

# Los Coches del Jefe

Informe previo

Jose López

2020-11-21

## ABSTRACT

En el presente trabajo se pretende realizar un análisis exploratorio de la base de datos **TTERRENO.SAV** con el objetivo de comprender mejor el contenido de la misma, de manera que seamos capaces de seleccionar aquellas características que consideremos más relevantes a la hora de discriminar los vehículos, además de comprender qué tipos de vehículos son los que posee nuestro jefe.

## PLANTEAMIENTO

El archivo **TTERRENO.SAV** contiene información acerca de 125 vehículos de todo terreno clásicos. Tan clásicos, que sus precios vienen expresados en pesetas, que era la moneda oficial del Reino de España antes de la entrada en vigor del euro. En su afán como coleccionista, el dueño del *family office* en el que ud. trabaja los ha comprado todos y ha recogido, además, información relevante de ellos, expresada en las variables de las que tiene información en el documento **TTERRENO-VARIABLES.PDF** (ubicado en la carpeta *references*).

El dueño del *family office*, al que ud. reporta, quiere dividir los todo-terreno en las distintas propiedades que tiene. Dado que cuenta con hasta 10 lugares en los que podría conservar los vehículos, su trabajo consiste en asignar de la forma más eficiente y consistente los distintos vehículos, agrupándolos en virtud de las características que ud. considere oportunas. Debe por tanto informarse acerca de qué características pueden ser relevantes y cuáles no para el trabajo a realizar. Ser la persona de confianza del dueño tiene esas cosas, a veces no hay más remedio que entrar a estudiar cosas para las que ud. no se ha formado.

Una vez conocidas y justificadas cuáles son las características sobre las que va a trabajar, deberá explorar a qué datos se enfrenta, siempre con la mente puesta en la segmentación que deberá llevar a cabo más adelante.

Así pues, prepare un informe de 5 páginas justificando qué características ha considerado relevantes, por qué ha rechazado, si acaso, algunas otras (esto le llevará no más de una página) y describa, asimismo, en un máximo de 4 páginas adicionales y sin perder la perspectiva de la segmentación que más adelante llevará a cabo, las características de los vehículos que su jefe ha comprado.

*Continuará.*

## CÓDIGO UTILIZADO

[HTTPS://GITHUB.COM/JLOPEZGALDON/TARD/BLOB/MAIN/03 LOS COCHES DEL JEFE/RMD/01 LOS COCHES DEL JEFE.RMD](https://github.com/jlopezgaldon/tard/blob/main/03%20los%20coches%20del%20jefe/rmd/01%20los%20coches%20del%20jefe.Rmd)

## BASE DE DATOS

Tenemos un colección de 125 vehículos y 15 variables, estas son:

- **MARCA:** Marca del vehículo, siendo:
  - 1: ASIA MOTORS
  - 2: CHEVROLET
  - 3: DAHIATSU
  - 4: FORD
  - 5: JEEP
  - 6: KIA
  - 7: LADA
  - 8: LAND ROVER
  - 9: MERCEDES
  - 10: MITSUBISHI
  - 11: NISSAN
  - 12: OPEL
  - 13: SSANGYONG
  - 14: SUZUKI
  - 15: TATA
  - 16: TOYOTA
  - 17: UAZ
- **MODELO:** Modelo del vehículo en cuestión
- **PVP:** Precio de venta del vehículo
- **CILINDRO:** Número de cilindros
- **CC:** Centímetros cúbicos
- **POTENCIA:** Potencia del TT medida en cv
- **RPM:** Revoluciones por minuto
- **PESO:** Peso del vehículo en kg
- **PLAZAS:** Número de plazas del vehículo
- **CONS90:** Consumo en litros a 90km/h
- **CONS120:** Consumo en litros a 120km/h
- **CONSURB:** Consumo en ciudad
- **VELOCIDA:** Velocidad máxima
- **ACELERAC:** Aceleración de 0 a 100 km/h
- **ACL2:** Clasificación de los vehículos según su aceleración, siendo:
  - 1: Menor a 10 segundos
  - 2: Mayor a 10 segundos

## EDA

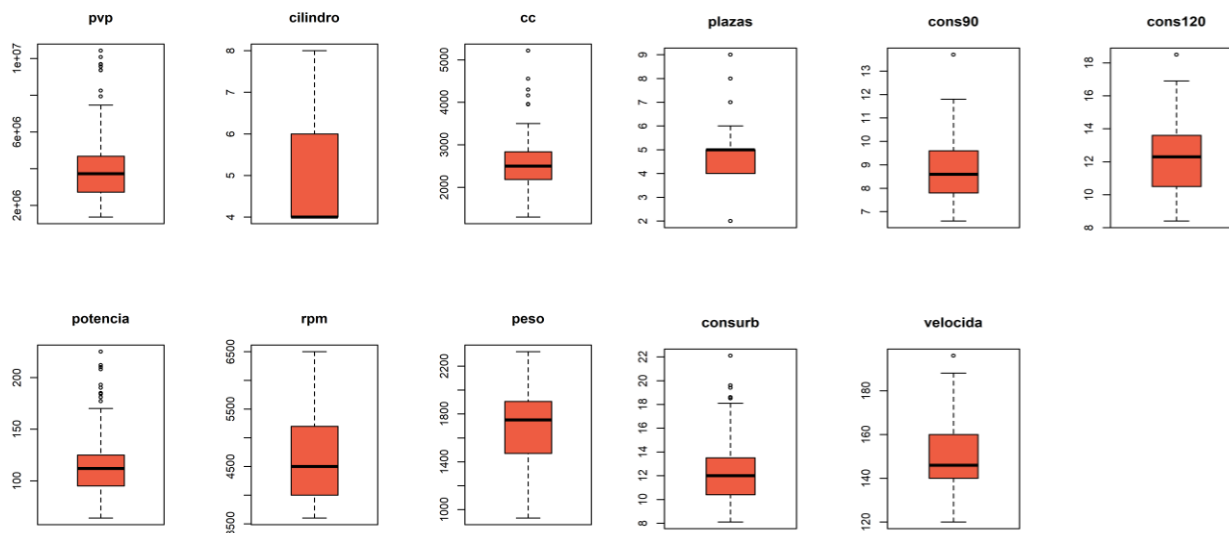
A continuación, realizaremos un análisis exploratorio de la base de datos para comprender mejor los datos.

Al realizar un resumen de los estadísticos principales podemos sacar las siguientes conclusiones:

- *Precio*: Son coches con un precio medio de 4.004.459 ptas, con una desviación de 1.936.691 ptas, por lo que los precios estarán comprendidos entre los 2 y 6 millones de ptas.
- *Motor*: La mayoría de vehículos montan motorizaciones comprendidas entre los 120 y 80 cv, con motores de 4 cilindros y cilindradas comprendidas entre los 2.000 y 3.000 centímetros cúbicos.
- *Capacidad*: Debido a que la mayoría de estos vehículos tienen 5 plazas, aunque algunos de ellos llegan hasta las 7 y ser vehículos pesados (ya que suelen superar la tonelada y media), los consumos consumos son relativamente elevados, hablamos de 12,5 l/100km en ciudad, y casi 9 litros por cada 100km en carretera.

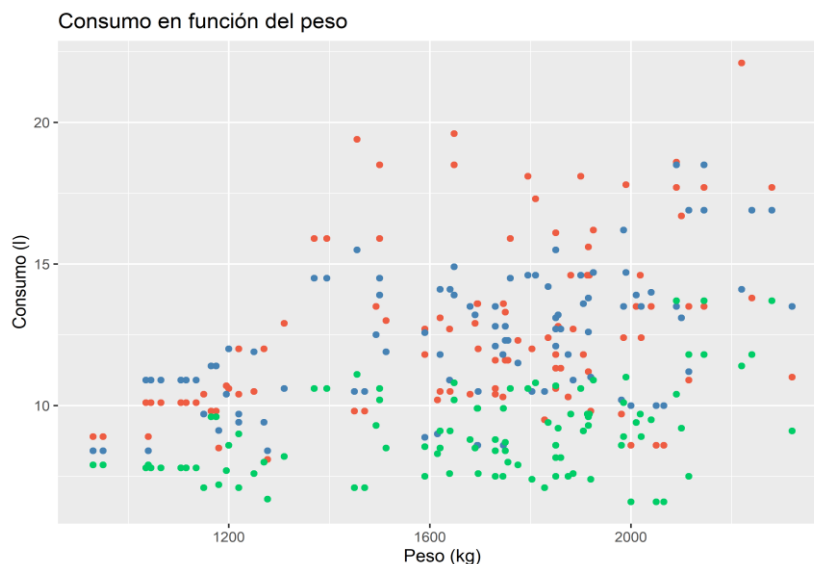
## Box-plot

Para comprender mejor las conclusiones extraídas anteriormente, realizaremos unos diagramas de cajas de las variables numéricas.



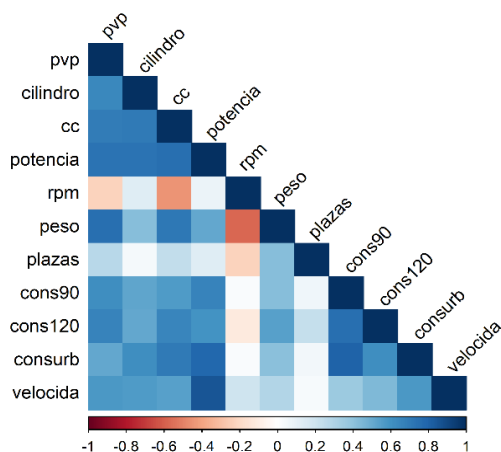
## Gráficos de dispersión

Vamos a estudiar los consumos de los vehículos en función de su peso



En el gráfico superior tenemos representados los consumos en función del peso. En rojo tenemos el consumo urbano, en verde el consumo a 90km/h, mientras que en azul tenemos el consumo a 120km/h, como habíamos comentado están correlacionadas las variables peso y consumo, ya que a más peso mayor consumo. Por lo que, el peso puede ser una buena variable clasificadora para los vehículos. Además, el consumo a 120km/h parece que se encuentra en la media entre el consumo a 90km/h y el consumo urbano, por lo que será un buen indicador del consumo medio.

## Correlaciones



Existen altas correlaciones entre los cc, la potencia y el número de cilindros, esto tiene sentido, ya que un vehículo de alta potencia suele estar asociado a motores de alto cubicaje y mayor número de cilindros, al tener mayor cubicaje el motor puede llevar a que el peso del vehículo aumente, esto produce que los consumos sean más elevados al igual que la velocidad máxima. Por lo tanto, podemos seleccionar la potencia del vehículo como componente explicador de los atributos del motor.

Por último, observamos que el número de plazas está bastante incorrelacionado con el resto de variables, salvo el peso, por lo que podría ser un buen atributo a la hora de clasificar los vehículos.

## CONCLUSIONES

Tras realizar el análisis exploratorio podemos concluir que los vehículos que ha adquirido el dueño del *family office* son las siguientes:

- Se trata de vehículos con un rango de precio entre los 2 y 6 millones de pesetas.
- Son vehículos con potencias comprendidas entre los 70 y 120cv, con motores de 4 cilindros en su mayoría y un cubicaje comprendido entre los 2.000 y 3.000 cc.
- Estos todoterrenos son pesados, ya que en media superan la tonelada y media. Además, la mayoría, presentan una capacidad de 5 plazas, llegando hasta las 9, lo que hace que la mayoría de ellos tengan una aceleración superior a los 10 segundos y baja velocidad punta.
- Son vehículos poco eficientes, ya que los consumos son bastantes elevados.

En cuanto a las variables que consideramos interesantes a la hora de clasificar los vehículos son:

- **PVP:** Nos ayuda a obtener una idea de la gama del vehículo (alta, media o baja)
- **POTENCIA:** Nos engloba todas las características relacionadas con el motor (cc, número de cilindros, aceleración y velocidad punta)
- **CONS120:** Tras analizar los consumos, podemos observar como el consumo en autovía es una buena medida para estimar el consumo medio de los TT, ya que se encuentra entre el consumo a 90km/h y el consumo urbano.
- **PLAZAS:** El número de plazas nos ayuda a entender la usabilidad del vehículo, así como características interesantes que nos están comprendidos en los datos pero que se pueden deducir con el número de plazas como el tamaño del vehículo.