

INFORME FINAL LOS COCHES DEL JEFE

Resumen ejecutivo

Tras realizar el análisis exploratorio previo y un primer intento de reparto de los coches en las distintas propiedades disponibles, en este informe se mejora esta clasificación y se explican las características de los tres grupos de vehículos resultantes y cómo han sido repartidos en los distintos garajes. El análisis cluster revela que existen tres grupos bien definidos de coches y a partir de ahí trabaja para repartirlos con un criterio de similitud (coches parecidos en garajes cercanos) empleando como criterio adicional algunas de las características distintivas de las localizaciones geográficas donde se encuentran las propiedades.

Previo a explicar el reparto y agrupación de vehículos es necesario señalar que con objeto de rellenar los registros vacíos en la base de datos se han aplicado distintas técnicas: algunos datos se han sustituido por aquellos de vehículos análogos, otros se han calculado como media de vehículos similares (aunque quizá poco riguroso esto al menos reduce el posible ruido de elegir técnicas de sustitución más arriesgadas) y por último algunos datos son introducidos a mano tras una pequeña investigación en línea.

Respecto a la selección de variables clasificadoras por una parte algunas se descartan por evidencias del sector (como los tipos de aceleración) y la correlación entre otras hace que, por ejemplo, el consumo a 90 sea prescindible pues es fácilmente explicado por el consumo urbano y el consumo a 120 km/h. Además se decide prescindir de la marca como variable para clusterizar aunque luego resulta de gran interés para estudiar los clusters formados.

Una vez tomadas estas decisiones previas más desarrolladas y explicadas en los informes precedentes¹ se procede al reparto en los distintos clusters. Para ello se garantiza que todas las variables sean numéricas y no presenten huecos tras lo cual se puede proceder a un análisis cluster jerarquizado (se elige este modelo porque no solo nos dividirá los clusters en grupos, si no que además aporta información respecto a la cercanía dentro de los propios grupos, algo de gran utilidad a la hora de proceder posteriormente al reparto). Es importante tipificar en este caso las variables debido a la enorme diferencia entre los valores medidos y las unidades empleadas para medirlos.

La decisión clave en este proceso es elegir el número de clusters en los que se dividirán las observaciones. Para ello en este estudio se han utilizado distintas técnicas enfocadas todas ellas en la determinación de dicho número. Observando distintos índices experimentales se aprecia que el número de clusters óptimos podría ser dos o tres. Lo cierto es que existe evidencia suficiente para la calidad de cualquiera de las dos elecciones por lo que se decide trabajar con tres clusters pues parece que esto aporta más especificidad y debido a que el reparto será en un número amplio de localizaciones esto parece una mejor opción (si el objetivo fuera, por ejemplo, minimizar los lugares donde almacenarlos sería más interesante elegir dos).

Una vez elegido el número de grupos se procede a la creación del cluster jerarquizado con el que se trabajará. Se actualiza además la base de datos añadiendo una nueva columna llamada "Grupo" que indica a qué grupo pertenece cada vehículo y otra llamada "Garaje" que indica el garaje asignado.

En las siguientes páginas se procede al estudio de dichos grupos viendo sus características principales y cómo se puede proceder a un reparto adecuado.

¹ Ver [Los coches del Jefe I](#) y [Los coches del Jefe II](#) para detalles.

En primer lugar se presenta el dendrograma (**Figura 1**) del reparto jerarquizado con tres clusters:

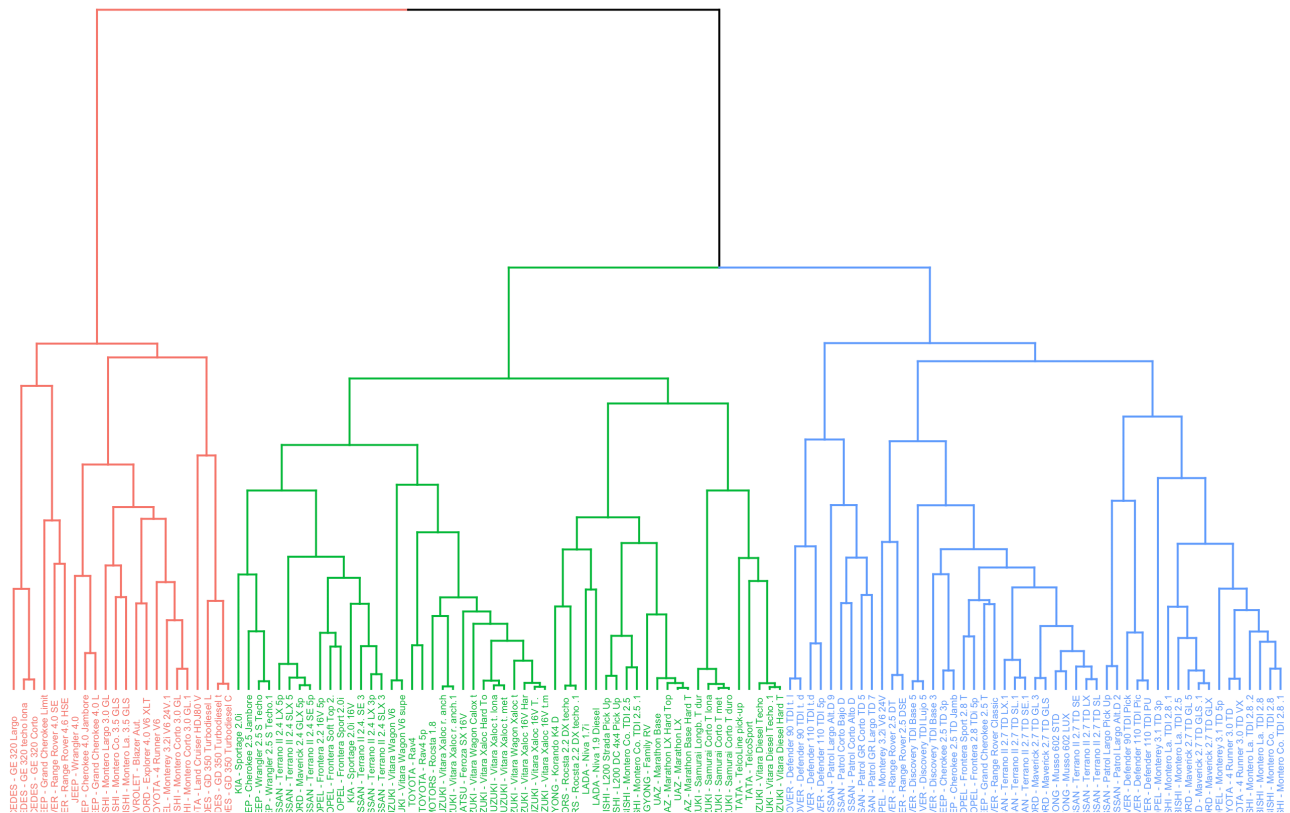


Figura 1. Reparto jerarquizado de los coches en los tres grupos.

A continuación se procede a explicar los tres grupos:

Grupo 1

Este grupo (representado en rojo en la gráfica) está formado por los coches de mayor lujo (y por tanto mayor precio). Abundan en este grupo compuesto por veintidós vehículos los modelos Mercedes todos los modelos de esta marca se encuentran en este cluster que contiene a su vez los Land Rover más lujosos. Los vehículos presentes tienen una potencia y velocidad (dos variables íntimamente ligadas con el precio) muy superior a la de los de los otros grupos. La potencia mínima en este grupo es bastante superior a la potencia media de los demás clusters. El número de cilindros es el mayor en la base de datos. Están en este grupo todos los vehículos de ocho cilindros siendo el resto de seis.

En temas de consumo son coches muy contaminantes con altos consumos (este dato resultará útil a la hora de distribuir los coches), siendo especialmente alto el consumo urbano (algo por otra parte habitual en coches todoterreno).

Grupo 2

Son los vehículos de menor calidad de la base de datos con la que se trabaja (representados en verde en la gráfica). El grupo está compuesto por un total de 54 coches. Son modelos baratos, todos ellos prácticamente de cuatro cilindros. Son con mucha diferencia los de menor potencia, cilindrada y peso lo cual repercute directamente en su precio. Su consumo es ligeramente inferior al de los vehículos del grupo 3 y drásticamente menor que el de los del grupo 1. En estos coches se alcanza el mayor número de revoluciones por minuto.

Grupo 3

En este grupo (representado en azul en la gráfica) está formado por coches de alta calidad pero que no llegan a los altos estándares del grupo 1. En él se encuentran los 49 autos restantes. Son coches que logran prácticamente las mismas revoluciones por minuto que los del grupo 1 pero alcanzando una velocidad mucho menor. Su potencia es un punto medio entre los modelos del grupo 1 y los del grupo 2.

Tras este pequeño análisis es el momento de proceder al reparto de los vehículos entre las distintas propiedades. Para ello se intentan considerar las propiedades especiales de las distintas localizaciones, prestando atención a aspectos como las retenciones de tráfico por exceso de contaminación de París o la adecuación de las características del todoterreno al medio destinando modelos más potentes a zonas montañosas y los vehículos menos potentes a zonas llanas y costeras.

El grupo 1 se puede almacenar empleando dos garajes (e incluso sobrarían ocho plazas). Para dicho grupo se eligen los dos garajes presentes en Suiza. Esta decisión se fundamenta en dos motivos, los coches con grandes consumos pagarán menos impuestos por sus emisiones y del mismo modo tiene sentido almacenar los coches más lujosos en Suiza. Además las dos propiedades en Suiza se encuentran fuera de núcleos urbanos y próximas a los Alpes por lo que resulta interesante tener vehículos de mucha potencia para circular por estas carreteras. Así los Mercedes y Mitsubishi de este grupo se almacenan en el garaje próximo a Lausana y los doce vehículos restantes en la propiedad más próxima a Basilea.

El grupo 2 está compuesto por 54 vehículos por lo que se requerirán cuatro garajes para almacenarlos. Como se comentó anteriormente estos vehículos se caracterizan por ser los de menor consumo por lo que es claro que una parte de ellos debe almacenarse en París. El criterio de reparto en este caso será enviar los todoterrenos de menor consumo urbano de este grupo a París. Así en uno de los dos garajes se almacenarán los Suzuki (exceptuando los Vitara Wagon que presentan un mayor consumo). En el otro se almacenan el resto de vehículos de menor consumo urbano de este grupo. Así quedan 24 coches por repartir, siguiendo el criterio del consumo los quince siguientes se colocarán en San Remo pues es el único garaje restante situado en un núcleo urbano. Finalmente los nueve coches restantes se almacenan en Córcega pues son vehículos de poca potencia y esta es la zona con un relieve menos pronunciado.

En el grupo 3 se encuentran los 49 vehículos restantes por lo que en un principio serían necesarios cuatro garajes aunque en el garaje de Córcega y en el de Suiza se podrían almacenar hasta seis y ocho vehículos respectivamente por lo que podría plantearse dejar un garaje libre. Esta última idea se descarta en aras de una clasificación más precisa pero puede plantearse en caso de que el jefe desee desprenderse de alguna propiedad próximamente.

Así, atendiendo a criterios geográficos se asignan al garaje andorrano los vehículos de mayor potencia en este cluster por encontrarse en el Pirineo. En el garaje más próximo a Cannes se almacenan los vehículos con mayor potencia restantes que además resultan ser los Land Rover y Opel de este grupo. Los 15 siguientes se almacenan en Niza (razonablemente cerca de estos últimos porque realmente no existen grandes diferencias con los anteriores).

Los seis restantes son enviados a La Rochelle (aunque parece consiste enviarlos a Córcega pues son próximos en potencia a los allí almacenados).

Conclusión

En este estudio se ha realizado el reparto de los 125 vehículos en torno a las propiedades propuestas. Es importante destacar que se ha intentado mantener un equilibrio a la hora repartir los coches entre un criterio de utilidad (coches con poco consumo urbano en ciudades o coches con mucha potencia en los Alpes) con criterios de similitud propios de un coleccionista (dentro de las localizaciones los coches se agrupan en los distintos garajes por similitud en marca y modelo y por proximidad en el dendograma). A partir de estos criterios se han empleado los diez garajes disponibles para almacenar los coches que en un principio se dividieron en tres grupos principales. De no sentirse usted conforme con el reparto propuesto podría plantearse un reparto que en lugar de situar coches similares próximos entre ellos los coches aparecieran repartidos de manera uniforme de forma que en cada propiedad se encontraran vehículos muy distintos entre ellos abarcando en mayor o menor medida todo el rango de características posibles, es decir, desde modelos más sencillos a modelos de lujo. Sin embargo este planteamiento me pareció personalmente poco eficiente. Si usted desea solo mantener los coches almacenados como una inversión, una colección de activos, parece más inteligente agruparlos por similitud. Si por el contrario desea usted disfrutar de su amplia colección este reparto permite disponer de los coches mejor adaptados a las circunstancias en las distintas localizaciones. La base de datos presente en el código adjunto se actualiza con los registros vacíos ya completados y añadiendo dos columnas, una con el grupo al que pertenece cada vehículo y otra con el nombre de la localización al que ha sido asignada. Además como se señalaba anteriormente, si usted deseara deshacerse de alguna de las propiedades esta clasificación permite una redistribución sencilla de los vehículos.

Código

En el código adjunto que se puede consultar en el siguiente [enlace](#) se puede observar tanto el trabajo previo con los NA's y el análisis exploratorio cómo los métodos empleados para elegir las variables a emplear en la clasificación y las distintas propuestas de cluster con un mayor desarrollo en el análisis cluster jerarquizado sobre el que se redacta este informe.