

ACTIVIDAD 1

Tipo actividad: (Lección 4: Taller de agrupamiento)

Tiempo de ejecución: 3 horas	
PLANTEAMIENTO DE LA SESIÓN	Materiales
En esta sesión se desarrollará un taller sobre los temas vistos en la unidad. Se trabajará con el conjunto de datos de vinos, el cual tiene algunas características (todas numéricas) de diferentes tipos de vinos.	Dataset de vinos (wine) https://www.kaggle.com/datasets/harrywang/wine-dataset-for-clustering

Nota para el docente: adjunto encontrará un cuaderno de jupyter de guía para la actividad donde está previamente desarrollada:

https://drive.google.com/file/d/1i8iyRPZevnAqLsvK-MLXOAnNOxSiGJSP/view?usp=drive_link

Para el desarrollo del taller, descargue el dataset de vinos y desarrolle las siguientes tareas:

1. Cargue el dataset en memoria utilizando un cuaderno de jupyter.
2. Encuentre la distribución de cada variable (característica) del dataset. Se puede realizar creando un histograma para cada columna del dataset.
3. Utilizando el algoritmo UMAP, reduce las dimensiones del conjunto de datos a 2 componentes, estas componentes serán renombradas a comp_1 y comp2.
4. Utilizando librerías como scikit-learn ejecute el algoritmo de k-means para la característica inercia. Repítalo utilizando k grupos siendo k enteros entre 2 y 10, luego grafique el número de clusters vs la cantidad de datos por cluster. Lleve a los estudiantes a que identifiquen el “codo” en el gráfico como el número óptimo de clusters.
5. Repita el proceso de clustering empleando otros algoritmos. Como guía para el desarrollo puede basarse en el cuaderno de jupyter adjunto.