

```
In [16]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from datetime import datetime
import dateutil.parser
```

```
pd.set_option('display.max_columns', 15)
pd.set_option('expand_frame_repr', True)

sns.set_palette('hls')
%matplotlib inline
```

```
In [17]: chocolate_bars = pd.read_csv('chocolate_bars.csv', index_col='id', parse_dates=True)
chocolate_bars
```

Out[17]:

	manufacturer	company_location	year_reviewed	bean_origin	bar_name	cocoa_percent	nu
id							
2454	5150	U.S.A.	2019	Tanzania	Kokoa Kamili, batch 1	76.0	
2458	5150	U.S.A.	2019	Dominican Republic	Zorzal, batch 1	76.0	
2454	5150	U.S.A.	2019	Madagascar	Bejofo Estate, batch 1	76.0	
2542	5150	U.S.A.	2021	Fiji	Matasawalevu, batch 1	68.0	
2546	5150	U.S.A.	2021	Venezuela	Sur del Lago, batch 1	72.0	
...
1205	Zotter	Austria	2014	Blend	Raw	80.0	
1996	Zotter	Austria	2017	Colombia	APROCAFA, Acandi	75.0	
2036	Zotter	Austria	2018	Blend	Dry Aged, 30 yr Anniversary bar	75.0	
2170	Zotter	Austria	2018	Congo	Mountains of the Moon	70.0	
2170	Zotter	Austria	2018	Belize	Maya Mtn	72.0	

2530 rows × 10 columns

```
In [3]: chocolate_bars.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2530 entries, 2454 to 2170
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   manufacturer    2530 non-null    object  
 1   company_location 2530 non-null    object  
 2   year_reviewed    2530 non-null    int64   
 3   bean_origin      2530 non-null    object  
 4   bar_name          2530 non-null    object  
 5   cocoa_percent     2530 non-null    float64 
 6   num_ingredients  2443 non-null    float64 
 7   ingredients       2443 non-null    object  
 8   review            2530 non-null    object  
 9   rating            2530 non-null    float64 
dtypes: float64(3), int64(1), object(6)
memory usage: 217.4+ KB
```

```
In [18]: cb1= chocolate_bars.dropna()
cb1
```

Out[18]:

	manufacturer	company_location	year_reviewed	bean_origin	bar_name	cocoa_percent	nu
id							
2454	5150	U.S.A.	2019	Tanzania	Kokoa Kamili, batch 1	76.0	
2458	5150	U.S.A.	2019	Dominican Republic	Zorzal, batch 1	76.0	
2454	5150	U.S.A.	2019	Madagascar	Bejofo Estate, batch 1	76.0	
2542	5150	U.S.A.	2021	Fiji	Matasawalevu, batch 1	68.0	
2546	5150	U.S.A.	2021	Venezuela	Sur del Lago, batch 1	72.0	
...
1205	Zotter	Austria	2014	Blend	Raw	80.0	
1996	Zotter	Austria	2017	Colombia	APROCAFA, Acandi	75.0	
2036	Zotter	Austria	2018	Blend	Dry Aged, 30 yr Anniversary bar	75.0	
2170	Zotter	Austria	2018	Congo	Mountains of the Moon	70.0	
2170	Zotter	Austria	2018	Belize	Maya Mtn	72.0	

2443 rows × 10 columns

In [19]:

```
cb2= cb1.drop_duplicates()
cb2
```

Out[19]:

	manufacturer	company_location	year_reviewed	bean_origin	bar_name	cocoa_percent	nu
id							
2454	5150	U.S.A.	2019	Tanzania	Kokoa Kamili, batch 1	76.0	
2458	5150	U.S.A.	2019	Dominican Republic	Zorzal, batch 1	76.0	
2454	5150	U.S.A.	2019	Madagascar	Bejofo Estate, batch 1	76.0	
2542	5150	U.S.A.	2021	Fiji	Matasawalevu, batch 1	68.0	
2546	5150	U.S.A.	2021	Venezuela	Sur del Lago, batch 1	72.0	
...
1205	Zotter	Austria	2014	Blend	Raw	80.0	
1996	Zotter	Austria	2017	Colombia	APROCAFA, Acandi	75.0	
2036	Zotter	Austria	2018	Blend	Dry Aged, 30 yr Anniversary bar	75.0	
2170	Zotter	Austria	2018	Congo	Mountains of the Moon	70.0	
2170	Zotter	Austria	2018	Belize	Maya Mtn	72.0	

2443 rows × 10 columns



In [8]: cb2.describe()

Out[8]:

	year_reviewed	cocoa_percent	num_ingredients	rating
count	2443.000000	2443.000000	2443.000000	2443.000000
mean	2014.485878	71.496725	3.041343	3.21009
std	3.957507	5.156974	0.913728	0.42837
min	2006.000000	42.000000	1.000000	1.000000
25%	2012.000000	70.000000	2.000000	3.000000
50%	2015.000000	70.000000	3.000000	3.25000
75%	2018.000000	74.000000	4.000000	3.50000
max	2021.000000	100.000000	6.000000	4.00000

In [9]:

```
cd=cb2.dtypes[cb2.dtypes=='object'].index
cb2[cd].describe()
```

Out[9]:

	manufacturer	company_location	bean_origin	bar_name	ingredients	review
count	2443	2443	2443	2443	2443	2443
unique	542	67	62	1567	21	2403
top	Soma	U.S.A.	Venezuela	Madagascar	B,S,C	spicy, cocoa
freq	56	1118	246	52	999	4

In [12]:

```
#1; Average rating by country of origin
avr= cb2.groupby('bean_origin').rating.mean()
avr
```

Out[12]:

bean_origin	
Australia	3.250000
Belize	3.243243
Blend	3.085069
Bolivia	3.180380
Brazil	3.259740
	...
U.S.A.	3.217742
Uganda	3.097222
Vanuatu	3.062500
Venezuela	3.239837
Vietnam	3.287671
Name: rating, Length: 62, dtype: float64	

In [14]:

```
avr2=avr.reset_index()
avr2
```

Out[14]:

	bean_origin	rating
0	Australia	3.250000
1	Belize	3.243243
2	Blend	3.085069
3	Bolivia	3.180380
4	Brazil	3.259740
...
57	U.S.A.	3.217742
58	Uganda	3.097222
59	Vanuatu	3.062500
60	Venezuela	3.239837
61	Vietnam	3.287671

62 rows × 2 columns

In [64]:

```
#2; Bars reviewed by country
bars_reviewed= cb2.groupby('bean_origin').bar_name.count()
bars_reviewed
```

Out[64]:

bean_origin	
Australia	3
Belize	74
Blend	144
Bolivia	79
Brazil	77
...	
U.S.A.	31
Uganda	18
Vanuatu	12
Venezuela	246
Vietnam	73

Name: bar_name, Length: 62, dtype: int64

In [65]:

```
bars_reviewed2= cb2.groupby('bean_origin').bar_name.count().reset_index()
bars_reviewed2
```

Out[65]:

	bean_origin	bar_name
0	Australia	3
1	Belize	74
2	Blend	144
3	Bolivia	79
4	Brazil	77
...
57	U.S.A.	31
58	Uganda	18
59	Vanuatu	12
60	Venezuela	246
61	Vietnam	73

62 rows × 2 columns

In [66]:

```
bars_reviewed2.rename(columns={'bar_name':'total_bars'},inplace=True)
bars_reviewed2
```

Out[66]:

	bean_origin	total_bars
0	Australia	3
1	Belize	74
2	Blend	144
3	Bolivia	79
4	Brazil	77
...
57	U.S.A.	31
58	Uganda	18
59	Vanuatu	12
60	Venezuela	246
61	Vietnam	73

62 rows × 2 columns

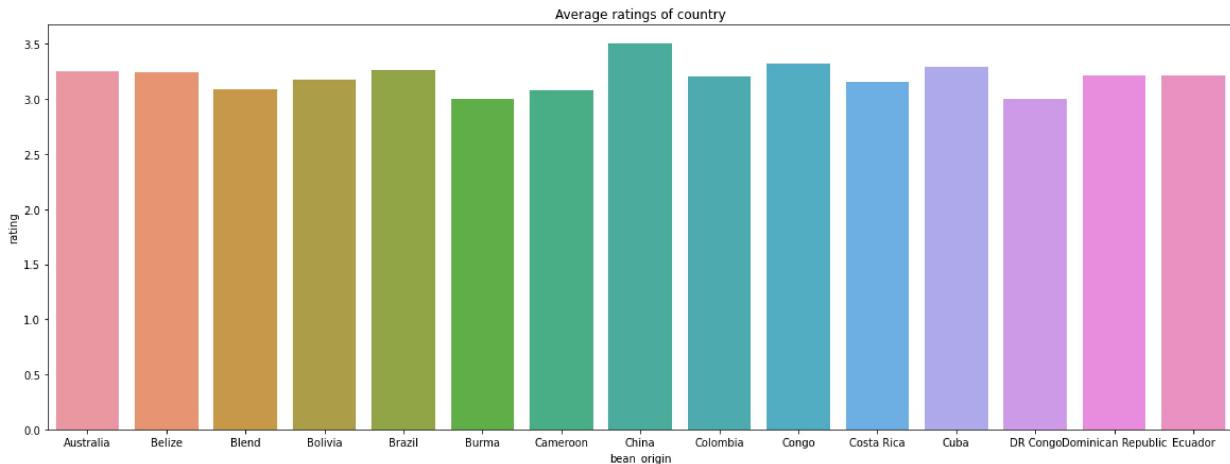
In [67]:

```
#3: Graph for question 1
plt.figure(figsize=(20,7))
plt.title('Average ratings of country')
sns.barplot(x='bean_origin', y='rating', data=avr2[0:15])
```

Out[67]:

```
<AxesSubplot:title={'center':'Average ratings of country'}, xlabel='bean_origin', ylabel='rating'>
```

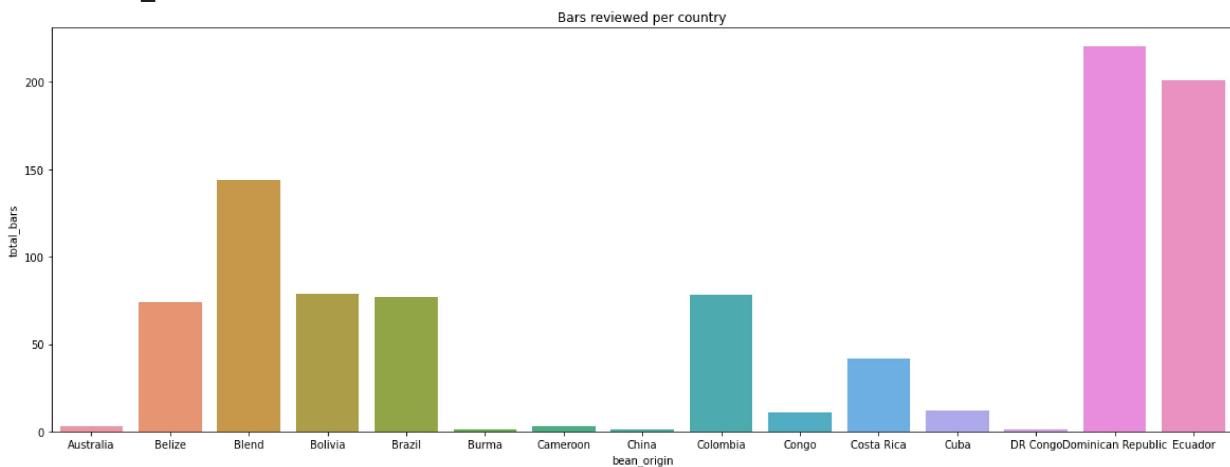
Group 9 chocolate bars



In [68]: #3b; Graph for question 2

```
plt.figure(figsize=(20,7))
plt.title('Bars reviewed per country')
sns.barplot(x='bean_origin', y='total_bars', data=bars_reviewed2[0:15])
```

Out[68]: <AxesSubplot:title={'center':'Bars reviewed per country'}, xlabel='bean_origin', ylabel='total_bars'>



In [28]: #4; cacao bean origin an indicator for quality?

```
cbr= cb2.loc[(cb2.rating > 3.5)]
cbr
```

Out[28]:

id	manufacturer	company_location	year_reviewed	bean_origin	bar_name	cocoa_percent	nu
2454	5150	U.S.A.	2019	Madagascar	Bejofo Estate, batch 1	76.0	
797	A. Morin	France	2012	Peru	Peru	63.0	
1011	A. Morin	France	2013	Ecuador	Equateur	70.0	
1015	A. Morin	France	2013	Venezuela	Chuao	70.0	
1019	A. Morin	France	2013	Peru	Chanchamayo Province	63.0	
...
2048	Zoto (Chocolatoa)	Belgium	2018	Nicaragua	El Castillero, batch ca1705, 3 turns	70.0	
647	Zotter	Austria	2011	Peru	Peru	70.0	
875	Zotter	Austria	2012	Dominican Republic	Loma Los Pinos, Yacao region, D.R.	62.0	
879	Zotter	Austria	2012	Dominican Republic	Santo Domingo	70.0	
1996	Zotter	Austria	2017	Colombia	APROCAFA, Acandi	75.0	

406 rows × 10 columns



In [34]:

```
cbc=cb2.loc[(cb2.cocoa_percent > 75)]
cbc
```

Out[34]:

	manufacturer	company_location	year_reviewed	bean_origin	bar_name	cocoa_percent	num
id							
2454	5150	U.S.A.	2019	Tanzania	Kokoa Kamili, batch 1	76.0	
2458	5150	U.S.A.	2019	Dominican Republic	Zorzal, batch 1	76.0	
2454	5150	U.S.A.	2019	Madagascar	Bejofo Estate, batch 1	76.0	
2546	5150	U.S.A.	2021	Uganda	Semuliki Forest, batch 1	80.0	
705	Adi aka Fijiana (Easy In Ltd)	Fiji	2011	Fiji	Vanua Levu, Toto-A	80.0	
...
1824	Zart Pralinen	Austria	2016	Tanzania	Kakao Kamili, Kilombero Valley	85.0	
1880	Zart Pralinen	Austria	2016	Trinidad	San Juan Estate, Gran Couva	78.0	
1716	Zokoko	Australia	2016	Solomon Islands	Guadalcanal	78.0	
879	Zotter	Austria	2012	Bolivia	El Ceibo Coop	90.0	
1205	Zotter	Austria	2014	Blend	Raw	80.0	

284 rows × 10 columns

In [36]:

```
cbf= cb2.loc[
    (cb2.cocoa_percent >=75)
    &(cb2.rating > 3.5)
]
cbf
```

Out[36]:

	manufacturer	company_location	year_reviewed	bean_origin	bar_name	cocoa_percent	num
id							
2454	5150	U.S.A.	2019	Madagascar	Bejofo Estate, batch 1	76.0	
636	Akesson's (Pralus)	U.K.	2011	Indonesia	Bali (west), Sukrama Family, Melaya area	75.0	
572	AMMA	Brazil	2010	Brazil	Monte Alegre, 3 diff. plantations	75.0	
331	Askinosie	U.S.A.	2009	Philippines	Davao	77.0	
24	Bonnat	France	2006	Blend	Carribean, Trinite	75.0	
...
2326	Taste Artisan	U.S.A.	2019	Peru	Piura	75.0	
1117	Videri	U.S.A.	2013	Blend	Dark, Central and S. America	90.0	
1916	Wm	U.S.A.	2016	Ghana	Ghana, 2013, batch 129	75.0	
1716	Zokoko	Australia	2016	Solomon Islands	Guadalcanal	78.0	
1996	Zotter	Austria	2017	Colombia	APROCAFA, Acandi	75.0	

75 rows × 10 columns

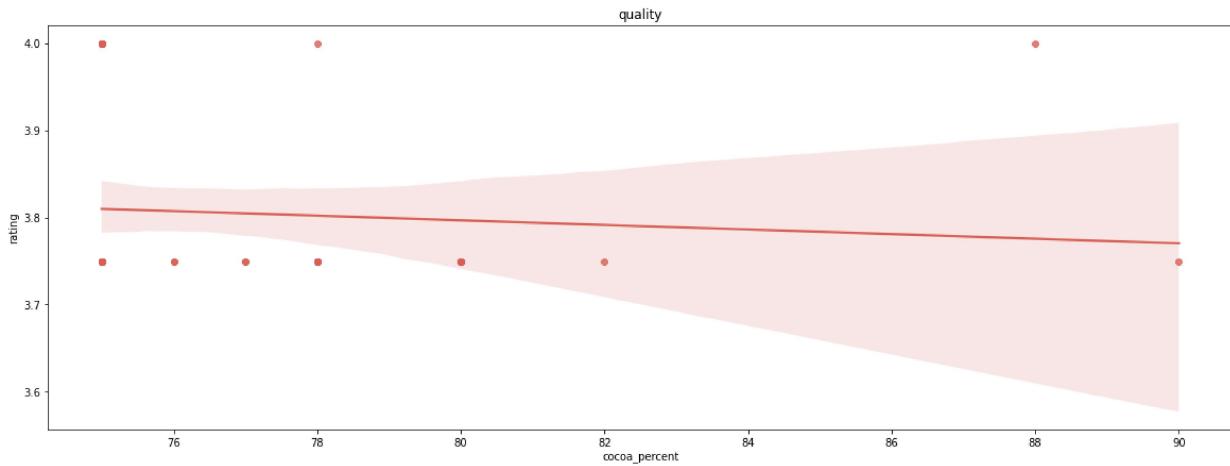
In [37]:

```
plt.figure(figsize=(20,7))
plt.title('quality')
sns.regplot(x= 'cocoa_percent', y= 'rating', data=cbf)
```

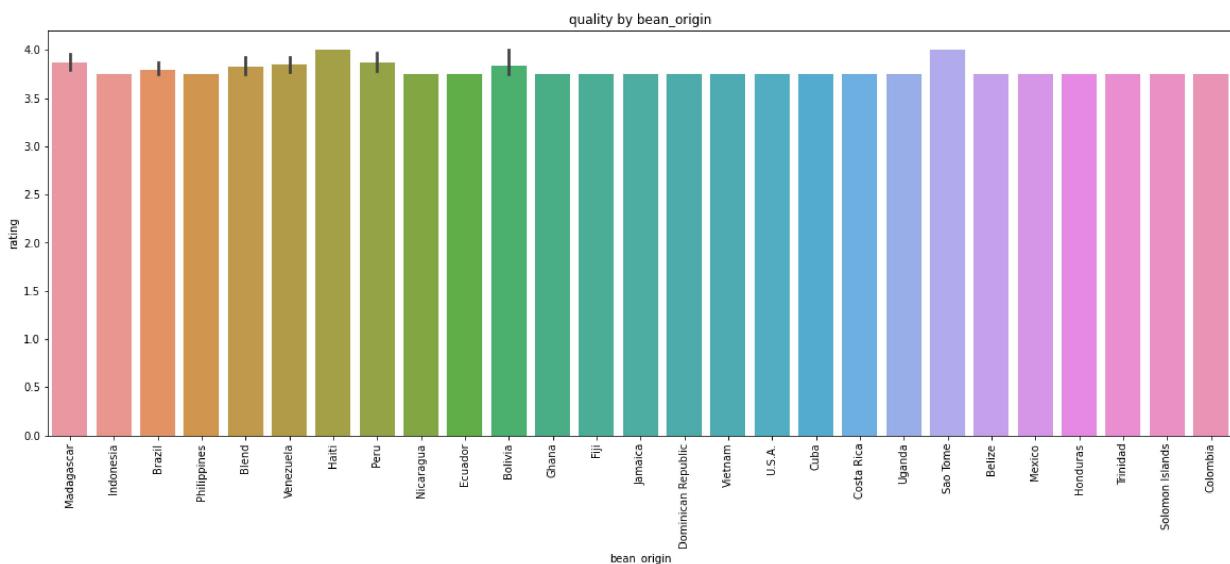
Out[37]:

```
<AxesSubplot:title={'center':'quality'}, xlabel='cocoa_percent', ylabel='rating'>
```

Group 9 chocolate bars

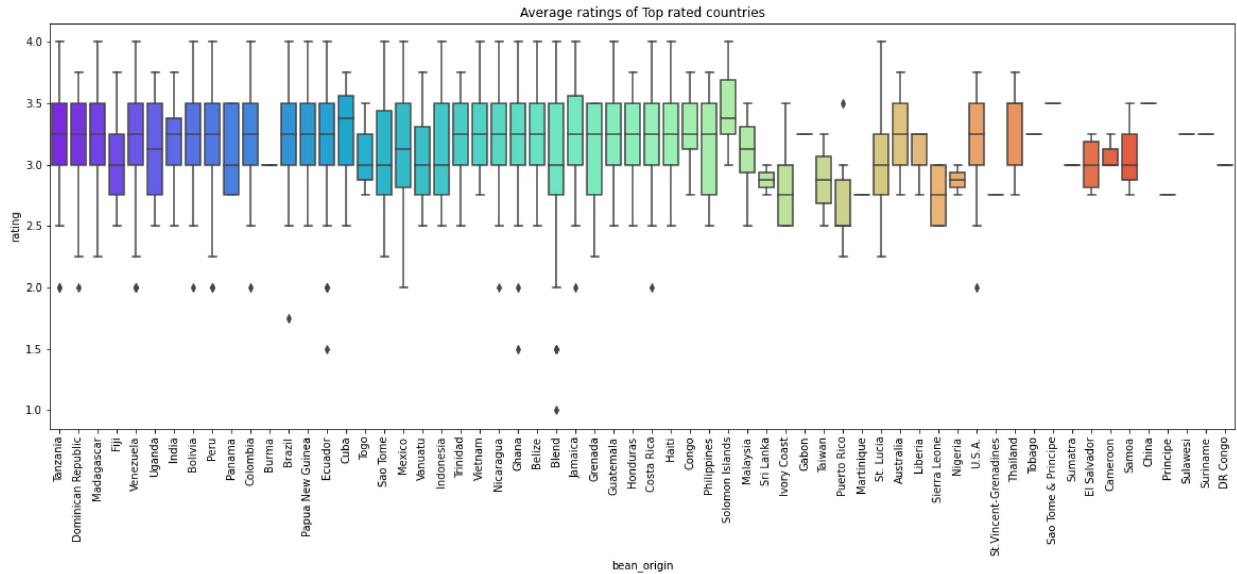


```
In [44]: plt.figure(figsize=(20,7))
plt.title('quality by bean_origin')
sns.barplot(x= cbf['bean_origin'], y= cbf['rating'])
plt.xticks(rotation=90);
```



```
In [69]: plt.figure(figsize=(20,7))
plt.title('Average ratings of Top rated countries')
sns.boxplot(x='bean_origin', y='rating', data=cb2, palette= 'rainbow')
plt.xticks(rotation=90);
```

Group 9 chocolate bars



```
In [71]: cbv= cb2[cb2.bean_origin== 'Venezuela']
cbv
```

Out[71]:

	manufacturer	company_location	year_reviewed	bean_origin	bar_name	cocoa_percent	num_i
id							
2546	5150	U.S.A.	2021	Venezuela	Sur del Lago, batch 1	72.0	
1015	A. Morin	France	2013	Venezuela	Chuao	70.0	
1315	A. Morin	France	2014	Venezuela	Carenero, Criollo	70.0	
1315	A. Morin	France	2014	Venezuela	Sur del Lago, Criollo	70.0	
1319	A. Morin	France	2014	Venezuela	Puerto Cabello, Criollo	70.0	
...
457	Willie's Cacao	U.K.	2009	Venezuela	Rio Caribe	72.0	
593	Willie's Cacao	U.K.	2010	Venezuela	Hacienda Las Trincheras	72.0	
1912	Wm	U.S.A.	2016	Venezuela	Guasare, Zulia Prov., 2015, batch 124	74.0	
2346	Wm	U.S.A.	2019	Venezuela	Porcelana	70.0	
741	Woodblock	U.S.A.	2011	Venezuela	Ocumare	70.0	

246 rows × 10 columns

In [72]: `cbv.describe()`

	year_reviewed	cocoa_percent	num_ingredients	rating
count	246.000000	246.000000	246.000000	246.000000
mean	2012.756098	71.762195	3.162602	3.239837
std	3.823326	4.507290	0.946521	0.458923
min	2006.000000	58.000000	2.000000	2.000000
25%	2010.000000	70.000000	2.000000	3.000000
50%	2013.000000	70.000000	3.000000	3.250000
75%	2015.000000	75.000000	4.000000	3.500000
max	2021.000000	91.000000	5.000000	4.000000

In [70]: #5b
cbr.describe()

	year_reviewed	cocoa_percent	num_ingredients	rating
count	406.000000	406.000000	406.000000	406.000000
mean	2014.073892	70.996305	3.029557	3.818350
std	3.991900	3.597065	0.880019	0.111563
min	2006.000000	50.000000	2.000000	3.750000
25%	2011.000000	70.000000	2.000000	3.750000
50%	2014.000000	70.000000	3.000000	3.750000
75%	2017.000000	72.000000	3.000000	4.000000
max	2021.000000	90.000000	5.000000	4.000000

In [73]: #6; Compare the average rating of bars with and without Lecithin
cb2.ingredients.unique()

Out[73]: array(['B,S,C', 'B,S,C,L', 'B,S', 'B,S,C,V', 'B,S,C,V,L', 'B,S,C,V,L,Sa', 'B,S,C,V,Sa', 'B,S,V,L', 'B,S*', 'B', 'B,S*,C', 'B,S,L', 'B,S,V', 'B,S*,C,L', 'B,S*,C,Sa', 'B,S*,Sa', 'B,S,C,Sa', 'B,S*,V,L', 'B,C', 'B,S*,C,V', 'B,S,C,L,Sa'], dtype=object)

In [59]: cb1= cb2.loc[
 (cb2.ingredients.isin(['B,S,C,L', 'B,S,C,V,L', 'B,S,C,V,L,SA', 'B,S,V,L', 'B,S,L']),
]
 cb1

Out[59]:

	manufacturer	company_location	year_reviewed	bean_origin	bar_name	cocoa_percent	num_ingredients
id							
797	A. Morin	France	2012	Bolivia	Bolivia	70.0	
797	A. Morin	France	2012	Peru	Peru	63.0	
1011	A. Morin	France	2013	Panama	Panama	70.0	
1015	A. Morin	France	2013	Colombia	Colombie	70.0	
1011	A. Morin	France	2013	Madagascar	Madagascar, Criollo	70.0	
...
697	Zokoko	Australia	2011	Bolivia	Alto Beni	68.0	
701	Zokoko	Australia	2011	Papua New Guinea	Tokiala	66.0	
701	Zokoko	Australia	2011	Bolivia	Trinidad, Baures	72.0	
1780	Zokoko	Australia	2016	Blend	Goddess Blend	65.0	
1716	Zokoko	Australia	2016	Solomon Islands	Guadalcanal	78.0	

489 rows × 10 columns

In [60]: `cb1.describe()`

Out[60]:

	year_reviewed	cocoa_percent	num_ingredients	rating
count	489.000000	489.000000	489.000000	489.000000
mean	2012.274029	69.945808	4.361963	3.152352
std	4.277582	6.179987	0.514010	0.486813
min	2006.000000	42.000000	3.000000	1.000000
25%	2009.000000	66.000000	4.000000	2.750000
50%	2012.000000	70.000000	4.000000	3.250000
75%	2015.000000	74.000000	5.000000	3.500000
max	2021.000000	91.000000	5.000000	4.000000

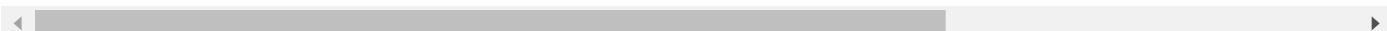
In [61]: `cbn1= cb2.loc[(cb2.ingredients.isin(['B,S,C', 'B,S', 'B,S,C,V', 'B,S,C,V,SA', 'B,S*', 'B', 'B,S']))]`

```
[ ]  
cbnl
```

Out[61]:

	manufacturer	company_location	year_reviewed	bean_origin	bar_name	cocoa_percent	nu
id							
2454	5150	U.S.A.	2019	Tanzania	Kokoa Kamili, batch 1	76.0	
2458	5150	U.S.A.	2019	Dominican Republic	Zorzal, batch 1	76.0	
2454	5150	U.S.A.	2019	Madagascar	Bejofo Estate, batch 1	76.0	
2542	5150	U.S.A.	2021	Fiji	Matasawalevu, batch 1	68.0	
2546	5150	U.S.A.	2021	Venezuela	Sur del Lago, batch 1	72.0	
...
1205	Zotter	Austria	2014	Blend	Raw	80.0	
1996	Zotter	Austria	2017	Colombia	APROCAFA, Acandi	75.0	
2036	Zotter	Austria	2018	Blend	Dry Aged, 30 yr Anniversary bar	75.0	
2170	Zotter	Austria	2018	Congo	Mountains of the Moon	70.0	
2170	Zotter	Austria	2018	Belize	Maya Mtn	72.0	

1938 rows × 10 columns



In [62]: `cbnl.describe()`

Out[62]:

	year_reviewed	cocoa_percent	num_ingredients	rating
count	1938.000000	1938.000000	1938.000000	1938.000000
mean	2015.051600	71.887513	2.693498	3.227038
std	3.680506	4.786860	0.626360	0.411253
min	2006.000000	55.000000	1.000000	1.500000
25%	2012.000000	70.000000	2.000000	3.000000
50%	2015.000000	70.000000	3.000000	3.250000
75%	2018.000000	74.000000	3.000000	3.500000
max	2021.000000	100.000000	4.000000	4.000000