

print(f"Status: {respons

USING BeautifulSoup to par

SOUP = BeautifulSoup(respons

in the particul of the particul of

Trabajar con EDA

8

g

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

```
    import pandes as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm

      unl = 'https://naw.githubusercontent.com/lorey/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: futurewarming: pandas.util.testing is deprecated. Use the functions in the public API at pandas.testing instead. import pandas.util.testing as tm
                                                                capital continent currency_code \
          alpha_2 alpha_3
AD AVD
                                   area capital
ess.e Andorra la vella
                AE ARE 82880.0
AF AFG 647500.0
                                                            Abu Dhabi
Kabul
                                                                                                            AED
AFN
                                                           St. John's
The Valley
                       ATG 443.0
AIA 102.0
          currency_name eqivalent_fips_code fips geoname_id
Euro Nam AN PARTER
                                                                           3041565 ca
200557 ar-AE,fa,en,h1,ur
1149361 fa-AF,p5,u2-AF,tk
1576296 en.ac
                                                                                                     languages \
                     pollar
          Andorra
United Arab Emirates
Afghanistan
Antigua and Barbuda
Anguilla
         postal_code_format postal_code_regex tld
AD### ^(?:AD)*(\d{3})$ .ad
```

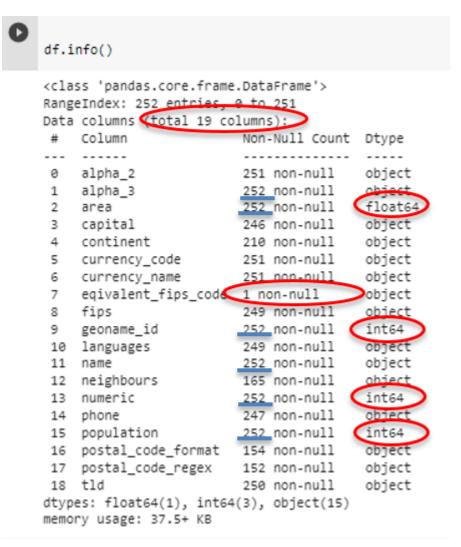
Luego, observemos los datos básicos que nos suministra Pandas:

Figura 2 Datos de Pandas

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

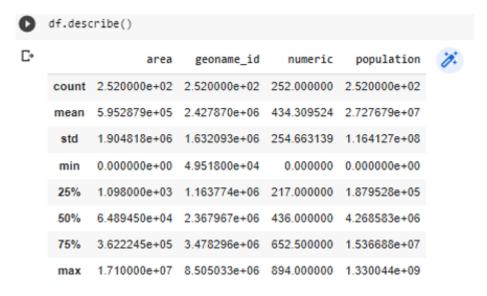


Figura 3 Identificación de datos



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





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

Figura 5 Representación gráfica de los datos.

<matplotlib.axes._subplots.AxesSubplot at 0x7f9fc17fe510>

