

# Tarea N<sup>o</sup> 3 Los coches del jefe

*Paúl Córdova*

*18 de noviembre de 2018*

## Resumen Ejecutivo

El presente informe tiene por objetivo clasificar de forma homogénea a vehículos todo terreno clásicos en diez grupos. Para ello, se emplearon como variables relevantes al objeto de estudio: la potencia del vehículo, el peso y las revoluciones por minuto; con dicha información se aplicó la técnica de clusterización “kmeans” y “jerárquica”, teniendo como resultado la conformación de grupos por cada modelo que ostenta el coleccionista (Anexo A).

## Introducción

El presente informe alberga como objetivo la agrupación de los vehículos clásicos todo terreno (4x4), en base a características relevantes que fundamenten la segmentación. Dado el requerimiento de establecer una determinada división de los vehículos, este trabajo busca emplear la forma más eficiente y consistente para dicha agrupación.

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.

Para comenzar con este proceso fue preciso determinar una visión general del objeto de estudio. Donde se establece que la creación de los vehículos TT 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.

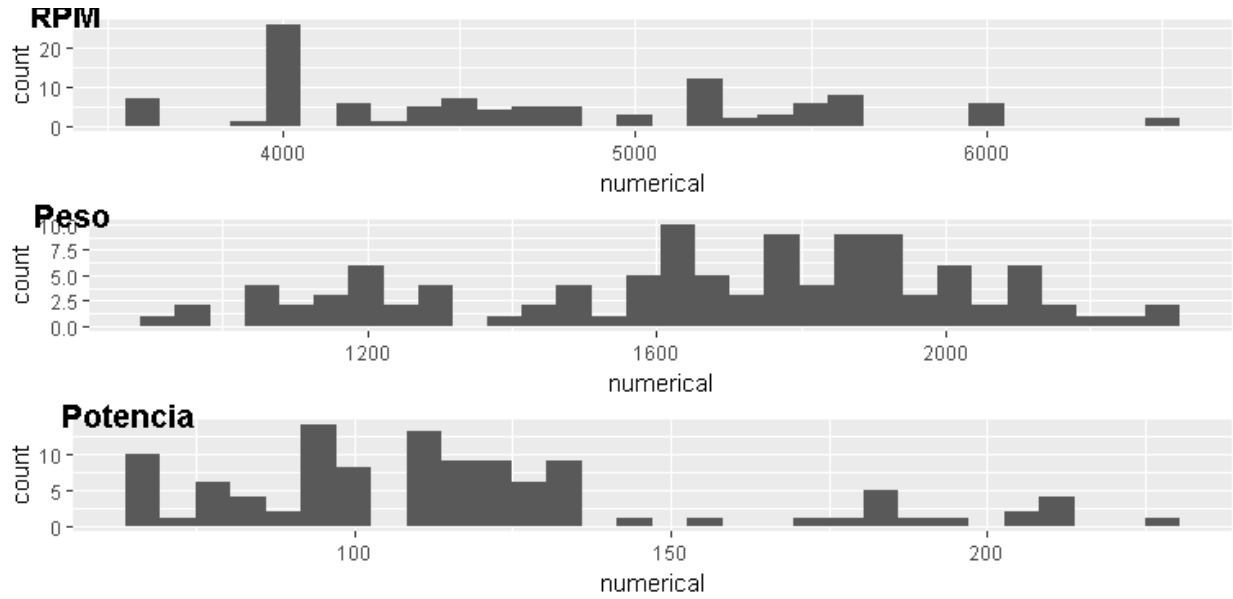


Figure 1: Histograma de variables

Al realizar un análisis de aquellas variables que presentaron valores NA, se tuvo como resultado que unicamente 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 lo que conlleva el análisis exploratorio de manera gráfica, se visualizaron las tres variables: revoluciones por minuto (rpm), peso del vehículo y la potencia mediante histogramas. Se destacó la homogeneidad en la distribución del peso de los todo terreno, la concentración de la potencia en menores caballos de vapor y la amplitud en la distribución de las revoluciones por minuto, especialmente concentrada en las 4000 rpm.

A continuación, se procedió a realizar la agrupación de los vehículos 4x4 mediante dos técnicas, la primera se enfoca en la determinación de centroides y la segunda se realiza por medio de la agrupación jerárquica.

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

En la Tabla 1, se denota que el Grupo 1, 2 y 8 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.

Con la finalidad de comprobar la robustez de la agrupación realizada, se procedió a realizar el modelo mediante la agrupación jerárquica. Al comparar los resultados, no se presentan grandes diferencias en las dimensiones de cada grupo comparados con los resultados de la técnica anterior.

Al analizar la Tabla 2, cabe destacar dos aspectos importantes que nos se presentaron en la primera agrupación. Se destaca que el grupo con menor cantidad de vehículos todo terreno fue el Grupo 7 (cuatro vehículos); mientras que el grupo con mayor cantidad de automotores fue el Grupo 9 con un total de 19 vehículos 4x4.

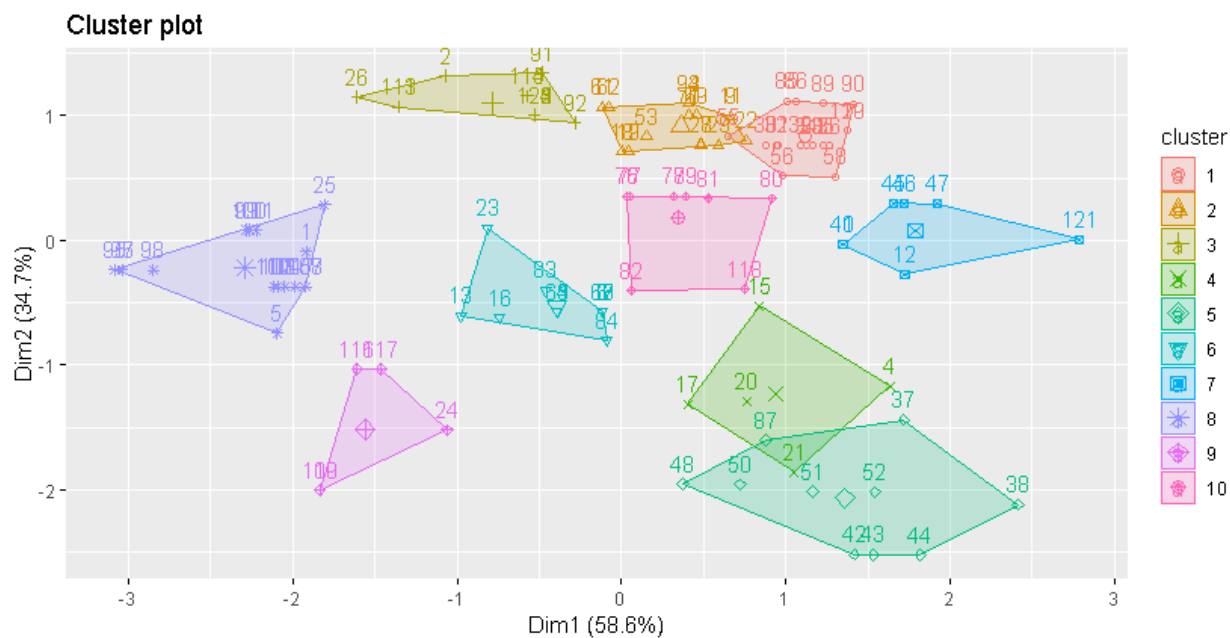


Figure 2: Gráfico Agrupación por Kmeans

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

Clúster - Agrupación	Vehículos
Grupo 1	17
Grupo 2	16
Grupo 3	12
Grupo 4	5
Grupo 5	10
Grupo 6	12
Grupo 7	7
Grupo 8	16
Grupo 9	5
Grupo 10	9

Table 2: Tabla de agrupación de los vehículos 4x4 (jerarquía)

Clúster - Agrupación	Vehículos
Grupo 1	16
Grupo 2	12
Grupo 3	11
Grupo 4	13
Grupo 5	17
Grupo 6	7
Grupo 7	4
Grupo 8	5
Grupo 9	19
Grupo 10	5

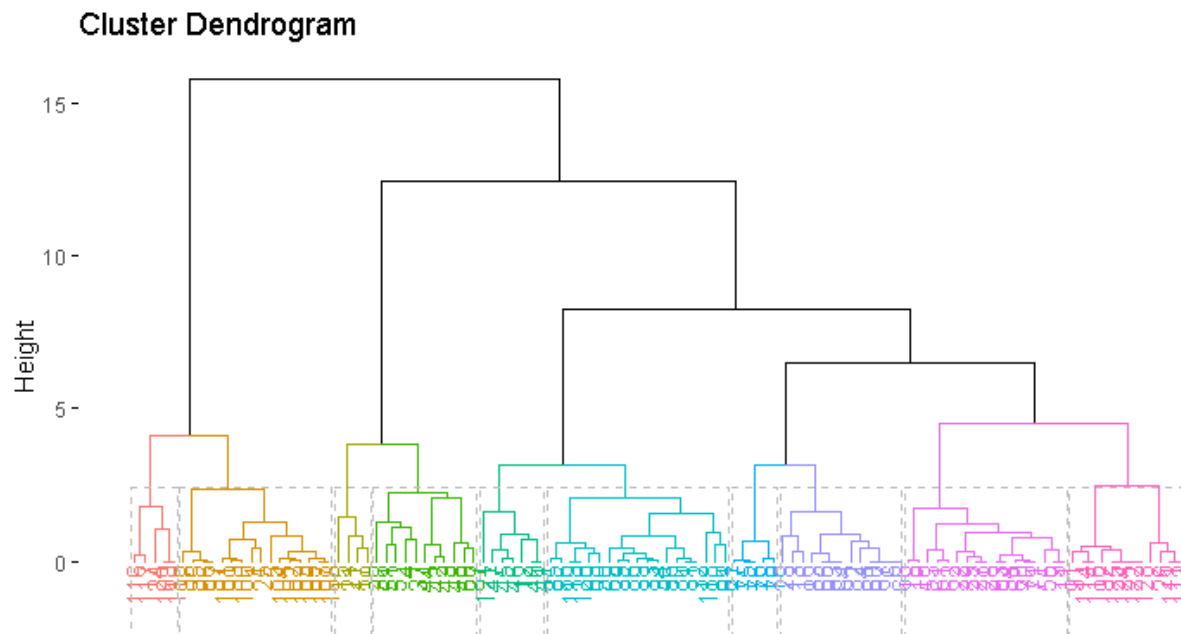


Figure 3: Gráfico Dendrograma Agrupación por Jerarquización

## 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” y “jerárquica” otorgan una adecuada aproximación para la segmentación de los vehículos; donde se comprobó la robustez de los resultados al comparar ambas metodologías.
- Dado que el objetivo del trabajo radicó en la determinación de 10 grupos que agruparan a los vehículos 4x4 se formuló el Anexo A.

## 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 and RStudio Version 1.1.456 – © 2009-2018 RStudio, Inc.

## Anexos

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

##	Grupo
## Rocsta.1.8.....	8
## Rocsta.2.2.DX.techo.	3
## Blazer.Aut.....	4
## Feroza.SX.16V.....	8
## Maverick.2.4.GLX.5p.	6
## Maverick.2.7.TD.GL.5	2
## Maverick.2.7.TD.GLX.	2
## Explorer.4.0.V6.XLT.	7

## Wrangler.2.5.S.Techo	6
## Wrangler.4.0.....	4
## Cherokee.2.5.Jambore	6
## Cherokee.4.0.Jambore	4
## Cherokee.2.5.TD.3p..	2
## Cherokee.2.5.TD.Jamb	2
## Grand.Cherokee.4.0.L	4
## Grand.Cherokee.Limit	4
## Grand.Cherokee.2.5.T	2
## Sportage.2.0i.....	6
## Sportage.2.0i.16V...	9
## Niva.1.7i.....	8
## Niva.1.9.Diesel.....	3
## Defender.90.TDI.t..l	2
## Defender.90.TDI.Pick	2
## Defender.90.TDI.t..d	2
## Defender.110.TDI.Pic	1
## Defender.110.TDI.PU.	1
## Defender.110.TDI.t..d	1
## Defender.110.TDI.5p.	1
## Discovery.TDI.Base.3	1
## Discovery.TDI.Base.5	1
## Discovery.TDI.Lujo.5	1
## Range.Rover.4.0.SE..	5
## Range.Rover.4.6.HSE.	5
## Range.Rover.Classic.	1
## Range.Rover.2.5.DT..	7
## Range.Rover.2.5.DSE.	7
## GE.320.techo.lona...	5
## GE.320.Corto.....	5
## GE.320.Largo.....	5
## GD.350.Turbodiesel.t	7
## GD.350.Turbodiesel.C	7
## GD.350.Turbodiesel.L	7
## Montero.Corto.3.0.GL	5
## Montero.Largo.3.0.GL	5
## Montero.Co..3.5.GLS.	5
## Montero.La..3.5.GLS.	5
## Montero.Co..TDI.2.5.	2
## Montero.La..TDI.2.5.	1
## Montero.Co..TDI.2.8.	1
## Montero.La..TDI.2.8.	1
## L200.Strada.Pick.Up.	2
## L200.DC.4x4.Pick.Up.	2
## Terrano.II.2.4.LX.3p	6
## Terrano.II.2.4.SLX.3	6
## Terrano.II.2.4..SE.3	6
## Terrano.II.2.4.LX.5p	6
## Terrano.II.2.4.SLX.5	6
## Terrano.II.2.4.SE.5p	6
## Terrano.II.2.7.TD.LX	2
## Terrano.II.2.7.TD.SL	2
## Terrano.II.2.7.TD.SE	2
## Patrol.Corto.Bajo.D.	10

## Patrol.Corto.Alto.D.	10
## Patrol.Largo.Pick.Up	10
## Patrol.Largo.Alt.D.2	10
## Patrol.Largo.Alt.D.9	10
## Patrol.GR.Corto.TD.5	10
## Patrol.GR.Largo.TD.7	10
## Frontera.Soft.Top.2.	10
## Frontera.Sport.2.0i.	6
## Frontera.2.2.16V.5p.	6
## Frontera.Sport.2.8.T	1
## Frontera.2.8.TDi.5p.	1
## Monterey.3.2i.V6.24V	5
## Monterey.3.1.TD.3p..	1
## Monterey.3.1.TD.5p..	1
## Korando.K4.D.....	3
## Family.RV.....	3
## Musso.602.STD.....	2
## Musso.602.LUX.....	2
## Samurai.Corto.T.lona	8
## Samurai.Corto.T.met.	8
## Samurai.Corto.T.duro	8
## Samurai.Longb..T.dur	8
## Vitara.Xaloc.t..lona	8
## Vitara.Xaloc.t..met.	8
## Vitara.Xaloc.Hard.To	8
## Vitara.Xaloc.16V.T..	8
## Vitara.Xaloc.16V.t.m	8
## Vitara.Xaloc.16V.Har	8
## Vitara.Xaloc.r..anch	8
## Vitara.Wagon.Xaloc.t	8
## Vitara.Wagon.Calox.t	8
## Vitara.Wagon.V6.....	9
## Vitara.Wagon.V6.supe	9
## Vitara.Diesel.Techo.	3
## Vitara.Diesel.Hard.T	3
## TelcoLine.pick.up...	3
## TelcoSport.....	3
## Rav4.....	9
## Rav4.5p.....	9
## X4.Runner.V6.....	10
## X4.Runner.3.0.TD.....	1
## X4.Runner.3.0.TD.VX..	1
## Land.Cruiser.HDJ80.V	7
## Marathon.Base.....	3
## Marathon.LX.....	3
## Marathon.Base.Hard.T	3
## Marathon.LX.Hard.Top	3