GUIA CIENCIA DE DATOS 5 NOVIEMBRE

PRESENTADO POR:

BRAYAN DAVID ROLDAN VARGAS

CIENCIA DATOS

ETITC

NOVIEMBRE

2024

1) Configuración entorno virtual :

```
!python3 -m venv pruebas_envi

[10] !source pruebas_envi/bin/activate

!pip install -r pruebas_envi/requirements.txt

Requirement already satisfied: pandas==2.2.3 in /usr/local/lib/python3.10/dist-packages (from -r pruebas_envi/requireme
Requirement already satisfied: numpy==2.1.3 in /usr/local/lib/python3.10/dist-packages (from -r pruebas_envi/requiremen
Requirement already satisfied: scikit-learn==1.5.2 in /usr/local/lib/python3.10/dist-packages (from -r pruebas_envi/req
Collecting matplotlib==3.9.2 (from -r pruebas_envi/requirements.txt (line 4))
  Using cached matplotlib-3.9.2-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (11 kB)
Requirement already satisfied: seaborn==0.13.2 in /usr/local/lib/python3.10/dist-packages (from -r pruebas_envi/require
Requirement already satisfied: plotly==5.3.1 in /usr/local/lib/python3.10/dist-packages (from -r pruebas_envi/requireme
Collecting bokeh==3.6.0 (from -r pruebas_envi/requirements.txt (line 7))
  Using cached bokeh-3.6.0-py3-none-any.whl.metadata (12 kB)
Collecting dash==2.18.2 (from -r pruebas_envi/requirements.txt (line 8))
  Using cached dash-2.18.2-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas==2.2.3->-
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas==2.2.3->-r pruebas_
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.10/dist-packages (from pandas==2.2.3->-r prueba
Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn==1.5.2->-r pr
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn==1.5.2->-r p
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn==1.5.
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2->-r 
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2->-r prue
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2->-r
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2->-r
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2->-r p
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2->-r pruebas
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2->-r 
Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from plotly==5.3.1->-r prueb
```
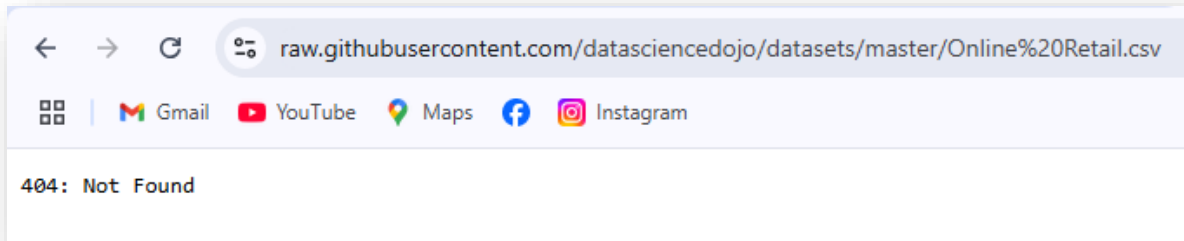
2)instalación manual librerías pip:

```
!pip install pandas==2.2.3
!pip install numpy==2.1.3
!pip install matplotlib==3.9.2
!pip install seaborn==0.13.2
!pip install plotly==5.3.1
!pip install bokeh==3.6.0
!pip install dash==2.18.2

Requirement already satisfied: pandas==2.2.3 in /usr/local/lib/python3.10/dist-packages (2.2.3)
Requirement already satisfied: numpy>=1.22.4 in /usr/local/lib/python3.10/dist-packages (from pandas==2.2.3) (2.1.3)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas==2.2.3) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas==2.2.3) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.10/dist-packages (from pandas==2.2.3) (2024.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas==2.2.3) (1.16.0)
Requirement already satisfied: numpy==2.1.3 in /usr/local/lib/python3.10/dist-packages (2.1.3)
Requirement already satisfied: matplotlib==3.9.2 in /usr/local/lib/python3.10/dist-packages (3.9.2)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2) (1.3.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2) (4.54.1)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2) (1.4.7)
Requirement already satisfied: numpy>=1.23 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2) (2.1.3)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2) (24.1)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2) (10.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib==3.9.2) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib==3.9.2) (1.16.0)
Requirement already satisfied: seaborn==0.13.2 in /usr/local/lib/python3.10/dist-packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in /usr/local/lib/python3.10/dist-packages (from seaborn==0.13.2) (2.1.3)
Requirement already satisfied: pandas>=1.2 in /usr/local/lib/python3.10/dist-packages (from seaborn==0.13.2) (2.2.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in /usr/local/lib/python3.10/dist-packages (from seaborn==0.13.2) (3.9.2)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn==0.13.2) (1.3.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn==0.13.2) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn==0.13.2) (4.54.1)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn==0.13.2) (1.4.7)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn==0.13.2) (24.1)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.10/dist-packages (from matplotlib!=3.6.1,>=3.4->seaborn==0.13.2) (10.4.0)
```

3) la url no se encontró por el momento:

Por lo tanto, no se pudo completar este punto de la actividad

```python
import pandas as pd
import numpy as np
url = "https://raw.githubusercontent.com/datasciencedojo/datasets/master/Online%20Retail.csv"
data=pd.read_csv(url, encoding ='latin 1')
print(data.head())
```

```
---------------------------------------------------------------------------
HTTPError                                 Traceback (most recent call last)
<ipython-input-45-1d58104d6890> in <cell line: 4>()
      2 import numpy as np
      3 url = "https://raw.githubusercontent.com/datasciencedojo/datasets/master/Online%20Retail.csv"  # Aquí es tu ruta al archivo
----> 4 data=pd.read_csv(url, encoding ='latin 1')
      5 print(data.head())

                              12 frames
/usr/lib/python3.10/urllib/request.py in http_error_default(self, req, fp, code, msg, hdrs)
    641 class HTTPDefaultErrorHandler(BaseHandler):
    642     def http_error_default(self, req, fp, code, msg, hdrs):
--> 643         raise HTTPError(req.full_url, code, msg, hdrs, fp)
    644
    645 class HTTPRedirectHandler(BaseHandler):

HTTPError: HTTP Error 404: Not Found
```

4) Creación repositorio en GITHUB:

5) Quedan pendientes las siguientes actividades para el próximo commit:

- Cargar el conjunto de datos de e-commerce desde una URL pública.
- Verificar valores faltantes y explorar las estadísticas del conjunto de datos.
- Filtrar datos para mostrar solo ventas exitosas.
- Crear un gráfico de las ventas por país utilizando 'matplotlib o 'seaborn '.
- Crear un histograma para visualizar la distribución de la cantidad de productos vendidos.
- Crear un gráfico interactivo de dispersión con 'plotly '.
- Crear un gráfico interactivo de líneas con 'bokeh' para mostrar ventas diarias.
- Crear una visualización básica usando 'dash' para mostrar transacciones por país. Presentar los resultados del análisis y recibir retroalimentación. Redactar un resumen de la experiencia y lo aprendido durante la sesión.