



UNIVERSIDAD  
DE GRANADA

# PRÁCTICA 2

ANÁLISIS RELACIONAL MEDIANTE SEGMENTACIÓN

MAURICIO LUQUE JIMÉNEZ

MAULUJIM@CORREO.UGR.ES

SUBGRUPO DE PRÁCTICAS AI

18 DE DICIEMBRE DE 2024

INTELIGENCIA DE NEGOCIO

**Sumario**

Introducción.....3

Primer Caso de Estudio: Agrupamiento por zonas geográficas.....4

Segundo Caso de Estudio: Agrupamiento por Ofertas.....7

## INTRODUCCIÓN

En esta práctica, se van a abordar tres casos de estudio principales. Cada uno orientado a un sector y a una manera distinta de enfocar el análisis sobre los alojamientos que puede haber en una ciudad a lo largo del año. Podría decirse que estos casos de estudio van a responder a tres preguntas diferentes: dónde, cuánto y cuándo. El primero de ellos va a comparar los servicios de los distintos alojamientos según su ubicación geográfica: se van a dividir los alojamientos por sus respectivas zonas y se va a intentar buscar semejanzas entre los alojamientos de una misma zona, así como encontrar las principales diferencias con los alojamientos de diferentes zonas. El segundo caso de estudio va a estar relacionado con la influencia (o no) de ofertas o de promociones especiales a la hora de anunciar los alojamientos: qué características comparten aquellos sitios que recurren a ofertas. Por último, el tercer caso de estudio va a comparar el tráfico en alojamientos en función de si estamos en temporada “alta” o temporada “baja”.

## PRIMER CASO DE ESTUDIO: AGRUPAMIENTO POR ZONAS GEOGRÁFICAS

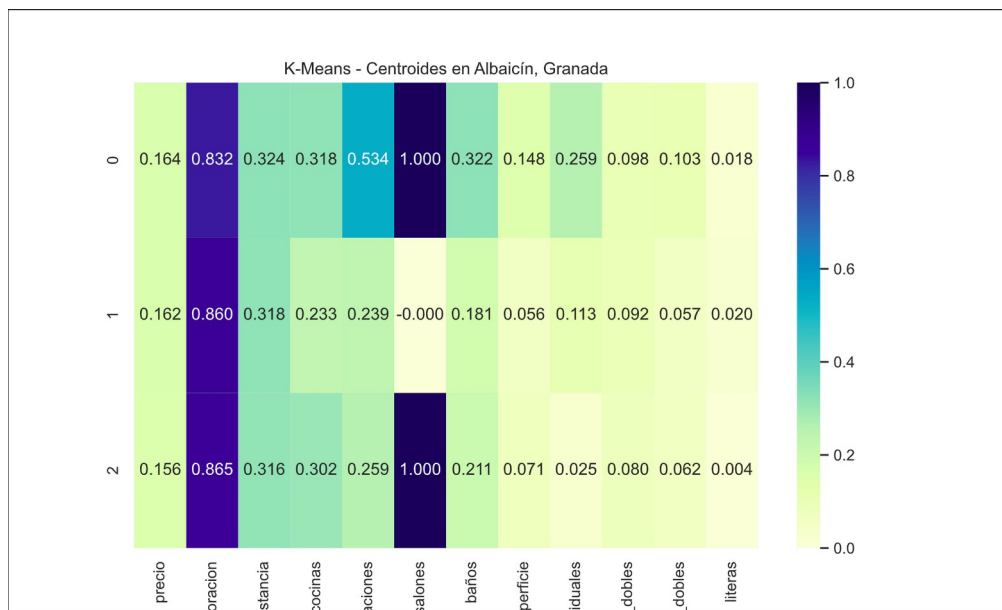
Para este estudio, se ha realizado una primera segmentación en función de la zona de cada alojamiento. Para ello, el conjunto de datos seleccionado ha sido aquel que agrupa los resultados de las búsquedas por alojamiento. Después de esto, se han aplicado cinco algoritmos de clustering para buscar semejanzas internas y diferencias externas entre elementos de otras zonas. Por ejemplo, una hipótesis inicial puede ser que en zonas más céntricas de Granada, donde el turismo es constante y los alojamientos son más sitios para dormir que casas en las que estar, las cocinas no son tan importantes, mientras que en zonas más alejadas del centro, qué sí pueden servir para hacer algo más de vida, sí requieran ese servicio. Del mismo modo, se podría pensar que, si bien los alojamientos céntricos tienen menos superficie, pueden llegar a ser más caros (algo totalmente lógico).

Para este caso de estudio, vamos a tener en cuenta las siguientes variables:

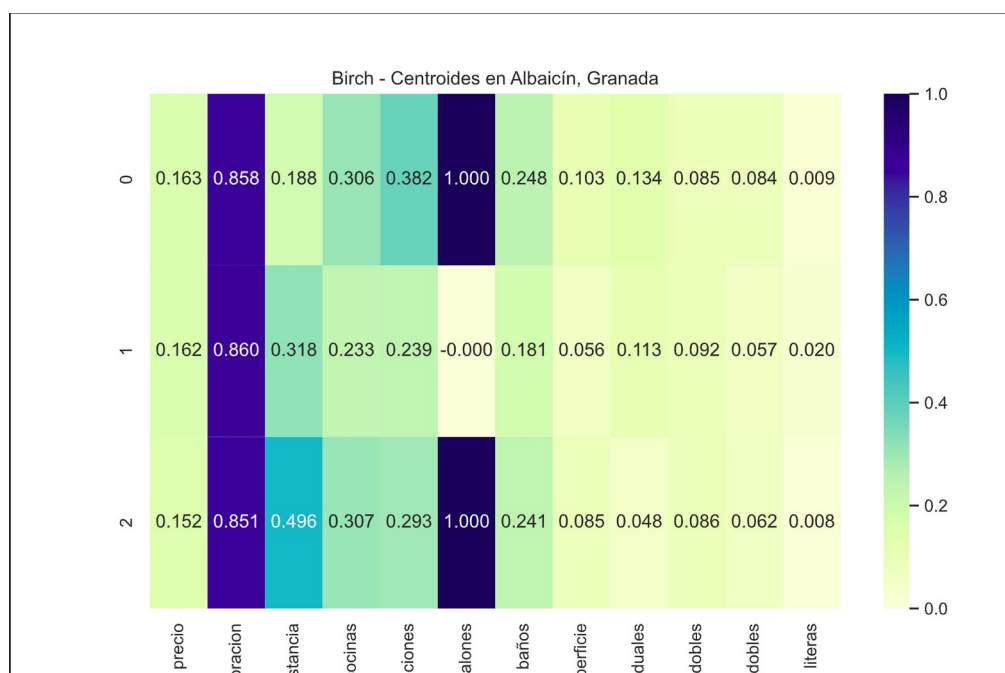
- Precio
- Distancia
- Valoración
- Localización
- Cocinas
- Habitaciones
- Salones
- Baños
- Superficie
- Camas individuales
- Camas dobles
- Camas grandes dobles
- Literas

Se han elegidos estos atributos puesto que se puede decir que son los más representativos a la hora de analizar qué puede buscar un cliente en un alojamiento. Una vez hemos seleccionado las variables a considerar (ligero preprocesamiento mediante para rellenar valores perdidos y renombrar las columnas seleccionadas) y se han separado los alojamientos por zonas, se han aplicado cinco algoritmos de clustering para cada subconjunto de alojamientos, para ver qué características se ven dentro de cada conjunto.

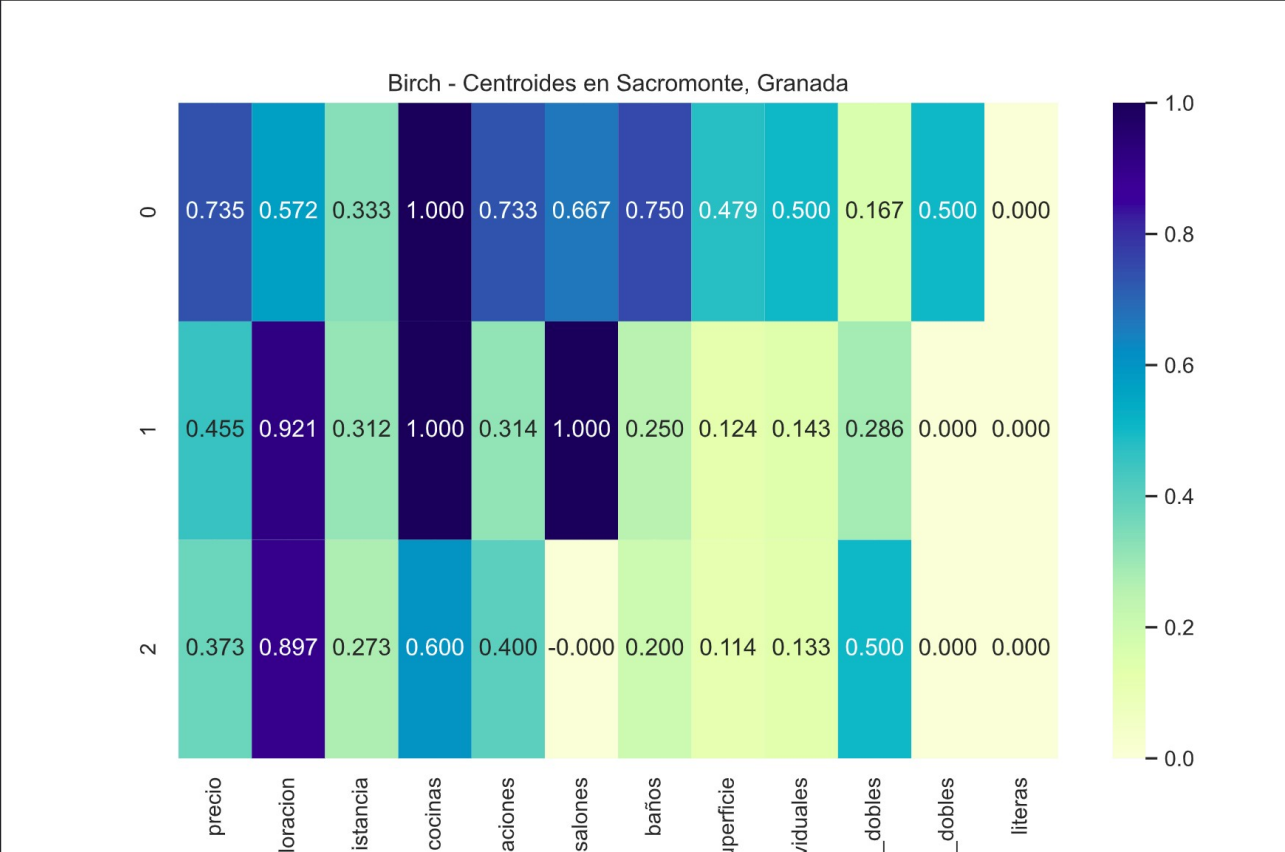
Centrándonos en un ejemplo de los muchos que surgen, vemos que, si aplicamos el algoritmo K-Means a los alojamientos del Albaicín con la orden de formar tres agrupaciones distintas, vemos que, en general, todos los alojamientos tienen un precio (bajo) y una valoración (alta) muy similar, vemos que hay diferencias notables en el número de habitaciones o en el de salones, si bien ambos valores están en la mitad baja del rango normalizado. También se ve una ligera diferencia, por ejemplo, en el número de camas individuales, si bien no se puede decir que destaque por ser un número alto.



Si en vez de K-Means aplicamos, por ejemplo, Birch, vemos como hay ligeras diferencias, como en la variable distancia, que no es tan homogénea como en la imagen anterior, sino que encuentra agrupaciones más diferentes, pero que en cualquier caso siguen quedándose en la mitad de menor valor del rango normalizado. A su vez, encuentra una diferencia ligeramente menor en cuanto al número de habitaciones se refiere.



Poniendo ahora como ejemplo la zona del Sacromonte, vemos que la diferencia de servicios es notable. Los precios son considerablemente más caros que en el Albaicín, y, confirmando la hipótesis inicial, servicios como las cocinas son totalmente diferentes. Hay más variedad en la oferta de salones, por ejemplo, y también hay más casos de camas individuales o dobles. Así que sí, se puede deducir que en función de la zona de cada alojamiento, la oferta de servicios es distinta.



## SEGUNDO CASO DE ESTUDIO: AGRUPAMIENTO POR OFERTAS

Para este segundo caso de estudio, la pregunta inicial es: ¿qué características comparten los alojamientos que presentan ofertas? ¿Hay algún patrón que se cumple, o es pura casualidad? Para ello, se ha hecho una primera división entre aquellos alojamientos que no incluyen oferta alguna y aquellos que sí. En segundo lugar, se han seleccionado tres tipos de oferta: “Oferta a finales de año”, “Oferta por tiempo limitado” y “Oferta de Navidad”. No se han tenido en cuenta aquellas ofertas que no estén directamente relacionadas con el alojamiento, sino que incluyan un detalle extra como el taxi desde el aeropuerto.

Viendo un posible ejemplo de aplicación del algoritmo de clustering Birch, notamos que, más allá de la obviedad de que los precios bajos iban a ser una constante, vemos que en general el valor de distancia es muy pequeño. Eso podría llevarnos a pensar que en general los alojamientos que presentan ofertas son aquellos que realmente necesitan imponerse a la competencia, algo que lógicamente es más común en el centro turístico de la ciudad que en las afueras. Además, se puede observar una muy clara diferenciación entre distintos tipos de alojamientos en función de sus menciones especiales (“preferente”, “plus”, o “destacado”, entre otros), lo cual también podría estar relacionado con el hecho de querer añadir una distinción entre tanta competencia.

