

# Cluster coches del Jefe

Sergio Casares

23/11/2020

## Contents

<b>1. Introducción al trabajo</b>	<b>2</b>
<b>2. Selección a priori de variables y número de clústers</b>	<b>2</b>
<b>3. Desarrollo del trabajo</b>	<b>3</b>
3.1. Análisis Factorial de Correspondencias . . . . .	3
3.2. Distancia Euclídea . . . . .	3
3.3. Agrupación en k-medios . . . . .	4
3.4. Silueta . . . . .	5
<b>4. Conclusiones</b>	<b>5</b>
<b>5. Anexos: Coches por Grupo</b>	<b>6</b>

# 1. Introducción al trabajo

Asignar de la forma más eficiente y consistente los distintos vehículos, agrupándolos en virtud de las características que ud. considere oportunas.

El objetivo de este trabajo será el de satisfacer la necesidad de nuestro jefe de ‘clusterizar’ los coches adquiridos en un número determinado de grupos con el fin de que estos puedan ser guardados en hasta un máximo de 10 garajes que tiene nuestro jefe.

## 2. Selección a priori de variables y número de clústers

### Selección de variables:

A priori (y con base en el *análisis de datos*) eliminamos las siguientes variables:

- Aceleración (acelerac) -> Por su alta proporción de valores nulos 46 / 125 (38 %), además de su alta correlación con *potencia* y *velocidad* (por encima de 0.8).
- Tiempo de aceleración (acel2) -> El tiempo de aceleración es una variable que aporta poca información a la hora de poder agrupar los coches (solo 3 coches tienen un valor de 1 y 122 tienen un valor de 2), por lo que optamos por eliminar dicha variable del análisis.
- Potencia (potencia)-> Realizando una matriz de correlaciones, observamos que la correlación entre **velocidad** y **potencia** es de 0.85, lo que indica que la relación lineal entre ambas variables es muy alta y debido a su gran similitud (y que puede generar multicolinealidad), se decide a eliminar la **potencia** (ya que también presenta una correlación alta con el **precio**, **cilindradas**, **centímetros cúbicos** y **consumo urbano**), todas por encima del 0.7 de correlación.

A priori (y con base en el *análisis financiero y de costes*) eliminamos las siguientes variables:

La colección se centra en todo terrenos clásicos (lo que indica que su finalidad es más de coleccionista que de uso personal y diario), desconocemos la finalidad de nuestro jefe (coleccionar para tenerlos almacenados en sus garajes o a lo mejor tiene la visión futura de venderlos a un mayor precio (a modo de inversión a largo plazo)).

- “cons90”, “cons120”, “consurb” -> entiendo que el consumo en gasolina no debe ser un factor discriminante a tener en cuenta a la hora de segmentar ni a la hora de su potencial venta (a nadie le interesa lo que consuma el coche, el que compra este tipo de coches no se fija en lo que le costará la gasolina).
- “plazas” -> por el mismo motivo que el anterior, es un coche de coleccionista, no de uso diario, el coleccionista se fija en el coche, no en el número de plazas que tenga mi todo-terreno clásico.
- “rpm” -> existen variables que se valoran más a la hora de comprar coches de colección, como los cc y el número de cilindros del motor.

### Selección final de variables:

Decidimos seleccionar con su precio de venta al público (pvp), el número de cilindros (cilindro), su velocidad máxima (velocida), su peso (peso) y los centímetros cúbicos del motor (cc).

### Selección del número de clústers:

Aunque el análisis el análisis debe realizarse por similitudes entre observaciones, pero si tenemos en cuenta la estrategia y el enfoque económico a la hora de hacer el reparto de los coches, deberíamos tener las siguientes condiciones en cuenta:

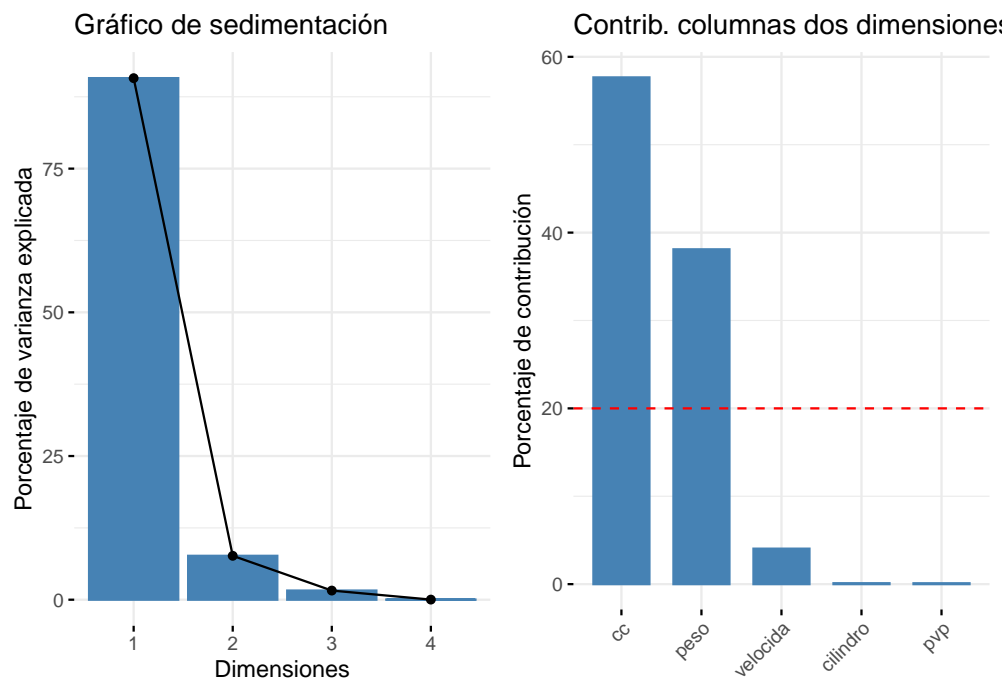
- El peso del vehículo afecta a la hora realizar el transporte (unos coches necesitarán un transporte con más cuidado o con camiones más pesados, lo que encarece el desplazamiento).
- Hay que tener en cuenta el número de coches que puede desplazar el camión (entre 6 y 10), para que cada clúster tenga un número múltiplo de coches y así evitar desplazamientos del camión ineficientes (que no esté lleno).

Por lo tanto, el número de clúster será aquél que minimice el coste de transporte, favoreciendo agrupaciones donde el tamaño de cada grupo pueda hacer que los camiones de desplazamientos vayan lo suficientemente llenos.

### 3. Desarrollo del trabajo

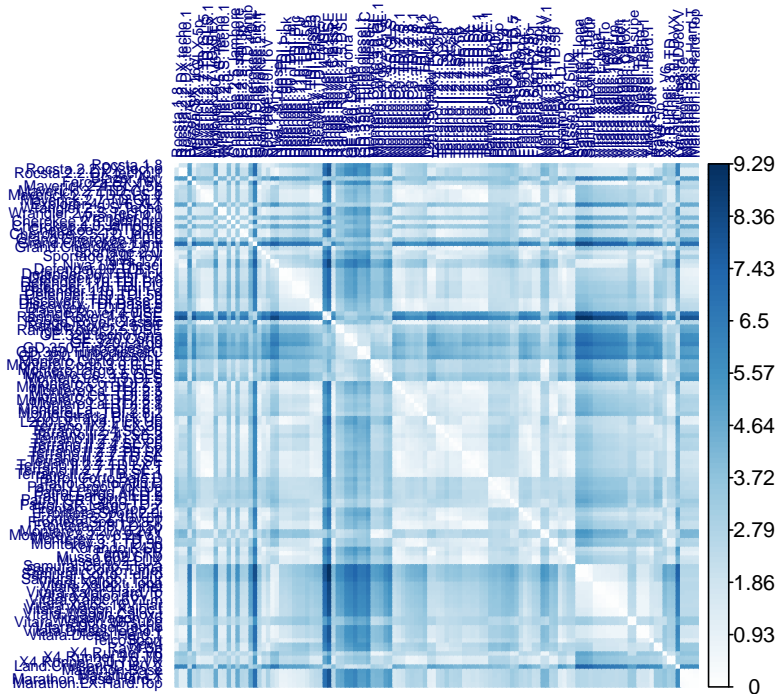
#### 3.1. Análisis Factorial de Correspondencias

De este breve análisis podemos discernir que dos dimensiones son más que suficientes para poder explicar la varianza de toda la información disponible, además, los centímetros cúbicos y el peso son las dos variables que más aportan a explicar la varianza total, lo que nos podría indicar que la agrupación de los coches pudiera seguir alguna de estas dos variables como método decisorio.



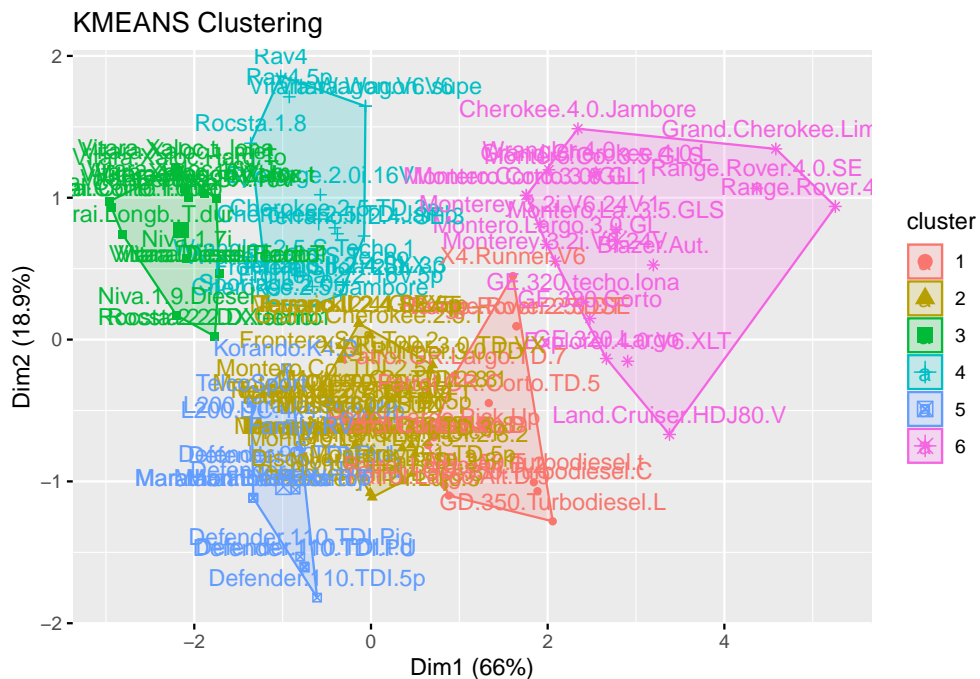
#### 3.2. Distancia Euclídea

En este gráfico se puede observar la distancia euclídes (distancia en el mapa de coordenadas entre dos valores) de cada coche (individualmente) con el resto de coches. A simple vista se pueden observar ciertas asociaciones entre diferentes grupos de coches, lo que nos indica que vamos por uen camino para poder segmentar.



### 3.3. Agrupación en k-medios

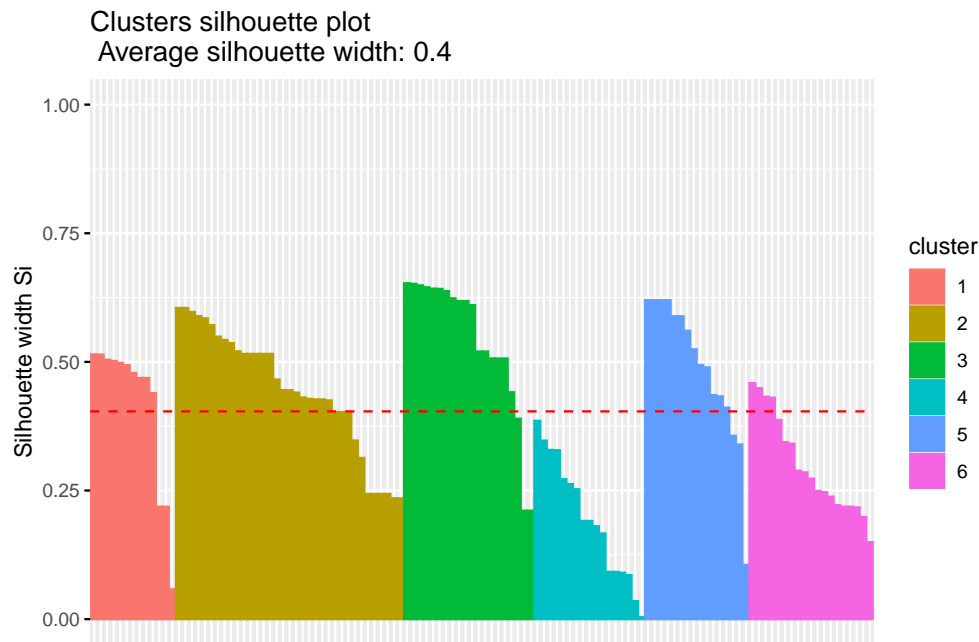
Realizamos la agrupación de los diferentes coches en función de sus coordenadas en el gráfico que componen ambas dimensiones, elegimos 6 grupos ya que previamente hemos comprobado que es el que más se ajusta a nuestros requisitos de eficiencia y consistencia.



### 3.4. Silueta

Silueta de nuestra distribución de coches con una media del ancho de silueta del 0.4.

##	cluster	size	ave.sil.width
## 1	1	13	0.41
## 2	2	35	0.44
## 3	3	20	0.54
## 4	4	17	0.19
## 5	5	16	0.49
## 6	6	19	0.30



## 4. Conclusiones

De este breve resumen acerca del análisis factorial de correspondencias y agrupación y ‘clusterización’ se pueden obtener las siguientes conclusiones:

- Las variables que hemos utilizado para realizar este trabajo son (pvp, cilindro, velocidad, peso, cc), las demás fueron descartadas por su poco valor analítico o su poco interés estratégico y económico.
- Del AFC llegamos a la conclusión de que dos dimensiones son más que suficientes para explicar la varianza total y que los *centímetros cúbicos* y el *peso* son las variables que más información aportan y las que se deberían tomar en consideración a la hora de realizar una agrupación de los coches.
- Del análisis de agrupación podemos comprobar que 6 son los grupos necesarios que hacen que se puedan cumplir con los requisitos exigidos (eficiencia y consistencia en la decisión).
- Un camión de transporte de coches transporta entre 6 y 10 vehículos (al tener 125 vehículos implica que se necesitan entre 13 y 21 desplazamientos para transportar todos), nuestra agrupación en 6 grupos hace que tan solo se necesiten 14 desplazamientos para transportar todos los vehículos, lo cual hace evidente que nuestra agrupación es consistente y eficiente.

## 5. Anexos: Coches por Grupo

### Coches del grupo 1:

##	Range.Rover.2.5.DT	Range.Rover.2.5.DSE	GD.350.Turbodiesel.t
##	1	1	1
##	GD.350.Turbodiesel.C	GD.350.Turbodiesel.L	Patrol.Corto.Bajo.D
##	1	1	1
##	Patrol.Corto.Alto.D	Patrol.Largo.Pick.Up	Patrol.Largo.Alt.D.2
##	1	1	1
##	Patrol.Largo.Alt.D.9	Patrol.GR.Corto.TD.5	Patrol.GR.Largo.TD.7
##	1	1	1
##	X4.Runner.V6		
##	1		

### Coches del grupo 2:

##	Maverick.2.4.GLX.5p	Maverick.2.7.TD.GL.5	Maverick.2.7.TD.GLS.1
##	2	2	2
##	Maverick.2.7.TD.GLX	Grand.Cherokee.2.5.T	Discovery.TDI.Base.3
##	2	2	2
##	Discovery.TDI.Base.5	Discovery.TDI.Lujo.5	Range.Rover.Classic
##	2	2	2
##	Montero.Co..TDI.2.5	Montero.Co..TDI.2.5.1	Montero.La..TDI.2.5
##	2	2	2
##	Montero.Co..TDI.2.8	Montero.Co..TDI.2.8.1	Montero.La..TDI.2.8
##	2	2	2
##	Montero.La..TDI.2.8.1	Montero.La..TDI.2.8.2	Terrano.II.2.4.LX.5p
##	2	2	2
##	Terrano.II.2.4.SLX.5	Terrano.II.2.4.SE.5p	Terrano.II.2.7.TD.LX
##	2	2	2
##	Terrano.II.2.7.TD.SL	Terrano.II.2.7.TD.SE	Terrano.II.2.7.TD.LX.1
##	2	2	2
##	Terrano.II.2.7.TD.SL.1	Terrano.II.2.7.TD.SE.1	Frontera.Soft.Top.2.
##	2	2	2
##	Frontera.Sport.2.8.T	Frontera.2.8.TDi.5p	Monterey.3.1.TD.3p
##	2	2	2
##	Monterey.3.1.TD.5p	Musso.602.STD	Musso.602.LUX
##	2	2	2
##	X4.Runner.3.0.TD	X4.Runner.3.0.TD.VX	
##	2	2	

### Coches del grupo 3:

##	Rocsta.2.2.DX.techo	Rocsta.2.2.DX.techo.1	Feroza.SX.16V
##	3	3	3
##	Niva.1.7i	Niva.1.9.Diesel	Samurai.Corto.T.lona
##	3	3	3
##	Samurai.Corto.T.met	Samurai.Corto.T.duro	Samurai.Longb..T.dur
##	3	3	3
##	Vitara.Xaloc.t..lona	Vitara.Xaloc.t..met	Vitara.Xaloc.Hard.To
##	3	3	3

##	Vitara.Xaloc.16V.T.	Vitara.Xaloc.16V.t.m	Vitara.Xaloc.16V.Har
##	3	3	3
##	Vitara.Wagon.Xaloc.t	Vitara.Wagon.Calox.t	Vitara.Diesel.Techo
##	3	3	3
##	Vitara.Diesel.Techo.1	Vitara.Diesel.Hard.T	
##	3	3	

#### Coches del grupo 4:

##	Rocsta.1.8	Wrangler.2.5.S.Techo	Wrangler.2.5.S.Techo.1
##	4	4	4
##	Cherokee.2.5.Jambore	Cherokee.2.5.TD.3p	Cherokee.2.5.TD.Jamb
##	4	4	4
##	Sportage.2.0i	Sportage.2.0i.16V	Terrano.II.2.4.LX.3p
##	4	4	4
##	Terrano.II.2.4.SLX.3	Terrano.II.2.4..SE.3	Frontera.Sport.2.0i
##	4	4	4
##	Frontera.2.2.16V.5p	Vitara.Wagon.V6	Vitara.Wagon.V6.supe
##	4	4	4
##	Rav4	Rav4.5p	
##	4	4	

#### Coches del grupo 5:

##	Defender.90.TDI.t..1	Defender.90.TDI.Pick	Defender.90.TDI.t..d
##	5	5	5
##	Defender.110.TDI.Pic	Defender.110.TDI.PU	Defender.110.TDI.t.d
##	5	5	5
##	Defender.110.TDI.5p	L200.Strada.Pick.Up	L200.DC.4x4.Pick.Up
##	5	5	5
##	Korando.K4.D	Family.RV	TelcoSport
##	5	5	5
##	Marathon.Base	Marathon.LX	Marathon.Base.Hard.T
##	5	5	5
##	Marathon.LX.Hard.Top		
##	5		

#### Coches del grupo 6:

##	Blazer.Aut.	Explorer.4.0.V6.XLT	Wrangler.4.0
##	6	6	6
##	Cherokee.4.0.Jambore	Grand.Cherokee.4.0.L	Grand.Cherokee.Limit
##	6	6	6
##	Range.Rover.4.0.SE	Range.Rover.4.6.HSE	GE.320.techo.lona
##	6	6	6
##	GE.320.Corto	GE.320.Largo	Montero.Corto.3.0.GL
##	6	6	6
##	Montero.Corto.3.0.GL.1	Montero.Largo.3.0.GL	Montero.Co..3.5.GLS
##	6	6	6
##	Montero.La..3.5.GLS	Monterey.3.2i.V6.24V	Monterey.3.2i.V6.24V.1
##	6	6	6
##	Land.Cruiser.HDJ80.V		
##	6		