

¿Qué coche de ocasión me compro?

Una aproximación con R

A group of people, mostly men in suits and a few women, are standing on a stage. Several of them are holding small, dark-colored awards or plaques. They are all smiling and looking towards the camera. The background is a light blue screen with some faint, illegible text. The overall atmosphere is professional and celebratory.

Antonio Vidal
Data Scientist en FINSA



@avidalvi



antoniovidalvidal

¿Cómo se busca
un coche de
ocasión?



Volkswagen Tiguan de segunda mano

Tenemos 484 coches Volkswagen Tiguan de segunda mano para esta búsqueda. 72 coches Volkswagen Tiguan son de particulares y 412 de concesionario. Una de las **mejores ofertas** de Volkswagen Tiguan de segunda mano puedes verla aquí.

Buscar

Provincia ▼

Precio:

Desde ▼

Hasta ▼

Kms:

Desde ▼

Hasta ▼

Cambio ▼

Puertas ▼

Combustible ▼

Potencia

Desde ▼

Hasta ▼

Año:

Desde ▼

Hasta ▼

Plazas ▼

Vendedor ▼

Color:



Tipo:



Pequeño



Mediano



Grande



Familiar



Monovolumen



Todoterreno



Deportivo



Comercial

Guardar búsqueda

Recibirás un aviso si hay anuncios nuevos de: Volkswagen + Tiguan

484 resultados para **Volkswagen Tiguan de segunda mano**

Ordenar por:

precio de menor a mayor ▼



Volkswagen Diesel
Tiguan 2.0TDI Advance 5p.
8.900 €

coche nacional, un solo propietario, muy cuidado tanto exterior como interior itv hasta 09/2019 se entrega revisado de todo. (mejor...

140 cv 220.000 km
2009 Sevilla



Volkswagen Diesel
Tiguan 2.0TDI Sport 5p.
8.990 €

aviso a navegantes nos acaba de entrar. nacional, estrenado en 12/07. en perfecto estado tanto por fuera como por dentro con...

140 cv 225.000 km
2007 Barcelona



Volkswagen Gasolina
Tiguan 2.0TSI +Motion 5p.
9.750 €

está impecable de plancha y pintura y interior como nuevo no fumador

170 cv 181.000 km
2008 Barcelona



Volkswagen Diesel
Tiguan 2.0TDI 5p.
9.900 €

coche con todas las revisiones pasadas en concesionario oficial (última revisión 21 de septiembre de 2018). itv hasta febrero...

140 cv 255.000 km
2010 Navarra



Volkswagen Diesel
Tiguan 2.0TDI Country 5p.
9.900 €



Volkswagen Diesel
Tiguan 2.0TDI +Motion 5p.
10.490 €

volkswagen tiguan 2.0 t.d.i 140 c.v 4 motion (4x4) del 2010 y 142.000 km. con climatizador bizona.

Publicidad

HONDA
The Power of Dreams

CIVIC 170€
ELEGANCE NAVI AL MES

Entrada 6.120,10 € · 36 cuotas 170 € · Cuota final 11.957 € · TAE 8,29%

AHORA CON 5 AÑOS DE MANTENIMIENTO GRATUITO

¿Qué
obtaines?

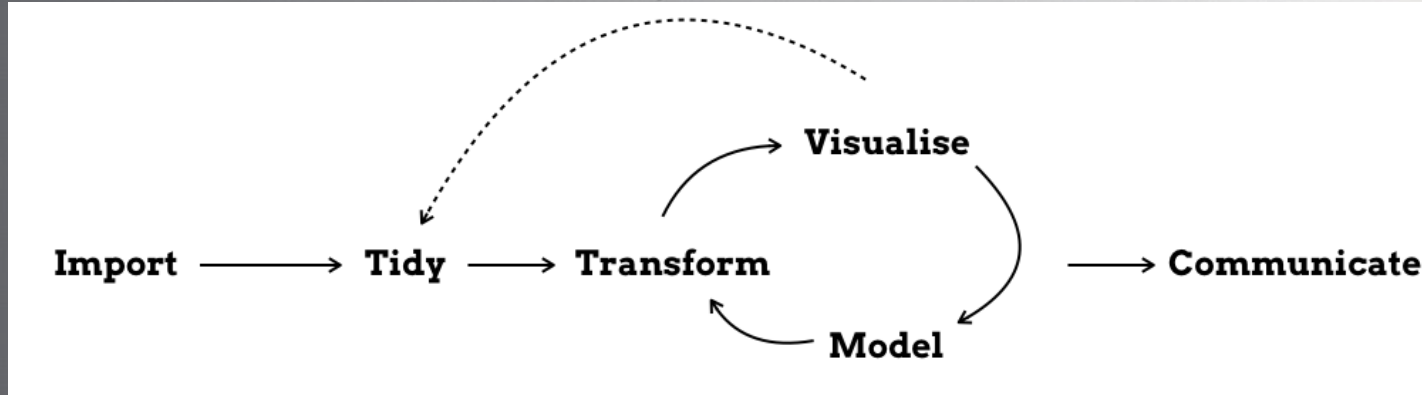


¿Podemos hacer
algo diferente con
estos datos y ML?





Proceso de análisis de datos (en R)



Basado en el libro "R for Data Science"

<http://r4ds.had.co.nz/>

Adaptado en:

<https://sinarueeger.github.io/2018/10/09/workflow/>



Entorno de desarrollo



Usando un entorno de desarrollo
y ejecución limpio

Ctrl+Shift+F10

```
# initialization -----  
setwd(here::here("R_tiguan_app"))  
rm(list = ls())
```





Cargar paquetes imprescindibles

tidyverse

```
> library(tidyverse)
-- Attaching packages ----- tidyverse 1.2.1 --
v ggplot2 3.0.0      v purrr  0.2.5
v tibble  1.4.2      v dplyr  0.7.4
v tidyr   0.8.0      v stringr 1.3.1
v readr   1.1.1      v forcats 0.3.0
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```


Capturando los datos: rvest

```
home.page <- try(html_session(url = page.url, encoding = "UTF-8",  
                             user_agent(kUserAgent), timeout(10)))
```

Volkswagen Tiguan de segunda mano

Tenemos 498 coches Volkswagen Tiguan de segunda mano para esta búsqueda. 72 coches Volkswagen Tiguan son de particulares y 426 de concesionario. Una de las mejores ofertas de Volkswagen Tiguan de segunda mano puedes verla aquí.

volkswagen tiguan,				Buscar	
Provincia	Precio: Desde	Hasta	Kms: Desde	Hasta	Cambio
Combustible	Potencia: Desde	Hasta	Año: Desde	Hasta	Puertas
Color:	Tipo:				

Guardar búsqueda

Recibirás un aviso si hay anuncios nuevos de: Volkswagen + Tiguan

498 resultados para Volkswagen Tiguan de segunda mano

Ordenar por: Seleccionar...



Volkswagen Diesel
Tiguan 2.0TDI +Motion 5p.

11.900 €

precio sujeto a financiación mínima de 10.000€. precio sin financiación 12900€ airbag, abs, cierre centralizado, control...



Volkswagen Diesel
Tiguan 2.0TDI BMT T1 5p.

16.600 €

abs, cierre centralizado, climatizador automático, control de distancia de aparcamiento frontal, control de distancia de aparcamiento...



Volkswagen Diesel
Tiguan 2.0TDI BMT T1 5p.

17.900 €

precio sujeto a financiación mínima de 10.000€. precio sin financiación 17900€ airbag, abs, cierre centralizado, control...



Volkswagen Diesel
Tiguan 2.0TDI Sport 5p.

31.590 €

precio sujeto a financiación mínima de 10.000€. precio sin financiación 31590€ airbag, abs, cierre centralizado, control...

SelectorGadget

.cc_precio

Clear (40)

Toggle Position

XPath

?

X



Capturando los datos: rvest

```
modelo <- html_nodes(x = captura, css = "strong.hidden-xs") %>% html_text() %>% trimws()
combustible <- html_nodes(x = captura, css = ".cc_card_data_hidden-xs+.hidden-xs") %>% html_text()
combustible <- combustible[seq(from = 1, to = length(combustible), by = 2)]
version <- html_nodes(x = captura, css = ".inherit_styles") %>% html_text()
version <- version[seq(from = 2, to = length(version), by = 2)] %>% trimws()
puertas <- html_nodes(x = captura, css = ".hidden-xs+strong") %>% html_text()
puertas <- ifelse(grepl(pattern = "5p.", puertas), 5, 3)
precio <- html_nodes(x = captura, css = ".cc_precio") %>% html_text() %>%
  gsub(pattern = "[\\0-9]", replacement = "") %>% as.integer()
descripcion <- html_nodes(x = captura, css = ".cc_line_height") %>% html_text()
CV <- html_nodes(x = captura, css = ".cc_card_data_tech span:nth-child(1)") %>% html_text() %>%
  gsub(pattern = "[\\0-9]", replacement = "") %>% as.integer()
km <- html_nodes(x = captura, css = ".cc_card_data_tech span:nth-child(2)") %>% html_text() %>%
  gsub(pattern = "[\\0-9]", replacement = "") %>% as.integer()
matriculacion <- html_nodes(x = captura, css = ".cc_card_data_tech span:nth-child(3)") %>% html_text() %>%
  gsub(pattern = "[\\0-9]", replacement = "") %>% as.integer()
provincia <- html_nodes(x = captura, css = ".cc_card_data_tech span:nth-child(4)") %>% html_text()
enlace <- html_nodes(x = captura, css = ".cc_carCard") %>% html_nodes(css = "a") %>% html_attr("href")
coches.com.id <- sapply(X = enlace, FUN = function(x){strsplit(x = x, split = "id=")[[1]][2]})
# km0 model
km0 <- grepl(pattern = "km0", x = enlace, ignore.case = TRUE)
```

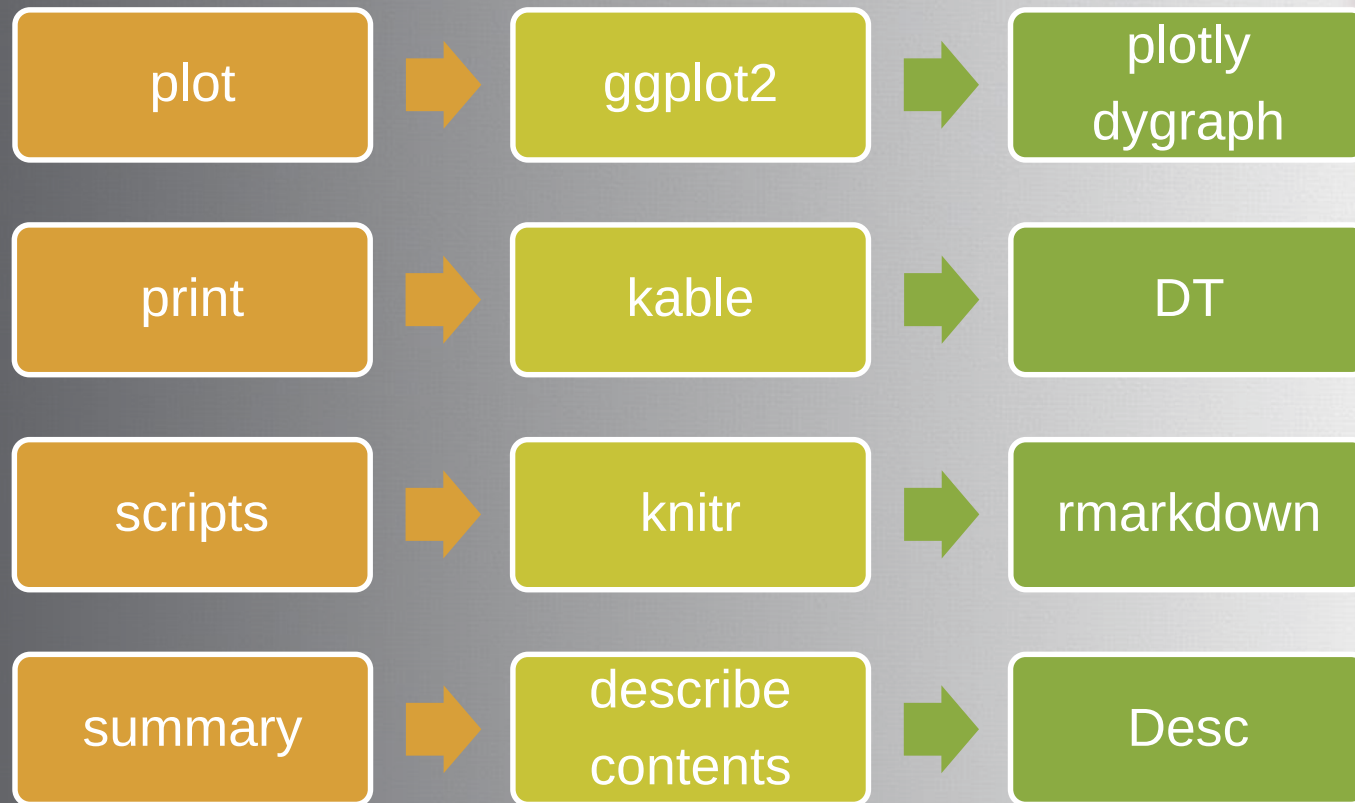
Programando scripts:

- taskscheduleR
- cronR

Tipos de datos

- Variables con el tiempo:
 - Km recorridos
 - Antigüedad
 - Precio
- Constantes respecto al tiempo:
 - Versión
 - Potencia
 - Combustible

Explorando datos



Datos en texto

tidytext:

procesado rápido de textos

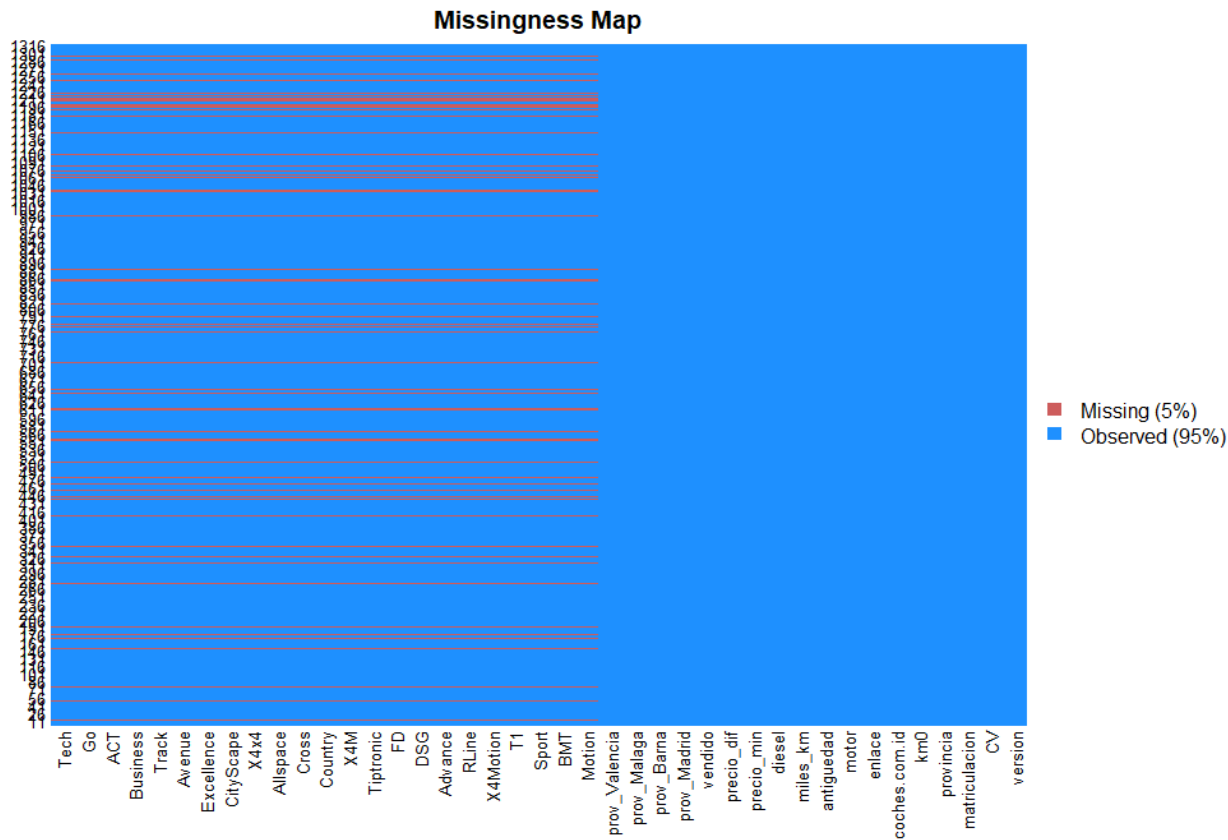


forcats:

gestionar factores



Explorando datos faltantes



Explicando el proceso con RMarkdown



Datos capturados del Tiguan

EXPLORACIÓN Y TRANSFORMACIÓN

Antonio Vidal

21/10/18

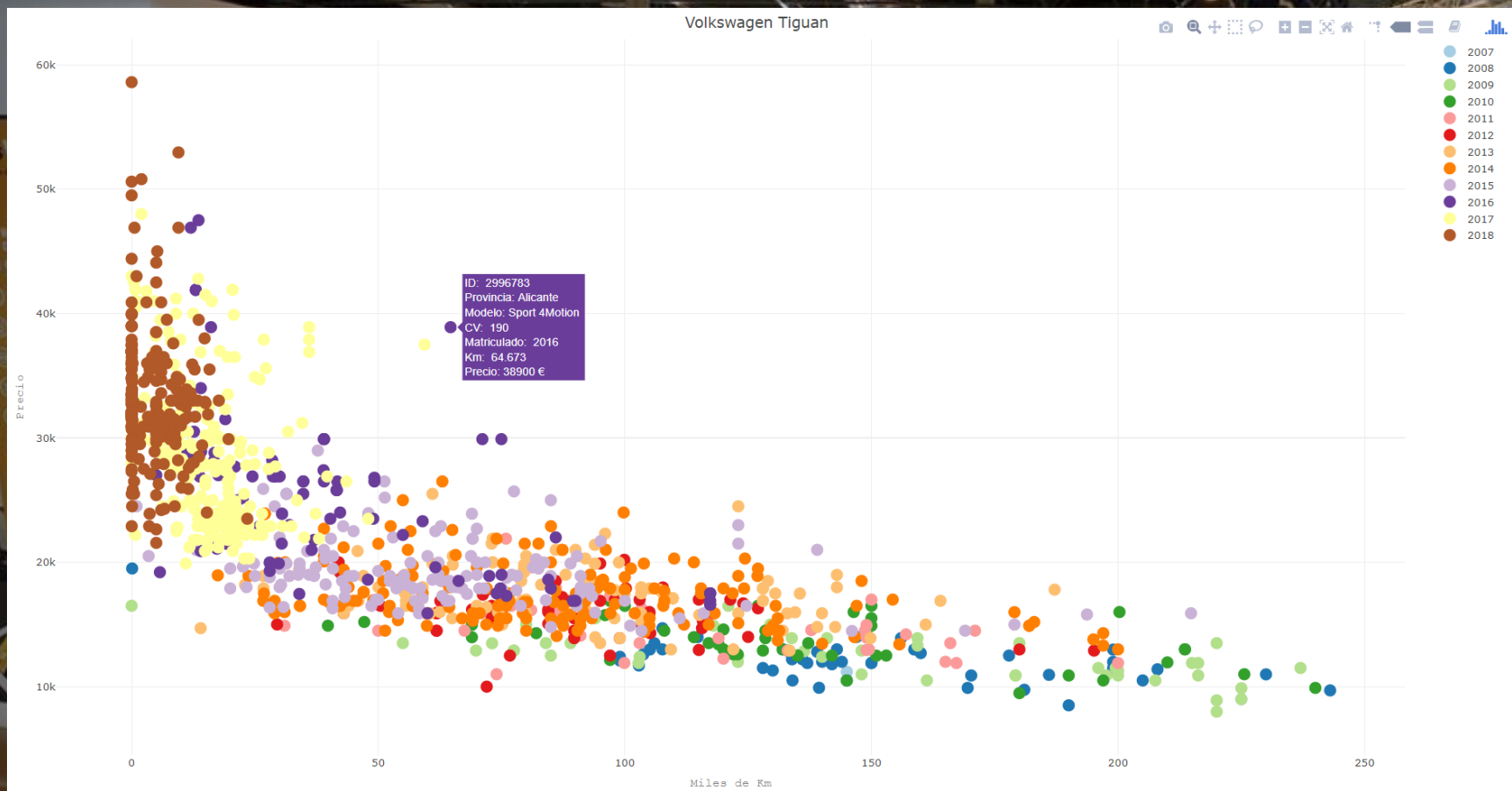
- 1 Objetivo
- 2 Datos originales
- 3 Preparación de los datos
 - 3.1 Variables que ya no contienen información
 - 3.2 Análisis de datos faltantes
 - 3.3 Visualización de los datos “variables”
- 4 Características de los datos
 - 4.1 Exploración visual de los datos

1 Objetivo

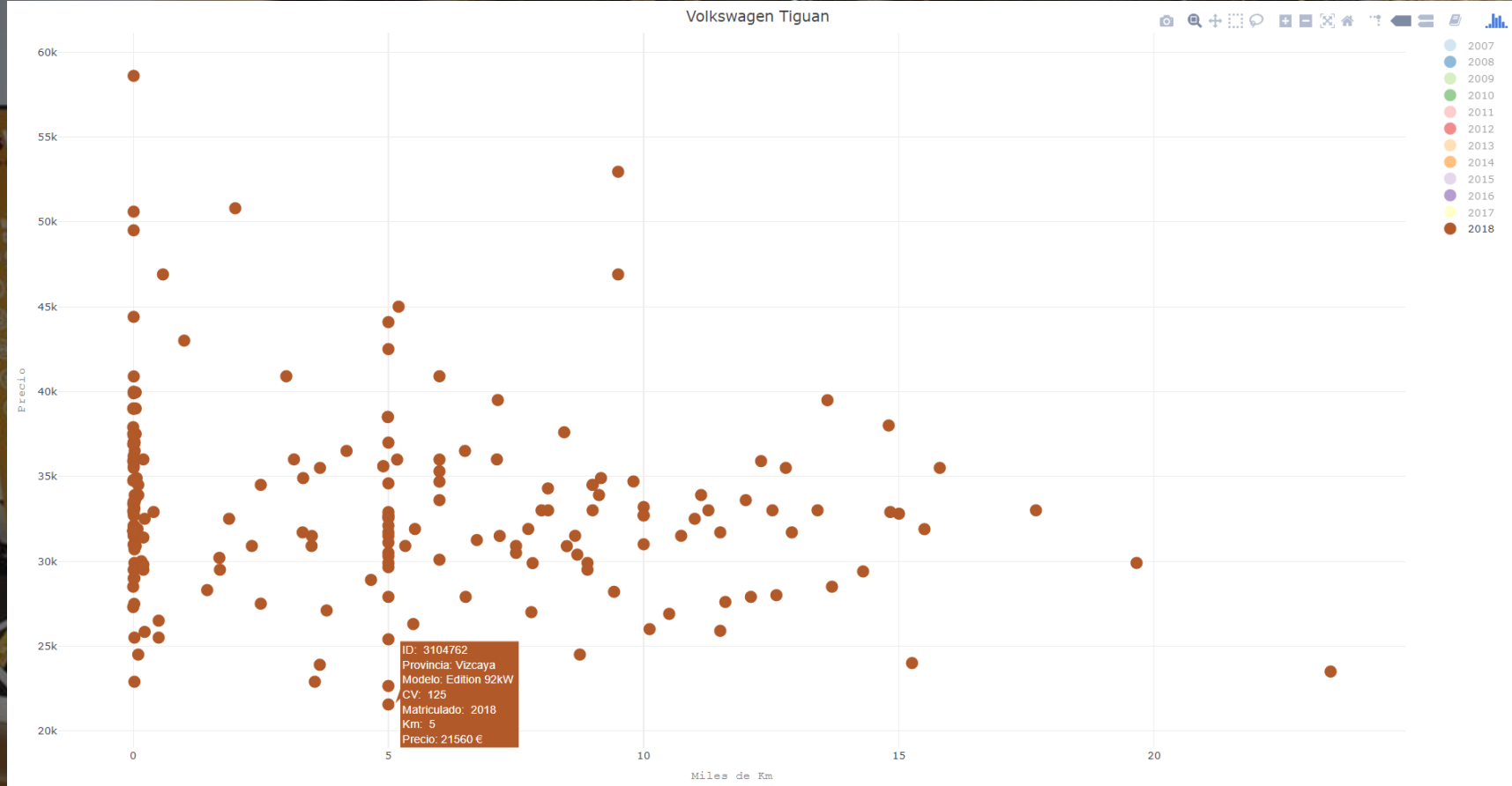
En este notebook se exploran los datos obtenidos del portal coches.com

El modelo de coche capturado es el **Volkswagen Tiguan**. La página inicial de la descarga es [esta](#)

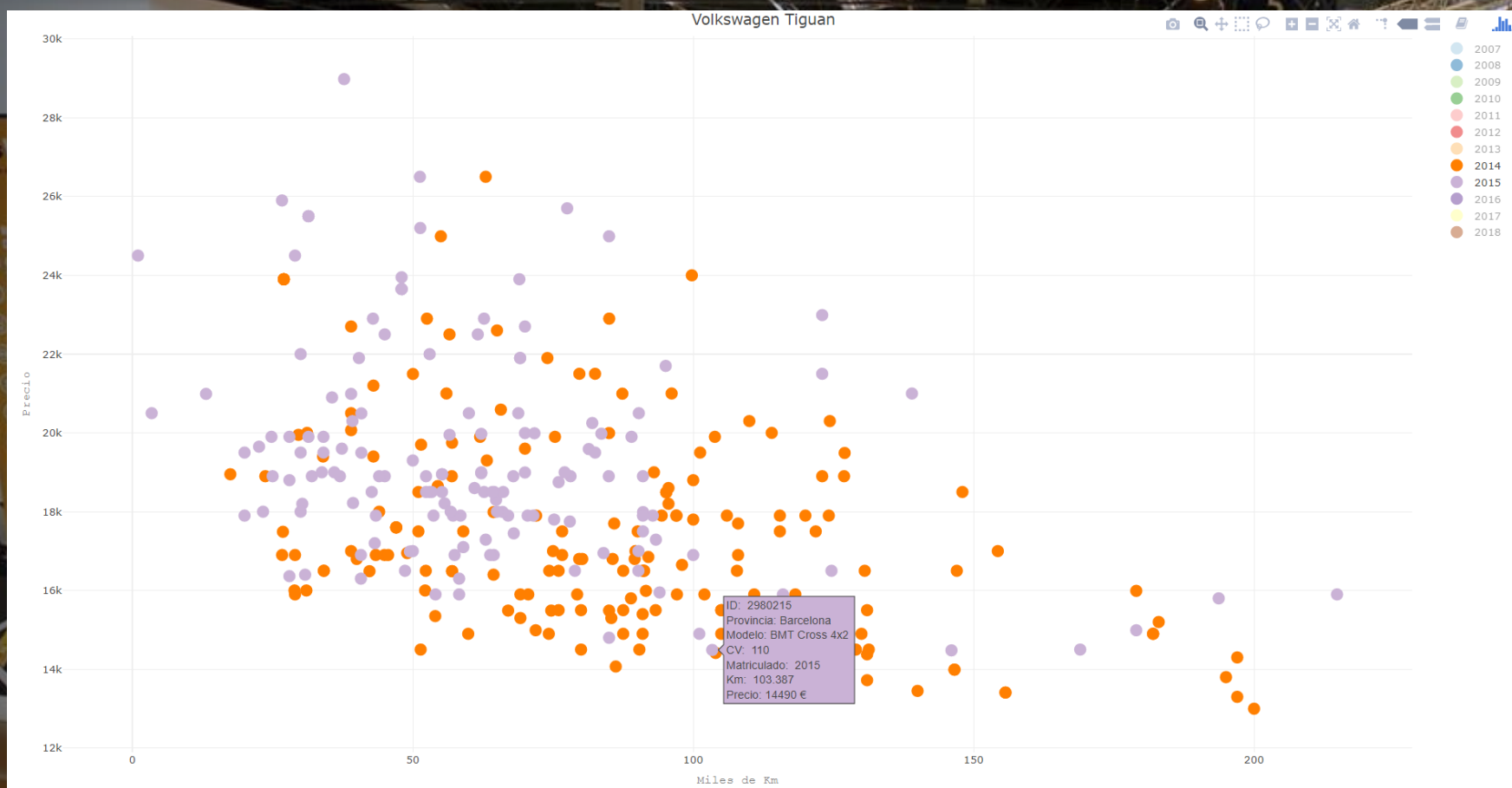
2 Datos originales



Mapa de coches



Coches del 2018

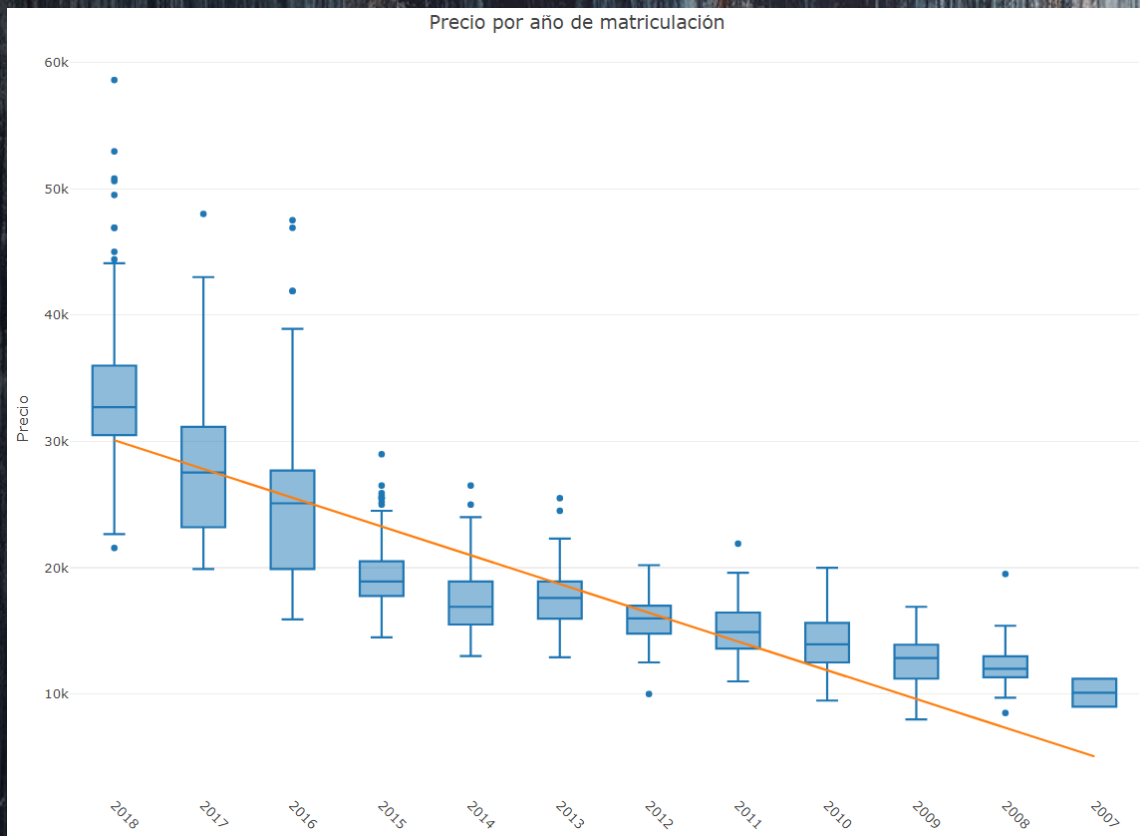


Coches del 2014-2015

¿Cuál es la
relación entre el
precio, los km y
la fecha de
matriculación?



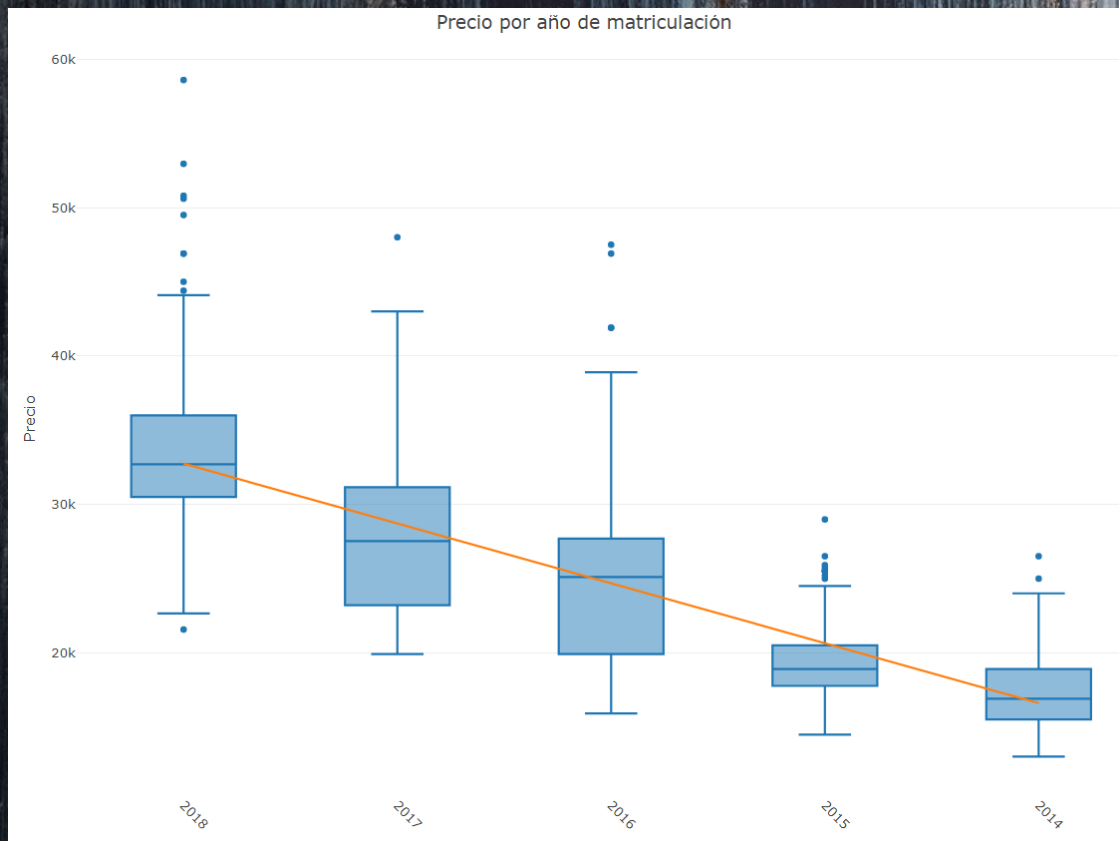
Relación entre precio y matriculación



La relación lineal indica que cada año los coches se deprecian 2275€

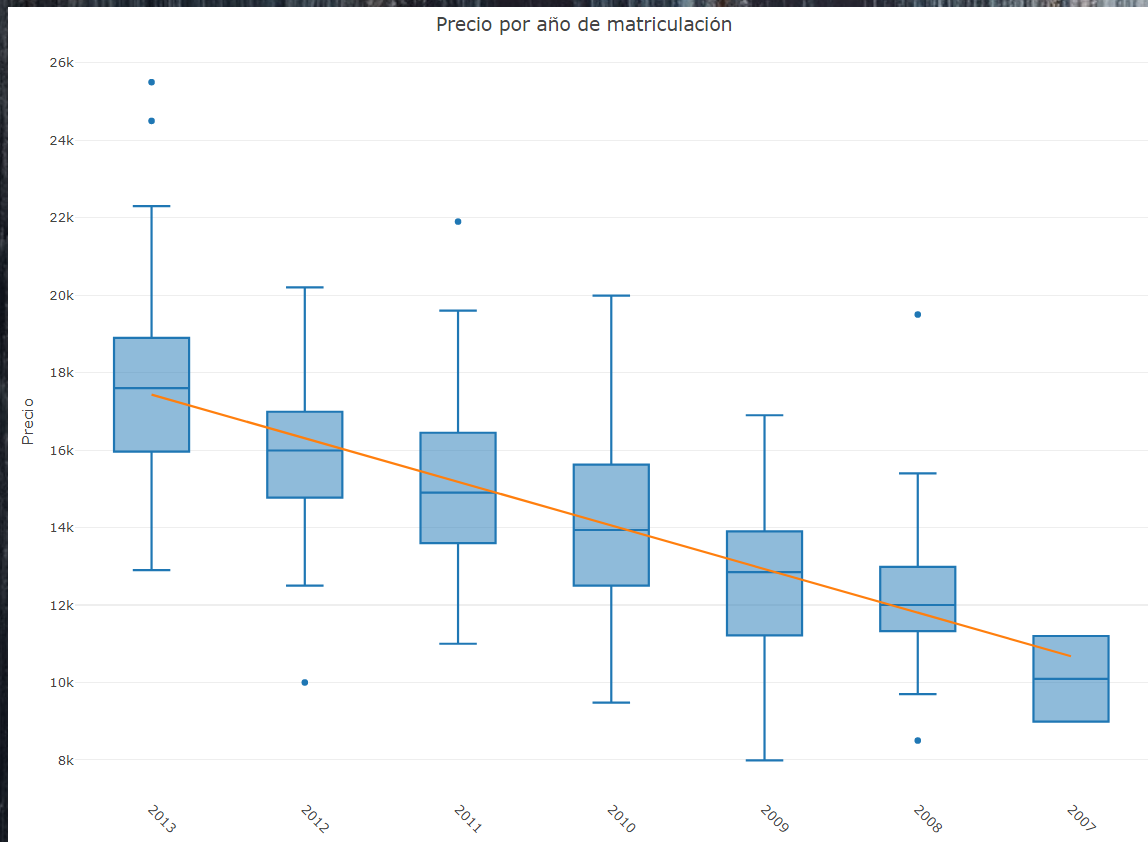


Relación entre precio y matriculación



En los primeros 5 años los coches se deprecian 4036€ por año

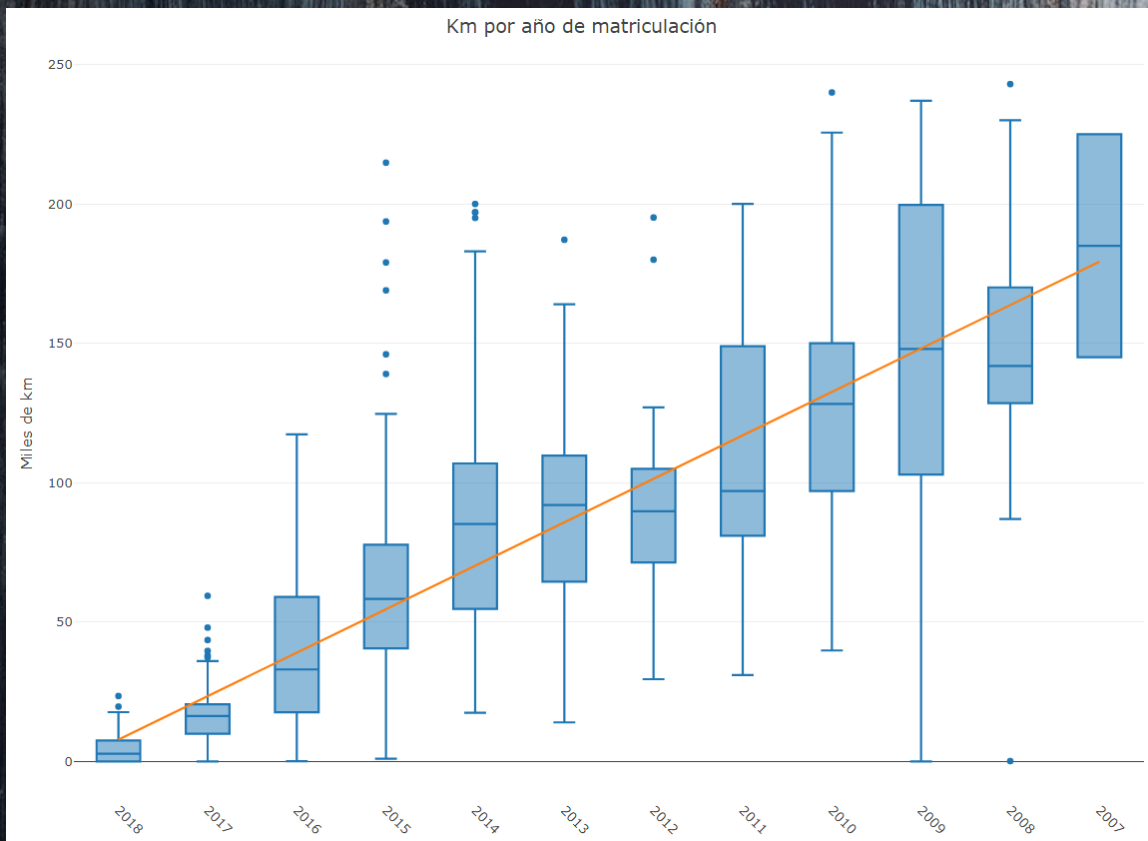
Relación entre precio y matriculación



A partir del 5º año
los coches se
deprecian 1126€
por año



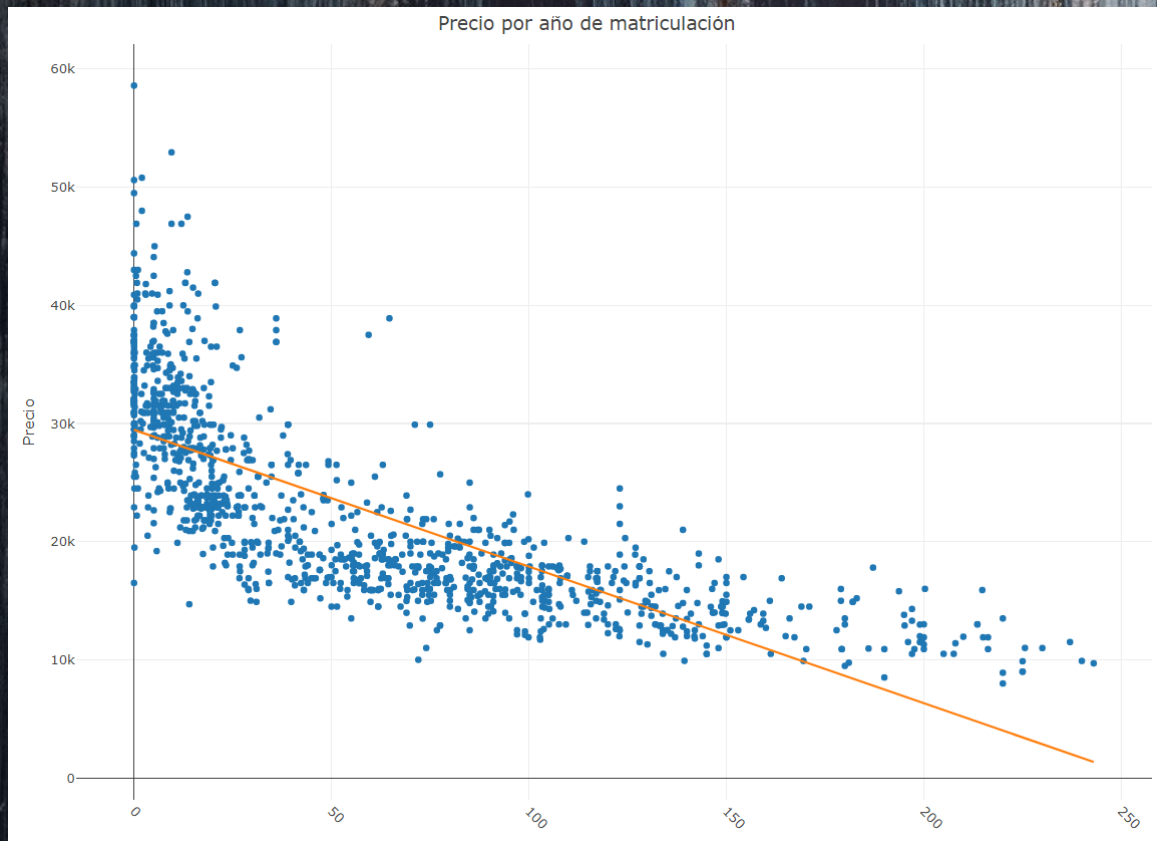
Uso medio por año



Estos coches
tienen una media
de uso de 15.000
km por año



Relación entre precio y km



Los coches se
deprecian 116€
por cada 1000km
recorridos
11,6c€ / km

Y ya tenemos nuestros primeros modelos ...

modelo	R2	MAE	RMSE	MAPE
lm-matriculacion	0.65	3925.04	5191.01	17.75
lm-km	0.60	4251.32	5544.05	19.36
lm-km y mat	0.69	3635.80	4883.23	16.26

Cómo mejorar estos resultados?

- Recogiendo más datos (otros portales)
- Incluyendo más variables
- Usando algoritmos más complejos





¿Cómo modelar rápido?

Usando un wrapper

- **caret** como interfaz de modelado
 - NZV, findLinearCombos
 - preparación de datos
 - definición como fórmula
 - PCA, ICA, centrado, escalado, Box-Cox, Yeo-Johnson, imputación
 - paralelización sencilla
 - búsqueda de hiperparámetros con varios tipos de validación
 - funciones de evaluación



¿Cómo modelar BBB?

H₂O:

- herramienta Java
- Implementación SOTA de algoritmos
 - Paralelos de forma nativa
 - Escalables en clústers de forma nativa
 - Aceleración GPU
- librería de acceso R/Python desarrolladas internamente
- GLM, RF, GBM, DL, XGBoost, K-Means, PCA, ...
- Validación cruzada
- AutoML
- Herramientas de exploración de resultados: PDP

H₂O.ai



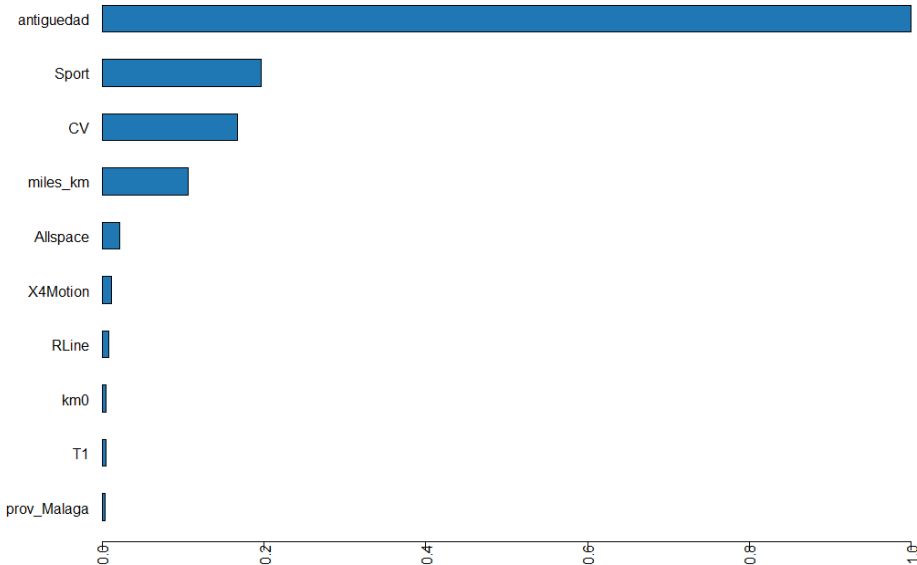
¿Cómo modelar BBB?

H₂O.ai

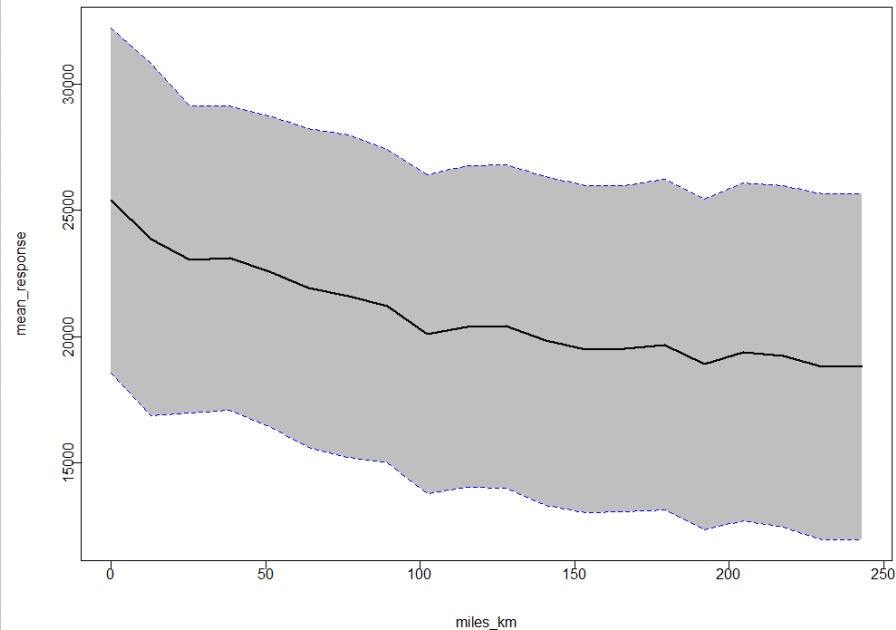


Importancia de las variables

Variable Importance: GBM

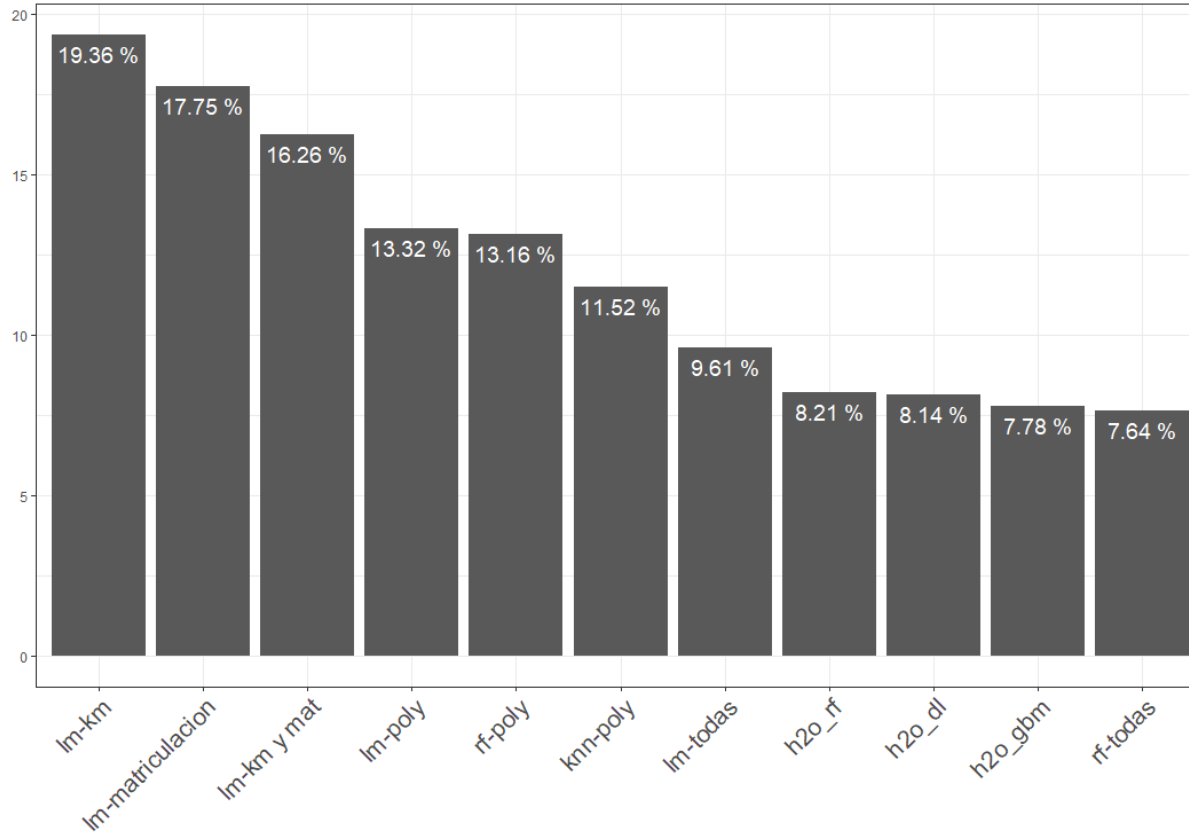


PDP del precio con los miles_km



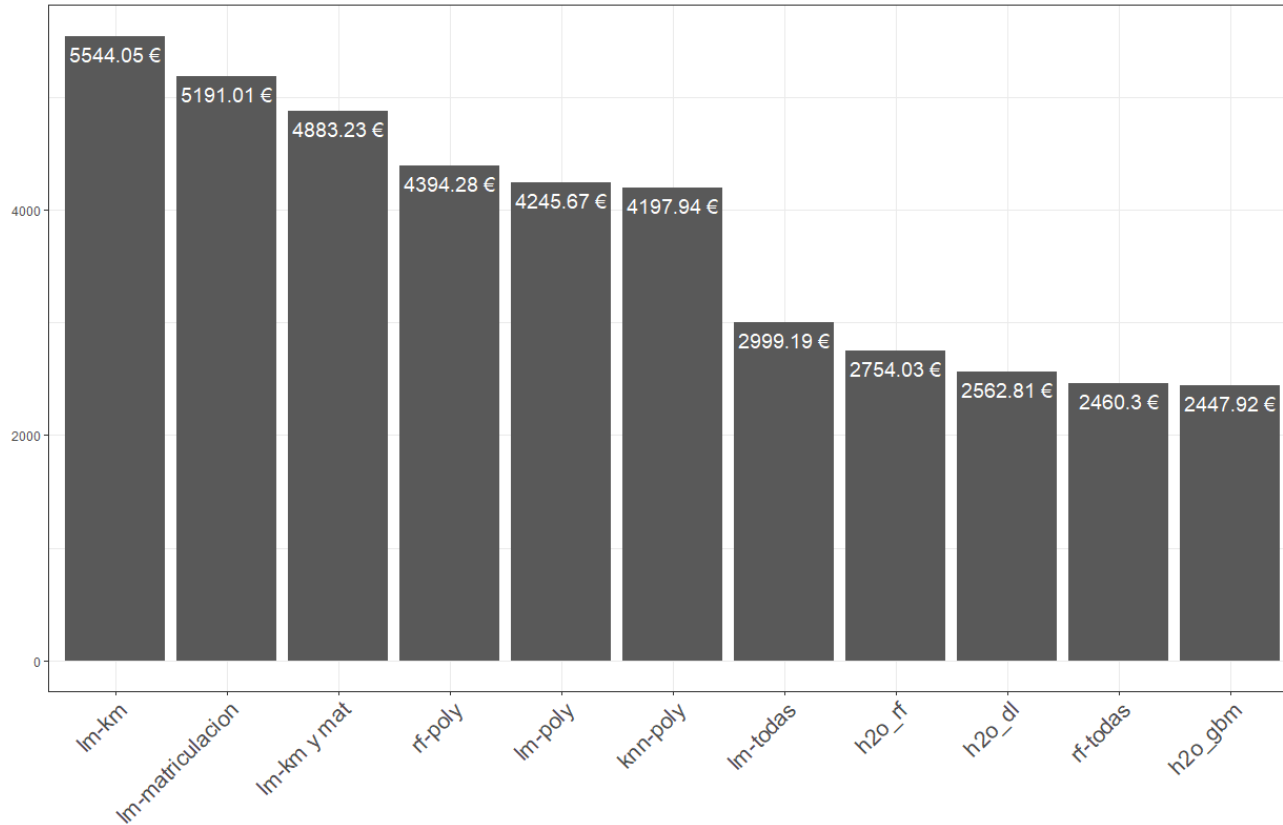
Hasta dónde se puede llegar?

MAPE de los diferentes modelos

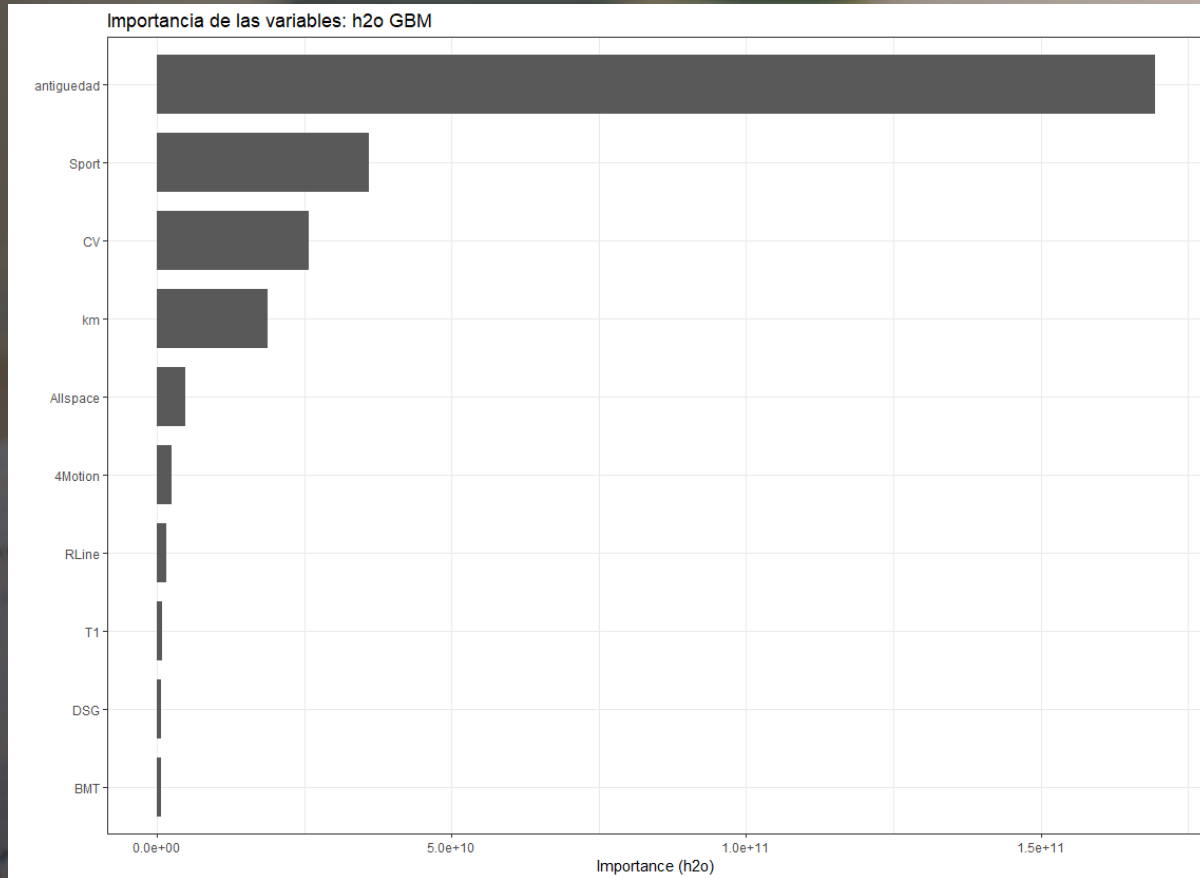


Hasta dónde se puede llegar?

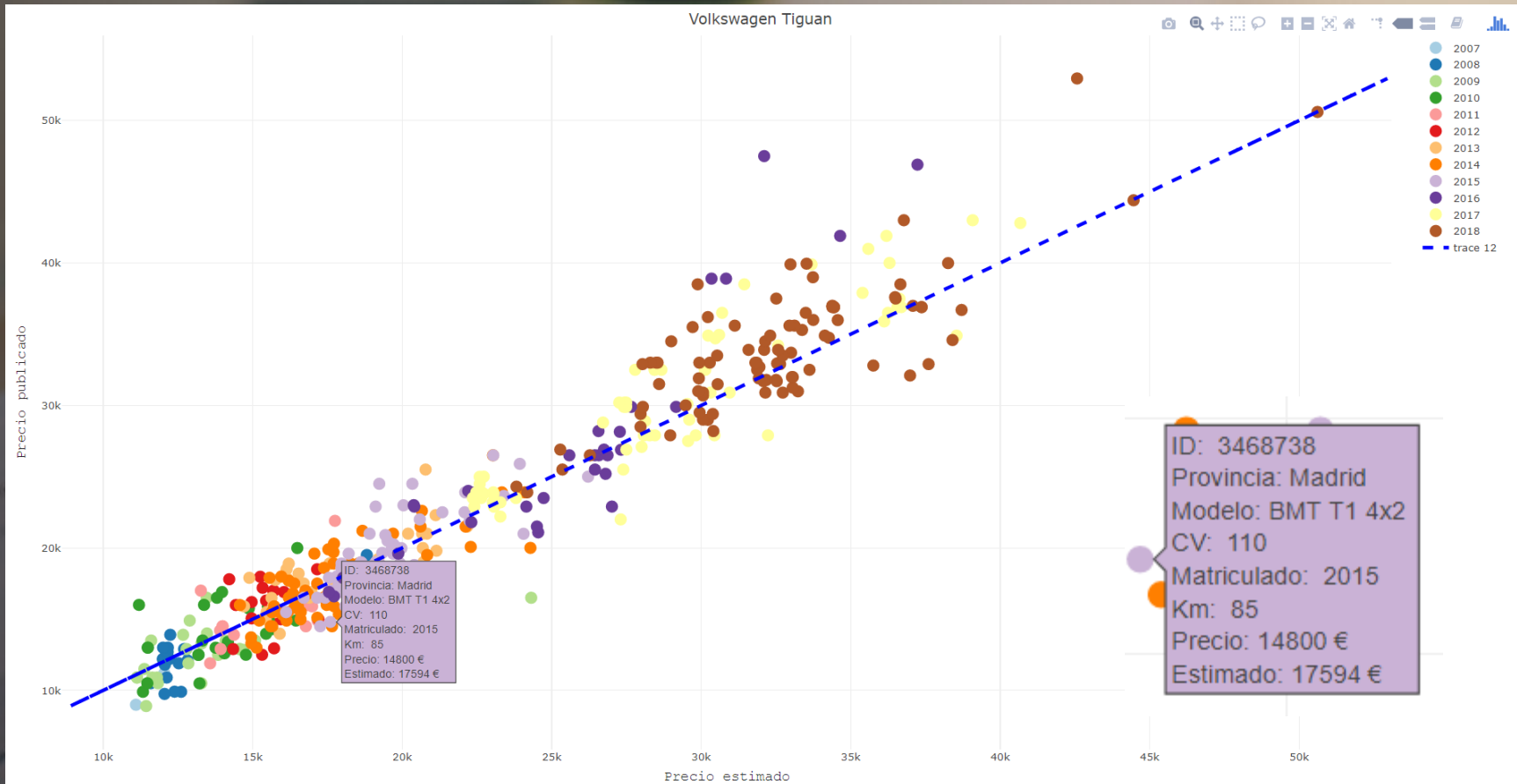
RMSE de los diferentes modelos



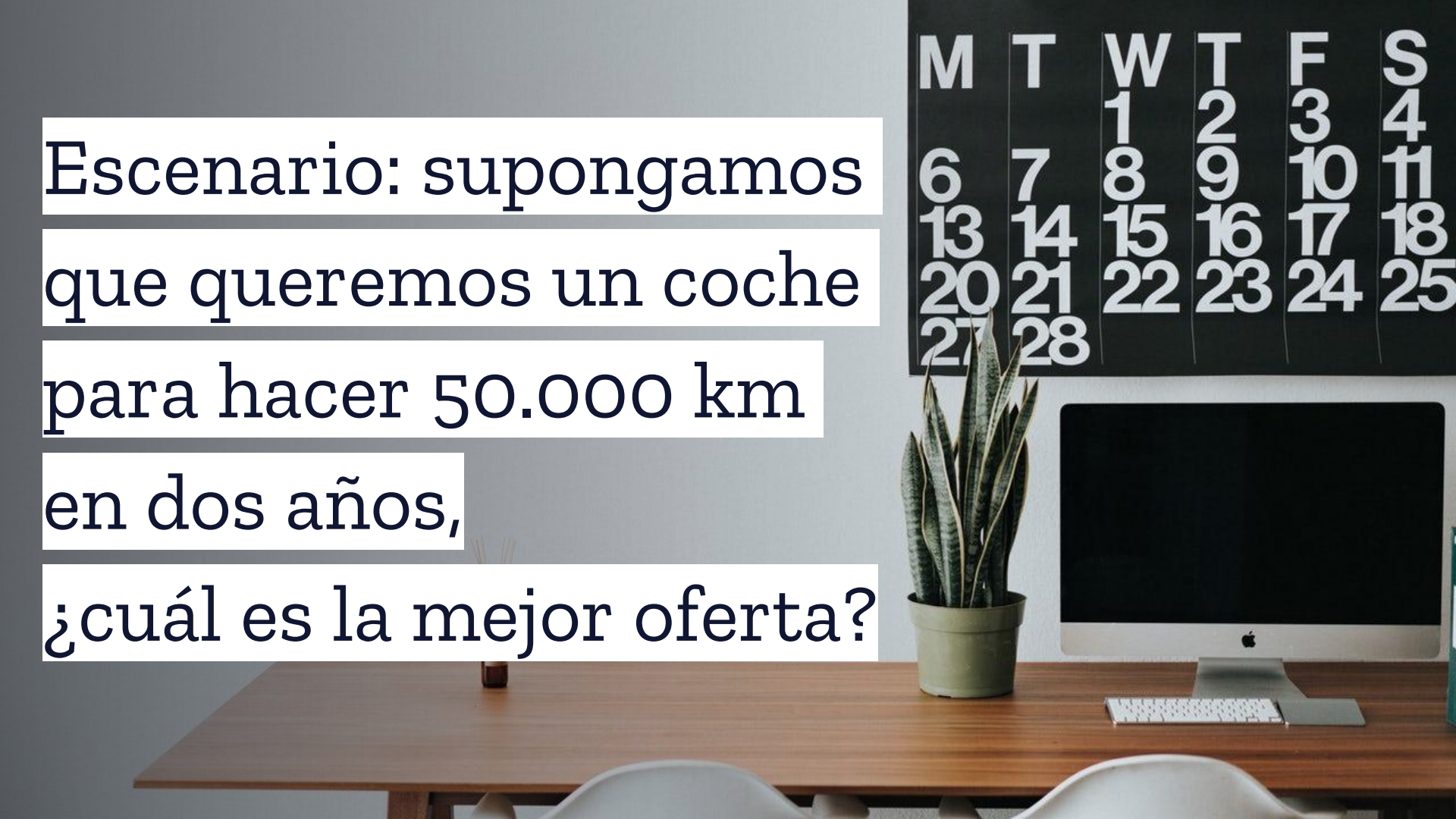
Importancia de las variables



¿Cuáles son las mejores ofertas?

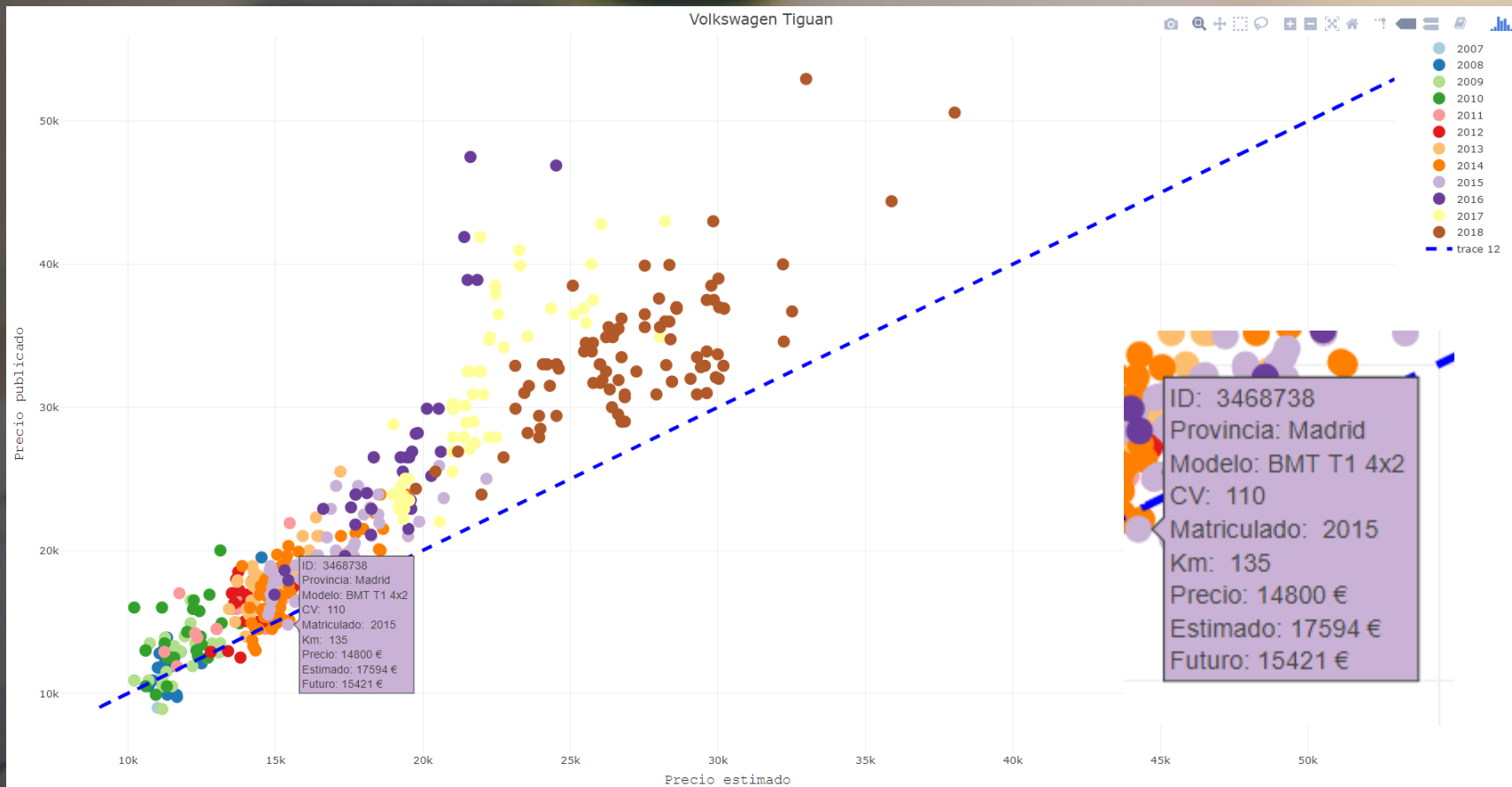


Escenario: supongamos
que queremos un coche
para hacer 50.000 km
en dos años,
¿cuál es la mejor oferta?



M	T	W	T	F	S
		1	2	3	4
6	7	8	9	10	11
13	14	15	16	17	18
20	21	22	23	24	25
27	28				

Simulando escenarios ...



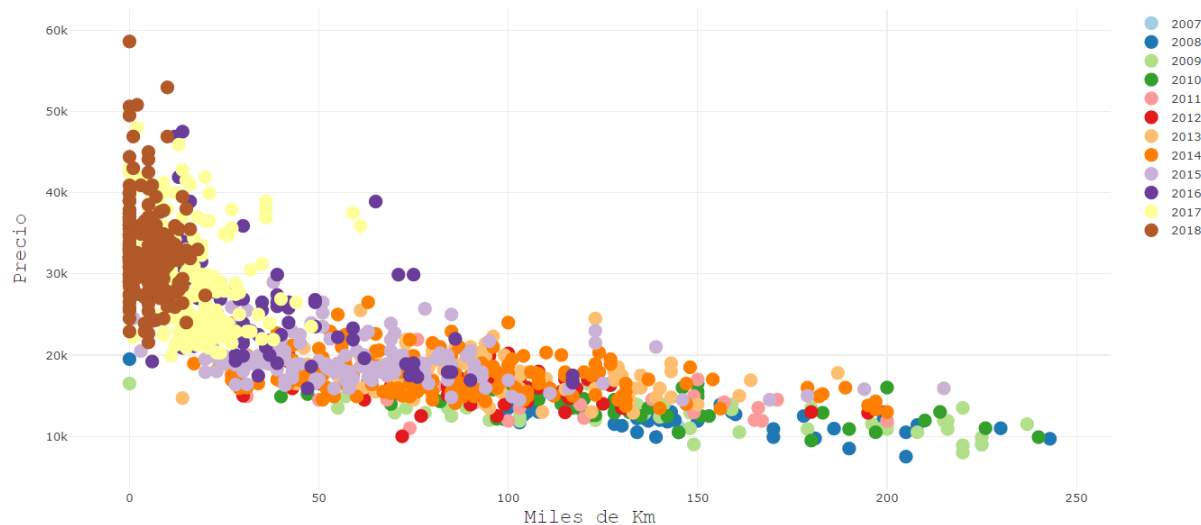
¿Convertirlo en un aplicación?



Volkswagen Tiguan



Exploración de los Tiguan capturados



Situación del coche

- ☒ Vendido
- ☒ A la venta

Último coche seleccionado

ID:

Modelo:

Motor:

CV:

Combustible:

Matriculación:

Km:

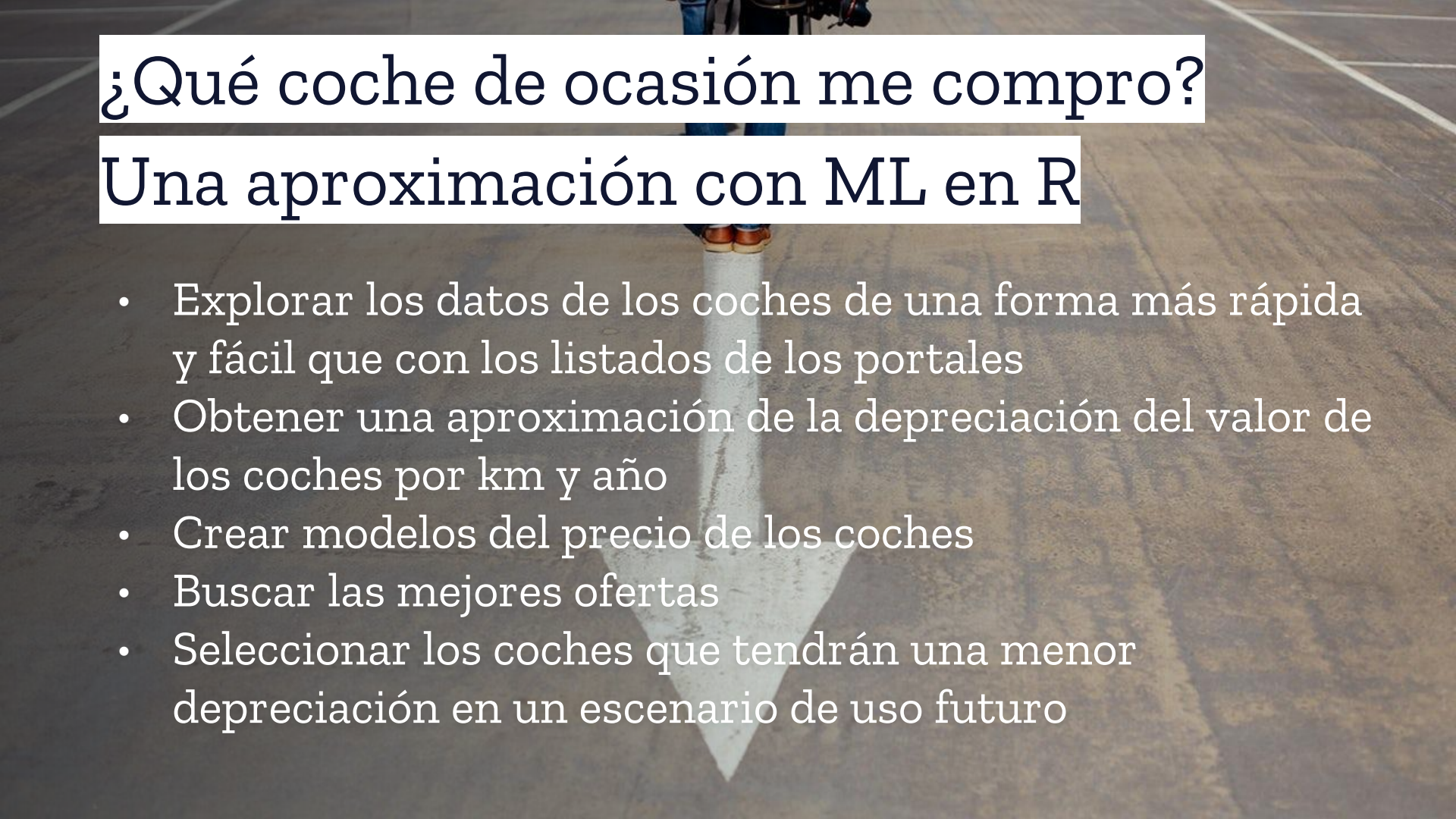
Precio:

Enlace: [Visitarlo](#)

Listado de coches

Show entries

Search:



¿Qué coche de ocasión me compro?

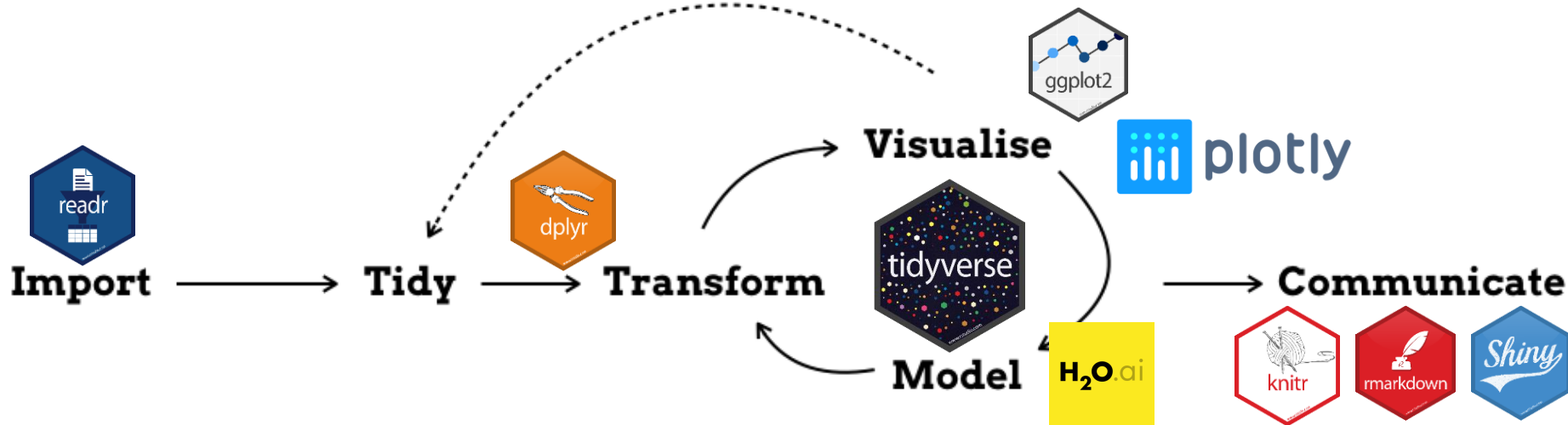
Una aproximación con ML en R

- Explorar los datos de los coches de una forma más rápida y fácil que con los listados de los portales
- Obtener una aproximación de la depreciación del valor de los coches por km y año
- Crear modelos del precio de los coches
- Buscar las mejores ofertas
- Seleccionar los coches que tendrán una menor depreciación en un escenario de uso futuro



¿Cuántos paquetes se han usado?

- Menos de 50
- Entre 50 y 100
- Más de 100



Muchas gracias!

Antonio Vidal
Data Scientist en FINSA



@avidalvi



antoniovidalvidal

