

Coches del Jefe - Parte 3

Hugo César Octavio del Sueldo

12/5/2020

Problema a solucionar

Finalmente, después de haber solucionado los problemas de selección de variables y tratamiento de valores perdidos, debe proceder a asignar los coches a las viviendas de su jefe. En un máximo de cuatro páginas, indique de qué forma va a proceder y cuáles son las características tanto de los grupos que ha creado como de los vehículos que asigna a cada vivienda.

Introducción

Esta es la tercera parte de la practica relacionada a los coches del jefe. En la primera parte de esta serie de entregas realizamos un analisis explotario de los datos con los que contabamos, analizando aquellas variables que nos resultaban interesantes para la agrupacion y posterior reparto de los coches del jefe en las localidades o residencias con las que el contaba. En la segunda parte, realizamos una propuesta de reparto de la coleccion de autos al jefe. Para esto, estudiamos el número adecuado de grupos en los que dividir la coleccion. Teniendo en cuenta de que el maximo numero de coches por residencia es de 15 y sabiendo de que en el caso de que hayamos propuesto grupos con mas coches, teníamos que escoger las residencias en las que guardarlos, atendiendo al criterio de distancia. En el siguiente **enlace** tiene un mapa de las residencias actuales. Ahora, como habrán observado arriba, debemos proceder a la asignación de los coches a la viviendas del jefe indicando de que forma procedemos finalmente y cuáles son las características tanto de los grupos que hemos creado como de los vehiculos que asignaremos a cada vivienda.

Fase 1

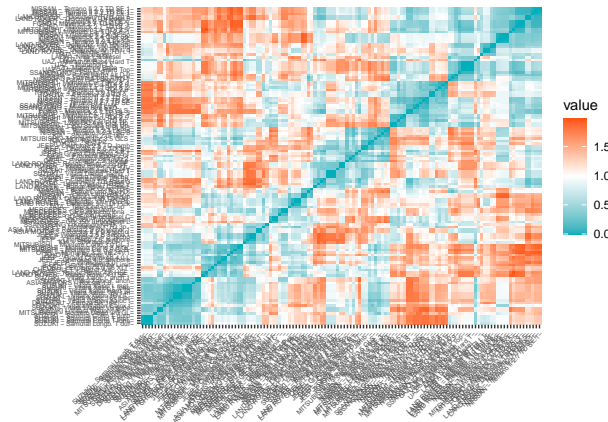
Una vez explicado el objetivo del trabajo, vamos a explicar cómo fue el tratamiento de los valores faltantes dentro del fichero utilizado. Aqui se utilizaron distintas tecnicas como reemplazar los pesos por pesos de coches equivalentes y para algunas marcas, como pueden ser Nissan y Ssanyong, se sustituyeron los consumos por los consumos medios de las marcas.

Respecto de las variables a utilizar para hacer la agrupación y reparto de los vehiculos del jefe, se descartaron las variables revoluciones por minuto, acelerac y acel2 ya que son prescindibles y no perderemos informacion a la hora de explicar los grupos si no estan estas variables.

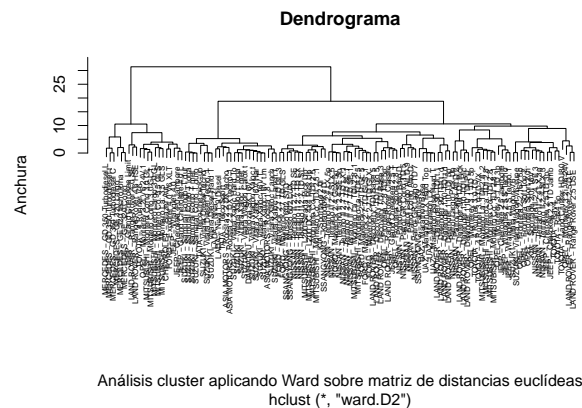
Una vez seleccionadas las variables a utilizar procederemos a tipificar las variables como paso previo necesario debido a que las variables que explican a los coches estan en distintas escalas metricas, entonces, para poder comparar unas con otras a traves de la tecnicas del analisis cluster debemos tipificarlas o escalarlas. ***

Fase 2

En la Fase 2 del analisis, realizamos algunas tecnicas para empezar a visualizar grupos de observaciones similares. Comenzamos con un grafico con las distancias de pearson que nos permite identificar que hay grupos de observaciones bien identificadas en base al tamaño de los cuadrados y de sus colores. Mientras mas azul, podemos observar que los coches son mas homogeneos y los rojos significa que son mas heterogeneos. Aqui para saber lo similares que son los vehiculos, lo analizamos en base a la posicion que tienen los coches en virtud de sus caracteristicas.



Luego, utilizando la tecnicas de la distancia euclidea, visualizamos una matriz de correlaciones y un dendrograma que nos permiten confirmar e interpretar mejor esas relaciones. Aqui, ya empezamos a observar que hay tres grupos de vehículos similares en base a su distancia euclidea.

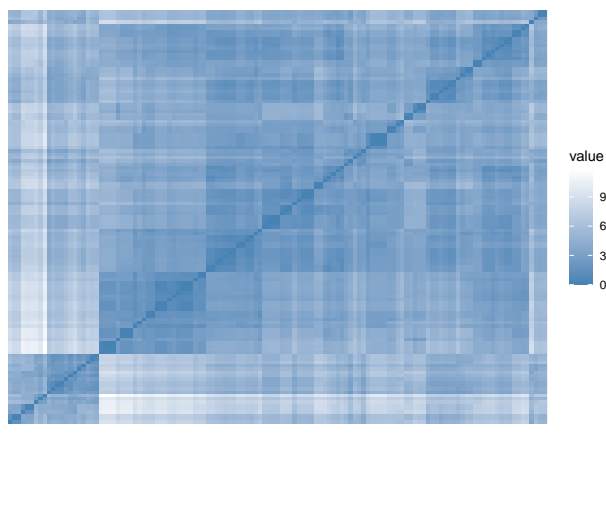


Fase 3

En la Fase 3 comenzamos el analisis cluster per se. Aqui comenzamos con distintas pruebas para encontrar el numero de clusters optimos para separar las observaciones. Inicializando con un analisis no jerarquizado *kmeans*, utilizando la distancia euclidea sobre los datos estandarizados, observamos que nos lanzaba solo un cluster como optimo. Entonces, probamos los valores de la silueta para un analisis jerarquico *hclust* con 4 clusters, opcion que terminamos descartando por considerarlas forzadas. Luego, realizamos distintas pruebas con datos aleatorios comparandolos con los datos sin aleatoriedad.

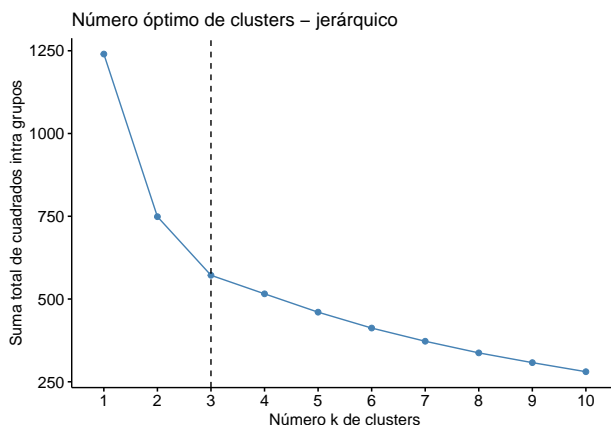
Continuando, analizamos la bondad del análisis cluster con el método Hopkins. Se trata de un contraste frente a la estructura aleatoria a través de una distribución uniforme del espacio de datos; la idea es contrastar una hipótesis de distribución uniforme / aleatoria de los datos frente a su alternativa (que no lo sea); de aceptarse la hipótesis nula, no existirían grupos de observaciones interesantes en el conjunto analizado. En nuestro caso, al conseguir un resultado de 0.145 nos permite rechazar la hipótesis de aleatoriedad y entonces avalamos la presencia de dos o más clusters en el conjunto de observaciones.

Además, utilizamos el método VAT (Virtual Assessment of cluster Tendency) que permite inspeccionar visualmente la posibilidad de agrupamiento de los datos analizados. La intensidad del azul es proporcional a la disimilaridad entre las observaciones; cuanto más oscuro, menor la distancia entre las observaciones ($\text{dist}(x_i, x_j) \rightarrow 0$, con azul indicando $\text{dist}(x_i, x_j) = 0$) mientras que cuanto más claro, mayor la distancia entre ellas ($\text{dist}(x_i, x_j) \rightarrow 1$, con blanco indicando $\text{dist}(x_i, x_j) = 1$). Las observaciones de un mismo cluster se presentan de forma correlativa.



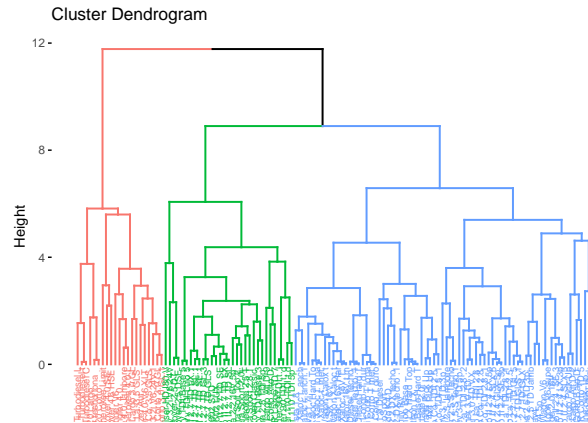
Fase 4

En la Fase 4, buscamos determinar el número óptimo de clusters. Observando distintos métodos como el método del hombro y haciendo uso de la librería NBClust, se aprecia que el número óptimo de clusters que se deben utilizar para la agrupación de los coches está entre dos y tres. Nosotros tomaremos la decisión de trabajar con tres clusters ya que nos facilita el trabajo de la repartición de los coches, algo que sería muy complicado de realizar solo con dos clusters.



Una vez elegido el numero de grupos que vamos a utilizar procedemos al estudio de dichos grupos analizando sus características principales y diferenciales respecto del resto de los grupos.

Dendrograma de los tres clusters



- Grupo 1: Este grupo esta compuesto por 73 coches, pintados de azul en la grafica, son modelos de 4 cilindros en su mayoria, mas baratos en media respecto del resto de los grupos. Al mismo tiempo, tienen menor potencia, cc y peso respecto del grupo 2 y 3. Ademas, su consumo es considerablemente inferior a los del grupo 2 y similares a los del grupo 3.
- Grupo 2: Este grupo esta compuesto por 21 coches, pintados de rojo en el dendrograma, son modelos de alta gama como los Mercedes Benz que se caracterizan por tener potencia, velocidad, consumo y numero de cilindros por encima de la media del resto de los grupos.
- Grupo 3: Este grupo esta compuesto por 31 coches, pintados de verde en la grafica, son modelos caracterizados por tener valores intermedios entre los del grupo 1 y los del grupo 2. Asi, logran superar en potencia, peso y velocidad a los coches del grupo 1 pero sin alcanzar los valores que caracterizan a los coches del grupo 2.

Ahora pasaremos al reparto de los grupos en las distintas casas del jefe. El grupo 1 formado por los 73 coches de gama baja, deberian ser asignados en 5 garages por tener la limitacion de 15 coches por garage como maximo, entonces procederemos a enviarlos a Monaco, Andorra y a Corse (en barco) por la poca distancia entre estas localidades, a su vez, creo que seria conveniente dejar estos coches de gama baja para estas localidades que, atendiendo a los gustos del jefe, le resultara una buena eleccion porque le gusta utilizar los coches de gama alta en otras localidades como Paris o en Suiza.

Siguiendo con el reparto y atendiendo a los criterios e interes del jefe, enviaremos los 21 coches de alta gama del grupo dos a las localidades de Suiza repartiendo 10 y 11 coches en cada una de estas localidades. Finalmente, repartiremos los coches del grupo 3, que tenian características intermedias entre el grupo 1 y 2, en las localidades de Paris donde repartiremos los coches mas caros de este grupo y luego los 5 modelos mas baratos como los SSANGYONG - Musso 602 STD o el FORD - Maverick 2.7 TD GL 3 los enviaremos a La Rochelle.