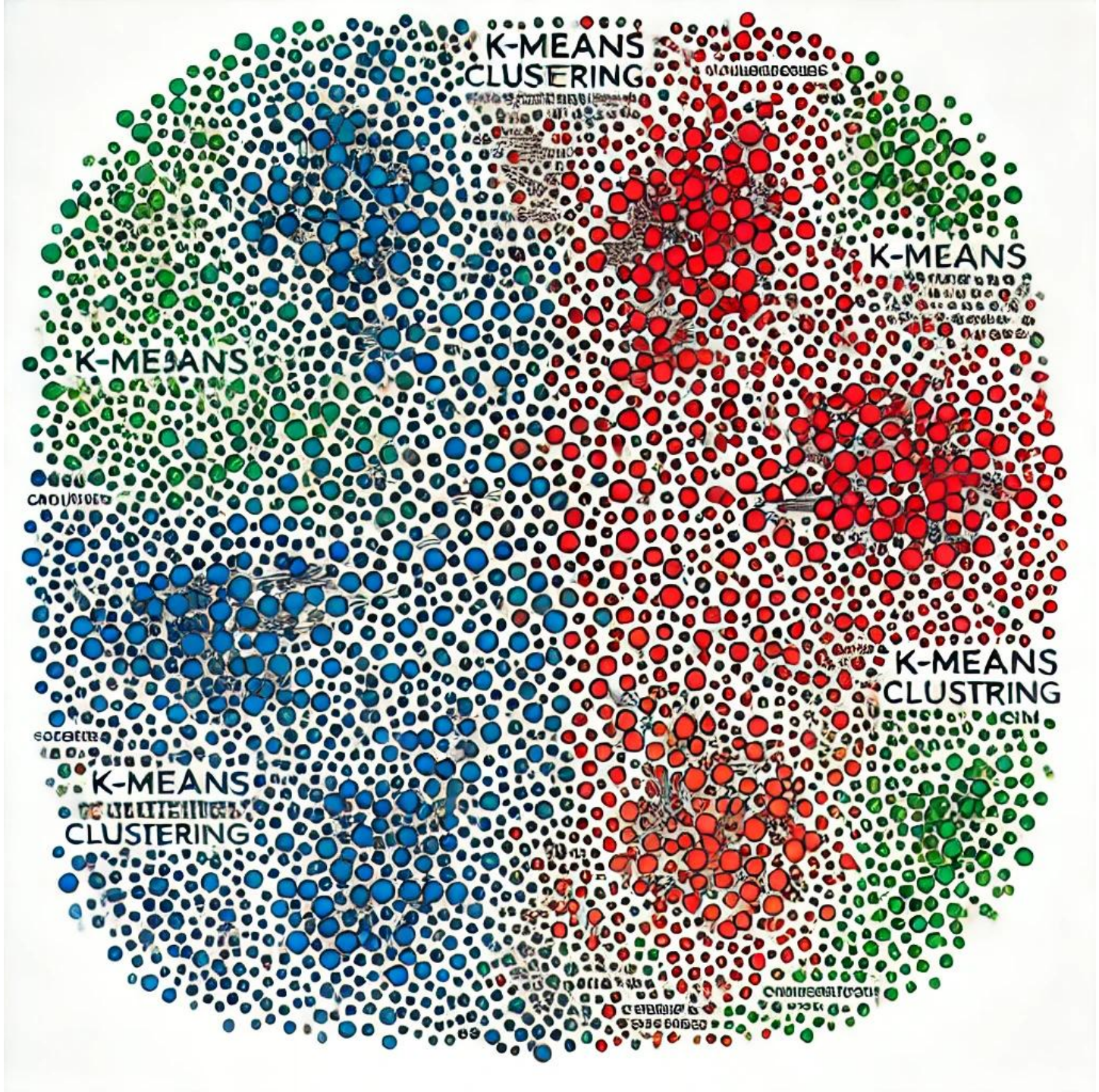
K-MEANS
CLUSTERING
K-MEANS
K-MEANS
K-MEANS
CLUSTERING
K-MEANS
CLUSTERING

K-Means es un algoritmo de aprendizaje no supervisado utilizado para **clustering** o agrupamiento de datos. Su objetivo es dividir un conjunto de datos en un número predefinido de grupos o clústeres basados en la similitud de las muestras. Cada clúster está definido por su **centroide**, que es el punto promedio de todas las muestras dentro de ese clúster.

Pasos básicos de K-Means:

1. **Inicialización**: Se eligen *K* centroides de forma aleatoria.
2. **Asignación**: Cada punto de datos se asigna al clúster cuyo centroide está más cercano, basado en una métrica de distancia, generalmente la **distancia euclidiana**.
3. **Actualización**: Se recalculan los centroides de cada clúster tomando el promedio de los puntos asignados a cada uno.
4. **Repetición**: Los pasos de asignación y actualización se repiten hasta que los centroides ya no cambien significativamente, o hasta alcanzar un número máximo de iteraciones.

Importancia del EDA:

- **Comprender los datos**: Ayuda a obtener una visión clara del comportamiento de los datos, sus características, y cualquier tendencia o patrón que pueda influir en el análisis posterior.
- **Preparación del modelado**: Identificar problemas que deben corregirse antes de aplicar cualquier modelo de machine learning, como valores perdidos, outliers o distribuciones sesgadas.
- **Generación de hipótesis**: Facilita la creación de hipótesis sobre los datos que pueden ser probadas más adelante con técnicas más avanzadas.

En resumen, el EDA es una etapa clave para analizar y explorar un conjunto de datos de forma exhaustiva antes de pasar a análisis más profundos o a la creación de modelos predictivos.