

```
import pandas as pd
df_coffee = pd.read_feather("data/coffee_ratings_full.feather")
df_coffee
```

	total_cup_points	species	owner
country_of_origin \			
0	90.58	Arabica	metad plc
Ethiopia			
1	89.92	Arabica	metad plc
Ethiopia			
2	89.75	Arabica	grounds for health admin
Guatemala			
3	89.00	Arabica	yidnekachew dabessa
Ethiopia			
4	88.83	Arabica	metad plc
Ethiopia			
...
...			
1333	78.75	Robusta	luis robles
Ecuador			
1334	78.08	Robusta	luis robles
Ecuador			
1335	77.17	Robusta	james moore United
States			
1336	75.08	Robusta	cafe politico
India			
1337	73.75	Robusta	cafe politico
Vietnam			
		farm_name	lot_number
mill \			
0		metad plc	None metad
plc			
1		metad plc	None metad
plc			
2	san marcos barrancas	"san cristobal cuch	None
None			
3	yidnekachew dabessa	coffee plantation	None
wolensu			
4		metad plc	None metad
plc			
...	
...			
1333		robustasa	Lavado 1 our own
lab			
1334		robustasa	Lavado 3 own
laboratory			
1335		fazenda cazengo	None cafe
cazengo			
1336		None	None

None					
1337			None		None
None					
		ico_number		company	
altitude \					
0		2014/2015		metad agricultural developmet plc	
1950-2200					
1		2014/2015		metad agricultural developmet plc	
1950-2200					
2		None		None	1600 -
1800 m					
3		None		yidnekachew debessa coffee plantation	
1800-2200					
4		2014/2015		metad agricultural developmet plc	
1950-2200					
...		
...					
1333		None		robustasa	
None					
1334		None		robustasa	
40					
1335		None		global opportunity fund	795
meters					
1336	14-1118-2014-0087			cafe politico	
None					
1337		n/a		cafe politico	
None					
	...	color	category_two_defects	expiration \	
0	...	Green	0.0	April 3rd, 2016	
1	...	Green	1.0	April 3rd, 2016	
2	...	None	0.0	May 31st, 2011	
3	...	Green	2.0	March 25th, 2016	
4	...	Green	2.0	April 3rd, 2016	
...	
1333	...	Blue-Green	1.0	January 18th, 2017	
1334	...	Blue-Green	0.0	January 18th, 2017	
1335	...	None	6.0	December 23rd, 2015	
1336	...	Green	1.0	August 25th, 2015	
1337	...	None	9.0	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			
...		...			
1333		Specialty Coffee Association			

1334	Specialty Coffee Association
1335	Specialty Coffee Association
1336	Specialty Coffee Association
1337	Specialty Coffee Association

	certification_address \
0	309fcf77415a3661ae83e027f7e5f05dad786e44
1	309fcf77415a3661ae83e027f7e5f05dad786e44
2	36d0d00a3724338ba7937c52a378d085f2172daa
3	309fcf77415a3661ae83e027f7e5f05dad786e44
4	309fcf77415a3661ae83e027f7e5f05dad786e44
...	...
1333	ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1334	ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1335	ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1336	ff7c18ad303d4b603ac3f8cff7e611ffc735e720
1337	ff7c18ad303d4b603ac3f8cff7e611ffc735e720

	certification_contact	unit_of_measurement \
0	19fef5a731de2db57d16da10287413f5f99bc2dd	m
1	19fef5a731de2db57d16da10287413f5f99bc2dd	m
2	0878a7d4b9d35ddb0fe2ce69a2062cceb45a660	m
3	19fef5a731de2db57d16da10287413f5f99bc2dd	m
4	19fef5a731de2db57d16da10287413f5f99bc2dd	m
...
1333	352d0cf7f3e9be14dad7df644ad65efc27605ae2	m
1334	352d0cf7f3e9be14dad7df644ad65efc27605ae2	m
1335	352d0cf7f3e9be14dad7df644ad65efc27605ae2	m
1336	352d0cf7f3e9be14dad7df644ad65efc27605ae2	m
1337	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
...
1333	NaN	NaN	NaN
1334	40.0	40.0	40.0
1335	795.0	795.0	795.0
1336	NaN	NaN	NaN
1337	NaN	NaN	NaN

[1338 rows x 43 columns]

#veri setindeki satır sayısı =1338 , seçmeli örneklemin büyüklüğü= 5 tane

#aralık = 1338 // 5 -> tam kısmı verir sadece // = 267. baştan başlayarak her 267. satırı örneklem olarak seçmeli

```
sample_size = 5
pop_size=len(df_coffee) #df_coffee.shape[0]
pop_size
```

```
1338
```

```
interval = pop_size // sample_size #aralık
interval
```

```
267
```

```
df_coffee.iloc[interval::interval] #sistematiik örnekleme
```

	total_cup_points	species	owner \
267	83.92	Arabica	federacion nacional de cafeteros
534	82.92	Arabica	consejo salvadoreño del café
801	82.00	Arabica	lin, che-hao krude 林哲豪
1068	80.50	Arabica	cqi taiwan icp cqi 台灣合作夥伴
1335	77.17	Robusta	james moore

	country_of_origin	farm_name \
267	Colombia	None
534	El Salvador	santa josefita
801	Taiwan	you siang coffee farmtainan, taiwan 台灣台南優香咖啡
1068	Taiwan	王秋金
1335	United States	fazenda cazengo

	lot_number	mill
ico_number \		
267	None	None 01-
1969		
534	1-198	beneficio cuzcachapa 09-030-
273		
801	None	you siang coffee farmtainan, taiwan 台灣台南優香咖啡
Taiwan		
1068	1	non
None		
1335	None	cafe cazengo
None		

	company	altitude ...	color \
267	federacion nacional de cafeteros	None ...	None
534	soc. coop. cuzcachapa de r.l.	1350 ...	Green
801	red on tree co., ltd.	600m ...	Green
1068	王秋金	50 ...	Blue-Green

1335	global opportunity fund	795 meters	...	None
	category_two_defects		expiration	
certification_body \				
267	1.0	March 11th, 2016		
Almacafé				
534	1.0	August 28th, 2018	Salvadoran Coffee	
Council				
801	0.0	July 22nd, 2015	Specialty Coffee	
Association				
1068	0.0	December 8th, 2018	Blossom Valley	
International				
1335	6.0	December 23rd, 2015	Specialty Coffee	
Association				

	certification_address \
267	e493c36c2d076bf273064f7ac23ad562af257a25
534	3d4987e3b91399dbb3938b5bdf53893b6ef45be1
801	36d0d00a3724338ba7937c52a378d085f2172daa
1068	fc45352eee499d8470cf94c9827922fb745bf815
1335	ff7c18ad303d4b603ac3f8cff7e611ffc735e720

	certification_contact	unit_of_measurement \
267	70d3c0c26f89e00fdae6fb39ff54f0d2eb1c38ab	m
534	27b21e368fb8291cbea02c60623fe6c98f84524d	m
801	0878a7d4b9d35ddb0fe2ce69a2062cceb45a660	m
1068	de73fc9412358b523d3a641501e542f31d2668b0	m
1335	352d0cf7f3e9be14dad7df644ad65efc27605ae2	m

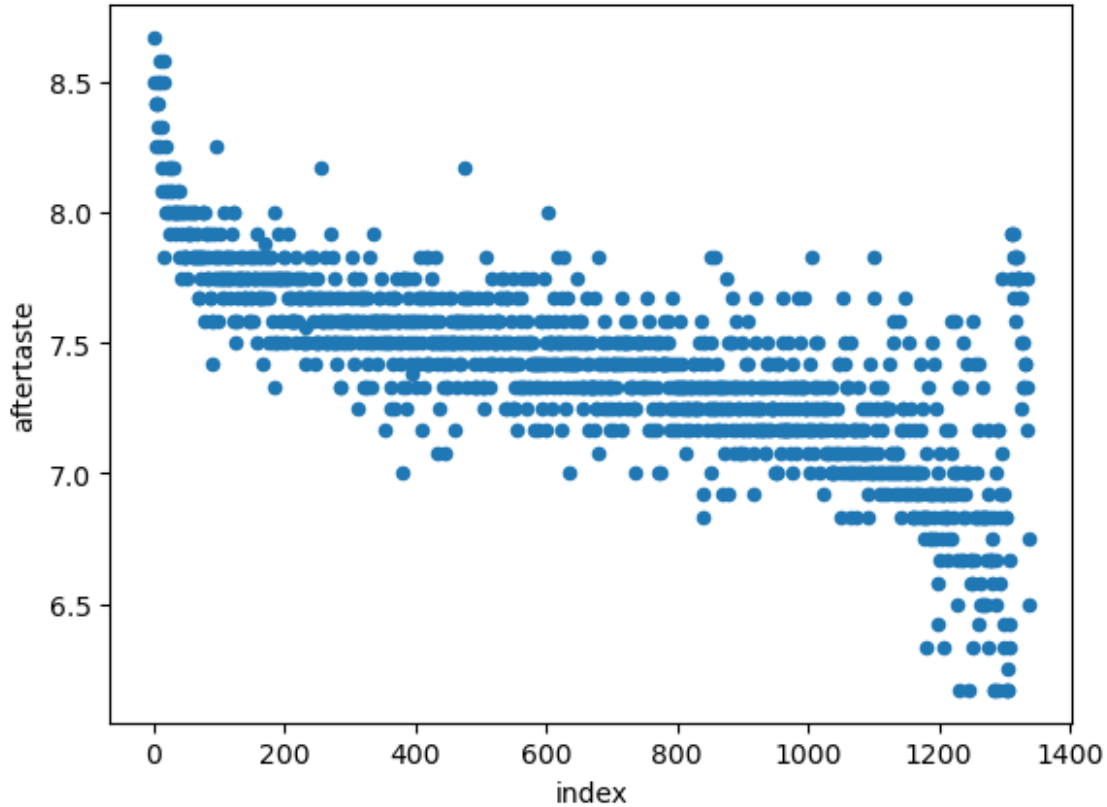
	altitude_low_meters	altitude_high_meters	altitude_mean_meters
267	NaN	NaN	NaN
534	1350.0	1350.0	1350.0
801	600.0	600.0	600.0
1068	50.0	50.0	50.0
1335	795.0	795.0	795.0

[5 rows x 43 columns]

#bizim dfde başta güzel kahveler aşapda kötüler var o yüzden. önce grafik izip örüntü var mı yok mu bakılır

```
df_coffee_id = df_coffee.reset_index()
df_coffee_id.plot(x='index',y='aftertaste', kind='scatter')
```

<Axes: xlabel='index', ylabel='aftertaste'>



#bizim dfde başta güzel kahveler aşağıda kötüler var o yüzden. üstte 267 giderken sonuçta yukarıdakiler iyi aşağıdakiler kötü kalıyor. bunu önlemek için bu yapılır, rastgele dağıtılır

```
shuffled = df_coffee.sample(frac=1)
shuffled = shuffled.reset_index(drop=True).reset_index()
shuffled.plot(x='index',y='aftertaste', kind='scatter')
#shuffled = df_coffee.sample(frac=1)
```

`df_coffee.sample(frac=1):`
Bu satır, `df_coffee` veri çerçevesindeki tüm satırları rastgele sıralar.
`frac=1` parametresi, veri çerçevesinin tamamının (%100'ünün) alınacağını belirtir.
Amaç: Veri çerçevesindeki "iyi kahvelerin" yukarıda, "kötü kahvelerin" aşağıda kalması durumunu engellemek ve tüm satırların rastgele sıralanmasını sağlamaktır.

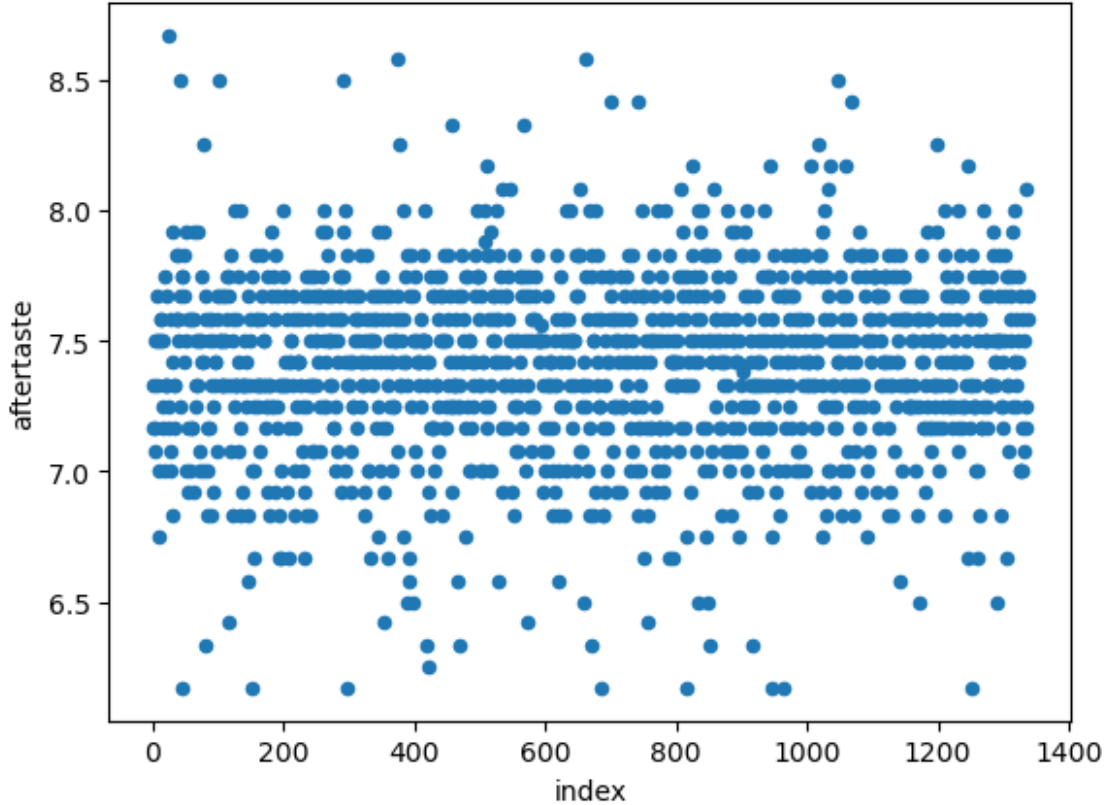
python
Kodu kopyala

```
shuffled = shuffled.reset_index(drop=True).reset_index()
reset_index(drop=True):
```

`reset_index` fonksiyonu, mevcut satır indekslerini sıfırlar ve veri çerçevesine yeni bir sıralı indeks ekler.
`drop=True`: Orijinal indekslerin veri çerçevesine eklenmemesini sağlar.
`reset_index()`:
İkinci `reset_index()` işlemi, yeni bir "index" sütunu oluşturur.

Bu sütun, her bir satır için yeni bir sıralı indeks içerir. Bu "index" sütunu daha sonra grafik çiziminde x eksenini kullanarak kullanılacaktır.

```
<Axes: xlabel='index', ylabel='aftertaste'>
```



```
#tabakalı ve ağırlıklı rastgele örnekleme
#alt grupları içeren bir popülasyonu örneklemeye olanak sağlayan
tekniktir tabakalı örnekleme
#her kahve çeşidinden aynı oranda ya da aynı sayıda veri seçilsin ->
tabakalı örnekleme kullanılır
top_counts = df_coffee['country_of_origin'].value_counts() #bu df
olarak isterse ? ->
top_counts
#Tabakalı ve Ağırlıklı Rastgele Örnekleme
Tabakalı Örnekleme (Stratified Sampling):
```

Tanım: Tabakalı örnekleme, popülasyonun alt gruplarına (tabakalarına) göre örneklem almayı sağlayan bir tekniktir. Bu yöntem, örneklemin her alt grubu (tabaka) doğru temsil etmesini sağlar.

Nasıl Çalışır: Popülasyon, belirli özelliklere göre alt gruplara ayrılır (örneğin, ülke, yaş grubu veya ürün türü). Sonra her alt gruptan aynı oranda veya aynı sayıda örnek seçilir.

Örnek: Kahve çeşitlerinin ülkelere göre dağılımını temsil etmek istiyorsanız, her ülke grubundan belirli sayıda veri seçebilirsiniz.

Böylece, örneğiniz her ülkenin kahve çeşitlerini dengeli bir şekilde içerir.

Ağırlıklı Rastgele Örnekleme (Weighted Random Sampling):

Tanım: Ağırlıklı rastgele örnekleme, belirli grupların diğerlerine göre daha fazla veya daha az temsil edilmesini sağlayarak örnekleme yapar. Her bir grubun popülasyon içindeki oranına göre ağırlık verilir.

Nasıl Çalışır: Her bir grubun popülasyondaki büyüklüğüne göre örnekleme yapılır, böylece büyük gruplardan daha fazla, küçük gruplardan daha az örnek alınır.

Örnek: Eğer belirli bir ülkeden çok fazla kahve verisi, diğer ülkelerden ise az sayıda veri varsa, her grubun popülasyondaki oranına göre veri seçimi yapılabilir.

country_of_origin	
Mexico	236
Colombia	183
Guatemala	181
Brazil	132
Taiwan	75
United States (Hawaii)	73
Honduras	52
Costa Rica	51
Ethiopia	44
Tanzania, United Republic Of	40
Uganda	36
Thailand	32
Nicaragua	26
Kenya	25
El Salvador	21
Indonesia	20
China	16
India	14
Malawi	11
United States	10
Peru	10
Myanmar	8
Vietnam	8
Haiti	6
Philippines	5
United States (Puerto Rico)	4
Panama	4
Ecuador	3
Laos	3
Burundi	2
Papua New Guinea	1
Rwanda	1
Zambia	1
Japan	1


```

Mauritius 1
Cote d'Ivoire 1
Name: count, dtype: int64

#list(top_counts.head(6).index)
#list(df_coffee.groupby('country_of_origin')
['country_of_origin'].agg('count').nlargest(6).index)
top_countries = df_coffee.groupby('country_of_origin')
['country_of_origin'].count().sort_values
top_countries

<bound method Series.sort_values of country_of_origin
Brazil 132
Burundi 2
China 16
Colombia 183
Costa Rica 51
Cote d'Ivoire 1
Ecuador 3
El Salvador 21
Ethiopia 44
Guatemala 181
Haiti 6
Honduras 52
India 14
Indonesia 20
Japan 1
Kenya 25
Laos 3
Malawi 11
Mauritius 1
Mexico 236
Myanmar 8
Nicaragua 26
Panama 4
Papua New Guinea 1
Peru 10
Philippines 5
Rwanda 1
Taiwan 75
Tanzania, United Republic Of 40
Thailand 32
Uganda 36
United States 10
United States (Hawaii) 73
United States (Puerto Rico) 4
Vietnam 8
Zambia 1
Name: country_of_origin, dtype: int64>

```

```
top_counted_countries = ['Mexico', 'Colombia',
                        'Guatemala', 'Brazil',
                        'Taiwan', 'United States (Hawaii)']
```

```
top_counted_subset = df_coffee['country_of_origin'].\
isin(top_counted_countries)
```

```
coffee_ratings_top = df_coffee[top_counted_subset]
coffee_ratings_top
```

	total_cup_points	species	owner \
2	89.75	Arabica	grounds for health admin
5	88.83	Arabica	ji-ae ahn
13	87.92	Arabica	grounds for health admin
22	87.17	Arabica	roberto licona franco
25	86.92	Arabica	nucoffee
...
1300	71.00	Arabica	ricardo aaron sampieri marini
1301	70.75	Arabica	kurt kappeli
1302	70.67	Arabica	volcafe ltda. - brasil
1306	68.33	Arabica	juan carlos garcia lopez
1309	59.83	Arabica	juan luis alvarado romero

	country_of_origin	farm_name
2	Guatemala	san marcos barrancas "san cristobal cuch
5	Brazil	None
13	United States (Hawaii)	arianna farms
22	Mexico	la herradura
25	Brazil	fazenda kaquend
...
1300	Mexico	la morena
1301	Mexico	various
1302	Brazil	None
1306	Mexico	el centenario
1309	Guatemala	finca el limon

	lot_number
mill \	
2	None

None		
5	None	
None		
13	None	
None		
22	None	la
herradura		
25	None	
None		
...	...	
...		
1300	None	tlamatoca, hutusco,
ver.		
1301	None	
f.i.e.c.h.		
1302	2017/2018 - Lot 2	
copag		
1306	None	la esperanza, municipio juchique de ferrer,
ve...		
1309	None	beneficio
serben		

	ico_number	company	altitude	...	color
\					
2	None	None	1600 - 1800 m	...	None
5	None	None	None	...	Bluish-Green
13	None	None	2000 ft	...	None
22	0	None	1320	...	Green
25	002/1251/0073	nucoffee	1250m	...	Green
...
1300	1104351023	None	1800	...	Green
1301	0016-2847-0001	globus coffee	1000 meters	...	Green
1302	None	volcafe ltda.	None	...	Green
1306	1104328663	terra mia	900	...	None
1309	11/853/165	unicafe	4650	...	Green

	category_two_defects	expiration	\
2	0.0	May 31st, 2011	
5	1.0	September 3rd, 2014	

13	2.0	May 31st, 2011
22	0.0	July 26th, 2013
25	2.0	December 2nd, 2012
...
1300	0.0	July 11th, 2013
1301	1.0	May 5th, 2015
1302	55.0	October 27th, 2018
1306	20.0	September 17th, 2013
1309	4.0	May 24th, 2013

	certification_body	\
2	Specialty Coffee Association	
5	Specialty Coffee Institute of Asia	
13	Specialty Coffee Association	
22	AMECAFE	
25	NUCOFFEE	
...	...	
1300	AMECAFE	
1301	Specialty Coffee Association	
1302	Brazil Specialty Coffee Association	
1306	AMECAFE	
1309	Asociacion Nacional Del Café	

	certification_address	\
2	36d0d00a3724338ba7937c52a378d085f2172daa	
5	726e4891cf2c9a4848768bd34b668124d12c4224	
13	36d0d00a3724338ba7937c52a378d085f2172daa	
22	59e396ad6e22a1c22b248f958e1da2bd8af85272	
25	567f200bcc17a90070cb952647bf88141ad9c80c	
...	...	
1300	59e396ad6e22a1c22b248f958e1da2bd8af85272	
1301	36d0d00a3724338ba7937c52a378d085f2172daa	
1302	3297cfa4c538e3dd03f72cc4082c54f7999e1f9d	
1306	59e396ad6e22a1c22b248f958e1da2bd8af85272	
1309	b1f20fe3a819fd6b2ee0eb8fdc3da256604f1e53	

	certification_contact	unit_of_measurement	\
2	0878a7d4b9d35ddb0fe2ce69a2062cceb45a660	m	
5	b70da261fcc84831e3e9620c30a8701540abc200	m	
13	0878a7d4b9d35ddb0fe2ce69a2062cceb45a660	ft	
22	0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7	m	
25	aa2ff513ffb9c844462a1fb07c599bce7f3bb53d	m	
...	
1300	0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7	m	
1301	0878a7d4b9d35ddb0fe2ce69a2062cceb45a660	m	
1302	8900f0bf1d0b2baf6807a73562c7677d57eb980	m	
1306	0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7	m	
1309	724f04ad10ed31dbb9d260f0dfd221ba48be8a95	ft	

altitude_low_meters altitude_high_meters altitude_mean_meters

2	1600.00	1800.00	1700.00
5	NaN	NaN	NaN
13	609.60	609.60	609.60
22	1320.00	1320.00	1320.00
25	1250.00	1250.00	1250.00
...
1300	1800.00	1800.00	1800.00
1301	1000.00	1000.00	1000.00
1302	NaN	NaN	NaN
1306	900.00	900.00	900.00
1309	1417.32	1417.32	1417.32

[880 rows x 43 columns]

```
coffee_ratings_samp = coffee_ratings_top.sample(frac=0.1,
random_state=2021) #sample kullanrak onluk basit rastgele bir örneğini
alalım
coffee_ratings_samp #88 tane seçer
```

	total_cup_points	species	owner \
1229	78.33	Arabica	pablo cervantes morelos
232	84.08	Arabica	carcafe ltda ci
697	82.42	Arabica	jose daniel cobilt castro
865	81.75	Arabica	diego manuel woolrich ramirez
155	84.58	Arabica	exportadora de cafe condor s.a
...
1282	74.83	Arabica	pablo enrique martinez gama
369	83.50	Arabica	gabriel bernardo rivass ross
852	81.83	Arabica	jacques pereira carneiro
713	82.33	Arabica	bourbon specialty coffees
806	82.00	Arabica	lin, che-hao krude 林哲豪

	country_of_origin	farm_name \
1229	Mexico	llano
hermoso		
232	Colombia	
None		
697	Mexico	cañada
fria		
865	Mexico	arroyo triste, arroyo triste, san jose vista
h...		
155	Colombia	
various		
...	...	
...		
1282	Mexico	la
orduña		
369	Mexico	la
corralera		

852	Brazil	sertao
farm		
713	Brazil	
None		
806	Taiwan	gao chun fang
高醇坊		

	lot_number		mill	\
1229	None	llano hermoso, xochitonalco huautla, oaxaca		
232	3-59-0503		neiva	
697	None		huatusco	
865	None	arroyo triste, arroyo triste, san jose vista h...		
155	None		trilladora boanza	
...	...			
1282	None		falcafe s.a. de c.v.	
369	None		dos puentes de finca kassandra	
852	None		armazens gerais cocarive	
713	None		None	
806	None		gao chun fang	高醇坊

	ico_number	\
1229	0	
232	3-59-0503	
697	1104558673	
865	2037240, 2037150, 1400213685	
155	3-68-0005	
...	...	
1282	1104362940	
369	2484	
852	002/1352/0159	
713	002/4542/0478	
806	Taiwan	

	company
altitude ... \	
1229	asociación agricola local de productores de ca...
1300	...
232	carcafe ltda
442	...
697	None
1350	...
865	None
1100	...
155	exportadora de cafe condor s.a 1800
msnm	...
...
.	
1282	None
1250	...
369	None

1400	...	
852		exportadora de cafés carmo de minas ltda
1250	...	
713		bourbon specialty coffees
None	...	
806		red on tree co., ltd. 600-700
m	...	

	color	category_two_defects	expiration	\
1229	Green	47.0	September 11th, 2013	
232	Green	3.0	November 9th, 2018	
697	Green	6.0	July 11th, 2013	
865	Green	1.0	September 4th, 2013	
155	Green	6.0	October 9th, 2013	
...	
1282	Green	30.0	August 1st, 2013	
369	Green	0.0	July 11th, 2013	
852	Bluish-Green	0.0	March 2nd, 2014	
713	Green	10.0	April 19th, 2016	
806	Bluish-Green	0.0	June 3rd, 2014	

	certification_body	\
1229	AMECAFE	
232	Almacafé	
697	AMECAFE	
865	AMECAFE	
155	Almacafé	
...	...	
1282	AMECAFE	
369	AMECAFE	
852	Specialty Coffee Association	
713	Brazil Specialty Coffee Association	
806	Specialty Coffee Association	

	certification_address	\
1229	3e18a5ae6f5e2aabca37e025f94e1974558bf5f0	
232	e493c36c2d076bf273064f7ac23ad562af257a25	
697	59e396ad6e22a1c22b248f958e1da2bd8af85272	
865	59e396ad6e22a1c22b248f958e1da2bd8af85272	
155	e493c36c2d076bf273064f7ac23ad562af257a25	
...	...	
1282	59e396ad6e22a1c22b248f958e1da2bd8af85272	
369	59e396ad6e22a1c22b248f958e1da2bd8af85272	
852	36d0d00a3724338ba7937c52a378d085f2172daa	
713	3297cfa4c538e3dd03f72cc4082c54f7999e1f9d	
806	36d0d00a3724338ba7937c52a378d085f2172daa	

	certification_contact	unit_of_measurement	\
1229	e3212d17882b7657b3fba559b4072e552604d5d1	m	
232	70d3c0c26f89e00fdae6fb39ff54f0d2eb1c38ab	m	

```

697    0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7    m
865    0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7    m
155    70d3c0c26f89e00fdae6fb39ff54f0d2eb1c38ab    m
...
1282    0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7    m
369    0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7    m
852    0878a7d4b9d35ddb0fe2ce69a2062cceb45a660    m
713    8900f0bf1d0b2baf6807a73562c7677d57eb980    m
806    0878a7d4b9d35ddb0fe2ce69a2062cceb45a660    m

```

```

      altitude_low_meters altitude_high_meters altitude_mean_meters
1229                1300.0                1300.0                1300.0
232                  442.0                  442.0                  442.0
697                 1350.0                 1350.0                 1350.0
865                 1100.0                 1100.0                 1100.0
155                 1800.0                 1800.0                 1800.0
...
1282                1250.0                1250.0                1250.0
369                 1400.0                 1400.0                 1400.0
852                 1250.0                 1250.0                 1250.0
713                  NaN                  NaN                  NaN
806                 600.0                 700.0                 650.0

```

```
[88 rows x 43 columns]
```

```

coffee_ratings_samp['country_of_origin'].value_counts(normalize=True)
#yüzde olarak gözükmelerini sağlar normalize. 88'in yani samplein
içindeki oranlarını verir. fena değil ama daha iyi olabilir veriler!

```

```

country_of_origin
Mexico                0.250000
Guatemala             0.204545
Colombia              0.181818
Brazil               0.181818
United States (Hawaii) 0.102273
Taiwan               0.079545
Name: proportion, dtype: float64

```

```
coffee_ratings_samp['country_of_origin'].value_counts()
```

```

country_of_origin
Mexico                22
Guatemala             18
Colombia              16
Brazil               16
United States (Hawaii) 9
Taiwan               7
Name: count, dtype: int64

```

```

#taiwan çok kötü gelmiş.
coffee_rating_strat =

```



```
coffee_ratings_top.groupby('country_of_origin').sample(frac=0.1,
random_state=2021)
coffee_rating_strat['country_of_origin'].value_counts(normalize=True)
```

```
country_of_origin
Mexico          0.272727
Colombia        0.204545
Guatemala       0.204545
Brazil          0.147727
Taiwan          0.090909
United States (Hawaii) 0.079545
Name: proportion, dtype: float64
```

#hepsinden eşit gelsin

```
coffee_rating_eq =
coffee_ratings_top.groupby('country_of_origin').sample(n=15,random_state=2021)
coffee_rating_eq['country_of_origin'].value_counts()
```

```
country_of_origin
Brazil          15
Colombia        15
Guatemala       15
Mexico          15
Taiwan          15
United States (Hawaii) 15
Name: count, dtype: int64
```

#ağırlıklı rastgele örnekleme

#her satırın örnekleme

```
import numpy as np # Numpy'ı import etmeyi unutmayın
```

```
coffee_ratings_weight = coffee_ratings_top.copy()
```

Taiwan için ağırlıklandırma koşulu

```
condition = coffee_ratings_weight['country_of_origin'] == 'Taiwan'
```

Taiwan olan satırlar için ağırlığı 2 yap, diğerleri için 1

```
coffee_ratings_weight['weight'] = np.where(condition, 2, 1)
```

DataFrame'i ağırlıklı rastgele örnekleme yaparak yeniden örnekleyin

```
coffee_ratings_weight = coffee_ratings_weight.sample(frac=0.1,
weights='weight')
```

Sonuçta, 'country_of_origin' sütunundaki değerlerin

normalleştirilmiş frekanslarını hesaplayın

```
coffee_ratings_weight['country_of_origin'].value_counts(normalize=True)
)
```

```
country_of_origin
Mexico          0.295455
```

```
Brazil          0.193182
Taiwan          0.181818
Colombia        0.159091
Guatemala       0.090909
United States (Hawaii) 0.079545
Name: proportion, dtype: float64
```

```
#cluster sampling
```

```
#katmanlıda her alt gruptan veri alınmalıdır. maliyeti azaltmak için  
de cluster sampling kullanılır
```

```
varieties_pop = list(df_coffee['variety'].unique())
varieties_pop
```

```
[None,
 'Other',
 'Bourbon',
 'Catimor',
 'Ethiopian Yirgacheffe',
 'Caturra',
 'SL14',
 'Sumatra',
 'SL34',
 'Hawaiian Kona',
 'Yellow Bourbon',
 'SL28',
 'Gesha',
 'Catuai',
 'Pacamara',
 'Typica',
 'Sumatra Lintong',
 'Mundo Novo',
 'Java',
 'Peaberry',
 'Pacas',
 'Mandheling',
 'Ruiru 11',
 'Arusha',
 'Ethiopian Heirlooms',
 'Moka Peaberry',
 'Sulawesi',
 'Blue Mountain',
 'Marigojipe',
 'Pache Comun']
```

```
import random
varieties_samp = random.sample(varieties_pop, k=3)
varieties_samp
```

```
['Marigojipe', 'Ethiopian Heirlooms', 'Catimor']
```

```

variety_condition=df_coffee['variety'].isin(varieties_samp)
coffee_ratings_cluster = df_coffee[variety_condition]
coffee_ratings_cluster

```

	total_cup_points	species	owner \
18	87.25	Arabica	yunnan coffee exchange
75	85.33	Arabica	yunnan coffee exchange
160	84.50	Arabica	yunnan coffee exchange
192	84.25	Arabica	松澤宏樹 koju matsuzawa
196	84.25	Arabica	yunnan coffee exchange
260	83.92	Arabica	sarawut premjit
300	83.75	Arabica	yunnan coffee exchange
338	83.58	Arabica	yunnan coffee exchange
381	83.42	Arabica	松澤宏樹 koju matsuzawa
444	83.17	Arabica	sunvirtue co., ltd.
447	83.17	Arabica	yunnan coffee exchange
448	83.17	Arabica	yunnan coffee exchange
537	82.92	Arabica	cqi taiwan icp cqi 台灣合作夥伴
590	82.75	Arabica	yunnan coffee exchange
591	82.75	Arabica	yunnan coffee exchange
609	82.67	Arabica	cqi taiwan icp cqi 台灣合作夥伴
680	82.42	Arabica	yunnan coffee exchange
711	82.33	Arabica	yunnan coffee exchange
895	81.58	Arabica	mdh
905	81.58	Arabica	enrique eduardo lopez aguilar
1030	80.83	Arabica	cqi taiwan icp cqi 台灣合作夥伴
1105	80.17	Arabica	sunvirtue co., ltd.

	country_of_origin	farm_name \
18	China	echo
75	China	pu'er city longke coffee co.ltd
160	China	mengliao ban'an coffee processing factory
192	Thailand	matsuzawa
196	China	mengliao county nayun town mangzhang coffee farm
260	Thailand	saveg coffee farm
300	China	pu'er specialty coffee cupping center
338	China	mengliao mang pan professional cooperative
381	Thailand	matsuzawa
444	Vietnam	apollo estate

447	China	chen
lin		
448	China	yunnan success trade
co.ltd		
537	Ethiopia	
-		
590	China	man ganna
estate		
591	China	pu'er lingdu yunlin coffee
co.ltd		
609	Laos	savan coffee
bean		
680	China	mengliao gao sanlin coffee
factory		
711	China	yun lan coffee co.
ltd		
895	Honduras	
various		
905	Mexico	las nubes - pante don
jaime		
1030	Laos	savan coffee
bean		
1105	Vietnam	apollo
estate		

	lot_number \
18	YNC-06114
75	YNC-16115
160	YNC-16018
192	MCCRNX115/16
196	YNC-16017
260	INT-2015
300	YNC-16063
338	YNC-16029
381	MCCFWXA15/16
444	Oriental Paris Natural Coffee
447	YNC-16056
448	YNC-16028
537	none
590	YNC-16057
591	YNC-16014
609	1
680	YNC-16047
711	YNC-16053
895	174.18
905	None
1030	201601
1105	Oriental Paris Civet Coffee

		mill	
ico_number \			
18		echo coffee mill	
None			
75		pu'er city longke coffee co.ltd	
None			
160		mengliao ban'an coffee processing factory	
None			
192		matsuzawa coffee	
None			
196		mengliao county nayun town mangzhang coffee farm	
None			
260		dry mill	
None			
300		pu'er specialty coffee cupping center	
None			
338		mengliao mang pan professional cooperative	
None			
381		matsuzawa coffee	
None			
444		yes	
None			
447		mojiang jinggong estate speciality development...	
None			
448		yunnan success trade co.ltd mill	
None			
537		None	
010/0357/0085			
590		puwen lvyou coffee co.,ltd	
None			
591		pu'er lingdu yunlin coffee co.ltd mill	
5300343627999			
609		non	
None			
680		mengliao gao sanlin coffee factory	
None			
711		yun lan coffee co. ltd mill	
None			
895		cafico	
None			
905		None	PC0-031106-
CF7			
1030		none	
None			
1105		yes	
None			
		company altitude ...	color
\			

18	yunnan coffee exchange	1450	...	Green
75	yunnan coffee exchange	1700	...	Green
160	yunnan coffee exchange	1680	...	Green
192	matsuzawa coffee	1200	...	Green
196	yunnan coffee exchange	1400	...	Green
260	intercof 1978 group company limited	1350	...	Bluish-Green
300	yunnan coffee exchange	1650	...	Green
338	yunnan coffee exchange	1100	...	Green
381	matsuzawa coffee	1200	...	Bluish-Green
444	sunvirtue co., ltd.	1550	...	None
447	yunnan coffee exchange	1200	...	Green
448	yunnan coffee exchange	1600	...	Green
537	belayneh kindie	2000	...	Bluish-Green
590	yunnan coffee exchange	1100	...	Green
591	yunnan coffee exchange	1500	...	Green
609	宸嶧國際	1300	...	Bluish-Green
680	yunnan coffee exchange	1600	...	Green
711	yunnan coffee exchange	1580	...	Green
895	molinos de honduras	1200	...	Blue-Green
905	finca santa cruz	1600	...	None
1030	宸嶧國際	1300	...	Green
1105	sunvirtue co., ltd.	1550	...	None
	category_two_defects	expiration \		
18	0.0	April 7th, 2017		
75	0.0	April 6th, 2017		
160	1.0	April 6th, 2017		
192	0.0	November 2nd, 2017		
196	1.0	April 6th, 2017		

260	1.0	October 18th, 2017
300	2.0	April 6th, 2017
338	4.0	April 6th, 2017
381	0.0	June 27th, 2017
444	0.0	May 8th, 2018
447	0.0	April 6th, 2017
448	3.0	April 5th, 2017
537	20.0	February 25th, 2017
590	0.0	April 6th, 2017
591	2.0	April 5th, 2017
609	0.0	June 2nd, 2017
680	0.0	April 6th, 2017
711	2.0	April 5th, 2017
895	3.0	January 19th, 2019
905	0.0	June 6th, 2013
1030	0.0	June 22nd, 2018
1105	2.0	May 9th, 2018

	certification_body	\
18	Yunnan Coffee Exchange	
75	Yunnan Coffee Exchange	
160	Yunnan Coffee Exchange	
192	Specialty Coffee Institute of Asia	
196	Yunnan Coffee Exchange	
260	Specialty Coffee Association	
300	Yunnan Coffee Exchange	
338	Yunnan Coffee Exchange	
381	Specialty Coffee Institute of Asia	
444	Specialty Coffee Association	
447	Yunnan Coffee Exchange	
448	Yunnan Coffee Exchange	
537	Blossom Valley International	
590	Yunnan Coffee Exchange	
591	Yunnan Coffee Exchange	
609	Blossom Valley International	
680	Yunnan Coffee Exchange	
711	Yunnan Coffee Exchange	
895	Instituto Hondureño del Café	
905	AMECAFE	
1030	Blossom Valley International	
1105	Specialty Coffee Association	

	certification_address	\
18	683fa6014608fc10ff681b0435b0b2dbe6df988f	
75	683fa6014608fc10ff681b0435b0b2dbe6df988f	
160	683fa6014608fc10ff681b0435b0b2dbe6df988f	
192	726e4891cf2c9a4848768bd34b668124d12c4224	
196	683fa6014608fc10ff681b0435b0b2dbe6df988f	
260	36d0d00a3724338ba7937c52a378d085f2172daa	

300	683fa6014608fc10ff681b0435b0b2dbe6df988f
338	683fa6014608fc10ff681b0435b0b2dbe6df988f
381	726e4891cf2c9a4848768bd34b668124d12c4224
444	36d0d00a3724338ba7937c52a378d085f2172daa
447	683fa6014608fc10ff681b0435b0b2dbe6df988f
448	683fa6014608fc10ff681b0435b0b2dbe6df988f
537	fc45352eee499d8470cf94c9827922fb745bf815
590	683fa6014608fc10ff681b0435b0b2dbe6df988f
591	683fa6014608fc10ff681b0435b0b2dbe6df988f
609	fc45352eee499d8470cf94c9827922fb745bf815
680	683fa6014608fc10ff681b0435b0b2dbe6df988f
711	683fa6014608fc10ff681b0435b0b2dbe6df988f
895	b4660a57e9f8cc613ae5b8f02bfce8634c763ab4
905	59e396ad6e22a1c22b248f958e1da2bd8af85272
1030	fc45352eee499d8470cf94c9827922fb745bf815
1105	36d0d00a3724338ba7937c52a378d085f2172daa

	certification_contact	unit_of_measurement	\
18	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
75	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
160	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
192	b70da261fcc84831e3e9620c30a8701540abc200	m	
196	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
260	0878a7d4b9d35ddb0fe2ce69a2062cceb45a660	m	
300	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
338	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
381	b70da261fcc84831e3e9620c30a8701540abc200	m	
444	0878a7d4b9d35ddb0fe2ce69a2062cceb45a660	m	
447	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
448	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
537	de73fc9412358b523d3a641501e542f31d2668b0	m	
590	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
591	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
609	de73fc9412358b523d3a641501e542f31d2668b0	m	
680	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
711	f6d87a6c04653c569d4911a66f89d5e30ce83b93	m	
895	7f521ca403540f81ec99daec7da19c2788393880	m	
905	0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7	m	
1030	de73fc9412358b523d3a641501e542f31d2668b0	m	
1105	0878a7d4b9d35ddb0fe2ce69a2062cceb45a660	m	

	altitude_low_meters	altitude_high_meters	altitude_mean_meters
18	1450.0	1450.0	1450.0
75	1700.0	1700.0	1700.0
160	1680.0	1680.0	1680.0
192	1200.0	1200.0	1200.0
196	1400.0	1400.0	1400.0
260	1350.0	1350.0	1350.0
300	1650.0	1650.0	1650.0

338	1100.0	1100.0	1100.0
381	1200.0	1200.0	1200.0
444	1550.0	1550.0	1550.0
447	1200.0	1200.0	1200.0
448	1600.0	1600.0	1600.0
537	2000.0	2000.0	2000.0
590	1100.0	1100.0	1100.0
591	1500.0	1500.0	1500.0
609	1300.0	1300.0	1300.0
680	1600.0	1600.0	1600.0
711	1580.0	1580.0	1580.0
895	1200.0	1200.0	1200.0
905	1600.0	1600.0	1600.0
1030	1300.0	1300.0	1300.0
1105	1550.0	1550.0	1550.0

[22 rows x 43 columns]

```
coffee_ratings_cluster.loc[:, 'variety'] =
coffee_ratings_cluster['variety'].astype('category').cat.remove_unused
_categories() #yanşırlıkla farklı category gelirse siler. cat ->
category olduğu için sr-tring olsaydı str yazılırdı
```

```
coffee_ratings_cluster.groupby('variety', observed=True).sample(n=1,
random_state=2021)
#observed benim varietymsen başkaları göstermemesi için
```

	total_cup_points	species	owner \
905	81.58	Arabica	enrique eduardo lopez aguilar
402	83.33	Arabica	afca
989	81.08	Arabica	wali ali

	country_of_origin	farm_name
lot_number \		
905	Mexico	las nubes - pante don jaime
None		
402	Tanzania, United Republic Of	mondul coffee estate
None		
989	United States	None
None		

	mill	ico_number	company
\			
905	None	PC0-031106-CF7	finca santa cruz
402	mondul coffee estate	C27	mondul coffee estate
989	juanana coffee association	n/a	conscious bean coffee

altitude ... color category_two_defects

expiration	\					
905	1600	...	None	0.0	June 6th, 2013	
402	1700	...	Green	6.0	December 12th, 2015	
989	5600 feet	...	None	0.0	March 3rd, 2016	

	certification_body
certification_address \	
905	AMECAFE
	59e396ad6e22a1c22b248f958e1da2bd8af85272
402	Africa Fine Coffee Association
	073285c0d45e2f5539012d969937e529564fa6fe
989	Specialty Coffee Association
	36d0d00a3724338ba7937c52a378d085f2172daa

	certification_contact	unit_of_measurement	\
905	0eb4ee5b3f47b20b049548a2fd1e7d4a2b70d0a7	m	
402	c4ab13415cdd69376a93780c0166e7b1a10481ea	m	
989	0878a7d4b9d35ddb0fe2ce69a2062cceb45a660	ft	

	altitude_low_meters	altitude_high_meters	altitude_mean_meters
905	1600.00	1600.00	1600.00
402	1700.00	1700.00	1700.00
989	1706.88	1706.88	1706.88

[3 rows x 43 columns]

#multistage sampling (çok aşamalı örneklendirme)
#alt gruplar da alt gruplara ayrılıyorsa

#comparing sampling methods

#simple random örnekleme : verilerin 1/3'ünü random şekilde ver

```
coffee_ratings_srs = coffee_ratings_top.sample(frac=1/3,
random_state=2021)
coffee_ratings_srs.shape
```

(293, 43)

#tabakalı : her alt grupta aynı sayıda

```
coffee_ratings_strat =
coffee_ratings_top.groupby['country_of_region'].sample(frac=1/3,
random_state=2021)
coffee_rating_strat.shape
```


TypeError Traceback (most recent call
last)

Cell In[51], line 2

```
1 #tabakalı : her alt grupta aynı sayıda
----> 2 coffee_ratings_strat =
coffee_ratings_top.groupby['country_of_region'].sample(frac=1/3,
random_state=2021)
3 coffee_rating_strat.shape
```

TypeError: 'method' object is not subscriptable

```
import random
```

```
top_countries_samp = random.sample(top_counted_countries, k=2)
```

```
top_condition = coffee_ratings_top['country_of_origin'].\
isin(top_countries_samp)
```

```
coffee_ratings_cls = coffee_ratings_top[top_condition]
```

```
coffee_ratings_cls.loc[:, 'country_of_origin'] = \
coffee_ratings_cls['country_of_origin'].\
astype('category').cat.remove_unused_categories()
```

```
coffee_ratings_clst = coffee_ratings_cls.\
groupby('country_of_origin').\
sample(n=len(coffee_ratings_top) // 1)
```

```
coffee_ratings_clst.shape
```


ValueError Traceback (most recent call last)

Cell In[106], line 16

```
8 coffee_ratings_cls = coffee_ratings_top[top_condition]
10 coffee_ratings_cls.loc[:, 'country_of_origin'] = \
11 coffee_ratings_cls['country_of_origin'].\
12 astype('category').cat.remove_unused_categories()
14 coffee_ratings_clst = coffee_ratings_cls.\
15 groupby('country_of_origin').\
--> 16 sample(n=len(coffee_ratings_top) // 1)
18 coffee_ratings_clst.shape
```

File

/opt/anaconda3/lib/python3.12/site-packages/pandas/core/groupby/groupby.py:5780, in GroupBy.sample(self, n, frac, replace, weights, random_state)

```
5777         assert frac is not None
5778         sample_size = round(frac * group_size)
-> 5780         grp_sample = sample.sample(
5781             group_size,
```

```
5782         size=sample_size,
5783         replace=replace,
5784         weights=None if weights is None else
weights_arr[grp_indices],
5785         random_state=random_state,
5786     )
5787     sampled_indices.append(grp_indices[grp_sample])
5789 sampled_indices = np.concatenate(sampled_indices)
```

```
File
/opt/anaconda3/lib/python3.12/site-packages/pandas/core/sample.py:152,
in sample(obj_len, size, replace, weights, random_state)
    149     else:
    150         raise ValueError("Invalid weights: weights sum to
zero")
--> 152 return random_state.choice(obj_len, size=size,
replace=replace, p=weights).astype(
    153     np.intp, copy=False
    154 )
```

```
File numpy/random/mtrand.pyx:1001, in
numpy.random.mtrand.RandomState.choice()
```

```
ValueError: Cannot take a larger sample than population when
'replace=False'
```