

Pre-requisitos

1. Este taller es **individual**.
2. La entrega de este taller consiste en un **reporte** por persona y unos **archivos de soporte**.
3. Cree el archivo de su **reporte** como un documento de texto en el que pueda fácilmente incorporar capturas de pantalla, textos y similares. La entrega se debe realizar en un archivo de formato PDF.
4. En su **reporte** indique claramente el número de la sección y el numeral que está respondiendo.
5. Incluya los cuadernos y/o scripts realizados como **archivos de soporte**.

1. Instale su ambiente de desarrollo

1. Instale Python **3.12**. Se recomienda emplear la distribución de pypi disponible en <https://www.python.org/downloads/>
Identifique claramente la ubicación (directorio) en la que queda instalado python.
Incluya la ubicación local de la distribución instalada en su **reporte**.
2. Instale Visual Studio Code disponible en
<https://code.visualstudio.com/>
Puede usar otro editor, pero el equipo docente tendrá mayor capacidad de atender sus dudas en Visual Studio Code.
3. En Visual Studio puede usar CTRL+SHIFT+P para seleccionar el intérprete de python. Seleccione el intérprete instalado de la versión 3.12 de Python.
4. Si no tiene una terminal abierta, abra una en el Menú Terminal, New Terminal.
5. En la terminal instale los paquetes numpy, scipy, pandas, matplotlib y seaborn con el comando (si está usando pypi)
`pip install numpy scipy pandas matplotlib seaborn`

6. Verifique la versión instalada de cada paquete con el comando

```
pip list
```

En su **reporte** incluya las versiones de los 5 paquetes instalados en el punto anterior.
Incluya también un pantallazo con la salida del comando en su **reporte** donde se vea claramente la versión instalada de cada paquete y su usuario en el equipo local.

7. Genere un archivo *requeriments* que almacene las versiones de los paquetes instalados con el comando

```
pip freeze > requirements.txt
```

Incluya el archivo resultado como **archivo de soporte**.

2. Exploración y limpieza de datos en python

1. Adjunto a este taller encontrará un cuaderno de nombre pydatos_pandas.ipynb, que le puede servir para refrescar el uso de Series y Dataframes en Pandas (si es la primera vez que ejecuta un cuaderno en Visual Studio, le pedirá instalar la extensión para cuadernos de Jupyter).
2. Adjunto a este taller encontrará también un cuaderno de nombre pydatos_limpieza.ipynb, que ilustra varias operaciones de manipulación y limpieza de datos, empleando un archivo de valores separados por comas (CSV) diabetes.csv adjunto.
3. Abra el cuaderno, revise y ejecute cada una de las celdas (si es la primera vez que ejecuta un cuaderno en Visual Studio, le pedirá instalar la extensión para cuadernos de Jupyter).
4. Seleccione **5 celdas** que hagan operaciones **diferentes**. En su **reporte** explique qué se hace en cada una de estas celdas (brevemente, dos a tres líneas por celda).
5. Ahora tome los datos BikePrices.csv que encontrará en Bloque Neón. Cargue los datos en un **nuevo cuaderno**. En este cuaderno realice **10 operaciones** de exploración y limpieza que considere relevantes. Para cada celda de código incluya una celda de markdown para explicar la operación realizada. Incluya el cuaderno como parte de su entrega (**archivo de soporte**).

3. Visualización de datos en python

1. Adjunto a este taller encontrará también un cuaderno de nombre pydatos_viz.ipynb, que ilustra varias visualizaciones de datos, empleando varios archivos CSV (diabetes, spx, titanic) adjuntos.
2. Abra el cuaderno, revise y ejecute cada una de las celdas (si es la primera vez que ejecuta un cuaderno en Visual Studio, le pedirá instalar la extensión para cuadernos de Jupyter).
3. Seleccione **5 celdas** que hagan operaciones **diferentes**. En su **reporte** explique qué se hace en cada una de estas celdas (brevemente, dos a tres líneas por celda). Incluya en esta elección 1 histograma y 1 diagrama de violín.
4. Ahora tome los datos BikePrices.csv que encontrará en Bloque Neón. Cargue los datos en un **nuevo cuaderno**. En este cuaderno realice **5 visualizaciones** que considere relevantes. Para cada celda de código incluya una celda de markdown para explicar la visualización y su **interpretación**. Incluya el cuaderno como parte de su entrega (**archivo de soporte**).