

# Tarea N° 3 Los coches del jefe 2

*Paúl Córdova*

*22 de noviembre de 2018*

## Resumen Ejecutivo

El presente informe tiene por objetivo clasificar de forma homogénea a vehículos todo terreno (4x4) clásicos en diez grupos correspondiente a diez residencias. Para ello, se aplicó la técnica de clusterización “kmeans” empleando como variables relevantes al objeto de estudio: la potencia del vehículo, el peso y las revoluciones por minuto; con dicha información se agrupó cada clúster y cada residencias en base a su distancia; permitiendo alcanzar eficiencia en la distribución al procurar valores mínimos en las plazas sobrantes en cada ubicación.

## Introducción

Como preámbulo se tiene que el coleccionista ostenta 125 vehículos todo terreno (TT) y posee información relevante de cada uno de ellos. Además, se precisa el requerimiento de agrupar a los vehículos en 10 clusters, correspondiente al número de propiedades. Dado el requerimiento de establecer una determinada división de los vehículos, este trabajo busca emplear la forma más eficiente y consistente para dicha agrupación.

Para comenzar con este proceso fue preciso determinar una visión general del objeto de estudio. Donde se establece que la creación de dichos vehículos surgió en tiempo de guerra, ante la necesidad de acceder a terrenos de difícil acceso (zonas rurales, alta montaña, entre otros); posteriormente su uso se difundió a los civiles, sin embargo mantuvieron las características para las cuales fueron concebidos.

Entre los principales aspectos a destacar en los vehículos 4x4 se encuentran: tracción en las cuatro ruedas, fuerza de torque, motor potente, chasis con refuerzos; dadas estas facetas con base a la información recopilada de cada vehículo se procedió a seleccionar las variables relevantes para el análisis:

- El peso: relacionado a la durabilidad y resistencia de los vehículos todo terreno; además, se encuentra relacionado a la fortaleza de sus componentes.
- La potencia: porque caracteriza la energía liberada por este tipo de vehículos que requieren de una mayor fuerza de torque para superar trayectos difíciles.
- Las revoluciones por minuto (rpm): su influencia radica en que a mayor número de vueltas en los pistones se incrementa la potencia del vehículo.

Mientras que no se tomaron en consideración las siguientes variables: el precio de venta al público (porque no se presenta como una característica de similitud), el número de asientos (por considerarse limitada en la segmentación de los todo terreno); el consumo de 90 km/h, el consumo de 120 km/h y el consumo urbano (debido a que la mayoría de los vehículos 4x4 albergan un alto consumo de combustible); el número de cilindros (porque su distribución es limitada a tres valores); los centímetros cúbicos (por su relación estrecha con la potencia y dicha variable ya se encuentra seleccionada); la velocidad máxima, la aceleración y el tiempo de aceleración (porque en estos vehículos se privilegia la potencia sobre la velocidad).

## Desarrollo

En un inicio, al examinar las variables relevantes se destaca que en los vehículos todo terreno la potencia media se encontró entre 117.5 caballos de vapor y las revoluciones por minuto oscilaron entre los 3600 y 6500. A su vez, el peso mínimo de los vehículos fue de 930 kilogramos y el máximo fue de 2320 kilogramos. Esta aproximación nos permite identificar que las variables se encuentran en diversas escalas, por lo cual se estandarizaron los datos para tener una homogeneidad en las medidas para realizar el método de agrupación.

Al realizar un análisis de aquellas variables que presentaron valores NA, se tuvo como resultado que únicamente la variable del peso de los todo terreno presentó dos observaciones sin registros; por lo cual, se optó por omitir dichos vehículos de la segmentación.

Es preciso destacar la existencia de 14 vehículos duplicados; ante ello, se procedió a no tomarlos en consideración en el análisis (manteniendo los vehículos no duplicados). Sin embargo, para su posterior inclusión se consolidarán en el mismo grupo correspondiente al modelo del vehículo que correspondan.

En esta primera instancia se empleó la función de agrupación “kmeans”, mediante la cual se pudo establecer el análisis gráfico de la división por 10 grupos; además, se destacó el escaso solapamiento entre los grupos y la semejanza en las dimensiones de cada clúster (ver Figura 1).

En la Tabla 1, se denota que el Grupo 1 y 2 contienen una mayor cantidad de automóviles todo terreno; mientras que los Grupos 4, 7 y 9 mantuvieron una menor cantidad. No obstante, al comparar la relación entre el grupo con mayor cantidad de vehículos con referencia al de menor se establece una relación de 3 a 1.

Posterior a la aplicación del método *kmeans* para la clasificación de los automóviles todo terreno, se procede a incluir los coches duplicados en cada uno de los clúster que corresponda; para ello, se identificó el modelo del 4x4 y agrupó en la categoría correspondiente (ver Tabla 1). Cabe destacar que para el modelo se emplearon 109 observaciones, al incluir los duplicados a la clasificación aumentó el número a 122; no obstante tres observaciones con valores perdidos quedaron pendientes de ser clasificadas.

Table 1: Tabla de agrupación de los vehículos 4x4 (kmeans)

Clúster - Agrupación	Modelo sin duplicados	Modelo con duplicados
Grupo 1	17	20
Grupo 2	16	20
Grupo 3	12	14
Grupo 4	5	5
Grupo 5	10	12
Grupo 6	12	13
Grupo 7	7	7
Grupo 8	16	7
Grupo 9	5	5
Grupo 10	9	9
Total	109	122

Ante el requerimiento de dividir la colección en diversos grupos es preciso tomar en consideración las siguientes directrices: los vehículos se deben distribuir entre las 10 propiedades existentes, cada residencia puede alojar máximo 15 automóviles y el criterio de distribución debe ser conforme a un criterio de distancia.

En primera instancia, en base a la visualización procedente del modelo de clasificación se realizó un conglomerado de los grupos (grupos macro) que contengan a los grupos del modelo; siendo el criterio de agrupación la cercanía entre grupos en el plano (ver Figura 1).

En este sentido, se obtuvo la siguiente clasificación de grupos macro y sus respectivos subgrupos según su cercanía:

- Grupo Macro 1: Grupo 1 y Grupo 2
- Grupo Macro 2: Grupo 4, Grupo 5 y Grupo 10
- Grupo Macro 3: Grupo 6 y Grupo 8
- Grupo Macro 4: Grupo 3
- Grupo Macro 5: Grupo 7
- Grupo Macro 6: Grupo 9

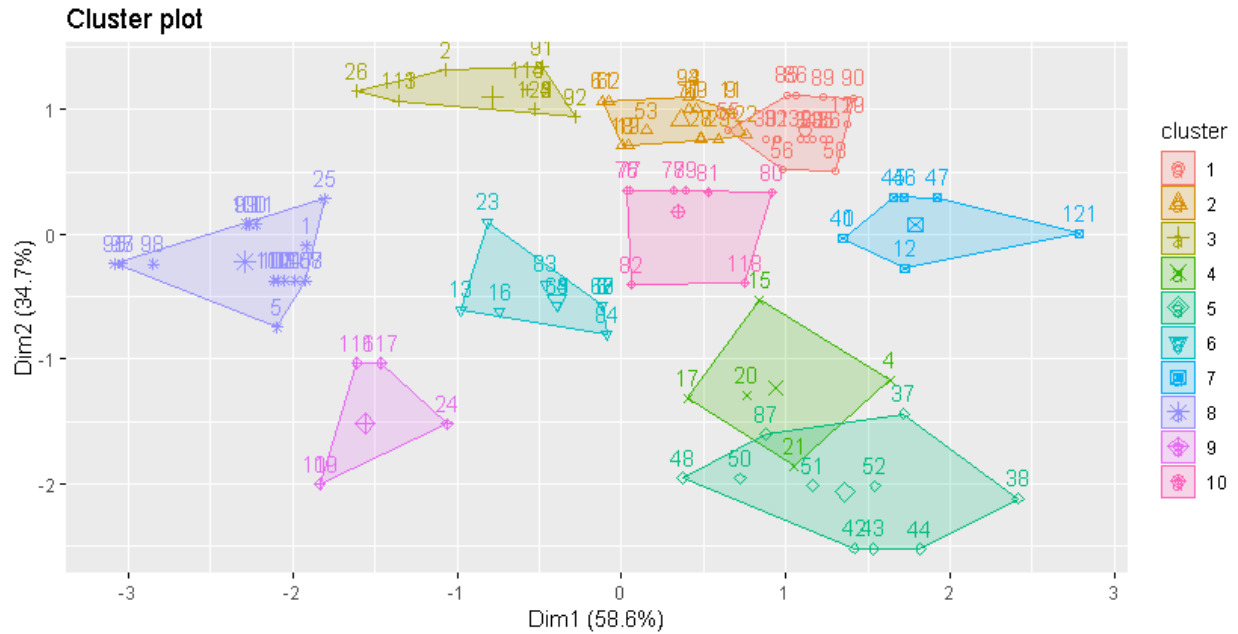


Figure 1: Gráfico Agrupación por Kmeans

Simultáneamente, se realizó el mismo procedimiento en las residencias mediante la agrupación por cercanía territorial (ver Figura 2):

- Territorio 1: Casa 3, Casa 5 y Casa 10
- Territorio 2: Casa 1 y Casa 2
- Territorio 3: Casa 6 y Casa 7
- Territorio 4: Casa 4
- Territorio 5: Casa 9
- Territorio 6: Casa 8

A partir de las distinciones propuestas y relacionando los criterios de distancias cercanas entre grupos y a nivel territorial se determina la siguiente distribución de los vehículos 4x4:

- 1) En Sanremo (casa 10) se ubican 15 vehículos (grupo 2), en Vence (casa 3) se disponen 5 vehículos (grupo 2) y 10 vehículos (grupo 1) y en Saint-Raphaël (casa 5) se colocan 10 vehículos (grupo 1).
- 2) En la residencia de Le Perreux-sur-Marne (casa 1) se disponen 5 vehículos (grupo 4) y 10 vehículos (grupo 5); además en Bougival (casa 2) se disponen 2 vehículos (grupo 5) y 9 vehículos (grupo 10).
- 3) En Colombier (casa 6) se ubican 15 vehículos (grupo 8) y en Montenol (casa 7) se colocan 2 vehículos (grupo 8) y 13 vehículos (grupo 6).
- 4) En la residencia de Palasca (casa 4) se disponen 14 vehículos (grupo 3).
- 5) En la residencia de Mérens-les-Vals (casa 9) se disponen 7 vehículos (grupo 7).
- 6) En la residencia de Yves (casa 8) se disponen 9 vehículos (grupo 9).



Figure 2: Gráfico ubicación geográfica residencias

Para un entendimiento de la distribución y las plazas restantes se dispuso la Tabla 2, que plasma la relación entre los grupos y las residencias.

Table 2: Relación de Grupos Macro y Conglomerados Territorio

Agrupación clúster	Conglomerado territorio	N <sup>a</sup> plazas restantes
Grupo Macro 1 (40 coches)	Territorio 1 (45 plazas)	5
Grupo Macro 2 (26 coches)	Territorio 2 (30 plazas)	4
Grupo Macro 3 (30 coches)	Territorio 3 (30 plazas)	0
Grupo Macro 4 (14 coches)	Territorio 4 (15 plazas)	1
Grupo Macro 5 (7 coches)	Territorio 5 (15 plazas)	8
Grupo Macro 6 (5 coches)	Territorio 6 (15 plazas)	10

Por último se estableció que las observaciones que ostentaban valores faltantes y no pudieron ser clasificadas, se las consideró dentro del Grupo 9 (grupo con menor número de vehículos) y se ubicaron en la residencia de Yves (casa 8), dejando solamente 7 plazas restantes.

## Conclusiones

- Para la agrupación de las observaciones se consideraron como variables fundamentales: la potencia del automotor, el peso y las revoluciones por minuto; debiéndose su consideración a las características primordiales y uso de este tipo de vehículos todo terreno.
- La técnica de agrupación “kmeans” otorgó una adecuada aproximación para la segmentación de los vehículos en las 10 categorías requeridas. Este proceso permitió una distribución consensuada de los automotores en base al criterio de distancia por grupos y la distancia territorial.
- Se destaca que la agrupación fue eficiente al dejar pocas plazas vacías en las residencias.

## Bibliografía y Recursos digitales

1. Notas técnicas. (octubre 2018). *El análisis cluster*. Profesor: Juan Manuel López Zafra. CUNEF-Máster en Data Science para Finanzas.
2. R Program y RStudio Version 1.1.456 – © 2009-2018 RStudio, Inc.

## Anexos

### Anexo A: Listado de vehículo por marca, modelo y su respectiva agrupación

##	marca	modelo	cluster
## 1	ASIA MOTORS	Rocsta 1.8	8
## 2	ASIA MOTORS	Rocsta 2.2 DX techo	3
## 3	ASIA MOTORS	Rocsta 2.2 DX techo	3
## 4	CHEVROLET	Blazer Aut.	4
## 5	DAIHATSU	Feroza SX 16V	8
## 6	FORD	Maverick 2.4 GLX 5p	6
## 7	FORD	Maverick 2.7 TD GL 3	NA
## 8	FORD	Maverick 2.7 TD GLS	NA
## 9	FORD	Maverick 2.7 TD GL 5	2
## 10	FORD	Maverick 2.7 TD GLS	NA
## 11	FORD	Maverick 2.7 TD GLX	2
## 12	FORD	Explorer 4.0 V6 XLT	7
## 13	JEEP	Wrangler 2.5 S Techo	6
## 14	JEEP	Wrangler 2.5 S Techo	6
## 15	JEEP	Wrangler 4.0	4
## 16	JEEP	Cherokee 2.5 Jambore	6

## 17	JEEP Cherokee 4.0 Jambore	4
## 18	JEEP Cherokee 2.5 TD 3p	2
## 19	JEEP Cherokee 2.5 TD Jamb	2
## 20	JEEP Grand Cherokee 4.0 L	4
## 21	JEEP Grand Cherokee Limit	4
## 22	JEEP Grand Cherokee 2.5 T	2
## 23	KIA Sportage 2.0i	6
## 24	KIA Sportage 2.0i 16V	9
## 25	LADA Niva 1.7i	8
## 26	LADA Niva 1.9 Diesel	3
## 27	LAND ROVER Defender 90 TDI t. 1	2
## 28	LAND ROVER Defender 90 TDI Pick	2
## 29	LAND ROVER Defender 90 TDI t. d	2
## 30	LAND ROVER Defender 110 TDI Pic	1
## 31	LAND ROVER Defender 110 TDI PU	1
## 32	LAND ROVER Defender 110 TDI t.d	1
## 33	LAND ROVER Defender 110 TDI 5p	1
## 34	LAND ROVER Discovery TDI Base 3	1
## 35	LAND ROVER Discovery TDI Base 5	1
## 36	LAND ROVER Discovery TDI Lujo 5	1
## 37	LAND ROVER Range Rover 4.0 SE	5
## 38	LAND ROVER Range Rover 4.6 HSE	5
## 39	LAND ROVER Range Rover Classic	1
## 40	LAND ROVER Range Rover 2.5 DT	7
## 41	LAND ROVER Range Rover 2.5 DSE	7
## 42	MERCEDES GE 320 techo lona	5
## 43	MERCEDES GE 320 Corto	5
## 44	MERCEDES GE 320 Largo	5
## 45	MERCEDES GD 350 Turbodiesel t	7
## 46	MERCEDES GD 350 Turbodiesel C	7
## 47	MERCEDES GD 350 Turbodiesel L	7
## 48	MITSUBISHI Montero Corto 3.0 GL	5
## 49	MITSUBISHI Montero Corto 3.0 GL	5
## 50	MITSUBISHI Montero Largo 3.0 GL	5
## 51	MITSUBISHI Montero Co. 3.5 GLS	5
## 52	MITSUBISHI Montero La. 3.5 GLS	5
## 53	MITSUBISHI Montero Co. TDI 2.5	2
## 54	MITSUBISHI Montero Co. TDI 2.5	2
## 55	MITSUBISHI Montero La. TDI 2.5	1
## 56	MITSUBISHI Montero Co. TDI 2.8	1
## 57	MITSUBISHI Montero Co. TDI 2.8	1
## 58	MITSUBISHI Montero La. TDI 2.8	1
## 59	MITSUBISHI Montero La. TDI 2.8	1
## 60	MITSUBISHI Montero La. TDI 2.8	1
## 61	MITSUBISHI L200 Strada Pick Up	2
## 62	MITSUBISHI L200 DC 4x4 Pick Up	2
## 63	NISSAN Terrano II 2.4 LX 3p	6
## 64	NISSAN Terrano II 2.4 SLX 3	6
## 65	NISSAN Terrano II 2.4. SE 3	6
## 66	NISSAN Terrano II 2.4 LX 5p	6
## 67	NISSAN Terrano II 2.4 SLX 5	6
## 68	NISSAN Terrano II 2.4 SE 5p	6
## 69	NISSAN Terrano II 2.7 TD LX	2
## 70	NISSAN Terrano II 2.7 TD SL	2

## 71	NISSAN Terrano II 2.7 TD SE	2
## 72	NISSAN Terrano II 2.7 TD LX	2
## 73	NISSAN Terrano II 2.7 TD SL	2
## 74	NISSAN Terrano II 2.7 TD SE	2
## 75	NISSAN Patrol Corto Bajo D	10
## 76	NISSAN Patrol Corto Alto D	10
## 77	NISSAN Patrol Largo Pick Up	10
## 78	NISSAN Patrol Largo Alt.D 2	10
## 79	NISSAN Patrol Largo Alt.D 9	10
## 80	NISSAN Patrol GR Corto TD 5	10
## 81	NISSAN Patrol GR Largo TD 7	10
## 82	OPEL Frontera Soft Top 2.	10
## 83	OPEL Frontera Sport 2.0i	6
## 84	OPEL Frontera 2.2 16V 5p	6
## 85	OPEL Frontera Sport 2.8 T	1
## 86	OPEL Frontera 2.8 TDi 5p	1
## 87	OPEL Monterey 3.2i V6 24V	5
## 88	OPEL Monterey 3.2i V6 24V	5
## 89	OPEL Monterey 3.1 TD 3p	1
## 90	OPEL Monterey 3.1 TD 5p	1
## 91	SSANGYONG Korando K4 D	3
## 92	SSANGYONG Family RV	3
## 93	SSANGYONG Musso 602 STD	2
## 94	SSANGYONG Musso 602 LUX	2
## 95	SUZUKI Samurai Corto T lona	8
## 96	SUZUKI Samurai Corto T met	8
## 97	SUZUKI Samurai Corto T duro	8
## 98	SUZUKI Samurai Longb. T dur	8
## 99	SUZUKI Vitara Xaloc t. lona	8
## 100	SUZUKI Vitara Xaloc t. met	8
## 101	SUZUKI Vitara Xaloc Hard To	8
## 102	SUZUKI Vitara Xaloc 16V T.	8
## 103	SUZUKI Vitara Xaloc 16V t.m	8
## 104	SUZUKI Vitara Xaloc 16V Har	8
## 105	SUZUKI Vitara Xaloc r. anch	8
## 106	SUZUKI Vitara Xaloc r. anch	8
## 107	SUZUKI Vitara Wagon Xaloc t	8
## 108	SUZUKI Vitara Wagon Calox t	8
## 109	SUZUKI Vitara Wagon V6	9
## 110	SUZUKI Vitara Wagon V6 supe	9
## 111	SUZUKI Vitara Diesel Techo	3
## 112	SUZUKI Vitara Diesel Techo	3
## 113	SUZUKI Vitara Diesel Hard T	3
## 114	TATA TelcoLine pick-up	3
## 115	TATA TelcoSport	3
## 116	TOYOTA Rav4	9
## 117	TOYOTA Rav4 5p	9
## 118	TOYOTA 4 Runner V6	10
## 119	TOYOTA 4 Runner 3.0 TD	1
## 120	TOYOTA 4 Runner 3.0 TD VX	1
## 121	TOYOTA Land Cruiser HDJ80 V	7
## 122	UAZ Marathon Base	3
## 123	UAZ Marathon LX	3
## 124	UAZ Marathon Base Hard T	3

## 125	UAZ Marathon LX Hard Top	3
--------	--------------------------	---