



## ACTIVIDAD 1

### Tipo actividad: Ejemplo de análisis de datos del Titanic

En este taller guiado se realizará el análisis exploratorio de los datos del titanic empleando un cuaderno de Jupyter y Python. Para el desarrollo del taller se creará un entorno virtual de Python y se seguirá con el contenido del cuaderno. También se practicarán técnicas de visualización con librerías comunes en el análisis de datos.

#### Materiales

- Computador con conexión a internet

Para el desarrollo del taller, es necesario crear un entorno virtual de Python. La manera más fácil de hacerlo es crear una nueva carpeta e ingresar en ella con el explorador de archivos, en la figura 1 se muestra el ejemplo en un sistema Windows.

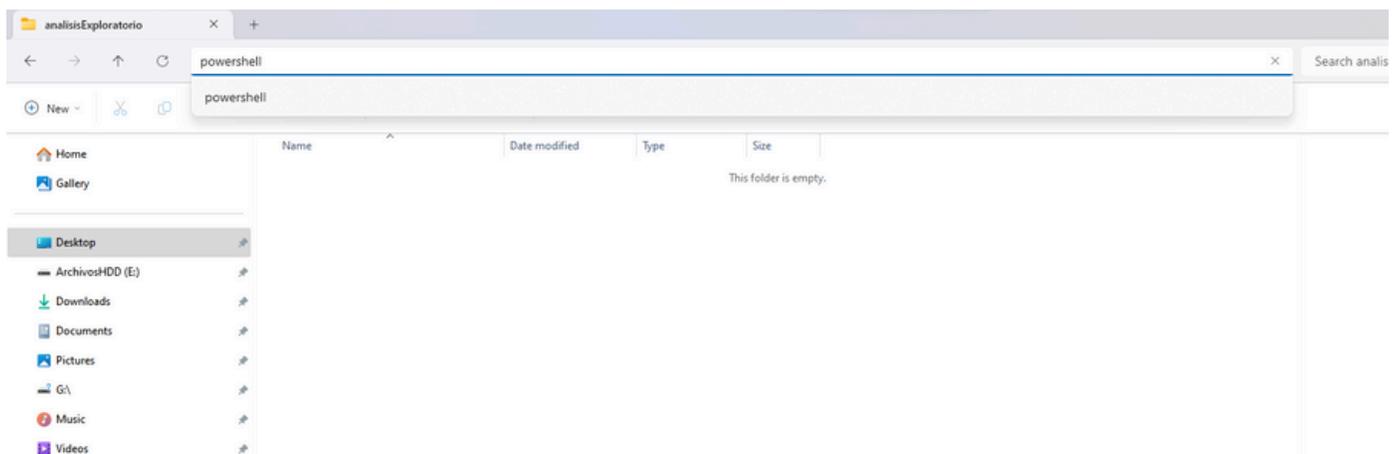




Figura 1. Vista de la carpeta al ingresar.

En la barra de direcciones escribir powershell tal como se muestra en la figura 1.

Se abrirá una ventana de terminal, para poder ejecutar instrucciones. Escribiremos:

```
python -m venv venv
```

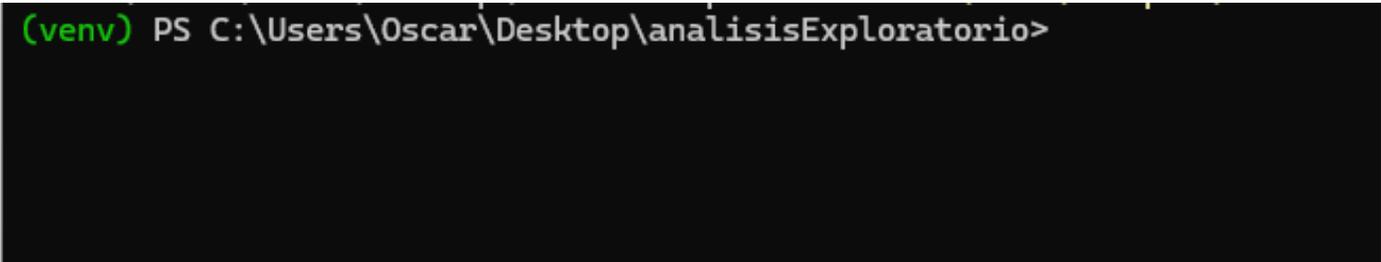
Esta instrucción creará un entorno virtual de Python en el que instalaremos las librerías necesarias para el análisis.

Luego activaremos el entorno virtual ejecutando la instrucción (en la consola)

```
.\venv\Scripts\Activate.ps1
```

(nótese el punto al principio, sin el, la instrucción fallará)

Si el proceso se ejecutó apropiadamente, en la consola se verá como la figura 2, en donde en color verde se verá el nombre del entorno (venv) lo que indica que está activo y recibiendo instrucciones para ese entorno.



```
(venv) PS C:\Users\Oscar\Desktop\ analisisExploratorio>
```

Figura 2: Entorno virtual activado.



Continuaremos instalando las librerías necesarias, digitando la instrucción:

```
pip install matplotlib pandas numpy seaborn jupyter
```

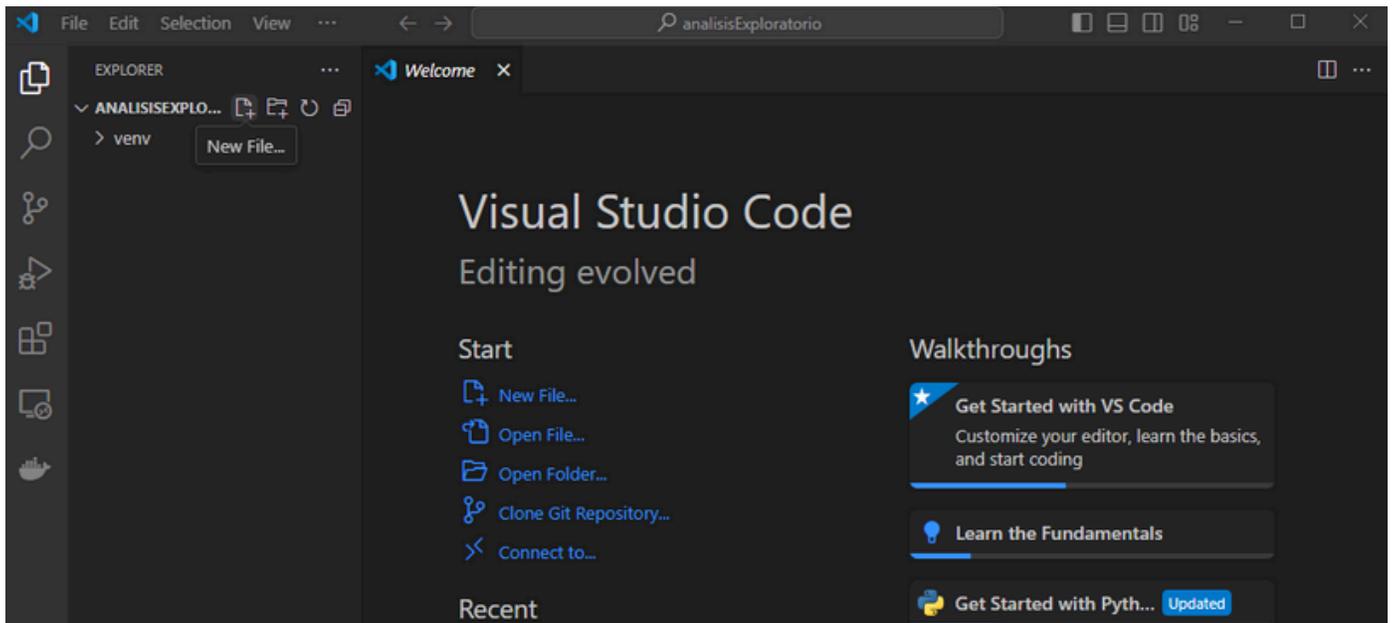
Esto iniciará la descarga e instalación de paquetes. Se verán varias barras de progreso como en la figura 3.

```
Collecting numpy
  Downloading numpy-1.26.4-cp39-cp39-win_amd64.whl (15.8 MB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 15.8/15.8 MB 72.6 MB/s eta 0:00:00
Collecting seaborn
  Downloading seaborn-0.13.2-py3-none-any.whl (294 kB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 294.9/294.9 KB ? eta 0:00:00
Collecting fonttools>=4.22.0
  Downloading fonttools-4.50.0-cp39-cp39-win_amd64.whl (2.2 MB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.2/2.2 MB 67.9 MB/s eta 0:00:00
Collecting kiwisolver>=1.3.1
  Downloading kiwisolver-1.4.5-cp39-cp39-win_amd64.whl (56 kB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 56.2/56.2 KB ? eta 0:00:00
Collecting contourpy>=1.0.1
  Downloading contourpy-1.2.0-cp39-cp39-win_amd64.whl (181 kB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 181.9/181.9 KB ? eta 0:00:00
Collecting pillow>=8
  Downloading pillow-10.2.0-cp39-cp39-win_amd64.whl (2.6 MB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.6/2.6 MB 84.2 MB/s eta 0:00:00
Collecting packaging>=20.0
  Downloading packaging-24.0-py3-none-any.whl (53 kB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 53.5/53.5 KB ? eta 0:00:00
Collecting importlib-resources>=3.2.0
  Downloading importlib_resources-6.3.0-py3-none-any.whl (35 kB)
Collecting python-dateutil>=2.7
  Downloading python_dateutil-2.9.0.post0-py2.py3-none-any.whl (229 kB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 229.9/229.9 KB ? eta 0:00:00
Collecting pyparsing>=2.3.1
  Downloading pyparsing-3.1.2-py3-none-any.whl (103 kB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 103.2/103.2 KB ? eta 0:00:00
```

Figura 3: consola instalando las bibliotecas de Python.

Al terminar podemos escribir en la consola la palabra 'code .' (con el punto) eso abrirá el editor de visual studio code. Crearemos un cuaderno llamado analisisexploratorio.ipynb desde el panel del lado izquierdo, el cual, se debió haber abierto en la carpeta que creamos (para eso se incluyó el punto en la instrucción code).

En la figura 4 se muestra el proceso de creación de un nuevo archivo que llamaremos analisisexploratorio.ipynb



Al crear el archivo Visual Studio Code detectará el entorno virtual creado y permitirá crear los programas con los que se va a trabajar en esta lección.



Finalmente debemos copiar el dataset del titanic (archivo .csv) en la carpeta que hemos creado con el fin de que todo el entorno y las dependencias estén correctamente dispuestas.

En la primera celda del cuaderno, crearemos el código que importa las bibliotecas:

```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

sns.set(style='white', context='notebook', palette='deep')
```

Posteriormente hacemos click en el botón triangular (signo de play) o haremos en el teclado la combinación de teclas shift y enter. Esto intentará ejecutar el programa y solicitará al usuario establecer el entorno de Python deseado. Saldrá un menú en el centro de la parte superior de la ventana (figura 5) en el que escogeremos la opción Python environments...

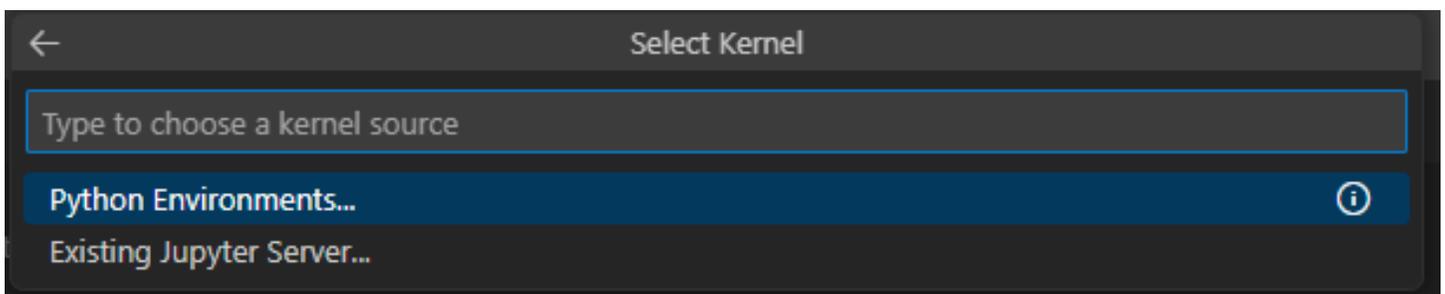


Figura 5: selección de entornos.



En caso de que la ventana se pierta, se puede hacer click en el botón select kernel de la parte superior derecha de la pantalla (se puede leer el texto con facilidad).

Escogeremos la opción recomendada (figura 6) la cual debe corresponder con el entorno virtual que creamos en pasos anteriores. Para validar que es el entorno que creamos, se debe visualizar la ruta

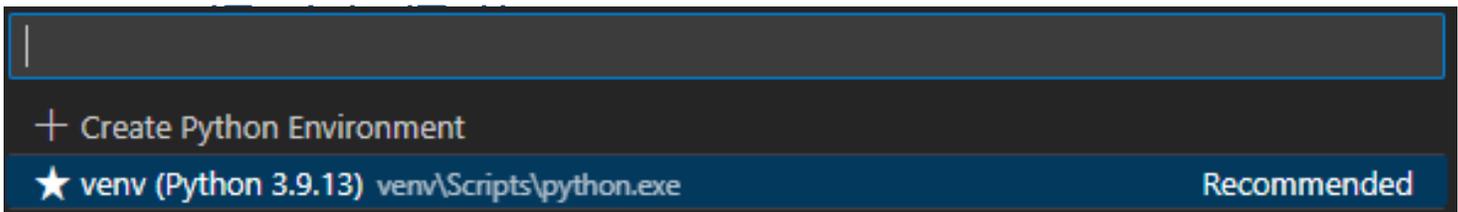


Figura 6: entorno a seleccionar.

Si todo sale bien, se debe ver un indicador verde en la parte inferior de la celda como el de la figura 7:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

sns.set(style='white', context='notebook', palette='deep')
```

✓ 1.4s

Figura 7 indicador de que el código se ejecutó correctamente en el entorno.



Crearemos una segunda celda en donde abriremos el archivo y visualizaremos las primeras 5 filas. Esto se hace con las instrucciones:

```
data = pd.read_csv('titanic.csv')  
  
data.head()
```

se debe ver la salida como la de la figura 8:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Figura 8: cabecera visible.

Se continuará con el taller en el cuaderno de referencia llamado `cuadernoexploracion.ipynb` disponible adjunto a la lección.