

Plan Formativo: Ciencia de Datos	Nivel de Dificultad
Módulo 4: Análisis exploratorio y programación estadística.	Bajo / medio
Tema: Gráficos	
Intención del aprendizaje o aprendizaje esperado:	
<ul style="list-style-type: none"> <li>● Presentar información de un set de datos utilizando librería Matplotlib para graficar la información.</li> </ul>	
Ejercicios planteados	
<p><b>Caso:</b></p> <p>Se tiene un fichero csv con información relativa a datos de un estudio realizado en la zona de Boston, donde se analizan diferentes variables, como son el índice de crimen, el número de habitaciones, el porcentaje de habitantes de clase baja y el valor medio de las casas de esa zona.</p> <div data-bbox="694 1500 866 1599" data-label="Image"> <p>casasboston.csv</p> </div> <p><b>Pasos:</b></p> <ul style="list-style-type: none"> <li>● Importar librerías a utilizar para el caso en cuestión (panda, matplotlib)</li> <li>● Leer csv correspondiente.</li> <li>● Renombrar campos.</li> <li>● Confeccionar los siguientes Gráficos             <ul style="list-style-type: none"> <li>○ <b>Histograma:</b> Un histograma es útil para ver la distribución de una variable, es decir, nos permite ver los valores más comunes. Para</li> </ul> </li> </ul>	

el caso que se presenta se desea ver la distribución de la cantidad media de habitaciones en el estudio realizado.

- **Dispersión:** sirve para representar la relación entre dos variables. Para nuestro caso se desea ver la relación entre índice de crimen y el valor mediano de las casas. En otras palabras, ¿el índice de crimen afecta el valor medio de las casas?
- **Barras:** Un gráfico de barras es útil para comparar una variable entre distintos grupos o categorías. Por ejemplo, si quisiéramos observar el valor medio de cada ciudad. Para nuestro caso se desea agrupar por ciudad y toma la media del valor mediano. Gráfica las 10 primeras ciudades.
- **Cajas:** Los diagramas de cajas son útiles para representar grupos de datos y compararlos entre ellos. Otro uso importante es que nos permiten identificar de forma sencilla si una variable tiene muchos *outliers* (valores atípicos) esto es, elementos que se alejan de los valores frecuentes de dicha variable. Para nuestro caso, se desea ver los valores atípicos de índice de crimen en los diferentes cuantiles de valor mediano.

#### Datos de apoyo al planteamiento

- Librería Matplotlib.

#### Preguntas guía

- Tipos de gráficos para la librería Matplotlib

#### Recursos Bibliográficos:

[1] Librería Matplotlib

<https://www.youtube.com/watch?v=2VeHtuqW3YY>

[https://programacion.net/articulo/introduccion a la libreria matplotlib de python 1599](https://programacion.net/articulo/introduccion%20a%20la%20libreria%20matplotlib%20de%20python%201599)

[2] Matplotlib

<https://matplotlib.org/>

[3] Estilos.

<https://interactivechaos.com/es/manual/tutorial-de-matplotlib/estilos>

[4] Gráficas, programación orientada a objetos en Python y Matplotlib

<https://www.youtube.com/watch?v=4AYPLe8yosY>

[3] Tablas de Contingencia

<https://www.youtube.com/watch?v=CEWJZxFpC8w>

<https://conceptosclaros.com/que-es-tabla-contingencias-par-que-sirve/>

[4] Coeficiente de Pearson

[https://www.uv.es/webgid/Descriptiva/31\\_coeficiente\\_de\\_pearson.html](https://www.uv.es/webgid/Descriptiva/31_coeficiente_de_pearson.html)