

Actividad 5

May 2, 2020

1 ACTIVIDAD 5

Realiza las siguientes actividades en el dataset 'Automobile_data.csv' subido a la carpeta compartida con la librería Pandas:

- Imprime las primeras y las últimas cinco filas.
- Limpia los datos y actualiza el archivo CSV.
- Encuentra el nombre de la compañía de automóviles más cara.
- Imprime todos los detalles de Toyota Cars.
- Cuenta el total de automóviles por compañía.
- Encuentra el kilometraje promedio de cada compañía de automoviles.
- Ordena todos los coches por columna Precio y guarda el resultado en un CSV

```
[1]: # Importamos la librerías necesarias

import pandas as pd
import numpy as np

# Asigno los datos a un dataframe

df=pd.read_csv('./Automobile_data.csv')
```

1.1 Imprime las primeras y las últimas cinco filas.

```
[2]: # Imprimimos las primeras 5 líneas
df.head(5)
```

```
[2]:
```

	index	company	body-style	wheel-base	length	engine-type	\
0	0	alfa-romero	convertible	88.6	168.8	dohc	
1	1	alfa-romero	convertible	88.6	168.8	dohc	
2	2	alfa-romero	hatchback	94.5	171.2	ohcv	
3	3	audi	sedan	99.8	176.6	ohc	
4	4	audi	sedan	99.4	176.6	ohc	

	num-of-cylinders	horsepower	average-mileage	price
--	------------------	------------	-----------------	-------

0	four	111	21	13495.0
1	four	111	21	16500.0
2	six	154	19	16500.0
3	four	102	24	13950.0
4	five	115	18	17450.0

```
[3]: # Imprimimos las últimas 5 líneas
df.tail(5)
```

```
[3]:      index  company body-style  wheel-base  length engine-type \
56      81  volkswagen    sedan        97.3   171.7         ohc
57      82  volkswagen    sedan        97.3   171.7         ohc
58      86  volkswagen    sedan        97.3   171.7         ohc
59      87      volvo    sedan       104.3   188.8         ohc
60      88      volvo    wagon       104.3   188.8         ohc

      num-of-cylinders  horsepower  average-mileage  price
56                four           85                27   7975.0
57                four           52                37   7995.0
58                four          100                26   9995.0
59                four          114                23  12940.0
60                four          114                23  13415.0
```

1.2 Limpia los datos y actualiza el archivo CSV.

Vemos los valores perdidos del dataframe

```
[4]: def ver_null(df):
      total = df.isnull().sum().sort_values(ascending = False)
      percent = (df.isnull().sum() / df.isnull().count()).sort_values(ascending =
      ↪False)
      missing_data = pd.concat([total, percent], axis = 1, keys = ['Total',
      ↪'Porcentaje'])
      return missing_data
ver_null(df).head(10)
```

```
[4]:      Total  Porcentaje
price           3      0.04918
average-mileage  0      0.00000
horsepower       0      0.00000
num-of-cylinders  0      0.00000
engine-type      0      0.00000
length           0      0.00000
wheel-base      0      0.00000
body-style       0      0.00000
company          0      0.00000
```

```
index          0      0.00000
```

Observamos que hay 3 valores en la columna precio (price) que están vacíos. Vamos a rellenarlos con la media de los valores de dicha columna

```
[5]: df_limpio = df.fillna(df.mean())
     ver_null(df_limpio)
```

```
[5]:
```

	Total	Porcentaje
price	0	0.0
average-mileage	0	0.0
horsepower	0	0.0
num-of-cylinders	0	0.0
engine-type	0	0.0
length	0	0.0
wheel-base	0	0.0
body-style	0	0.0
company	0	0.0
index	0	0.0

Hemos comprobado que no tiene valores vacíos. Vamos a actualizar el archivo CSV.

```
[6]: df_limpio.to_csv('./Automobile_data_limpio.csv')
```

1.3 Encuentra el nombre de la compañía de automóviles más cara.

```
[7]: company = df_limpio['company'][df_limpio['price']==df_limpio['price'].max()].
     ↪values[0]
     print("La compañía más cara es: {}".format(company))
```

La compañía más cara es: mercedes-benz

Hemos considerado que la más cara es la que tiene el coche más caro.

1.4 Imprime todos los detalles de Toyota Cars.

```
[8]: toyota = df_limpio[df_limpio['company'] == 'Toyota Cars']
     toyota
```

```
[8]: Empty DataFrame
     Columns: [index, company, body-style, wheel-base, length, engine-type, num-of-
     cylinders, horsepower, average-mileage, price]
     Index: []
```

Vemos que no existe ninguna compañía que se llama Toyota Cars. Como creo que es una errata del enunciado, probare con la compañía toyota.

```
[9]: toyota = df_limpio[df_limpio['company'] == 'toyota']
toyota
```

```
[9]:
```

	index	company	body-style	wheel-base	length	engine-type	num-of-cylinders	\
48	66	toyota	hatchback	95.7	158.7	ohc	four	
49	67	toyota	hatchback	95.7	158.7	ohc	four	
50	68	toyota	hatchback	95.7	158.7	ohc	four	
51	69	toyota	wagon	95.7	169.7	ohc	four	
52	70	toyota	wagon	95.7	169.7	ohc	four	
53	71	toyota	wagon	95.7	169.7	ohc	four	
54	79	toyota	wagon	104.5	187.8	dohc	six	

	horsepower	average-mileage	price
48	62	35	5348.0
49	62	31	6338.0
50	62	31	6488.0
51	62	31	6918.0
52	62	27	7898.0
53	62	27	8778.0
54	156	19	15750.0

```
[10]: toyota.describe()
```

```
[10]:
```

	index	wheel-base	length	horsepower	average-mileage	\
count	7.000000	7.000000	7.000000	7.000000	7.000000	
mean	70.000000	96.957143	167.571429	75.428571	28.714286	
std	4.320494	3.326087	10.479299	35.528660	5.089672	
min	66.000000	95.700000	158.700000	62.000000	19.000000	
25%	67.500000	95.700000	158.700000	62.000000	27.000000	
50%	69.000000	95.700000	169.700000	62.000000	31.000000	
75%	70.500000	95.700000	169.700000	62.000000	31.000000	
max	79.000000	104.500000	187.800000	156.000000	35.000000	

	price
count	7.000000
mean	8216.857143
std	3502.409048
min	5348.000000
25%	6413.000000
50%	6918.000000
75%	8338.000000
max	15750.000000

Ahí hemos impreso todos los detalles importantes de la compañía toyota (todos los datos suyos que tenemos así como algunas estadísticas).

1.5 Cuenta el total de automóviles por compañía.

```
[11]: df_limpio['company'].value_counts()
```

```
[11]: toyota          7
      bmw            6
      nissan         5
      mazda          5
      mercedes-benz  4
      audi           4
      volkswagen     4
      mitsubishi     4
      porsche        3
      chevrolet      3
      isuzu          3
      jaguar         3
      honda          3
      alfa-romero    3
      volvo          2
      dodge          2
      Name: company, dtype: int64
```

1.6 Encuentra el kilometraje promedio de cada compañía de automoviles.

```
[12]: df_limpio.groupby('company')['average-mileage'].mean()
```

```
[12]: company
      alfa-romero    20.333333
      audi          20.000000
      bmw           19.000000
      chevrolet     41.000000
      dodge         31.000000
      honda         26.333333
      isuzu         33.333333
      jaguar        14.333333
      mazda         28.000000
      mercedes-benz 18.000000
      mitsubishi    29.500000
      nissan         31.400000
      porsche       17.000000
      toyota        28.714286
      volkswagen    31.750000
      volvo         23.000000
      Name: average-mileage, dtype: float64
```

1.7 Ordena todos los coches por columna Precio y guarda el resultado en un CSV

```
[13]: df_limpio = df_limpio.sort_values('price', ascending=False)
df_limpio.head(3)
```

```
[13]:
```

	index	company	body-style	wheel-base	length	engine-type	\
	35	47	mercedes-benz	hardtop	112.0	199.2	ohcv
	11	14	bmw	sedan	103.5	193.8	ohc
	34	46	mercedes-benz	sedan	120.9	208.1	ohcv

	num-of-cylinders	horsepower	average-mileage	price
35	eight	184	14	45400.0
11	six	182	16	41315.0
34	eight	184	14	40960.0

Hemos considerado ordenarlos de mayor a menor precio. Si quisieramos ordenarlos en orden ascendente simplemente ponemos el flag ascending a True. Vamos a guardar el resultado en un nuevo csv.

```
[14]: df_limpio.to_csv('./Automobile_data_limpio_ordenado.csv')
```