# Aplicación de algoritmos de agrupamiento en Sklearn

Guía paso a paso para que un estudiante aplique algoritmos de agrupamiento en Scikit-learn a una base de datos tipo CSV descargada de plataformas como Kaggle o UCI Machine Learning Repository:

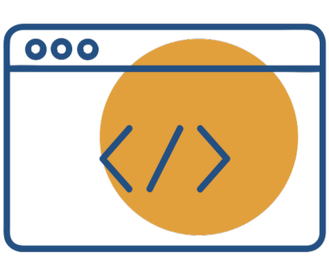
## 1. Descargar y explorar los datos

Descarga el archivo CSV de la plataforma deseada y examina su contenido para comprender la estructura de los datos y las características disponibles.

## 2. Preprocesamiento de datos

- Limpia los datos eliminando filas con valores faltantes o duplicados.
- Si es necesario, normaliza o estandariza las características para asegurarte de que todas estén en la misma escala.
- Divide los datos en características (X) si es un problema no supervisado.

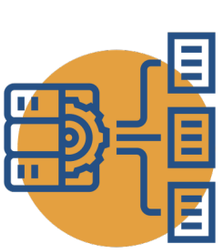
## 3. Selección del algoritmo de agrupamiento



- Elige el algoritmo de agrupamiento adecuado para tu problema. Algunas opciones comunes incluyen K- Means, Mean Shift, DBSCAN, y algoritmos de agrupamiento jerárquico.
- Importa el algoritmo correspondiente de Scikit-learn.

## 4. Entrenamiento del modelo

- Ajusta el modelo a los datos utilizando el método fit.
- Puedes ajustar diferentes hiperparámetros del modelo según sea necesario.



## 5. Visualización de los resultados

- Visualiza los resultados del agrupamiento utilizando técnicas de reducción de dimensionalidad como PCA o t-SNE.
- Grafica los datos en un espacio bidimensional y colorea los puntos según el grupo asignado por el algoritmo de agrupamiento.

## 6. Evaluación del agrupamiento (opcional)

- Si tienes etiquetas verdaderas disponibles, puedes evaluar la calidad del agrupamiento utilizando métricas como el índice de Rand ajustado, el índice de Jaccard ajustado o la pureza.
- Ten en cuenta que en la mayoría de los casos, el agrupamiento es un problema no supervisado y no hay etiquetas verdaderas disponibles.

## 7. Interpretación de los clusters

- Analiza los clusters resultantes para entender las características de los grupos y cómo se relacionan con las características originales.
- Examina los centroides o los puntos centrales de los clusters si estás utilizando K-Means.

Siguiendo estos pasos, un estudiante puede aplicar con éxito algoritmos de agrupamiento en Scikit-learn a una base de datos tipo CSV descargada de plataformas de aprendizaje automático como Kaggle o UCI Machine Learning Repository.