

```
import pandas as pd
```

```
df_sleep = pd.read_csv("data/msleep.csv")
```

```
df_sleep
```

	conservation	name	genus	vore	order
0	lc	Cheetah	Acinonyx	carni	Carnivora
1	NaN	Owl monkey	Aotus	omni	Primates
2	nt	Mountain beaver	Aplodontia	herbi	Rodentia
3	lc	Greater short-tailed shrew	Blarina	omni	Soricomorpha
4	domesticated	Cow	Bos	herbi	Artiodactyla
..
78	NaN	Tree shrew	Tupaia	omni	Scandentia
79	NaN	Bottle-nosed dolphin	Tursiops	carni	Cetacea
80	NaN	Genet	Genetta	carni	Carnivora
81	NaN	Arctic fox	Vulpes	carni	Carnivora
82	NaN	Red fox	Vulpes	carni	Carnivora

	sleep_total	sleep_rem	sleep_cycle	awake	brainwt	bodywt
0	12.1	NaN	NaN	11.9	NaN	50.000
1	17.0	1.8	NaN	7.0	0.01550	0.480
2	14.4	2.4	NaN	9.6	NaN	1.350
3	14.9	2.3	0.133333	9.1	0.00029	0.019
4	4.0	0.7	0.666667	20.0	0.42300	600.000
..
78	8.9	2.6	0.233333	15.1	0.00250	0.104
79	5.2	NaN	NaN	18.8	NaN	173.330
80	6.3	1.3	NaN	17.7	0.01750	2.000
81	12.5	NaN	NaN	11.5	0.04450	3.380
82	9.8	2.4	0.350000	14.2	0.05040	4.230

```
[83 rows x 11 columns]
```

```
# **p-değeri (p-value)**
```

```
# p-değeri, bir istatistiksel hipotezin rastgele bir örneklemden elde edilen verilere dayanarak doğruluğunun değerlendirilmesi için kullanılan bir metriktir. Bu değer, gözlemlenen ya da daha aşırı sonuçların, null hipotezi (H0) doğru olduğu durumda rastgele bir şans
```

sonucu olarak elde edilme olasılığını ifade eder. Genellikle, p -değerinin küçük olması ($p < 0.05$ gibi), gözlemlediğimiz sonuçların null hipotezi altında rastgele bir şans sonucu olma ihtimalinin düşük olduğunu gösterir ve bu durumda null hipotezi reddedilir.

****Null Hipotezi (H_0)****

Null hipotezi, genellikle iki grup arasında bir fark olmadığı veya bir tedavinin etkisiz olduğu gibi durumları ifade eden başlangıç varsayımıdır.

H_0 olarak gösterilir ve istatistiksel testlerde, bu hipotezin doğru olup olmadığını test ederiz. Null hipotezi, genellikle araştırma hipotezi

veya alternatif hipotezi (H_a veya H_1) ile karşılaştırılır.

****Hipotez Testinin Sonucu****

- **** $p < 0.05$ ****: Eğer p -değeri 0.05 'ten küçükse, bu, elde edilen sonucun null hipotezi altında rastgele bir şans sonucu ortaya çıkma ihtimalinin 5 'ten daha düşük olduğunu gösterir. Bu durumda, null hipotezi red (reddedilir) edilir ve alternatif hipotezin (araştırma hipotezi) desteklendiği sonucuna varılır. Bu, genellikle istatistiksel olarak anlamlı olarak tanımlanır.

- **** $p > 0.05$ ****: Eğer p -değeri 0.05 'ten büyükse, elde edilen sonuçların null hipotezi altında rastgele bir şans sonucu olarak ortaya çıkma ihtimali 5 veya daha yüksektir. Bu durumda, null hipotezi kabul edilir, yani yeterli kanıt olmadığı için null hipotezi reddedilemez.

Bu, genellikle bir çalışmada kullanılan standart bir eşik değerdir, ancak bazı durumlarda daha düşük veya daha yüksek p -değer eşikleri

kullanılabilir, bu tamamen araştırmanın bağlamına ve önemine bağlıdır. Örneğin, daha kritik tıbbi veya farmakolojik araştırmalarda daha

#düşük bir p -değer eşik değeri kullanılabilir.

#significance level (anlamlılık leveli) (α) H_0 hipotezini daha az yanlışlıkla red edebilmek için iki fonksiyonun kesişim alanıdır.

Anlamlılık düzeyi (significance level), istatistiksel testlerde sıklıkla karşılaştığımız ve genellikle α (α) ile sembolize edilen bir kavramdır. Anlamlılık düzeyi, istatistiksel bir test sırasında null hipotezi (H_0) yanlışlıkla reddetme ihtimalini sınırlamak için kullanılır. Bu değer, bir hipotezin test edilmesi sırasında belirlenen ve kabul edilebilir hata sınırını ifade eder.

Anlamlılık Düzeyi

Anlamlılık düzeyi, genellikle 5 (0.05) olarak belirlenir ama bu oran araştırmanın gereksinimlerine göre daha düşük ya da yüksek olabilir.

#Örneğin, çok kritik kararlar gerektiren tıbbi veya farmakolojik çalışmalarda (α) değeri daha düşük, örneğin 1 (0.01) olarak

ayarlanabilir.
Bu oran, bir hipotezi reddetmede yapılabilecek maksimum Type I hata oranını belirtir.

Type I Hata (Birinci Tür Hata)
Type I hata, gerçekte (H_0) doğru olduğu halde yanlışlıkla reddedilmesi durumudur. Yani, aslında geçerli olan null hipotezi, verilerde
rastlantısal varyasyonlar nedeniyle yanlışlıkla reddedilir. anlamlılık değeri, bu hatanın olasılığını sınırlandırmak için kullanılır.

İki Fonksiyonun Kesişim Alanı
İki fonksiyonun kesişim alanı ifadesi, genellikle iki olasılık yoğunluk fonksiyonu (örneğin, null hipotez altında bir dağılım ve alternatif
#hipotez altında bir dağılım) arasındaki kesişme bölgesini ifade eder. Bu kesişim, her iki hipotezin aynı anda mümkün olduğu durumları ve #bu durumların olasılıklarını gösterir.

#İstatistiksel testlerde, (α) seviyesi, bu iki dağılımın kesişim bölgesini sınırlayan bir eşik değer olarak işlev görür. #Eğer test istatistiği bu eşik değerın ötesine geçerse, (H_0) reddedilir. Bu, testin belirlenen anlamlılık düzeyinde ne kadar "anlamlı"
#olduğunu belirler ve bu da sonuçların istatistiksel olarak önemli olup olmadığını gösterir.

Anlamlılık düzeyi, istatistiksel karar verme sürecinde yanlış pozitif sonuçları kontrol altında tutmak için temel bir araçtır.

```
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]

```
#df_coffee[df_coffee["total_cup_points"] ==
df_coffee["total_cup_points"].max()]
df_coffee.sort_values(by="total_cup_points",ascending=False).iloc[0]
#head(1) de olur.seri döner. Seri ile df farkı seri tek kayıt df
birden fazla kayıt
```

total_cup_points	90.58
species	Arabica
owner	metad plc
country_of_origin	Ethiopia
farm_name	metad plc
lot_number	None
mill	metad plc
ico_number	2014/2015
company	metad agricultural developmet plc
altitude	1950-2200
region	guji-hambela
producer	METAD PLC
number_of_bags	300.0
bag_weight	60 kg
in_country_partner	METAD Agricultural Development plc
harvest_year	2014
grading_date	April 4th, 2015

owner_1	metad plc
variety	None
processing_method	Washed / Wet
aroma	8.67
flavor	8.83
aftertaste	8.67
acidity	8.75
body	8.5
balance	8.42
uniformity	10.0
clean_cup	10.0
sweetness	10.0
cupper_points	8.75
moisture	0.12
category_one_defects	0.0
quakers	0.0
color	Green
category_two_defects	0.0
expiration	April 3rd, 2016
certification_body	METAD Agricultural Development plc
certification_address	309fcf77415a3661ae83e027f7e5f05dad786e44
certification_contact	19fef5a731de2db57d16da10287413f5f99bc2dd
unit_of_measurement	m
altitude_low_meters	1950.0
altitude_high_meters	2200.0
altitude_mean_meters	2075.0

Name: 0, dtype: object

```
pts_vs_flavor_pop = df_coffee[["total_cup_points", "flavor"]] # bu iki
sütunu alarak yeni dataframe döndürür
pts_vs_flavor_pop
```

	total_cup_points	flavor
0	90.58	8.83
1	89.92	8.67
2	89.75	8.50
3	89.00	8.58
4	88.83	8.50
...
1333	78.75	7.58
1334	78.08	7.67
1335	77.17	7.33
1336	75.08	6.83
1337	73.75	6.67

[1338 rows x 2 columns]

```
pts_vs_flavor_pop_samp = pts_vs_flavor_pop.sample(10) # replace false o
yüzden yerine koymadan seçer. yani hep aynı veri çekilir
pts_vs_flavor_pop_samp
```

	total_cup_points	flavor
885	81.67	7.50
470	83.17	7.75
100	85.00	7.83
466	83.17	7.75
712	82.33	7.83
607	82.67	7.58
165	84.50	7.83
89	85.08	7.83
704	82.42	7.25
1137	79.92	7.33

```
cup_points_samp = df_coffee["total_cup_points"].sample(n=10) #
totalcuppoinsten random 10 veri
cup_points_samp
```

678	82.42
924	81.50
894	81.67
874	81.67
504	83.00
798	82.00
870	81.67
792	82.08
457	83.17
230	84.17

Name: total_cup_points, dtype: float64

```
import numpy as np
print(f"Popülasyon Ortalaması =
{np.mean(pts_vs_flavor_pop['total_cup_points'])}")
print(f"Örnek Ortalaması =
{np.mean(pts_vs_flavor_pop_samp['total_cup_points'])}")
print(f"Örnek Ortalaması = {np.mean(cup_points_samp)}")
```

Popülasyon Ortalaması = 82.15120328849028

Örnek Ortalaması = 82.99299999999998

Örnek Ortalaması = 82.335

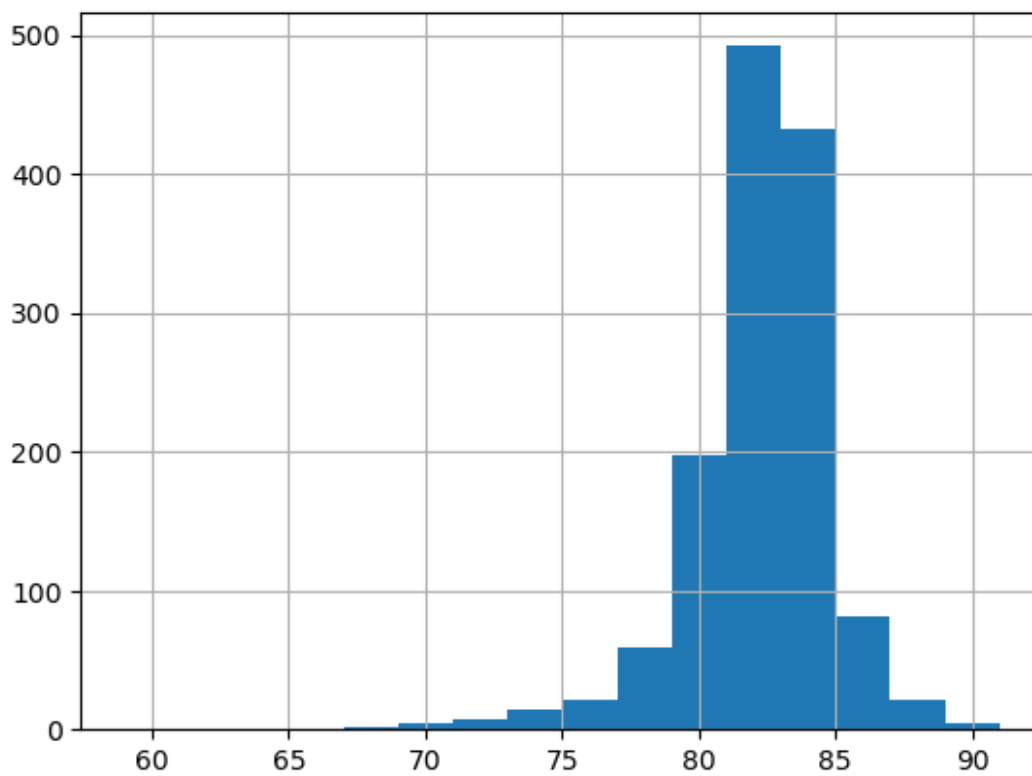
```
df_coffee["total_cup_points"].mean()
df_coffee.head()["total_cup_points"].mean()
```

89.616

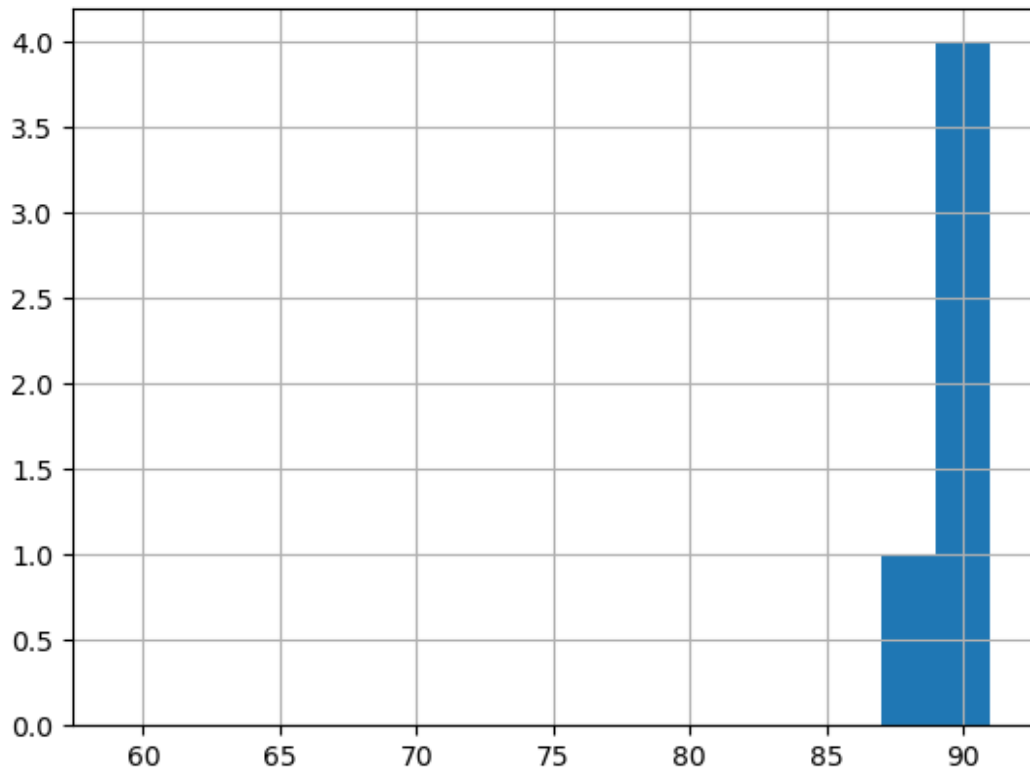
```
import matplotlib.pyplot as plt
```

```
df_coffee["total_cup_points"].hist(bins=np.arange(59,93,2)) # 59'dan
başlayıp 93'e kadar olan değerleri her bir adımda 2 artırarak bir dizi
```

<Axes: >

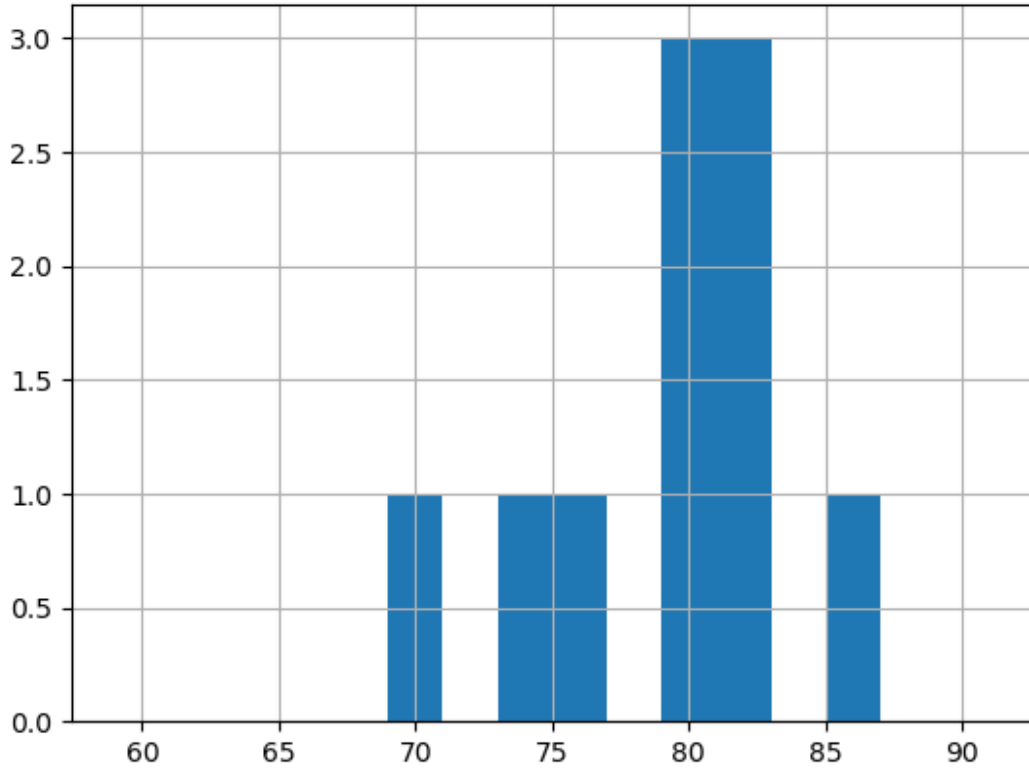


```
df_coffee.head()["total_cup_points"].hist(bins=np.arange(59,93,2))  
<Axes: >
```



```
df_coffee.sample(n=10)
["total_cup_points"].hist(bins=np.arange(59,93,2))
```

<Axes: >



```
randoms = np.random.beta(a=2,b=2,size=5000)
randoms
# **`np.random.beta(a=2, b=2, size=5000)`**: Bu fonksiyon, beta
dağılımı kullanarak rastgele sayılar üretir. Beta dağılımı, iki
pozitif parametre
# olan  $\alpha$  (alfa) ve  $\beta$  (beta) tarafından şekillendirilen ve  $[0, 1]$ 
aralığında değerler alan bir sürekli dağılımdır.
# `a=2` ve `b=2`: Bu parametreler, dağılımın şeklini belirler. Hem  $\alpha$ 
hem de  $\beta$  değerinin 2 olması, simetrik bir dağılım anlamına
gelir.
# Bu özel durumda, dağılım  $[0, 1]$  aralığında düzgün bir şekilde yüksek
sıklıkta orta değerlere yoğunlaşır.
# - `size=5000`: Bu, üretilecek rastgele sayıların sayısını belirtir.
Bu durumda, 5000 adet rastgele sayı üretilir.

# Beta dağılımı, özellikle başarı/kazanç oranları gibi oranlar üzerine
çalışmalarda sıkça tercih edilen bir dağılımdır, çünkü  $[0, 1]$ 
aralığını
#doğal bir şekilde kullanır ve parametrelerine bağlı olarak çok
çeşitli şekiller alabilir.

array([0.4663423 , 0.46606134, 0.16527482, ..., 0.46336641,
       0.43330784,
       0.56216631])
```

```
df_coffee.sample(n=5, random_state=15452) #aynı değerleri vermesi için
her seferinde random state verilir
```

	total_cup_points	species	owner \
27	86.83	Arabica	kabum trading company
1043	80.75	Arabica	cadexsa
544	82.92	Arabica	kawacom uganda ltd
596	82.75	Arabica	eileen koyanagi
1099	80.25	Arabica	gloria antonieta escobar urrutia

	country_of_origin	farm_name	
lot_number \			
27	Uganda	chebonet (23) women coffee	None
1043	Honduras	bethel	None
544	Uganda	sipi organic coffee project	None
596	United States (Hawaii)	None	None
1099	Nicaragua	finca el consuelo	None

	mill	ico_number \
27	kabum trading company	0
1043	cadexsa	13-63-318
544	kawacom	0
596	None	KP010914
1099	beneficio matagalpa coffee group, s.a.	017/117/018

	company	altitude	...	color
\				
27	kabum trading company	1950	...	Green
1043	cadexsa	1350	...	Green
544	kawacom uganda ltd	1400-1900	...	Green
596	kona pacific farmers cooperative	None	...	Blue-Green
1099	zeas escobar y compañía limitada	900 m.s.n.m.	...	Green

	category_two_defects	expiration \
27	1.0	June 26th, 2015
1043	2.0	May 21st, 2016
544	0.0	June 27th, 2015
596	0.0	February 5th, 2015
1099	2.0	February 20th, 2016

certification_body \

27	Uganda Coffee Development Authority
1043	Instituto Hondureño del Café
544	Uganda Coffee Development Authority
596	Specialty Coffee Association
1099	Asociación de Cafés Especiales de Nicaragua

	certification_address \
27	188fe373b511e21f614564bf86aa4774270d8e04
1043	b4660a57e9f8cc613ae5b8f02bfce8634c763ab4
544	188fe373b511e21f614564bf86aa4774270d8e04
596	36d0d00a3724338ba7937c52a378d085f2172daa
1099	fc561dd3c2eee024b032933e0a97b4aede0dc206

	certification_contact	unit_of_measurement \
27	b7614767a5343729bbde3a2777c60ce836aed928	m
1043	7f521ca403540f81ec99daec7da19c2788393880	m
544	b7614767a5343729bbde3a2777c60ce836aed928	m
596	0878a7d4b9d35ddb0fe2ce69a2062cceb45a660	ft
1099	f79a8d4dee92a80ff14025f03ea34fa316b2132f	m

	altitude_low_meters	altitude_high_meters	altitude_mean_meters
27	1950.0	1950.0	1950.0
1043	1350.0	1350.0	1350.0
544	1400.0	1900.0	1650.0
596	NaN	NaN	NaN
1099	900.0	900.0	900.0

[5 rows x 43 columns]