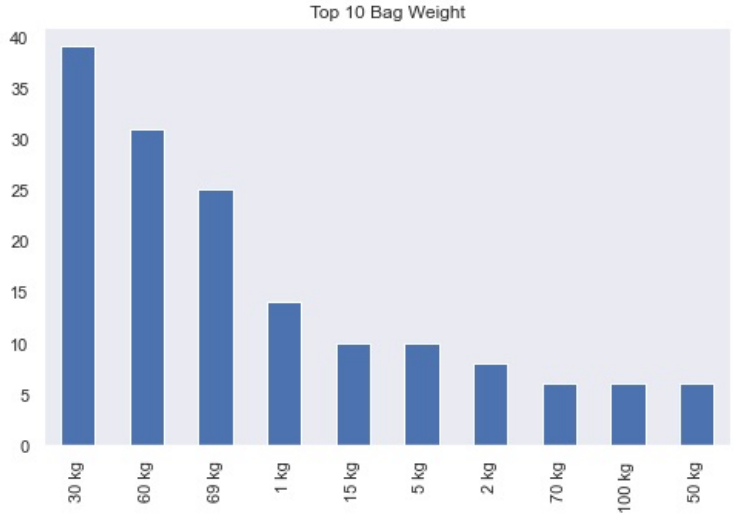

Top 10 ICO Number
5
4
3
2
1
0
non
11/441/50
3093
11/35
9-392-35
033/DE/503/002 and 033/DE/268/002
3/01/12857
017-053-0017
13-167-181
3-59-0654

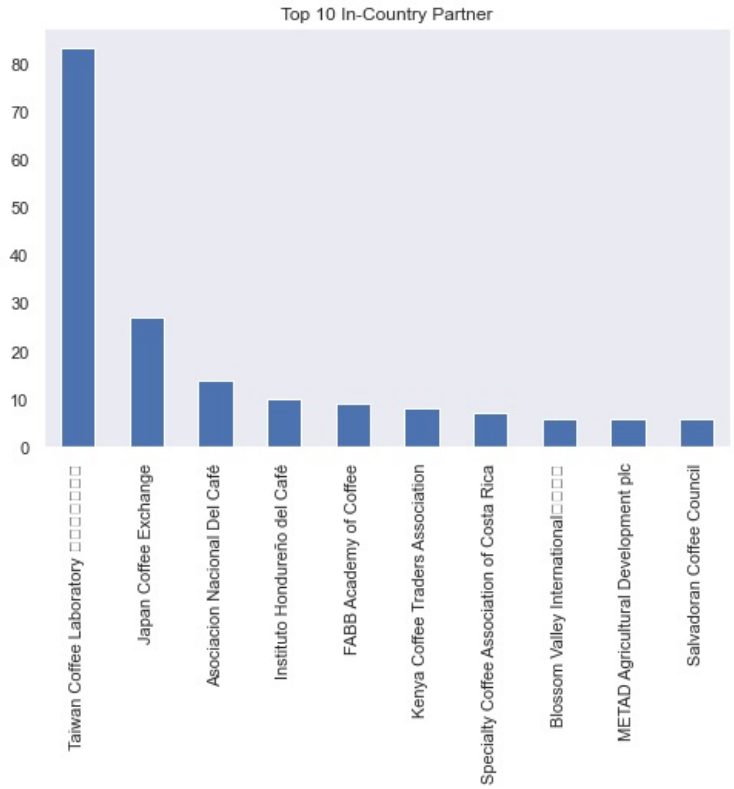

Top 10 Company
50
40
30
20
10
0
Taiwan Coffee Laboratory
Taiwu Coffee Cooperative
Coffee Quality Union
Doi Tung Development Project
Peter Schoenfeld, S.A.
Mr.Brown Cafe
Mercon Honduras
Marubeni Corporation
CECA S.A.
InterAmerican Coffee

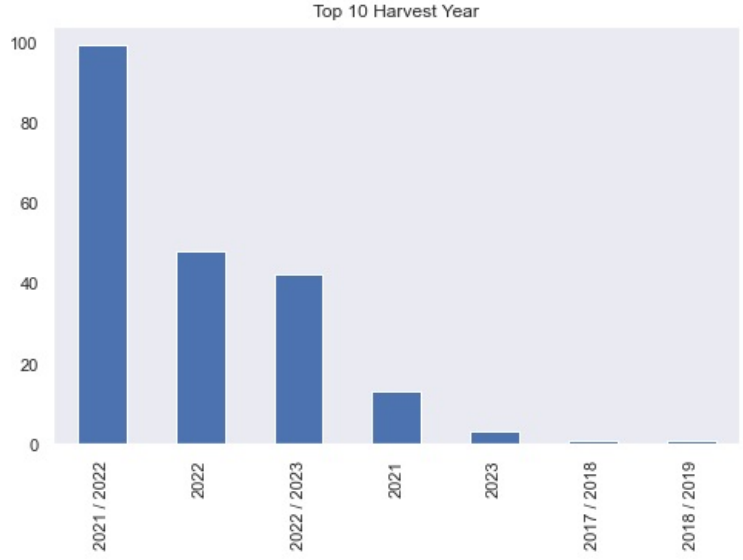

Top 10 Altitude
20
15
10
5
0
1200
1600
1300
1250
1400
1350
1100
1450
1368
1700

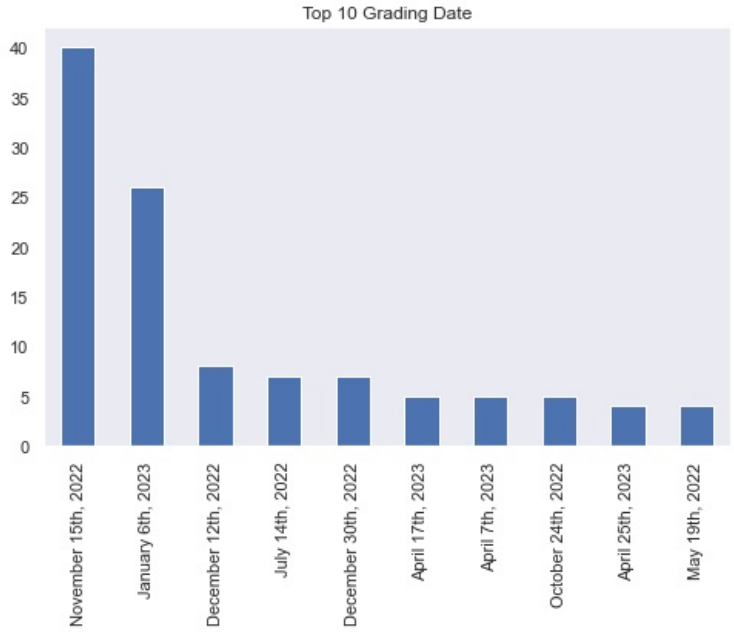

Top 10 Region
12
10
8
6
4
2
0
Chiayi
□□□
□□□
North of Thailand
Yunlin
□□□
□□□Shoufeng Township
Marcala
Quindio
Sul de Minas

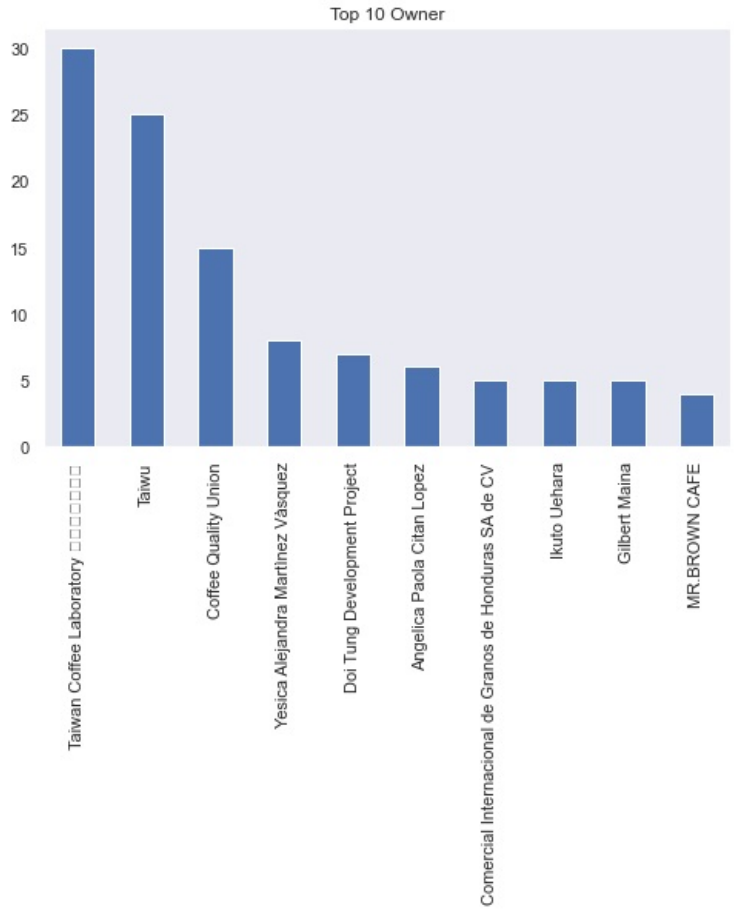

Top 10 Producer
7
6
5
4
3
2
1
0
Doi Tung Development Project
Roselia Yglesias
Fausto Nahun Maldonado
Martin Gutierrez
J.J. Borja Nathan, S.A.
Omar Acosta
Cafetalera Paquim, S.A.
□□□
Sandro Crispin Espiritu
NicaNoLA Coffee

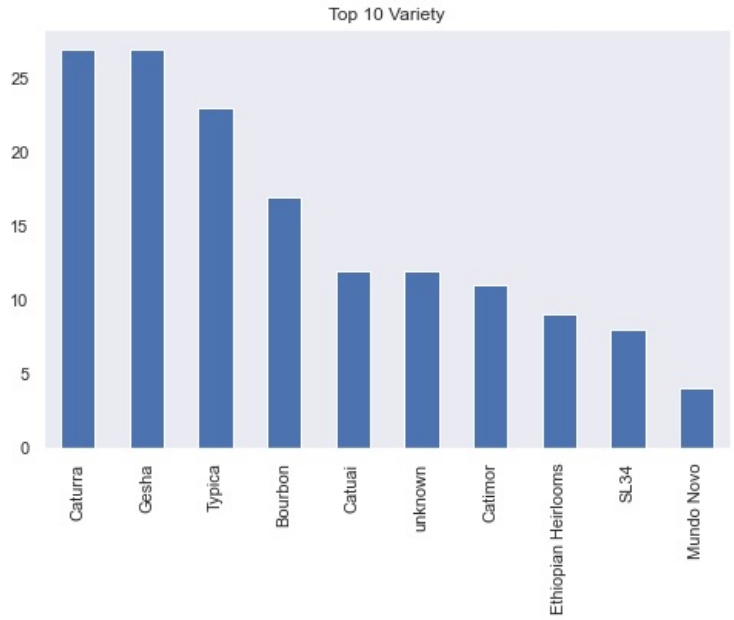

Top 10 Bag Weight
40
35
30
25
20
15
10
5
0
30 kg
60 kg
69 kg
1 kg
15 kg
5 kg
2 kg
70 kg
100 kg
50 kg

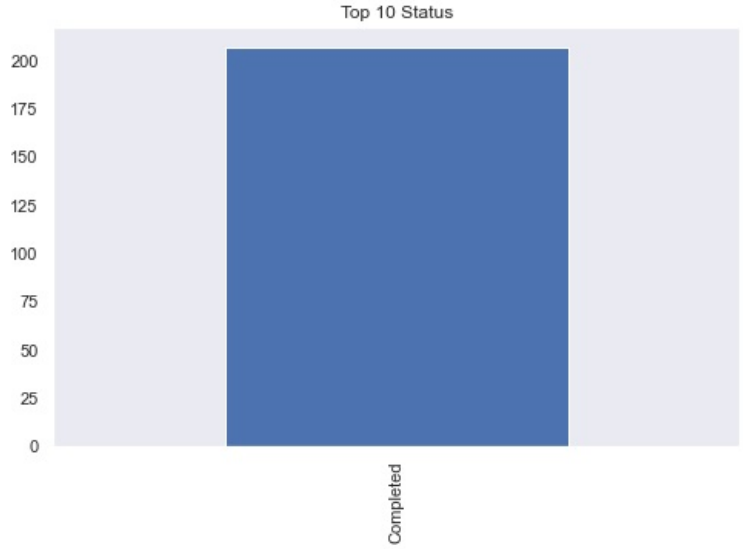

Top 10 In-Country Partner
80
70
60
50
40
30
20
10
0
Taiwan Coffee Laboratory □□□□□□
Japan Coffee Exchange
Asociacion Nacional Del Café
Instituto Hondureño del Café
FABB Academy of Coffee
Kenya Coffee Traders Association
Specialty Coffee Association of Costa Rica
Blossom Valley International□□□
METAD Agricultural Development plc
Salvadoran Coffee Council

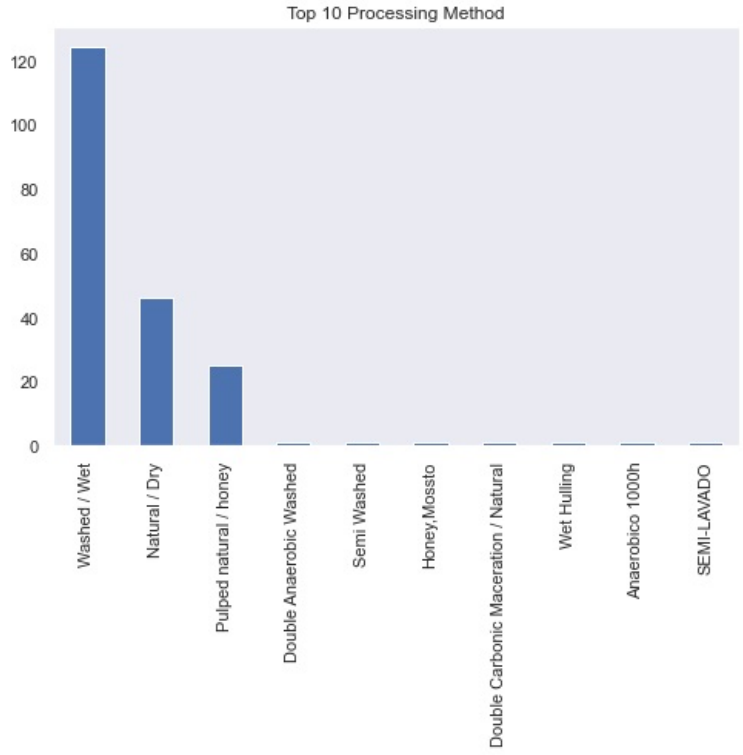

Top 10 Harvest Year
100
80
60
40
20
0
2021 / 2022
2022
2022 / 2023
2021
2023
2017 / 2018
2018 / 2019

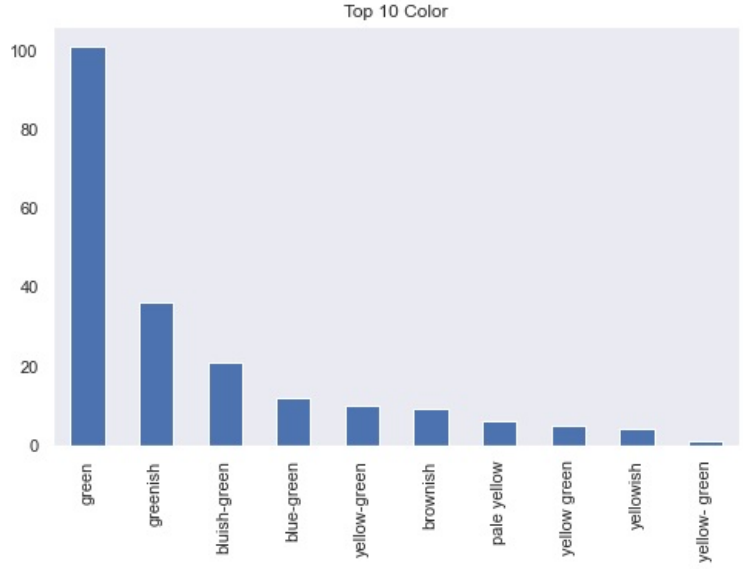

Top 10 Grading Date
40
35
30
25
20
15
10
5
0
November 15th, 2022
January 6th, 2023
December 12th, 2022
July 14th, 2022
December 30th, 2022
April 17th, 2023
April 7th, 2023
October 24th, 2022
April 25th, 2023
May 19th, 2022

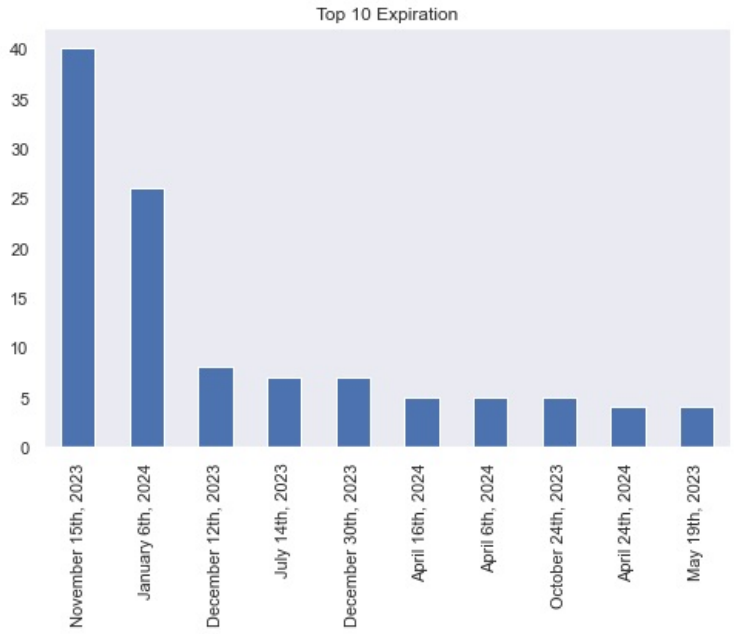

Top 10 Owner
30
25
20
15
10
5
0
Taiwan Coffee Laboratory □□□□□□
Taiwu
Coffee Quality Union
Yesica Alejandra Martínez Vásquez
Doi Tung Development Project
Angélica Paola Citan Lopez
Comercial Internacional de Granos de Honduras SA de CV
Ikuto Uehara
Gilbert Maina
MR.BROWN CAFE

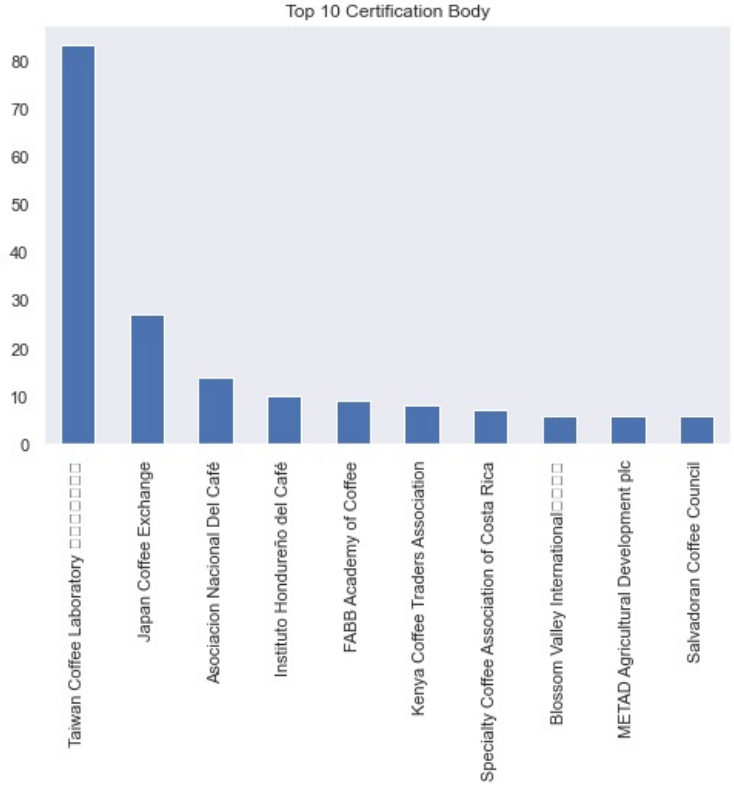

Top 10 Variety
25
20
15
10
5
0
Caturra
Gesha
Typica
Bourbon
Catuai
unknown
Catimor
Ethiopian Heirlooms
SL34
Mundo Novo

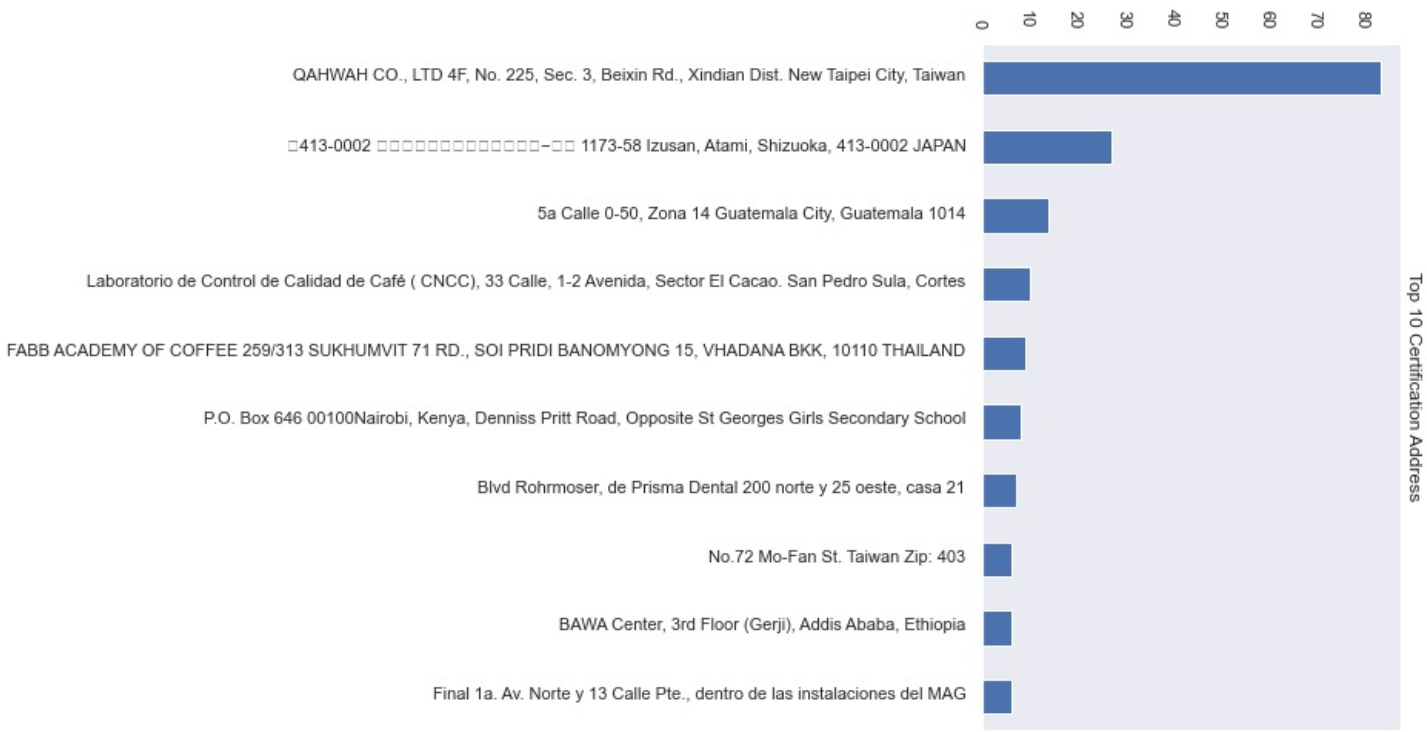

Top 10 Status
200
175
150
125
100
75
50
25
0
Completed

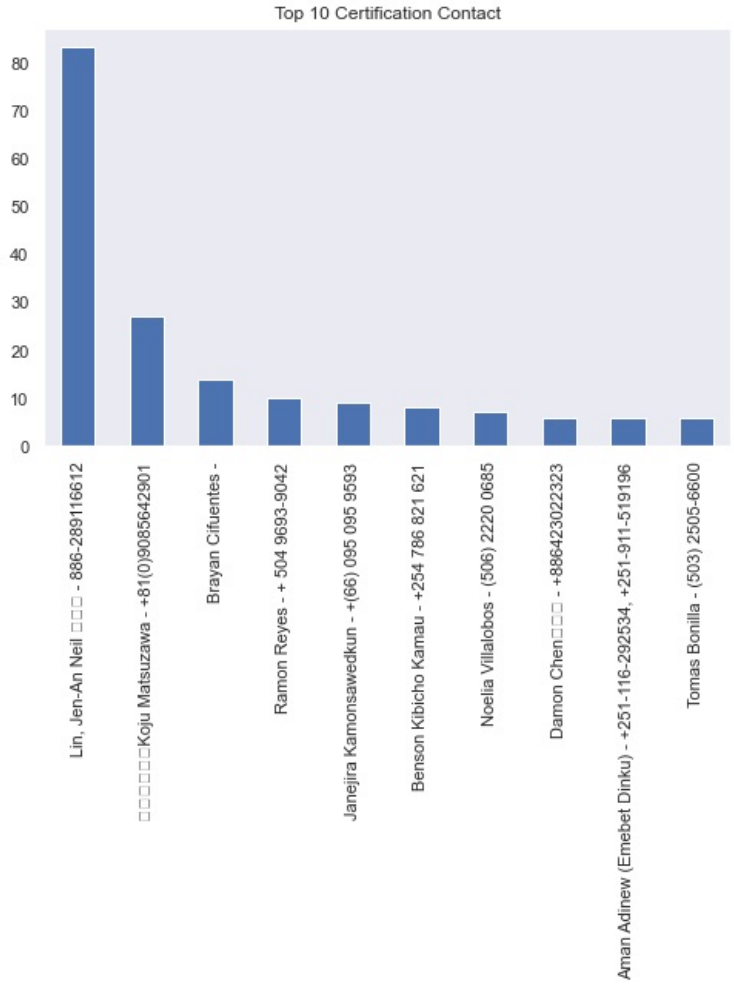

Top 10 Processing Method
120
100
80
60
40
20
0
Washed / Wet
Natural / Dry
Pulped natural / honey
Double Anaerobic Washed
Semi Washed
Honey,Mossto
Double Carbonic Maceration / Natural
Wet Hulling
Anaerobico 1000h
SEMI-LAVADO


Top 10 Color
100
80
60
40
20
0
green
greenish
bluish-green
blue-green
yellow-green
brownish
pale yellow
yellow green
yellowish
yellow- green


Top 10 Expiration
40
35
30
25
20
15
10
5
0
November 15th, 2023
January 6th, 2024
December 12th, 2023
July 14th, 2023
December 30th, 2023
April 16th, 2024
April 6th, 2024
October 24th, 2023
April 24th, 2024
May 19th, 2023


Top 10 Certification Body
80
70
60
50
40
30
20
10
0
Taiwan Coffee Laboratory □□□□□□
Japan Coffee Exchange
Asociacion Nacional Del Café
Instituto Hondureño del Café
FABB Academy of Coffee
Kenya Coffee Traders Association
Specialty Coffee Association of Costa Rica
Blossom Valley International□□□
METAD Agricultural Development plc
Salvadoran Coffee Council


0
10
20
30
40
50
60
70
80
QAHWAH CO., LTD 4F, No. 225, Sec. 3, Beixin Rd., Xindian Dist. New Taipei City, Taiwan
□413-0002 □□□□□□□□□□□□□□−□□ 1173-58 Izusan, Atami, Shizuoka, 413-0002 JAPAN
5a Calle 0-50, Zona 14 Guatemala City, Guatemala 1014
Laboratorio de Control de Calidad de Café (CNCC), 33 Calle, 1-2 Avenida, Sector El Cacao. San Pedro Sula, Cortes
FABB ACADEMY OF COFFEE 259/313 SUKHUMVIT 71 RD., SOI PRIDI BANOMYONG 15, VHADANA BKK, 10110 THAILAND
P.O. Box 646 00100Nairobi, Kenya, Denniss Pritt Road, Opposite St Georges Girls Secondary School
Blvd Rohrmoser, de Prisma Dental 200 norte y 25 oeste, casa 21
No.72 Mo-Fan St. Taiwan Zip: 403
BAWA Center, 3rd Floor (Gerji), Addis Ababa, Ethiopia
Final 1a. Av. Norte y 13 Calle Pte., dentro de las instalaciones del MAG
Top 10 Certification Address



In [21]:
```
# company wise count
df1['Company'].value_counts()#Return a Series containing counts of unique rows in the DataFrame.
```

Out[21]:
```
Taiwan Coffee Laboratory        51
Taiwu Coffee Cooperative        25
Coffee Quality Union            15
Doi Tung Development Project     7
Peter Schoenfeld, S.A.           6
                                ..
Taylor Winch Coffee Ltd          1
ECOM COLOMBIA                    1
Exportadora Café California      1
Coffee Quality Institute         1
marubeni                         1
Name: Company, Length: 72, dtype: int64
```

In [65]:
```
# Variety wise count
df1['Variety'].value_counts()#Return a Series containing counts of unique rows in the DataFrame.
```

Out[65]:
```
Caturra                                                          27
Gesha                                                            27
Typica                                                           23
Bourbon                                                          17
Catuai                                                           12
unknown                                                          12
Catimor                                                          11
Ethiopian Heirlooms                                               9
SL34                                                              8
Mundo Novo                                                        4
SL14                                                              3
Yellow Bourbon                                                    3
SHG                                                               3
Java                                                              3
Maragogype                                                        2
Parainema                                                         2
Pacamara                                                          2
Sarchimor                                                         2
SL28                                                              2
Santander                                                         1
Typica Gesha                                                      1
Catucai                                                           1
Yellow Catuai                                                     1
SL28,SL34,Ruiru11                                                 1
Caturra-Catuai                                                    1
Typica Bourbon Caturra Catimor                                    1
Caturra,Colombia,Castillo                                         1
Castillo,Caturra,Bourbon                                          1
unknow                                                            1
Bourbon, Catimor, Caturra, Typica                                 1
Pacas                                                             1
Gayo                                                              1
Castillo                                                          1
Lempira                                                           1
Red Bourbon,Caturra                                               1
MARSELLESA, CATUAI, CATURRA & MARSELLESA, ANACAFE 14, CATUAI      1
Typica + SL34                                                     1
Catimor,Catuai,Caturra,Bourbon                                    1
Bourbon Sidra                                                     1
BOURBON, CATURRA Y CATIMOR                                        1
Jember,TIM-TIM,Ateng                                              1
Castillo and Colombia blend                                       1
Catrenic                                                          1
Castillo Paraguaycito                                             1
Wolishalo,Kurume,Dega                                             1
Sl34+Gesha                                                        1
Red Bourbon                                                       1
Catuai and Mundo Novo                                             1
Name: Variety, dtype: int64
```

In [22]:
```
# color wise count
df1['Color'].value_counts()#Return a Series containing counts of unique rows in the DataFrame.
```

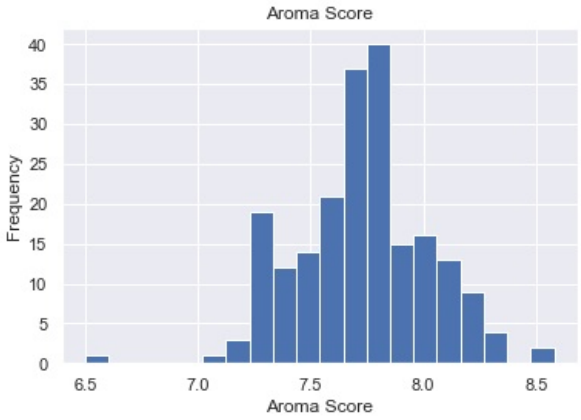
Out[22]:
```
green            101
greenish          36
bluish-green      21
blue-green        12
yellow-green      10
brownish           9
pale yellow        6
yellow green       5
yellowish          4
yellow- green      1
browish-green      1
yello-green        1
Name: Color, dtype: int64
```

In [56]:
```
df1.groupby('Country of Origin')['Country of Origin'].agg('count')

#A groupby operation involves some combination of splitting the
# object, applying a function, and combining the results.
```

Out[56]:
```
Country of Origin
Brazil                           10
Colombia                         19
Costa Rica                        8
El Salvador                       7
Ethiopia                         11
Guatemala                        21
Honduras                         13
Indonesia                         3
Kenya                             2
Laos                              3
Madagascar                        1
Mexico                            4
Myanmar                           1
Nicaragua                         7
Panama                            2
Peru                              4
Taiwan                           61
Tanzania, United Republic Of      6
Thailand                         12
Uganda                            3
United States (Hawaii)            5
Vietnam                           4
Name: Country of Origin, dtype: int64
```

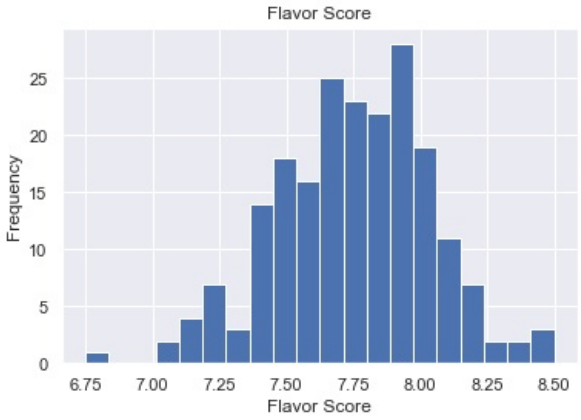
In [35]:
```
#Creating a GRAPH to analyze AROMA distribution
ax = df1['Aroma'].plot(kind='hist',bins=20,title= 'Aroma Score')
ax.set_xlabel("Aroma Score")
plt.show()
```



observation

- Aroma score,i.e,between 7.5 to 8.0, has the highest frequency.

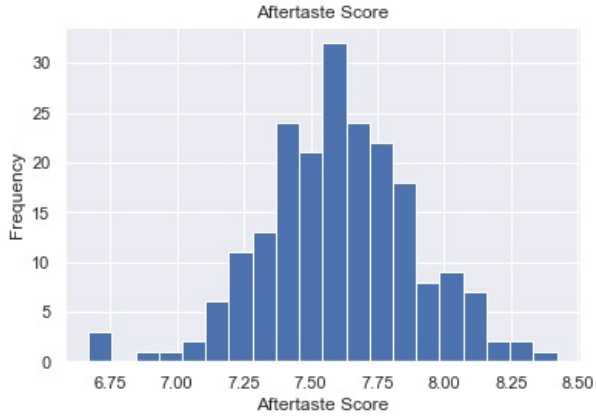
In [33]:
```
#Creating a GRAPH to analyze Flavour distribution
ax = df1['Flavor'].plot(kind='hist', bins=20, title= 'Flavor Score')
ax.set_xlabel("Flavor Score")
plt.show()
```



Observation

- Flavor score,i.e,between 7.75 to 8.0, has the highest frequency.

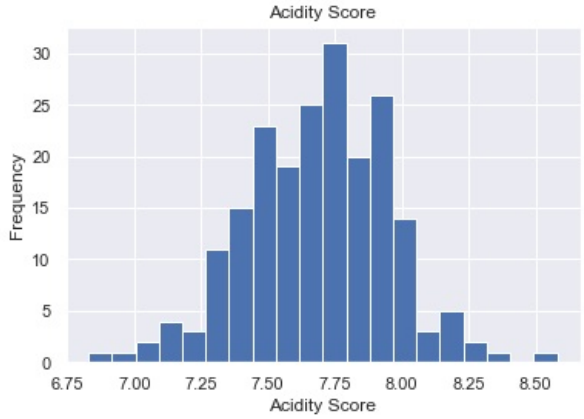
In [34]:
```
#Creating a GRAPH to analyze After Taste distribution
ax = df1['Aftertaste'].plot(kind='hist', bins=20, title= 'Aftertaste Score')
ax.set_xlabel("Aftertaste Score")
plt.show()
```



Observation

- Aftertaste score,i.e,between 7.50 to 7.75, has the highest frequency.

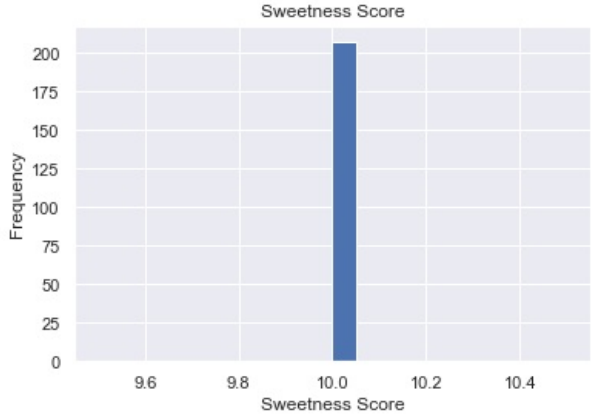
In [36]:
```
#Creating a GRAPH to analyze Acidity Analysis distribution
ax = df1['Acidity'].plot(kind='hist', bins=20, title= 'Acidity Score')
ax.set_xlabel("Acidity Score")
plt.show()
```



Observation

- Acidity score,i.e,between 7.50 to 8.0, has the highest frequency.

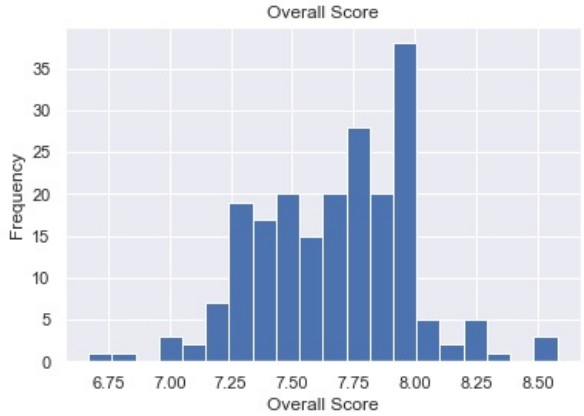
In [37]:
```
#Creating a GRAPH to analyze Sweetness Analysis distribution
ax = df1['Sweetness'].plot(kind='hist', bins=20, title= 'Sweetness Score')
ax.set_xlabel("Sweetness Score")
plt.show()
```



Observation

- Sweetness score is 10 for frequency 200+,this signifies all the flavours/coffee has sweetness score 10.

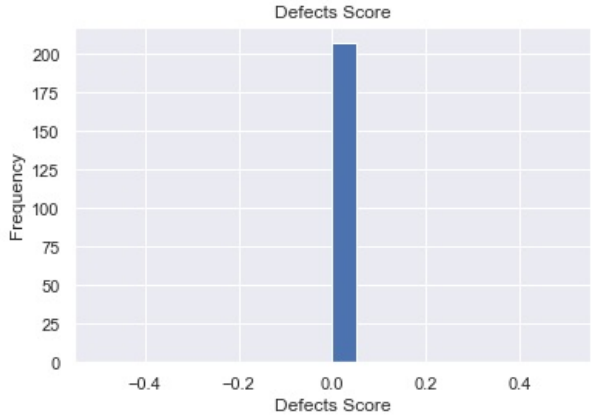
In [38]:
```
#Creating a GRAPH to analyze Overall Analysis distribution
ax = df1['Overall'].plot(kind='hist', bins=20, title= 'Overall Score')
ax.set_xlabel("Overall Score" )
plt.show()
```



Observation

- Overall score,i.e,between 7.50 to 8.0, has the highest frequency.

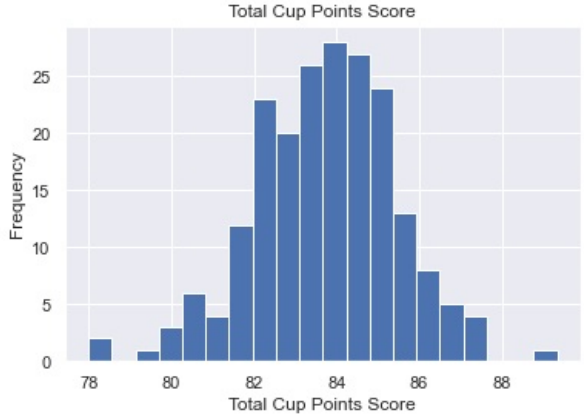
In [39]:
```
#Creating a GRAPH to analyze Defects Analysis distribution
ax = df1['Defects'].plot(kind='hist', bins=20, title= 'Defects Score')
ax.set_xlabel("Defects Score")
plt.show()
```



Observation

- Defects score is 0.0 for 200+ frequencies, this signifies all the rows have no defects.

In [40]:
```
#Creating a GRAPH to analyze Total Cup Points distribution
ax = df1['Total Cup Points'].plot(kind='hist', bins=20, title= 'Total Cup Points Score')
ax.set_xlabel("Total Cup Points Score")
plt.show()
```



Observation

- Total cup points score,i.e,from 82 to 86, has the highest frequency.

In [53]:
```
# Plotting a scatter plot of flavor and aftertaste scores
plt.figure(figsize=(8, 6))

# Plotting the scatter plot
plt.scatter(df1['Flavor'], df1['Aftertaste'])

# Adding labels and title
```

```
plt.xlabel('Flavor')
plt.ylabel('Aftertaste')
plt.title('Scatter Plot of Flavor and Aftertaste Scores')

# Displaying the plot
plt.show()
```

In [47]:
```
# Determining the countries Maximum and the Minimum total cup score
mean_country_cupscore = df1.groupby('Country of Origin')['Total Cup Points'].mean().reset_index().head(10)
mean_country_cupscore.sort_values('Total Cup Points', ascending=False)
```

Out[47]:

	Country of Origin	Total Cup Points
4	Ethiopia	84.960909
5	Guatemala	84.301429
1	Colombia	83.877368
2	Costa Rica	83.740000
8	Kenya	83.710000
7	Indonesia	83.693333
9	Laos	83.390000
6	Honduras	83.282308
0	Brazil	81.883000
3	El Salvador	81.532857

Observations

-Ethiopia has the highest total cup score with mean score 84.960909.
-El Salvador has the lowest total cup score with mean score 81.532857.