



**Escuela Superior  
de Ingeniería y Tecnología**  
Universidad de La Laguna

# Tratamiento Inteligente de Datos: Agrupamiento

Álvaro Fontenla León  
([alu0101437989@ull.edu.es](mailto:alu0101437989@ull.edu.es))

C/ Padre Herrera s/n  
38207 La Laguna  
Santa Cruz de Tenerife. España

T: 900 43 25 26

[ull.es](http://ull.es)



# Índice

<b>Problema planteado</b>	<b>2</b>
<b>Modelo de agrupamiento</b>	<b>2</b>
Método del codo	2
Ejecución del algoritmo	3
Visualización de los resultados	3
<b>Definición de grupos</b>	<b>4</b>
Grupo 1	4
Grupo 2	4
Grupo 3	4
Grupo 4	4
Grupo 5	4



## Problema planteado

Nos encontramos ante una base de datos de clientes de un centro comercial con las variables: ID del cliente, género, edad, sueldo anual y puntuación de gasto.

La tarea a realizar será la agrupación de los distintos tipos de clientes que visitan el centro comercial y su posterior evaluación.

## Modelo de agrupamiento

El modelo de agrupamiento utilizado es el k-means. Este método necesita un número de grupos para poder realizar de manera eficiente el agrupamiento, para ello, utilizaremos el método del codo.

### Método del codo

Se ejecuta el algoritmo de agrupamiento con diez iteraciones. Generamos la gráfica.

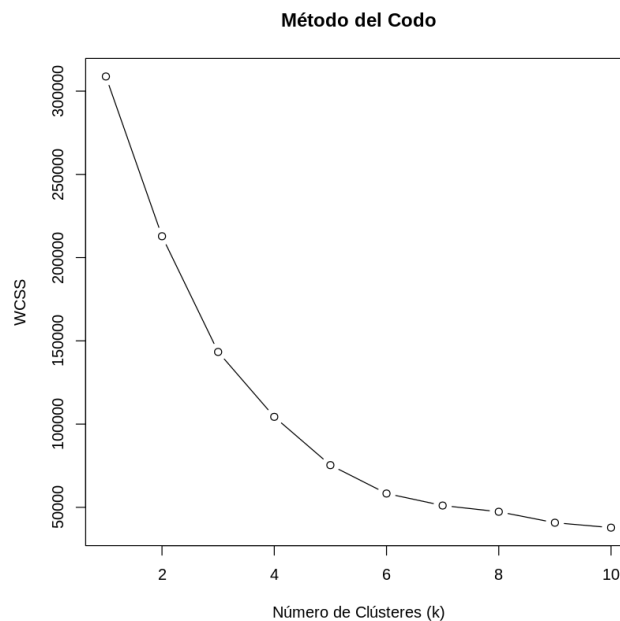


Figura 1: Método del codo.



Con este resultado he determinado que el número óptimo de grupos a utilizar será cinco.

## Ejecución del algoritmo

Una vez que se ha determinado el número óptimo de grupos, se procede a ejecutar el algoritmo sobre el conjunto de datos, utilizando las variables de ingresos anuales y puntuación de gasto.

Antes de ver los resultados, me aseguro de que el modelo es de buena calidad midiendo la cantidad de información retenida, la cual es del 96%. Esto nos indica que el modelo es de buena calidad.

## Visualización de los resultados

Como podemos ver en la figura 2, existen 5 grupos de clientes bien definidos con sus respectivos centroides.

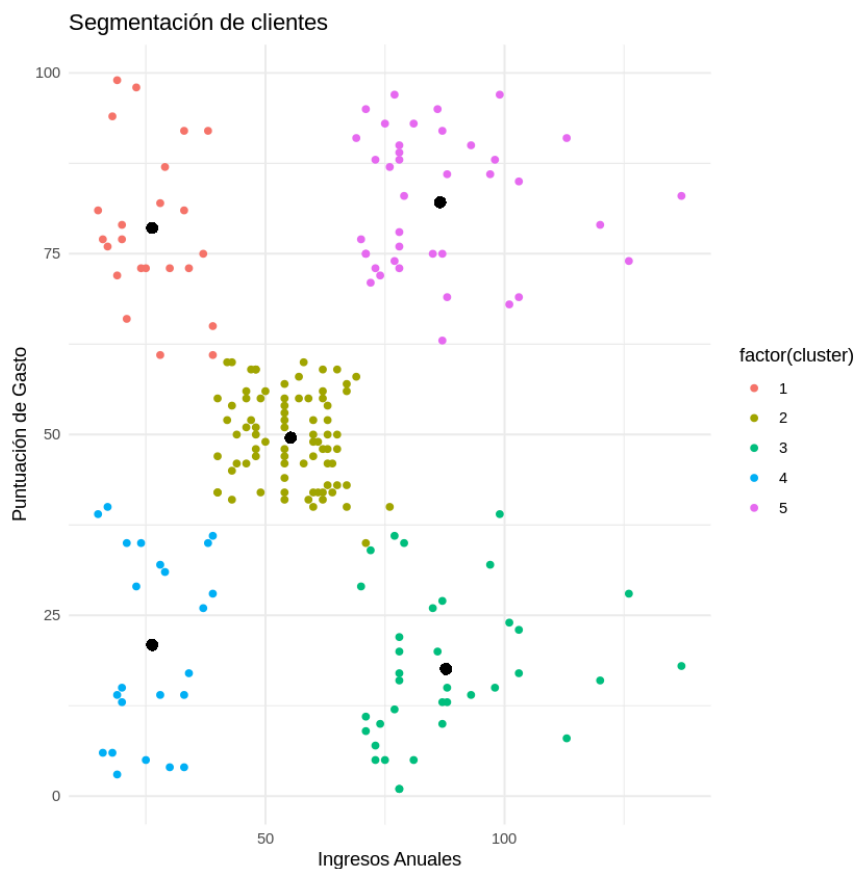


Figura 2: Agrupamiento k-means.



## Definición de grupos

### Grupo 1

Presenta un ingreso anual promedio cercano a los 26.30 y una distribución de puntuación de gasto más dispersa, con valores medios alrededor de 78.57. Esto indica un grupo con ingresos relativamente bajos pero con una puntuación de gasto considerablemente alta.

### Grupo 2

Tiene un ingreso anual promedio de aproximadamente 55.29 y una puntuación de gasto media de alrededor de 49.57, indicando un nivel de gasto moderado en relación con sus ingresos.

### Grupo 3

Se destaca por tener el ingreso anual más alto de todos los grupos, con un promedio de aproximadamente 87.75. Sin embargo, la puntuación de gasto promedio en este grupo es notablemente baja, alrededor de 17.58. Esto podría indicar un segmento de clientes con altos ingresos pero que son más conservadores en su gasto.

### Grupo 4

Similar al grupo 1, presenta un ingreso anual promedio cercano a los 26.30 y una distribución de puntuación de gasto más dispersa, con valores medios alrededor de 20.91. A pesar de tener ingresos similares al grupo 1, la puntuación de gasto en este grupo es considerablemente más baja.

### Grupo 5

Muestra un ingreso anual promedio más alto, alrededor de 86.54, junto con una puntuación de gasto significativamente más alta, con un promedio de aproximadamente 82.13. Esto sugiere un segmento de clientes con altos ingresos y unos hábitos de gastos más elevados.