

LABORATORIO ★ 10

INTRODUCCIÓN AL DATA SCIENCE

Instrucción :

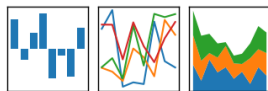
- *El presente ejercicio tiene una bonificación en décimas para la siguiente evaluación individual del curso. La bonificación estará en el rango 0 - 10 décimas. Se subirá solución del ejercicio propuesto dentro de una semana.*

Problema

Para este laboratorio deberás realizar una breve implementación del algoritmo de *clustering k-means* revisado en clases. Para ello se sugiere armar una base de datos **sencilla** (no más de diez datos y tres parámetros cada uno por ejemplo). Se deberá establecer un valor de k previamente y terminar la iteración del algoritmo en base a un criterio lógico y conveniente (2 iteraciones con la conformación de *clustering* idéntico). Posteriormente se pide mostrar una visualización de estos datos por medio del uso de la librería **Matplotlib**. Disponible en <https://matplotlib.org/>. Fuentes de información disponibles recomendadas son las siguientes:

- K. Arvai. (2022). *K-Means Clustering in Python: A Practical Guide*. Recuperado de <https://realpython.com/k-means-clustering-python/>
- Na8. (2018). *K-Means en Python paso a paso*. Recuperado de <https://bit.ly/3MT2Z9F>
- C. Maklin. (2018). *K-means Clustering Python Example*. Recuperado de <https://bit.ly/3LP3ksE>

pandas
 $y_{it} = \beta^t x_{it} + \mu_i + \epsilon_{it}$



NumPy



SciPy

matplotlib