

Introducción al Business Analytics

Week 08: Agrupando Observaciones

Departamento de Economía, Universidad Icesi

September 17, 2025

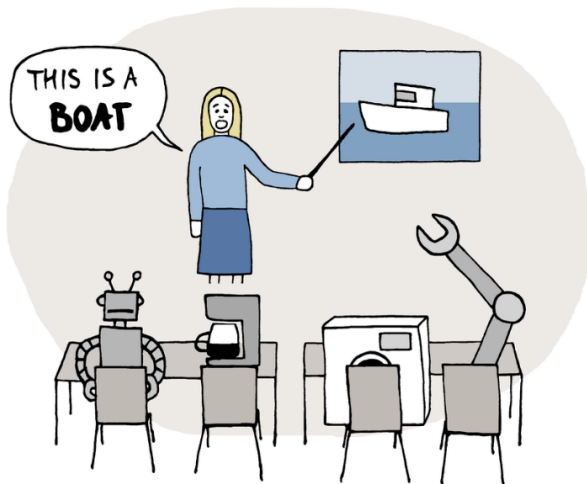
Objetivos de aprendizaje

Al finalizar la unidad, el estudiante será capaz de:

- Explicar en sus propias palabras qué es un clúster.
- Construir clústeres de observaciones en R e interpretar sus resultados.
- Interpretar un dendograma.
- Explicar en sus propias palabras la diferencia entre un clúster jerárquico y uno construido por el método de k-means.

¿Qué es el aprendizaje de maquinas?

MACHINE LEARNING



¿Qué es el aprendizaje de máquinas?

Definición simple:

Es cuando las computadoras aprenden de los datos para hacer predicciones o descubrir patrones, sin estar programadas paso a paso.

Ejemplos cotidianos:

- Netflix recomienda series según lo que viste antes.
- Un banco detecta transacciones sospechosas en tu tarjeta.
- El celular reconoce rostros en fotos automáticamente.

Idea clave:

El aprendizaje de máquinas es como dar **experiencia** a la computadora: cuanta más experiencia (datos) tenga, mejores serán sus decisiones o descubrimientos.

Roadmap

1 Aprendizaje Supervisado y no Supervisado

2 Clustering

3 Aplicación en R

4 Aplicación en R

Aprendizaje Supervisado vs. No Supervisado

- **Supervisado:**

- ▶ Usa datos con **etiquetas** (la respuesta ya está dada).
- ▶ Objetivo: **predecir** correctamente esa respuesta.
- ▶ Ejemplos: detectar spam, predecir precio de vivienda, estimar duración de un vuelo.

Aprendizaje Supervisado vs. No Supervisado

- **Supervisado:**

- ▶ Usa datos con **etiquetas** (la respuesta ya está dada).
- ▶ Objetivo: **predecir** correctamente esa respuesta.
- ▶ Ejemplos: detectar spam, predecir precio de vivienda, estimar duración de un vuelo.

- **No Supervisado:**

- ▶ Trabaja con datos **sin etiquetas**.
- ▶ Objetivo: **descubrir patrones** o grupos ocultos.
- ▶ Ejemplos: segmentar clientes, encontrar productos que se compran juntos.

Aprendizaje Supervisado vs. No Supervisado

• Supervisado:

- ▶ Usa datos con **etiquetas** (la respuesta ya está dada).
- ▶ Objetivo: **predecir** correctamente esa respuesta.
- ▶ Ejemplos: detectar spam, predecir precio de vivienda, estimar duración de un vuelo.

• No Supervisado:

- ▶ Trabaja con datos **sin etiquetas**.
- ▶ Objetivo: **descubrir patrones** o grupos ocultos.
- ▶ Ejemplos: segmentar clientes, encontrar productos que se compran juntos.

Idea clave:

En supervisado sabemos qué predecir;
en no supervisado buscamos *estructura oculta*.

Video: Aprendizaje Supervisado vs. No Supervisado

Recurso recomendado

▶ [Ver el video en YouTube](#)

Instrucción: Mira el video completo y toma nota de un ejemplo que te ayude a diferenciar entre aprendizaje supervisado y no supervisado.

Actividad interactiva en Kahoot

Pon a prueba tu comprensión

► Ingresar al Kahoot

Instrucciones:

- Ingresa al Kahoot desde tu celular o computador.
- Forma un equipo de **2 a 3 estudiantes**.
- Al registrarte, escribe el **primer nombre** y la **primera letra de tu apellido**, para identificar tu grupo.
- Responde todas las preguntas relacionadas con el video de **aprendizaje supervisado y no supervisado**.
- **Premio:** Los 3 mejores equipos recibirán una **bonificación en la nota del segundo parcial**.
 - ▶ La bonificación se divide entre los integrantes del grupo.
 - ▶ Ejemplo: si el grupo es de 2 personas, cada uno recibe una mayor proporción que si el grupo fuera de 3.

Roadmap

1 Aprendizaje Supervisado y no Supervisado

2 Clustering

3 Aplicación en R

4 Aplicación en R

Clustering: idea en 60 segundos

¿Qué es? Agrupar observaciones **parecidas** entre sí (y distintas de otras) *sin* etiquetas.

Clustering: idea en 60 segundos

¿**Qué es?** Agrupar observaciones **parecidas** entre sí (y distintas de otras) *sin etiquetas*.

¿**Para qué sirve?**

- Segmentar clientes / productos.
- Detectar patrones y simplificar datos.
- Explorar estructura oculta antes de modelar.

Clustering: idea en 60 segundos

¿Qué es? Agrupar observaciones **parecidas** entre sí (y distintas de otras) *sin etiquetas*.

¿Para qué sirve?

- Segmentar clientes / productos.
- Detectar patrones y simplificar datos.
- Explorar estructura oculta antes de modelar.

Tres sabores (intuición rápida)

- **k-means** (centroides): eliges k ; grupos “redondos”; sensible a escala.
- **Jerárquico** (dendrograma): no pides k al inicio; puedes “cortar” el árbol.

Clustering: idea en 60 segundos

¿Qué es? Agrupar observaciones **parecidas** entre sí (y distintas de otras) *sin etiquetas*.

¿Para qué sirve?

- Segmentar clientes / productos.
- Detectar patrones y simplificar datos.
- Explorar estructura oculta antes de modelar.

Tres sabores (intuición rápida)

- **k-means** (centroides): eliges k ; grupos “redondos”; sensible a escala.
- **Jerárquico** (dendrograma): no pides k al inicio; puedes “cortar” el árbol.

Buenas prácticas

- *Estandarizar* variables (escala comparable).
- Elegir k con *codo* / *silueta*; validar la calidad de clústeres.
- **Interpretar** cada clúster en clave de negocio (tú pones el nombre/acción).

Video: ¿Qué es Clustering?

Recurso recomendado

▶ [Ver el video en YouTube](#)

Instrucción: Mira el video completo y piensa en un ejemplo real de tu vida diaria donde podría aplicarse el clustering.

Actividad interactiva en Kahoot

Pon a prueba tu comprensión sobre Clustering

► Ingresar al Kahoot

Instrucciones:

- Ingresa al Kahoot desde tu celular o computador.
- Forma un equipo de **2 a 3 estudiantes**.
- Al registrarte, escribe el **primer nombre** y la **primera letra de tu apellido**, para identificar tu grupo.
- Responde todas las preguntas relacionadas con el video de **Clustering**.
- **Premio:** Los 3 mejores equipos recibirán una **bonificación en la nota del segundo parcial**.
 - La bonificación se divide entre los integrantes del grupo.
 - Ejemplo: si el grupo es de 2 personas, cada uno recibe una mayor proporción que si el grupo fuera de 3.

Roadmap

- 1 Aprendizaje Supervisado y no Supervisado
- 2 Clustering
- 3 Aplicación en R
- 4 Aplicación en R

Aplicación en R

En esta parte del curso trabajaremos directamente con la base de datos `spotify-t` para aplicar el algoritmo de K-means.

¿Qué haremos?

- Generar descriptivas básicas por *cluster*.
- Visualizar relaciones entre variables (ej. streams, popularidad, duración).
- Resumir características musicales (tempo, danceability, energy).
- Interpretar los resultados en clave de **segmentación**.

► [Ir a la Aplicación en R](#)