

The logo consists of a solid orange square. Inside the square, the word "CUNEF" is written in large, bold, white, sans-serif capital letters.

CUNEF

COLEGIO UNIVERSITARIO DE
ESTUDIOS FINANCIEROS

Técnicas de Agrupación y Reducción de la Dimensión

Práctica: Los coches del jefe 3

V. Miguel Sempere Navarro

En la presente práctica se va a resolver un problema que el jefe de la empresa nos ha encargado. El jefe es un amante de los coches clásicos y se ha comprado una colección con 125 coches. El problema que le surge es que no sabe cómo dividirlos entre sus propiedades, sin embargo, sabemos que cuenta con al menos 10 de ellas con espacio suficiente como para almacenarlos según la agrupación que se considere conveniente.

Para la resolución de este problema no hemos sido contratados. No obstante, ser la persona de confianza del jefe tiene sus pros y contras, por lo que haremos diferentes agrupaciones. A priori, se podría dar por válida una primera clasificación por tamaño, potencia, tipo, etcétera. Sin embargo, tenemos la suerte de contar con una base de datos de la cual se puede extraer mucha información.

Esta base de datos contiene muchas variables, entre ellas: PVP, cilindrada, potencia, RPM, peso, plazas, etcétera. Nada más descargarla, se ha optado por escalar algunas variables y eliminar las que no nos interesaban. De entre las tareas realizadas, se han reemplazado los consumos de Nissan y Ssanyong y se han rellenado otros datos que faltaban con valores orientativos o similares. Con esto se ha conseguido tener unos datos preparados para trabajar con ellos y minimizar así las probabilidades de realizar una agrupación o clasificación errónea.

Una vez ya se tienen los datos para trabajar, hay que intentar encontrar características que permitan realizar agrupaciones de coches. Una primera aproximación ha sido realizar un dendrograma gracias al cual se han podido observar distintos grupos de vehículos según su distancia euclídea. Haciendo una lectura rápida, ya se observa como los Mercedes crean un grupo y los Suzuki y Tata otro (seguro que es por PVP). Sin embargo, el reto que esto plantea es averiguar el número óptimo de “grupos” o “clusters” que permitan clasificar a todos los coches en unos u otros según algunas características específicas.

Por tanto, nos interesa saber cuál es el número óptimo de clusters; es decir, en cuántas propiedades del jefe dejaremos los coches clásicos que se ha comprado y cuales. Se ha estudiado y, como se observa en la siguiente ilustración, el número óptimo es 3:

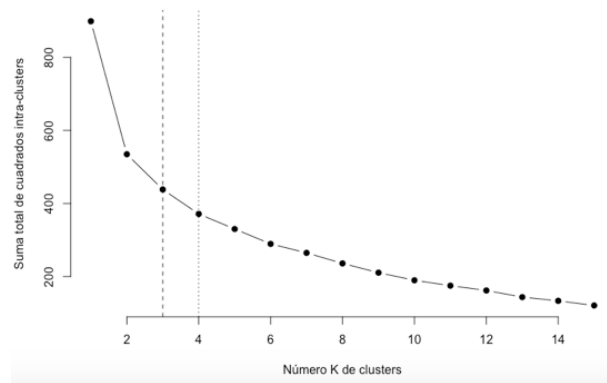


Ilustración 1. Número óptimo de clusters. Fuente: elaboración propia.

Sin embargo, gracias a nuestra dilatada experiencia en el sector, y dado el siguiente mapa con las localizaciones del jefe, vamos a pensar mejor el número de clusters:



Ilustración 2. Garajes del jefe. Fuente: Google.

Como se observa, el jefe tiene varias localizaciones posibles y, si bien por características, el número óptimo de clusters es 3, nosotros vamos a repartirlos de una forma más laxa a priori: en 4 o 5. Esto se debe a que, aunque estadísticamente sea peor, nos compensa tener al jefe contento y que en cuantas más propiedades tenga sus caprichos mejor, por lo que intentaremos realizar más agrupaciones de coches para adecuar los clusters a las diferentes localizaciones.

Pero primero debemos comprobar que es posible o al menos no es mucho más perjudicial cuantitativamente hablando:

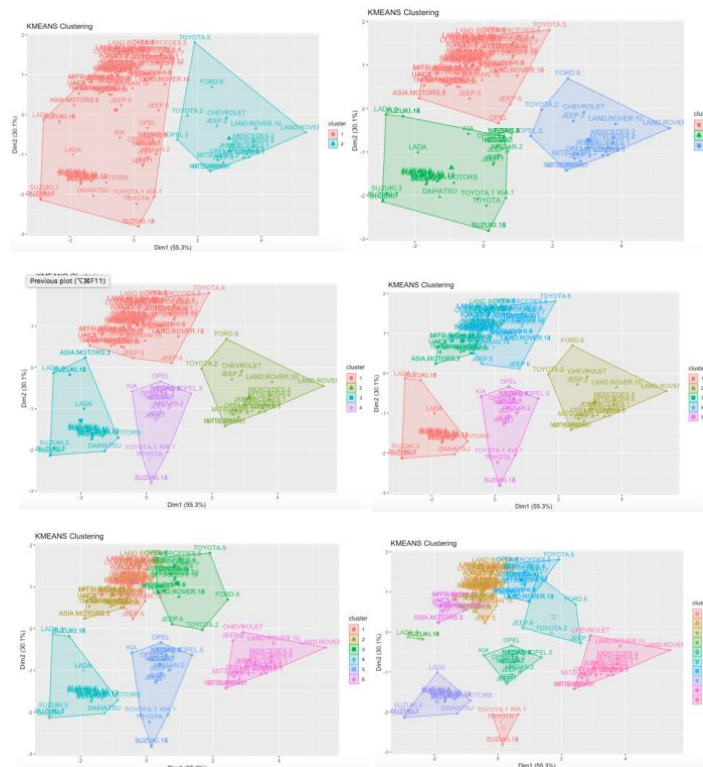


Ilustración 3. Organización según clusters. Fuente: elaboración propia.

Ampliando las imágenes, se pueden observar distintas organizaciones por número de clusters. Como se observa fácil, para 2, 3 y 4 clusters los grupos están claramente diferenciados. Sin embargo, a partir de 5 ya hay solapamientos. Así que, teniendo en cuenta la respuesta del número óptimo de clusters (3) y con buena visión de negocio, damos por válida una distribución en 4 grupos de coches. Teniendo en cuenta que hay varias localizaciones y un máximo de 15 coches por localización, vamos a realizar una agrupación de localizaciones por distancia:



Ilustración 4. Agrupación de localizaciones. Fuente: elaboración propia.

Las localizaciones son las siguientes, y en ellas, los diferentes clusters de coches, que se conforman por **A – Los más pesados; B – Los más caros, C – Los más ligeros, D – Los más pequeños:**

1. **Andorra y Le Rochelle** → A estos destinos irán los 30 coches más pesados (cluster A). Con esto conseguiremos reducir al máximo el gasto en gasolina, pues como se ha analizado, el peso tiene mucha relación con el consumo y queremos evitar a toda costa un gasto excesivo en ello. Irán 15 y 15 a cada uno.
2. **Cannes, Niza, San Remo y Suiza (Lausana y Saint-Ursanne)**→ A estos destinos se irán los coches con un PVP más alto (cluster B), pues un destino como la Costa Azul tiene un perfil de turista de poder adquisitivo muy alto y, del mismo modo, Suiza es uno de los países más ricos del mundo, por lo que nos interesa que los 60 coches más caros se encuentren repartidos por igual en esos 5 garajes.
3. **Córcega** → A este destino mandaremos los 15 coches más ligeros (cluster C). Se justifica con el transporte marítimo, pues la compañía de transporte nos cobra más por peso y volumen, primando el peso de la mercancía.
4. **París** → A la capital nos interesa enviar los 20 coches más pequeños (cluster D), pues con modelos biplaza será más fácil la movilidad por dentro en caso de ser necesario. Se repartirán entre los dos garajes de París con 10 unidades en cada uno.

En la presente práctica se ha abordado el reto de clasificar una serie de coches según sus características y, con la información de las diferentes localizaciones, se ha realizado una serie de agrupaciones tanto de los coches como de las localizaciones para hacer el mejor reparto posible. Y se ha conseguido, pues cuando el jefe quiera hacer turismo de lujo en Suiza y la Costa Azul tendrá coches de lujo; cuando tenga necesidad de buena movilidad en París, tendrá coches pequeños y, el resto de coches que tiene que distribuir, será muy barato el coste de transporte.