



```
# checking response.status_
if response.status_code != 2
    print(f"Status: {respons
else:
    print(f"Status: {respons

# using BeautifulSoup to par
soup = BeautifulSoup(respons

# finding Post images in the
img = soup.find_all("img")
```

Trabajar con EDA

Metodologías de visualización de datos

Servicio Nacional de Aprendizaje - SENA

Trabajar con EDA

Un ejemplo básico para trabajar con EDA es leer información de un archivo .CSV de un GitHub, que nos traerá datos geográficos, y es el que se utilizará para realizar los ejercicios respectivos y así poder manipular un poco los datos.

Figura 1 Datos de un archivo .CSV

```

1 import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm

url = 'https://raw.githubusercontent.com/lorrey/list-of-countries/master/csv/countries.csv'
df = pd.read_csv(url, sep=";")
print(df.head(5))

/usr/local/lib/python3.7/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is deprecated. Use the functions in the public API at pandas.testing instead.
import pandas.util.testing as tm

alpha_2 alpha_3 area capital continent currency_code \
0 AD AND 468.0 Andorra la Vella EU EUR
1 AE ARE 82800.0 Abu Dhabi AS AED
2 AF AFG 647500.0 Kabul AS AFW
3 AG ATG 443.0 St. John's NAN XCD
4 AI AIA 162.0 The Valley NAN XCD

currency_name equivalent_fips_code fips geoname_id languages \
0 Euro NAN AN 3041565 ca
1 Dirham NAN AE 290557 ar-AE,fa,en,hi,ur
2 Afghani NAN AF 1149261 fa-AF,ps,uz-AF,tk
3 Dollar NAN AC 1576396 en-AG
4 Dollar NAN AV 1579511 en-AI

name neighbours numeric phone population
0 Andorra ES,FR 20 376 84000
1 United Arab Emirates SA,OH 764 971 4975593
2 Afghanistan TH,CN,IR,TJ,PK,UZ 4 93 29121286
3 Antigua and Barbuda NAN 28 +1-268 86754
4 Anguilla NAN 660 +1-264 13254

postal_code_format postal_code_regex tld
0 AD### ~(?:AD)"(0{2})$ .ad
1 NAN NAN .ae
2 NAN NAN .af
3 NAN NAN .ag
4 NAN NAN .ai

```

Luego, observemos los datos básicos que nos suministra Pandas:

Figura 2 Datos de Pandas

```

1 print('Cantidad de Filas y columnas:',df.shape)
print('Nombre columnas:',df.columns)

Cantidad de Filas y columnas: (252, 19)
Nombre columnas: Index(['alpha_2', 'alpha_3', 'area', 'capital', 'continent', 'currency_code',
                        'currency_name', 'equivalent_fips_code', 'fips', 'geoname_id',
                        'languages', 'name', 'neighbours', 'numeric', 'phone', 'population',
                        'postal_code_format', 'postal_code_regex', 'tld'],
                        dtype='object')

```

Después, revisemos las columnas, nulos y los tipos de datos:

Figura 3 Identificación de datos

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 252 entries, 0 to 251
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   alpha_2               251 non-null   object
1   alpha_3               252 non-null   object
2   area                  252 non-null   float64
3   capital               246 non-null   object
4   continent              210 non-null   object
5   currency_code         251 non-null   object
6   currency_name         251 non-null   object
7   equivalent_fips_code  1 non-null     object
8   fips                  249 non-null   object
9   geoname_id            252 non-null   int64
10  languages              249 non-null   object
11  name                   252 non-null   object
12  neighbours            165 non-null   object
13  numeric                252 non-null   int64
14  phone                 247 non-null   object
15  population             252 non-null   int64
16  postal_code_format     154 non-null   object
17  postal_code_regex     152 non-null   object
18  tld                   250 non-null   object
dtypes: float64(1), int64(3), object(15)
memory usage: 37.5+ KB
```

Ahora, se observan en las columnas el total de filas sin nulos y también los tipos de datos. Se muestra la descripción estadística de datos numéricos.

Figura 4 Descripción estadística

```
df.describe()
```

	area	geoname_id	numeric	population
count	2.520000e+02	2.520000e+02	252.000000	2.520000e+02
mean	5.952879e+05	2.427870e+06	434.309524	2.727679e+07
std	1.904818e+06	1.632093e+06	254.663139	1.164127e+08
min	0.000000e+00	4.951800e+04	0.000000	0.000000e+00
25%	1.098000e+03	1.163774e+06	217.000000	1.879528e+05
50%	6.489450e+04	2.367967e+06	436.000000	4.268583e+06
75%	3.622245e+05	3.478296e+06	652.500000	1.536688e+07
max	1.710000e+07	8.505033e+06	894.000000	1.330044e+09

Por último, utilizando Pandas y la librería Matplotlib, obtendremos una gráfica poblacional entre España y Argentina.

Figura 5 Representación gráfica de los datos.

