

Universidad de El Salvador
Facultad Multidisciplinaria de Occidente
Mayra Cañas
Estadística Descriptiva con Python

Table of contents

Estadística Descriptiva Con Python	3
Exploración de datos	3
Medidas de tendencia central	5

Estadística Descriptiva Con Python

Se utiliza para presentar las características básicas de los datos de manera sencilla y comprensible. Esto incluye medidas como la media, mediana, moda, rango, varianza, desviación estándar, y más.

Exploración de datos

Cargamos los paquetes a utilizar:

```
#!pip install pandas
```

Cargamos la base de datos:

```
import pandas as pd

df = pd.read_csv(r"C:\Users\PC\Documents\clases\r\cars.csv")
df
```

	manufacturer_name	model_name	transmission	color	odometer_value	year_produced	engine
0	Subaru	Outback	automatic	silver	190000	2010	gasol
1	Subaru	Outback	automatic	blue	290000	2002	gasol
2	Subaru	Forester	automatic	red	402000	2001	gasol
3	Subaru	Impreza	mechanical	blue	10000	1999	gasol
4	Subaru	Legacy	automatic	black	280000	2001	gasol
...
38526	Chrysler	300	automatic	silver	290000	2000	gasol
38527	Chrysler	PT Cruiser	mechanical	blue	321000	2004	diese
38528	Chrysler	300	automatic	blue	777957	2000	gasol
38529	Chrysler	PT Cruiser	mechanical	black	20000	2001	gasol
38530	Chrysler	Voyager	automatic	silver	297729	2000	gasol

Tipos de datos:

```
#tipos de datos en la base
df.dtypes
```

```
manufacturer_name    object
model_name            object
```

```

transmission      object
color             object
odometer_value    int64
year_produced     int64
engine_fuel       object
engine_has_gas    bool
engine_type       object
engine_capacity   float64
body_type         object
has_warranty      bool
state            object
drivetrain        object
price_usd         float64
is_exchangeable   bool
location_region   object
number_of_photos  int64
up_counter        int64
feature_0         bool
feature_1         bool
feature_2         bool
feature_3         bool
feature_4         bool
feature_5         bool
feature_6         bool
feature_7         bool
feature_8         bool
feature_9         bool
duration_listed   int64
dtype: object

```

Generamos el conjunto completo de estadísticos descriptivos de la base de datos

```

#Estadísticos descriptivos fundamentales
df.describe()

```

	odometer_value	year_produced	engine_capacity	price_usd	number_of_photos	up_count
count	38531.000000	38531.000000	38521.000000	38531.000000	38531.000000	38531.000000
mean	248864.638447	2002.943734	2.055161	6639.971021	9.649062	16.306091
std	136072.376530	8.065731	0.671178	6428.152018	6.093217	43.286933
min	0.000000	1942.000000	0.200000	1.000000	1.000000	1.000000
25%	158000.000000	1998.000000	1.600000	2100.000000	5.000000	2.000000

	odometer_value	year_produced	engine_capacity	price_usd	number_of_photos	up_count
50%	250000.000000	2003.000000	2.000000	4800.000000	8.000000	5.000000
75%	325000.000000	2009.000000	2.300000	8990.000000	12.000000	16.000000
max	1000000.000000	2019.000000	8.000000	50000.000000	86.000000	1861.000000

Medidas de tendencia central

Las medidas de tendencia central son la media(promedio), mediana(dato central) y moda(dato que mas se repite de una serie)

La media quedaria como:

```
#Media utilizando siempre un dato cuantitativo
media = df['price_usd'].mean()
print(media)
```

6639.971021255613

La mediana se obtiene como:

```
mediana = df['price_usd'].median()
print(mediana)
```

4800.0