

# ANEXO clustering

Ana

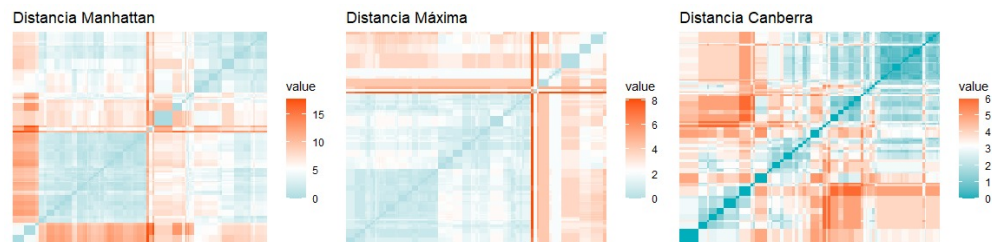
2025-06-06

## ANEXO CLUSTERING

### Clustering de Servicios

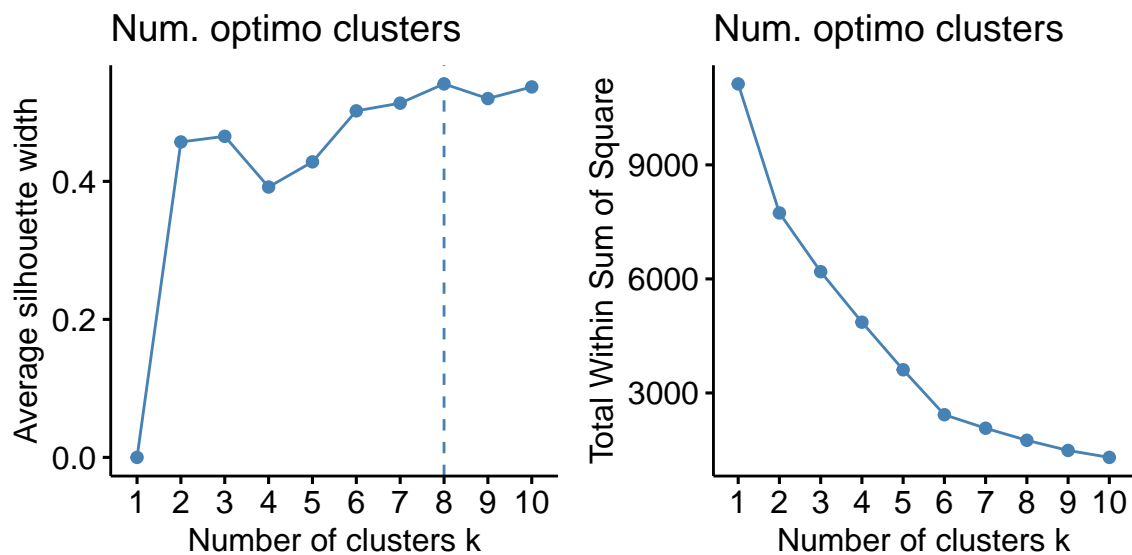
#### *Prueba de distancias*

A continuación, se muestran los resultados obtenidos con distintas métricas de distancias que fueron evaluadas pero finalmente descartadas en favor de la distancia euclídea, por ofrecer menor cohesión entre grupos o menor interpretabilidad.



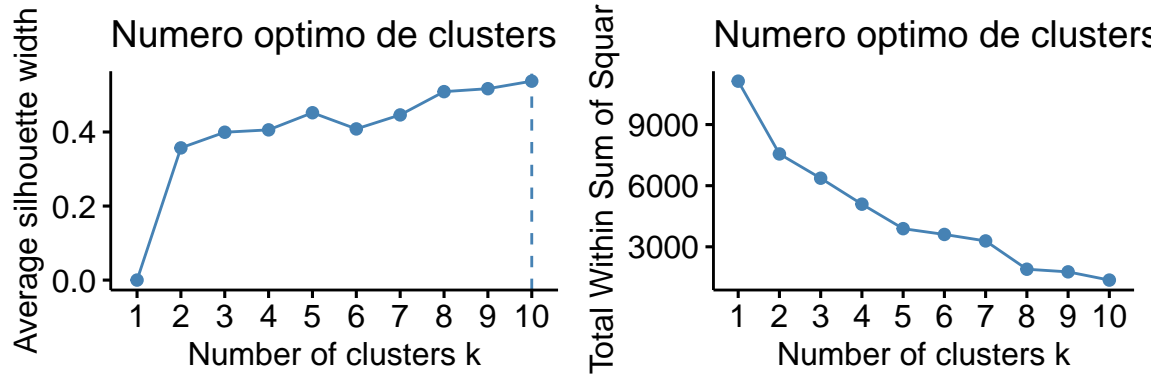
#### *Selección de la técnica de agrupación y del número de clusters*

-Método jerárquico de WARD En este caso, seleccionamos el resultado con **8 clusters**. Se trataba de la organización con mayor coeficiente de Silhouette, además de una suma de cuadrados residual muy reducida.



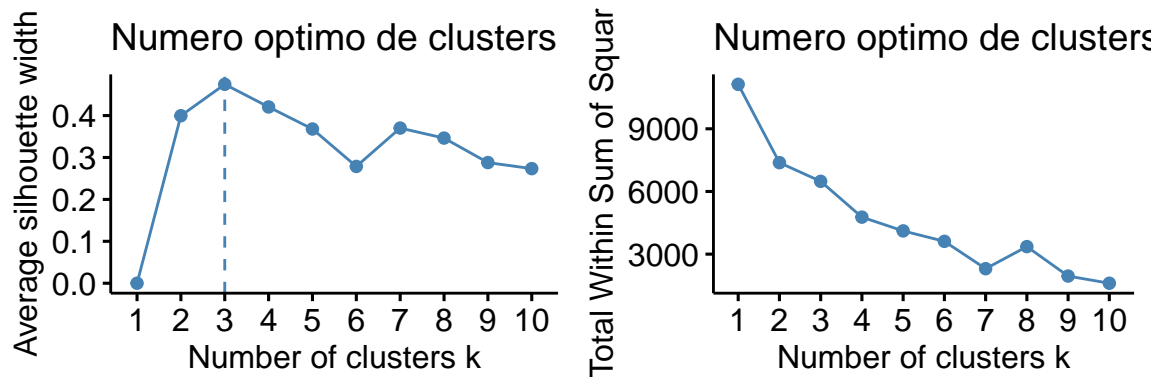
-Método de partición PAM

A través de la técnica PAM (k-medioides), obteníamos el máximo coeficiente y el mínimo SCR con 10 clusters. Sin embargo, la partición en **8 grupos** reducíamos la complejidad del análisis con índices de calidad muy similares.

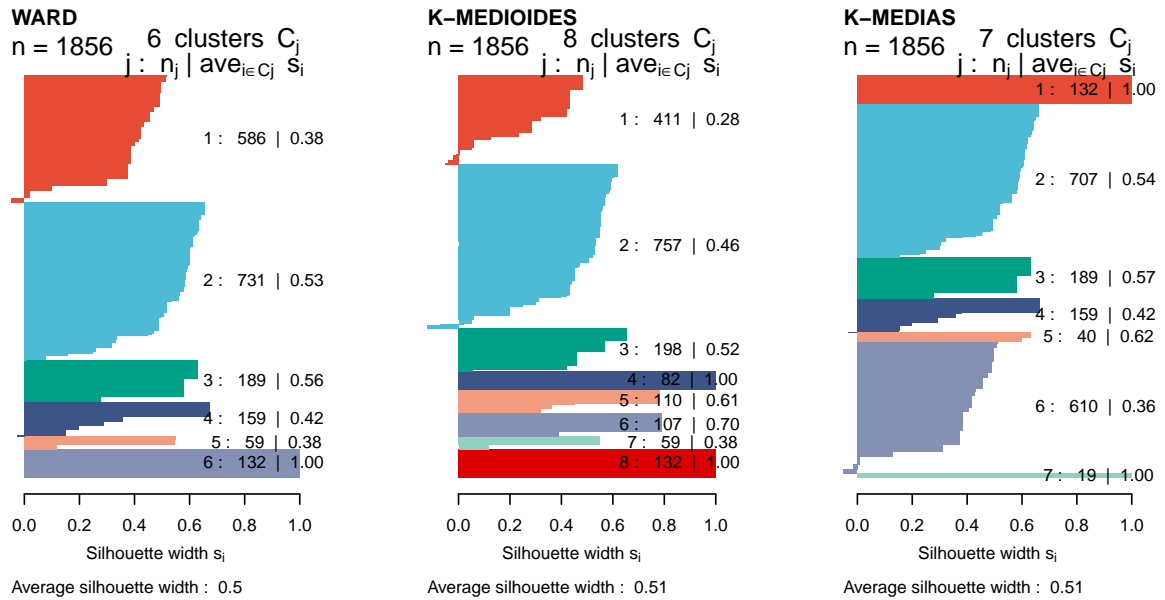


-Método de partición k-medias

Finalmente, con k-medias se maximizaba el coeficiente de Silhouette dividiendo los pisos en 3 grupos. No obstante, la SCR era todavía excesivamente elevada. La opción que mejor combinaba ambos parámetros era **k=7**.



Una vez decidido el número de divisiones en cada método, comparamos los clusters resultantes. Vimos que, aunque los resultados eran similares, las **particiones de k-medias** contenían generalmente menos inmuebles mal clasificados.



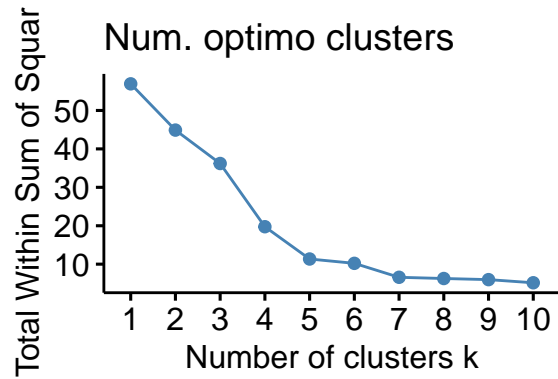
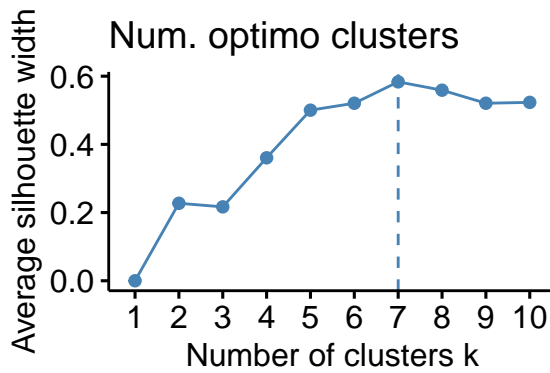
## Clustering de Confort

*Distancias y selección del método de agrupación* Dado que el conjunto que describe las comodidades ofrecidas por cada piso contiene tanto variables categóricas como numéricas, únicamente podíamos consultar las tendencias de agrupamiento con la distancia de Gower.

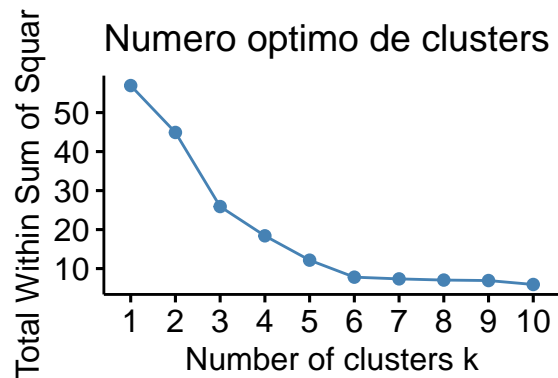
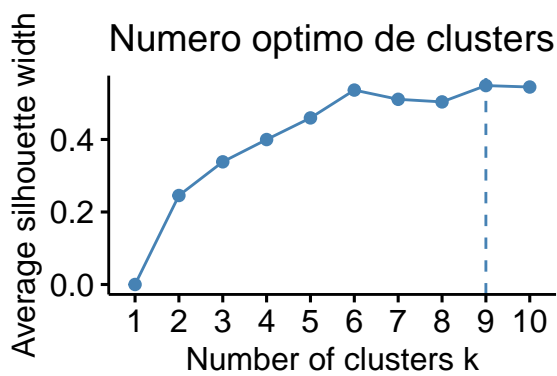


En el mapa de calor encontramos evidencias de la existencia de subgrupos en el conjunto de datos. Con el objetivo de detectar dichos clusters e interpretarlos, comenzamos el análisis probando las técnicas de Ward y k-medioides para determinar cuál generaba resultados más definidos. No aplicamos k-medias porque no habíamos utilizado la distancia euclídea.

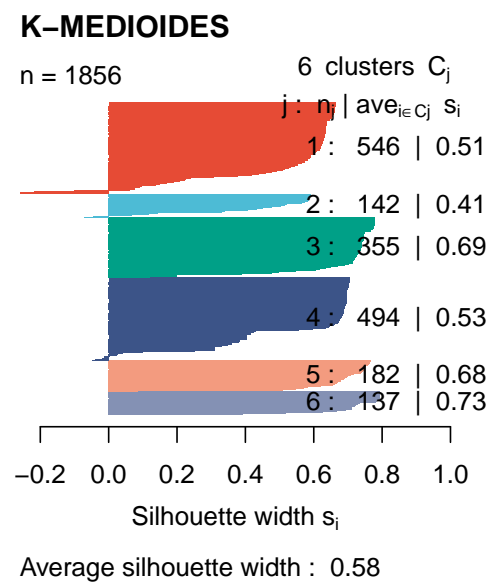
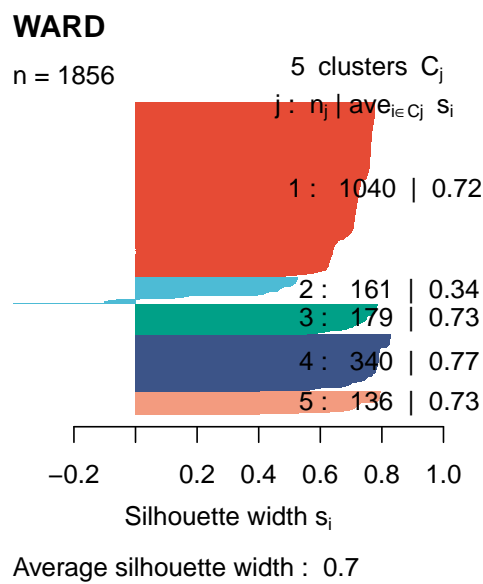
- Método de WARD (k=5)



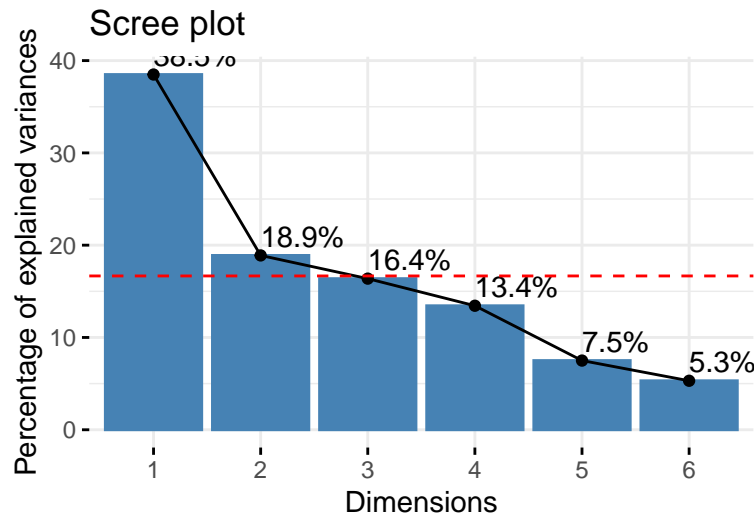
-Método PAM (k=6)



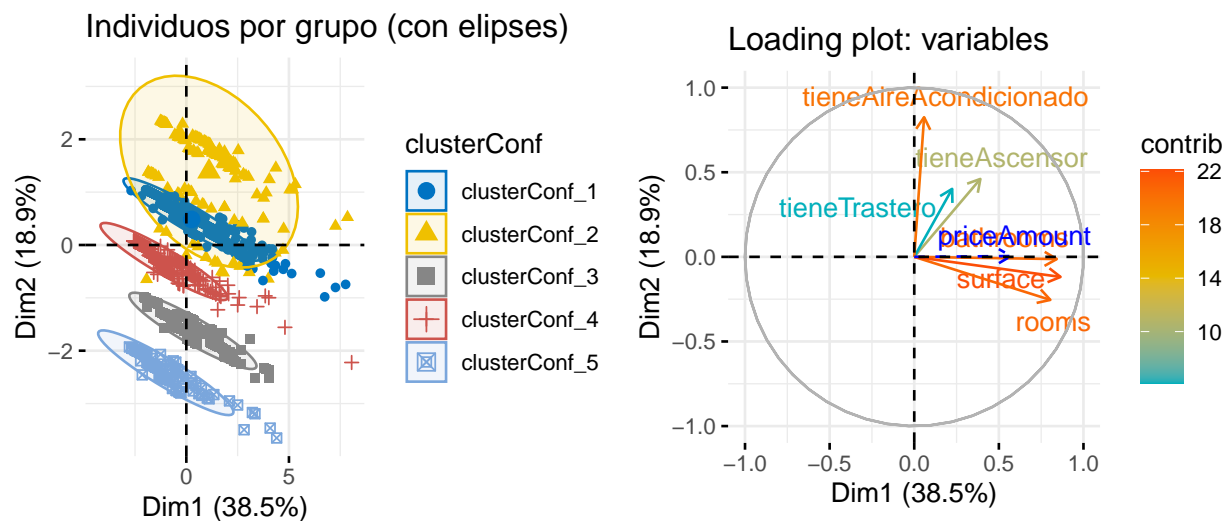
Finalmente, nos decantamos por el método *Ward* porque su coeficiente medio de Silhouette era notablemente superior.



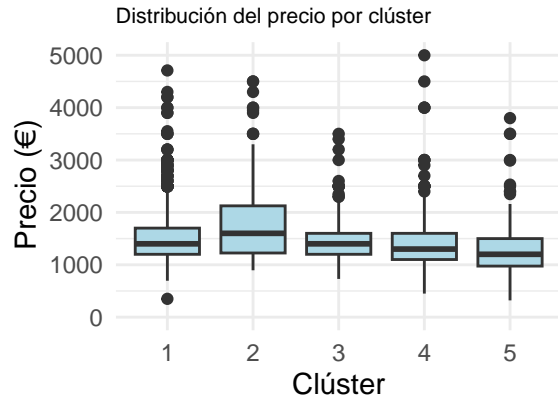
Se realizó un PCA de 2 componentes (casi el 60% de la variabilidad) para comprender mejor las características de cada Cluster.



- Primera componente (*Tamaño útil*): mayor superficie, más habitaciones y baños (tamaño del piso), positivamente relacionado con el precio. Los clusters se reparten de forma relativamente uniforme, aunque el grupo 1 tiene más viviendas con valores altos en esta variable latente.
  - Segunda componente (*Equipamiento interior*): más pisos con aire acondicionado, calefacción, ascensor... Entre otras infraestructuras que aportan valor añadido a la vivienda. En esta dimensión hay una segmentación más clara de los clusters. El cluster 2 destaca positivamente, seguido del 1. Los clusters 4 y 5 son los que peores condiciones ofrecen.



Clúster	Precio medio (€)
1	1577.748
2	1852.099
3	1476.313
4	1454.141
5	1346.397



Si comparamos el precio medio de cada cluster, vemos el orden es casi idéntico a la estructura que siguen a lo largo de la segunda componente; únicamente se intercambian los clusters 3 y 4, que a su vez son los que tienen un promedio más parecido. Por lo tanto, podríamos considerar el cluster como una variable ordinal que nos indica el la categoría de cada vivienda (muy relacionada con su precio). Esta división no coincide estrictamente con el tamaño de la vivienda, sino que está más basada en sus condiciones de habitabilidad. Los niveles serían los siguientes:

Table 1: Correspondencia entre nivel de calidad y clústers

Nivel	Calidad	Clusters_originales
1	Baja	5
2	Media	3-4
3	Alta	1
4	Lujo	2

A continuación, se muestran las frecuencias de pisos en cada una de las nuevas categorías.

