

The background of the slide features a complex network diagram with numerous nodes and connecting lines, rendered in a light blue color against a dark blue background. The nodes are small squares, and the lines are thin, creating a web-like structure that fills the entire slide.

Aprendizaje de Máquina **No Supervisado**

Sesión 2



¿Qué es la Clusterización?

1

Agrupación de datos

Método de aprendizaje no supervisado para agrupar datos en clusters.

2

Similitud

Organiza datos similares dentro del mismo cluster.

3

Sin etiquetas

Descubre relaciones ocultas sin necesidad de etiquetas previas.

Aplicaciones de la Clusterización

Segmentación de clientes

Personaliza estrategias de marketing agrupando clientes similares.

Medicina y biología

Identifica patrones en enfermedades y pacientes para tratamientos personalizados.

Detección de fraudes

Detecta transacciones sospechosas en finanzas y seguridad.

Análisis geográfico

Identifica patrones en datos ambientales y demográficos.

Recomendación de contenido

Personaliza sugerencias en plataformas como Netflix y Spotify.

Visión por computadora

Agrupar imágenes y videos para reconocimiento automático.

Algoritmo K-Means

K-Means es un algoritmo de clusterización basado en particiones. Se usa para dividir un conjunto de datos en K grupos (clusters). El número K es elegido por el usuario.

Divide datos

Divide datos en K grupos definidos por el usuario.

Centroides

Asigna cada dato al centroide más cercano.

Rápido

Es rápido y eficiente en grandes volúmenes de datos.



Ventajas y Desventajas de K-Means

Ventajas

Es rápido y eficiente en grandes volúmenes de datos.

Fácil de entender e implementar.

Funciona bien cuando los grupos tienen formas circulares o esféricas.

Desventajas

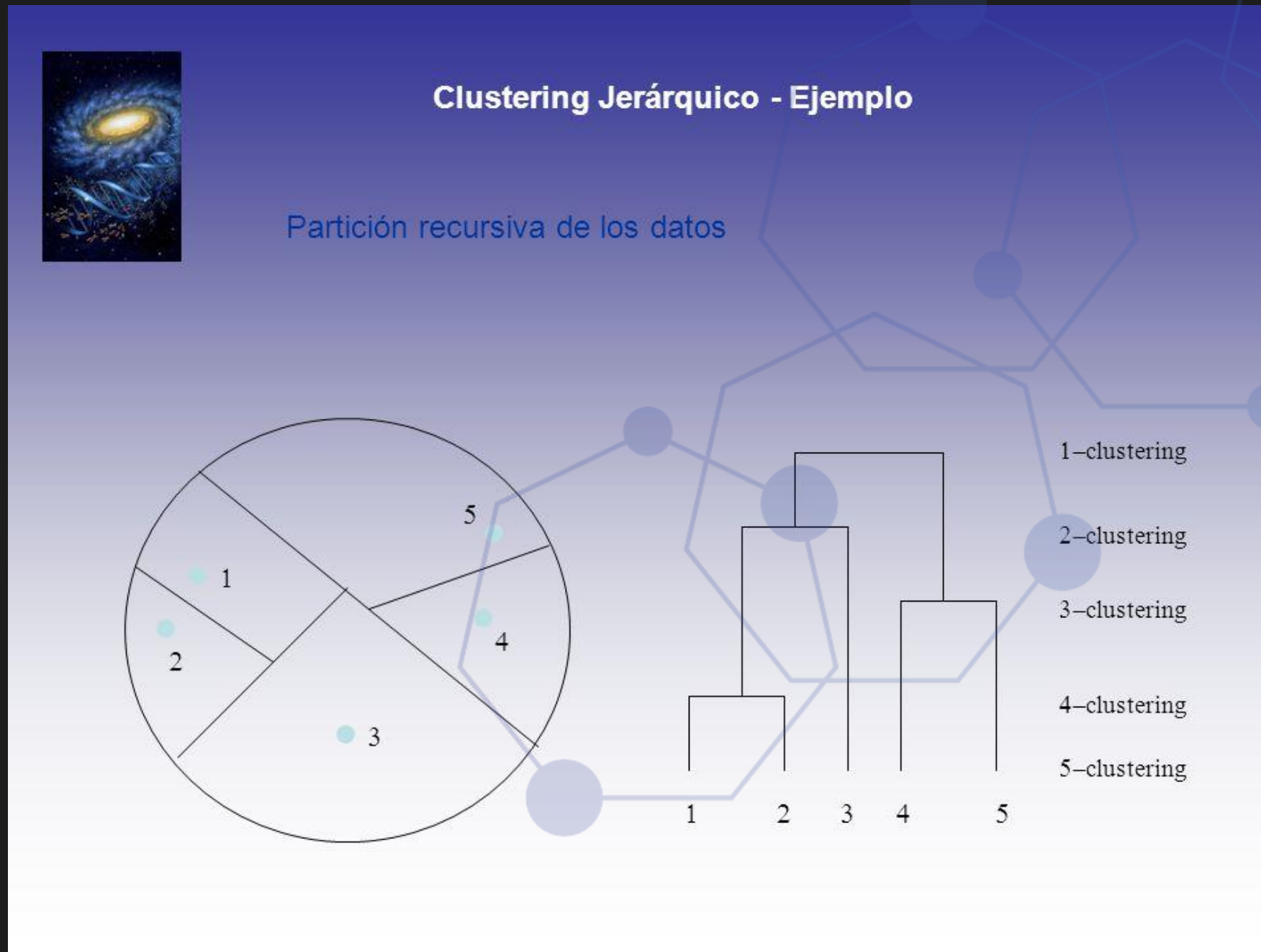
Debes definir el número de clusters K de antemano.

No funciona bien si los clusters tienen formas irregulares.

Sensible a outliers (datos atípicos).



Clustering Jerárquico



Fuente: slideplayer.es

Estructura de árbol

Agrupar datos en una estructura de árbol (dendrograma).

No requiere K

No requiere especificar el número de clusters.

Tipos

Aglomerativo (bottom-up) y
divisivo (top-down).

Ventajas y Desventajas del Clustering Jerárquico

Ventajas	Desventajas
No requiere definir el número de clusters de antemano.	Computacionalmente más costoso que K-Means.
Permite visualizar la relación entre los datos con un dendrograma.	Difícil de aplicar en grandes volúmenes de datos.

Algoritmo DBSCAN

Algoritmo DBSCAN

**Basado en
densidad**
Identifica
clusters como
regiones densas
de puntos.

Outliers

Deja fuera los
puntos aislados
(outliers).

No requiere K

No requiere
especificar el
número de
clusters.

Mezcla Gaussiana (GMM)

1

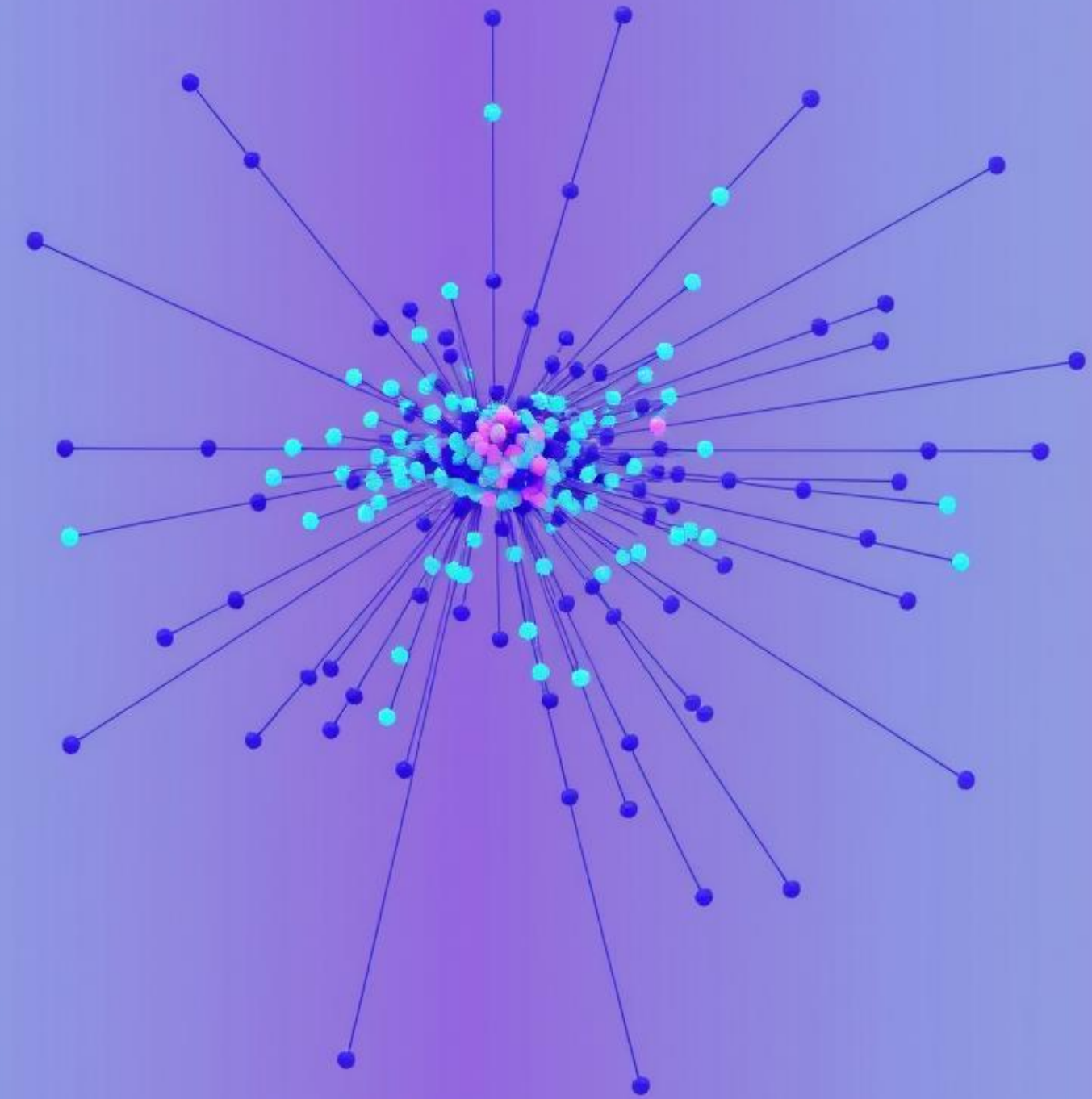
- Basado en estadística
- Datos provienen de distribuciones normales.

2

- Flexible
- Más flexible que K-Means.

3

- Superposición
- Funciona bien con clusters superpuestos.



Casos Prácticos



E-commerce

Segmentación de clientes para personalizar ofertas.



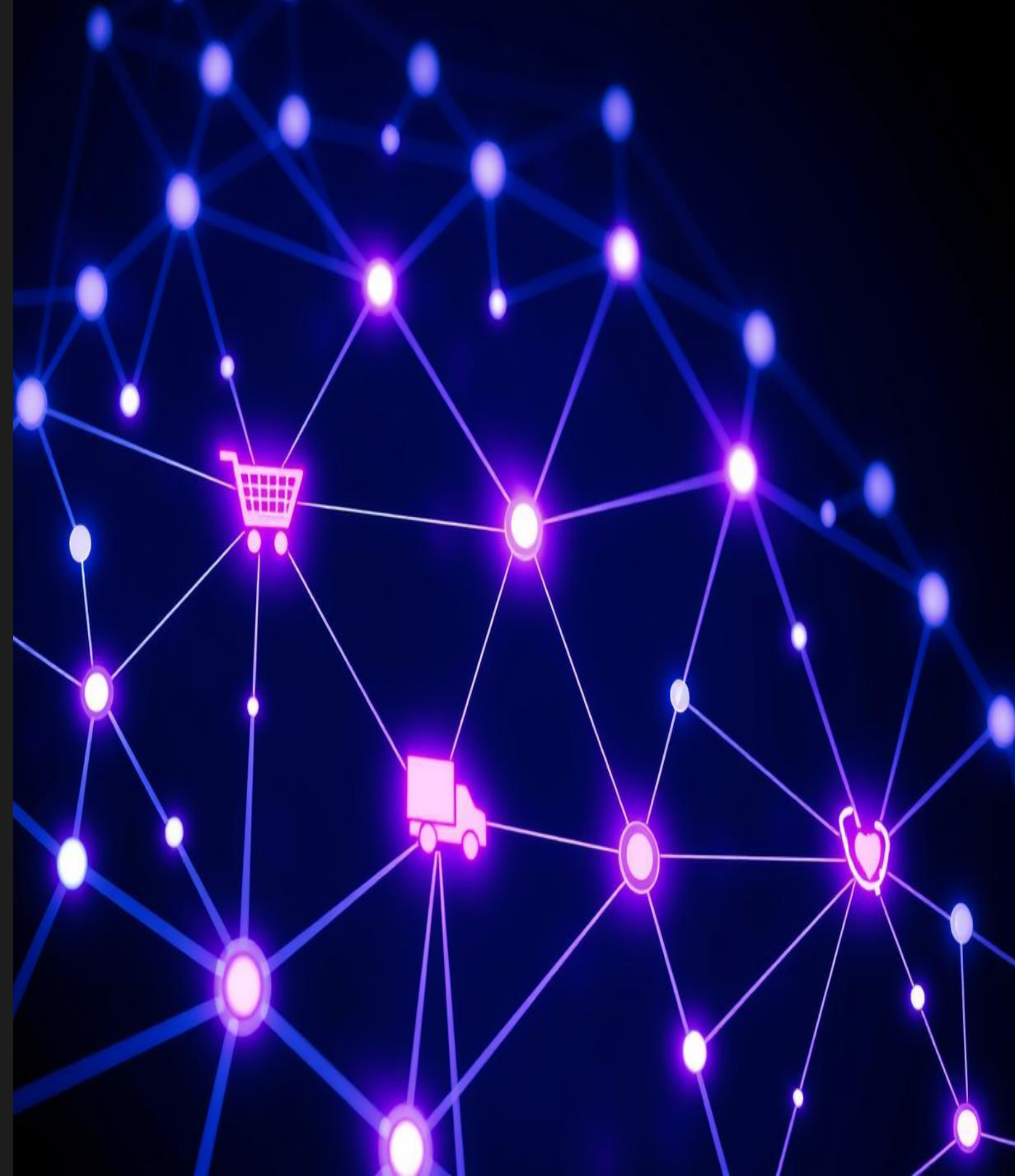
Logística

Optimización de rutas de entrega con DBSCAN.



Medicina

Diagnóstico y personalización de tratamientos.



Otros Casos



Streaming

Recomendación de contenido en Netflix, Spotify.



Bancos

Seguridad y detección de fraudes en bancos.

Actividad Práctica Guiada



SEGMENTACIÓN DE CLIENTES CON K-MEANS

Objetivo: Aplicar K-Means para agrupar clientes según su comportamiento de compra.

Requisitos:

- Tener instalado Python y Jupyter Notebook (o Google Colab).
- Instalar numpy, matplotlib y sklearn si aún no los tienes (pip install numpy matplotlib scikit-learn).

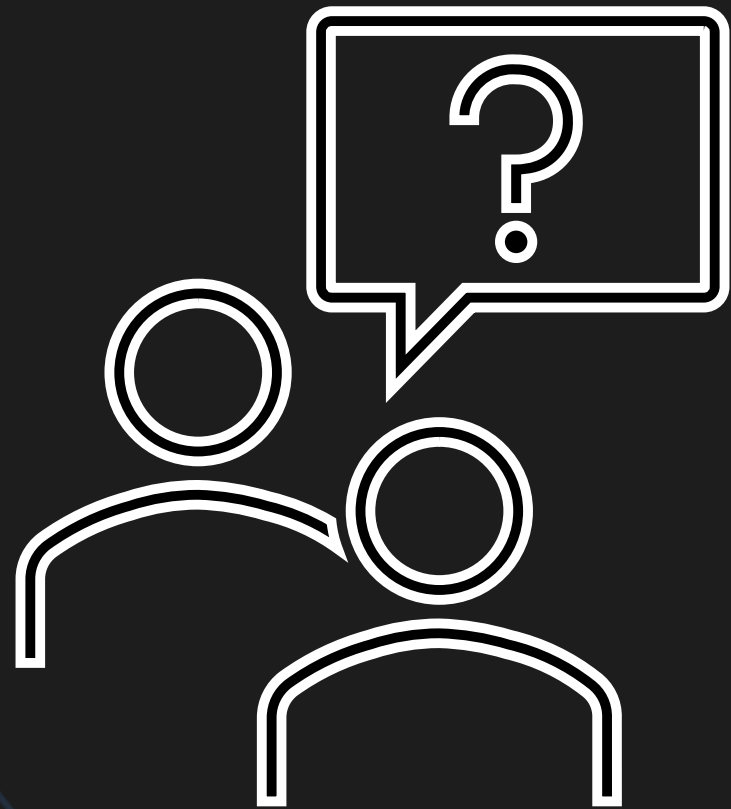
Pasos:

1. Importar librerías.
2. Crear Datos Simulados: 100 clientes con dos variables > Monto gastado en compras > Frecuencia de compra.
3. Aplicar K-Means.
4. Analizar los Clusters.

El detalle de la actividad se encuentra en la guía de estudio de la sesión.

Preguntas

Sección de preguntas



The background of the slide features a complex network diagram with numerous nodes and connecting lines, rendered in a light blue color against a dark blue background. The nodes are small squares, and the lines are thin and interconnected, creating a web-like structure that fills the entire slide.

Aprendizaje de Máquina

No Supervisado

Continúe con las
actividades
