

Índice general

Índice de Figuras	IV
Índice de Tablas	V
Resumen	VII
Abstract	VIII
1. Introducción	1
2. Objetivos	2
2.1. Objetivo general	2
2.2. Objetivos específicos	2
3. Marco teórico	3
3.1. Mercados de ocasión y su importación	3
4. Metodología	6
4.1. Lenguaje y entorno integrado de desarrollo	7
4.2. R Markdown	8
4.3. ¿Por qué utilizar R markdown para escribir un trabajo de fin de grado? .	9
5. Obtención y descripción de la base de datos	10
5.1. Fuente de obtención de datos	10
5.2. Recolección de la información	13
5.3. Priorización de datos	13
6. Limpieza de la base de datos	19
7. Análisis exploratorio	24
8. Formulación de modelos OLS	33
8.1. Bondad del ajuste del modelo	34
8.2. Modelización del precio respecto a la edad	35

9. Modelos de regresión multivariable	39
9.1. Modelización del precio respecto a la edad y el kilometraje	39
9.2. Modelización del precio respecto a la edad y las marcas	41
10. Conclusiones	45
A. Apéndice A:	47
A.1. Código de las librerías utilizadas en R	47
A.2. Creación del diagrama	47
A.3. Código del WebScraping en R	48
B. Apéndice B: Códigos implementados	50
B.1. Código del Fichero: capitulo05.Rmd	50
B.2. Código del Fichero: capitulo06.Rmd	51
B.3. Código del Fichero: capitulo07.Rmd	52
B.4. Código del Fichero: capitulo08.Rmd	53
B.5. Código del Fichero: capitulo09.Rmd	55
C. Apéndice C: Modelos de regresión OLS	57
C.1. Modelos del precio respecto a la edad	57
C.2. Modelos del precio respecto a la edad, km y la marca	59
D. Apéndice D: Objetivos de desarrollo sostenible	63
Bibliografía	70

Índice de figuras

4.1. Modelo metodológico	6
4.2. Posible formatos estáticos y dinámicos de salida.	9
5.1. Páginas web Grupo Vía Mobilis.	10
5.2. Servicios especializados del grupo Vía Mobilis.	11
5.3. Marcas de vehiculos ofertados en la páginas web.	13
5.4. Ejemplo de visualización de código y formato de página web.	14
5.5. Visualización herramienta Selector Gadget.	15
5.6. Visualización de la informacion que se encuentra en la página principal. .	15
5.7. Visualización de la informacion caracteristicas de las cabezas tractoras. .	16
5.8. Visualización de código y formato CSS de la página web.	16
6.1. Representación gráfica de los datos con outliers.	19
6.2. Representación gráfica de los datos con outliers.	20
6.3. Representación gráfica de los datos posterior a la aplicación del método de diferencia de rango intercuartílico.	21
6.4. Representación gráfica de los datos sin outliers posterior a la aplicación del método MAD.	22
6.5. Representación gráfica de los datos sin outliers posterior a la aplicación del método MAD.	22
7.1. Datos característicos de las diferentes marcas de cabezas tractoras.	27
7.2. Datos característicos de las diferentes países que ofrecen las cabezas tractoras.	28
7.3. Número de las cabezas tractoras según el país de venta y marca.	29
7.4. Gráfico de dispersión de los precios de las diferentes cabezas tractoras respecto al tramo de edad al que pertenece.	30
7.5. Representación de las potencias de los distintos grupos de cabezas tractoras.	31
7.6. Representación de la correlación respecto a la edad, kilometraje y precio.	32
8.1. Diferentes modelos de regresión del precio respecto a la edad.	36

8.2. Intervalo de confianza de todos los modelos de los diferentes modelos de regresión lineal.	37
8.3. Gráfico de dispersión de los datos de las cabezas tractoras a analizar y su respectivos modelos de regresión.	38
9.1. Modelización del precio respecto a la edad y el kilometraje.	40
9.2. Regresión multivariadas del precio respecto a los años y la marca.	41
9.3. Diferentes modelos de regresión lineal multivariada.	43
D.1. Objetivos de Desarrollo Sostenible.	64

Índice de tablas

5.1.	Número de celdas de cada variable sin valor (NA)	17
5.2.	Valores estadísticos de la base de datos bruta.	17
7.1.	Valores estadísticos de las variables a estudiar posterior a la implementación del método MAD.	24
7.2.	Número de observaciones marca.	25
7.3.	Número de cabezas tractoras según el país en que se encuentra.	25
7.4.	Número de cabezas tractoras por años que han transcurrido desde su primera matriculación.	25
8.1.	Ecuaciones de las diferentes modelizaciones	34
9.1.	One-Hot-encoding según la marca de cabeza tractora.	42
C.1.	Modelo lineal del precio respecto a la edad.	57
C.2.	Modelo exponencial del precio respecto a la edad.	58
C.3.	Modelo potencial del precio respecto a la edad.	58
C.4.	Modelo lineal del precio respecto a la edad, kilometraje y marca.	59
C.5.	Modelo exponencial del precio respecto a la edad, kilometraje y marca.	60
C.6.	Modelo potencial del precio respecto a la edad, kilometraje y marca.	61

Glosario de términos

Glosario de términos

- (CSS): Hoja de estilo en cascada
- (Ganvam): Grupo Autónomo Nacional de Vendedores de Automóviles, Camiones y Motocicletas
- (GUI): Interfaz de usuario gráfica
- (HTML): Lenguaje de Marcas de Hipertexto
- (IDE): Entorno de desarrollo integrado
- (IQR): Rango intercuartílico
- (MAD): Desviación absoluta media
- (NaN): No un número
- (OLS): Regresión de mínimos cuadrados ordinarios
- (Outliers): Valores atípicos
- (RSE): Error estándar de los residuos
- (Web scraping): Raspado web

Glossary of terms

- (CSS): Cascading Style Sheets
- (GUI): Graphical user interface
- (HTML): HyperText Markup Language
- (IDE): Integrated development environment
- (IQR): Interquartile range
- (MAD): Median absolute deviation
- (NaN): Not a number
- (OLS): Ordinary least squares regression
- (RSE): Residual standard error

Resumen

En este trabajo final de grado se busca analizar el comportamiento de los valores y depreciación en el mercado de la maquinaria de ocasión, específicamente el mercado de cabezas tractoras. Para ello, se empleó el lenguaje de programación R mediante el que se pretende alcanzar una técnica interactiva para la recolección de la información y construcción de modelos OLS de regresión.

El cálculo de la depreciación de la maquinaria con fines valorativos suele hacerse asumiendo una depreciación lineal a imitación de la amortización lineal comúnmente empleada en el ámbito contable y fiscal. Sin embargo, este patrón de depreciación puede no ser adecuado en todas las ocasiones.

En aquellos casos con un mercado secundario es posible contrastar si otro tipo de modelos de depreciación reflejan mejor el comportamiento del valor. Actualmente, existe una gran cantidad de información disponible en internet, principalmente de activos en mercados secundarios, que permiten obtener los datos para poder estudiar el comportamiento de el precio de estos activos en función de diversas variables.

En este trabajo se emplea una técnica para la recolección de la información conocida como WebScraping, la cual es capaz de obtener datos específicos, precisos, fiables y ajustados a las necesidades de cada caso de forma automatizada. Posteriormente se lleva a cabo el tratamiento de la base de datos bruta, para eliminar aquellos valores que se desvien mucho de las observaciones y generen valoraciones incorrectas.

Con el fin de obtener una visión general representativa de la depreciación que llegan a sufrir estos activos, se desarrollan diversos modelos de regresión mínimo cuadrática (lineal, exponencial y potencial) y se relaciona la antigüedad de estos activos con su valor y a su vez, se relaciona el efecto que puede llegar a tener la marca en el modelo de la cabeza tractora o el país de donde provenga la maquinaria.

Palabras claves

Cabezas tractoras, Depreciación, Mercado de ocasión, Modelos de regresión, R, Valores atípicos, WebScraping.

Abstract

In this final degree work we seek to analyze the behavior of the values and depreciation in the used machinery market, specifically the tractor heads market. For this purpose, the R programming language will be used to achieve an interactive technique for the collection of information and construction of OLS regression models.

The calculation of machinery depreciation for valuation purposes is usually made assuming straight-line depreciation in imitation of the straight-line depreciation commonly used in the accounting and tax fields. However, this depreciation pattern may not be appropriate in all cases.

In those cases with a secondary market, it is possible to test whether other types of depreciation models better reflect the behavior of the value. Currently, there is a large amount of information available on the Internet, mainly of assets in secondary markets, which allows to obtain the information to study the behavior of the price of these assets as a function of several variables.

In this work we use a technique for the collection of information known as WebScraping, which is capable of obtaining specific, accurate, reliable and adjusted to the needs of each case in an automated way. Subsequently, the raw database is processed to eliminate those values that deviate greatly from the observations and generate incorrect valuations.

In order to obtain a representative overview of the depreciation that these assets undergo, various minimum quadratic regression models (linear, exponential and potential) are developed and the age of these assets is related to their value and, in turn, the effect that the make and model of the tractor unit or the country of origin of the machinery may have is related.

Keywords

Depreciation, OLS regressions, Outliers, R, Secondary market, Tractor heads, WebScraping.

Capítulo 1

Introducción

Gestionar los recursos limitados puede suponer para empresas mejoras a nivel táctico, estratégico, operativo, ejecutivo, etc... pudiendo orientarlas a la obtención de mejores resultados. Con el pasar de los años, principalmente por las crisis económicas, ha aumentado la importancia en la adquisición de estos recursos, llegando a tener una gran influencia en el sector del transporte y más concretamente del transporte de mercancías por carreteras y en el sector de la construcción, por la caída en la demanda de operaciones de este tipo de servicios, provocando una fuerte recesión en los ingresos de las empresas del sector.

Por esa razón muchas de las empresas al no poder mantener sus operaciones y/o poder costear maquinarias para continuar sus actividades económicas, se vieron obligadas a recurrir a mercados secundarios. Siendo vital para las empresas el estudio económico profundo considerando aquellos posibles elementos condicionantes en sus operaciones.

No obstante, resulta complejo llevar a cabo estos estudios y poder llegar a explicar el funcionamiento y estado de la economía, debido a que, como indica ([León Merán et al., 2019](#)) el sistema económico es una red ampliamente compleja en la que existen una infinidad de variables que pueden condicionarla.

Es por ello que en este trabajo de fin de grado se pretende analizar la importancia que pueden llegar a tener diversas variables en la depreciación en el mercado de la maquinaria de ocasión, específicamente el mercado de cabezas tractoras. La selección de este mercado se produjo debido a la existencia de una extensa cantidad de cabezas tractoras ofrecidas en este mercado y, el acceso a una gran cantidad de información permite obtener una valoración mas fiable y robusta de estos activos a lo largo de su vida útil.

Sin embargo, resulta tedioso llevar a cabo la recolección de grandes cantidades de información, por lo que se emplea la técnica WebScraping con el lenguaje de programación R, al ser una técnica iterativa que facilita la extracción de manera fiable y precisa.

Una vez construida la base de datos se establecen diferentes modelos de regresión de mínimos cuadrados ordinarios que reflejen una visión general de la situación actual y pasada de estos activos, y permiten llevar a cabo el análisis del comportamiento del valor de las cabezas tractoras.

Capítulo 2

Objetivos

2.1. Objetivo general

El objetivo de este trabajo es analizar el comportamiento de los valores y depreciación de cabezas tractoras en el mercado secundario de maquinaria, para el período comprendido entre los meses de junio a diciembre del año 2022, desarrollando una técnica interactiva para la recolección de la información y modelos de regresión, considerando elementos condicionantes y criterios de elección.

2.2. Objetivos específicos

1. Seleccionar un método que permita llevar a cabo la recolección de los datos específicos, precisos, fiables y ajustados a las necesidades de cada caso de forma automatizada.
2. Automatizar la extracción de información de la fuente para lograr trabajar con agilidad la data histórica y la data actual.
3. Aprender a gestionar bases de datos, analizar el comportamiento de los valores y la depreciación en el mercado de la maquinaria de ocasión.
4. Extraer información de las bases aplicando gráficos descriptivos y modelos sencillos a través del lenguaje de programación R con el entorno RStudio.

Capítulo 3

Marco teórico

3.1. Mercados de ocasión y su importación

La crisis económica ha afectado a muchos sectores, llegando a tener una gran influencia en el sector de la construcción, teniendo como consecuencia una caída en la demanda de este tipo de servicios, provocando una fuerte recesión en los ingresos de las empresas del sector. Es por ello que muchas de las empresas al no poder mantener sus operaciones y/o poder costear maquinarias para continuar sus actividades económicas, se vieron obligadas a recurrir a mercados secundarios.

Como nos indica ([Ortí Lahoz et al., 1974](#)) estos mercados se hallan vinculados fuertemente tanto al nivel de renta de los demandantes, como a la existencia del mercado paralelo de bienes de primera mano. Ya que, a pesar de que los bienes de ocasión o de segunda mano pueden seguir prestando el mismo tipo de servicios que tenían asignados cuando eran nuevos, lo hacen bajo unas condiciones que sus compradores consideran menos eficiente.

Dentro de los mercados de ocasión, solemos encontrar maquinaria, la cual constituye el patrimonio de una empresa como activo no corriente. La adquisición de la maquinaria constituye un aspecto de vital importancia en su rentabilidad económica actual y futura, por tanto, como nos indica ([Nava Rosillón, 2009](#)) deberá basarse en un estudio económico profundo considerando elementos condicionantes y criterios de elección.

Como factores condicionantes que influyen en la selección de la maquinaria y su posible adquisición, pueden considerarse, la situación económica y financiera, el escenario económico y las perspectivas de trabajo. Sin embargo, las condiciones y la tipología de los trabajos, la capacidad de producción necesaria, la flexibilidad al cambio en las condiciones, la fiabilidad y el servicio postventa son las condiciones de contorno del problema de la elección de la máquina más conveniente.

Además, es fundamental considerar que estas pierden valor al emplearse en el proceso productivo y por el mero transcurso del tiempo. A esta merma de los activos de la empresa se denomina depreciación. Cuando se habla de depreciación, se suele hacer referencia a la pérdida de valor de un bien mueble o inmueble; esta valoración es realizada por expertos mediante un procedimiento donde se examinan y analizan variables que reflejen la imagen fiel del bien, es decir, que no oculten información.

Según (Piquerias, 2015), la cotización del equipo depende de diversos factores, como del número de años de servicio, de las horas trabajadas hasta el momento, de las que le restan para llegar a su obsolescencia, de la naturaleza de las tareas realizadas y de las condiciones en que se ha usado. La antigüedad es, en numerosas ocasiones, el factor que más influye en la cotización, por ser el dato más fiable. El abandono de la fabricación de determinados modelos es una circunstancia que hace bajar la cotización de los equipos. Si la máquina pertenece a una gran firma internacional, se garantiza cierta confianza en su valor de reventa.

Uno de los principales aspectos que dificultan la gestión operativa de las pequeñas y medianas empresas, consiste en conocer y determinar la depreciación que sufren los productivos que requieren para poder llevar a cabo su actividad económica. Dentro de las causas que la produce, según lo publicado por (Piquerias, 2015), podemos encontrar:

1. Depreciación material: Dentro de esta, podemos encontrar dos tipos, la depreciación funcional y la depreciación física. En la primera, la maquinaria pierde valor a medida que presta los servicios que le son propios, y con respecto a la segunda, la devaluación de los bienes de equipo se debe al transcurso del tiempo, a veces incluso más que si estuvieran trabajando con normalidad. Una adecuada política de mantenimiento reducirá o retrasará la desvalorización de las máquinas, pero nunca la eliminará.
2. Depreciación por agotamiento, caducidad o siniestro: un siniestro deprecia de forma brusca el valor del equipo.
3. Depreciación por obsolescencia: Es la merma que sufre una máquina cuando, incluso siendo nueva, queda anticuada por no ser competitiva frente a otras. Una de las principales causas es la competencia entre fabricantes. Esta puede deberse a:
 - Obsolescencia tecnológica: La innovación y los avances técnicos motivan la aparición continua de nuevas máquinas que cumplen la misma función de las existentes, pero con mayor eficiencia, menores costes, ofreciendo mayor seguridad, etc. Si el ahorro de costes es suficiente, a la empresa le convendrá renovar o cambiar el equipo anticuado antes de terminar su vida técnica.
 - Obsolescencia por variaciones en la demanda: Una máquina excelente para un determinado nivel de producción, puede no ser rentable en otro nivel.
 - Obsolescencia por alteración en la retribución de algún factor productivo: Ante consecuencias de la inflación como lo son las subidas del precio de la mano de obra o del precio del combustible, puede ser rentable aumentar la automatización o cambiar el tipo de máquina.

Si se conociera exactamente la depreciación de un equipo podríamos estimar en cada momento su cotización en el mercado o valor de reventa. Sin embargo, este valor fluctúa según las condiciones locales y circunstancias específicas de cada caso, de modo que el precio depende de lo que un comprador esté dispuesto a pagar.

Generalmente, se suelen utilizar tablas para la tasación de maquinarias, Es una herramienta muy fiable que transmite confianza y transparencia a todas las partes involucradas. Dentro de estas tablas, podemos encontrar las del “Grupo Autónomo Nacional de Vendedores de Automóviles, Camiones y Motocicletas”- Ganvam, la cual desde su creación

ha figurado como la organización más representativa de concesionarios oficiales, compraventas independientes, agencias concertadas, talleres independientes de vehículos nuevos y usados.

Estas tablas de valoración se actualizan cada tres meses en colaboración con Eurotax (líder en el mercado de valoración de vehículos de ocasión). Tienen en cuenta la depreciación del vehículo, pero no otros aspectos, como el kilometraje, posibles accesorios o equipamientos extras, ni su estado general. Por tanto, podemos decir que es un valor estimado que pueden utilizar los concesionarios, talleres, peritos profesionales y otras empresas relacionadas con la compraventa de vehículos y la tasación.

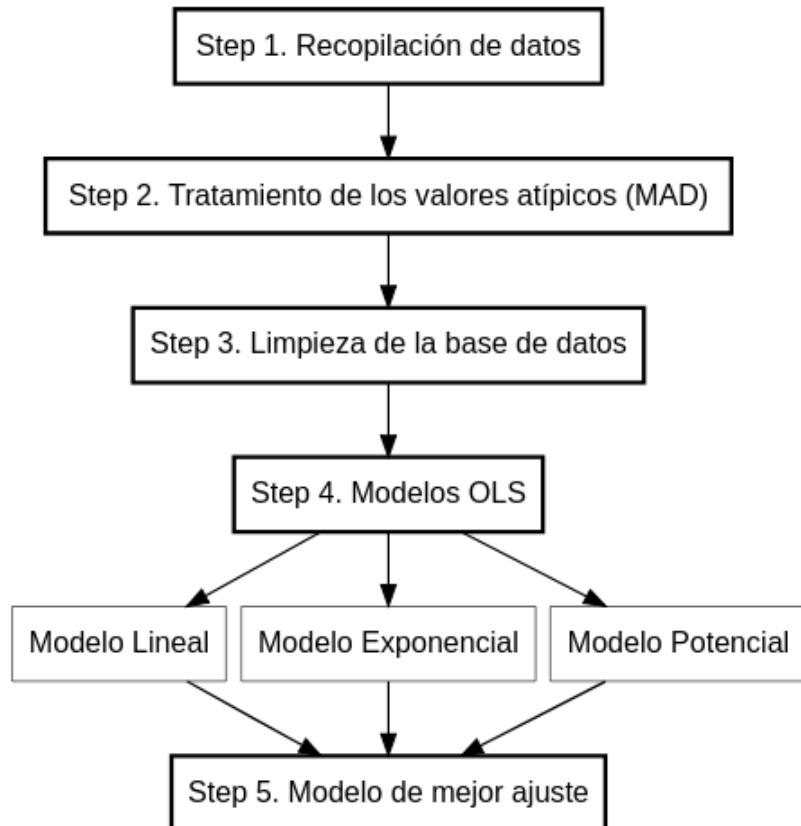
Las tasaciones mediante a tablas y los métodos aceptados, en ocasiones, suelen alejarse de la verdadera depreciación que sufre el activo, sin embargo, como nos indica ([Postigullo et al., 2018](#)) que, a partir de la evolución del valor de mercado de los activos en su vida útil, características técnicas, se puede hallar un modelo de amortización más apropiado y a partir de este, su depreciación.

Capítulo 4

Metodología

El presente Trabajo de Fin de Grado tiene como objetivo analizar el comportamiento de los valores y depreciación en el mercado secundario de cabezas tractoras. Para ello se seguirá el siguiente modelo:

Figura 4.1: Modelo metodológico.



Fuente: Elaboración propia

4.1. Lenguaje y entorno integrado de desarrollo

Para llevar a cabo la recolección y modelización de los datos, así como la evaluación del comportamiento de los valores y depreciación en el mercado de la maquinaria de ocasión se utilizará el lenguaje de programación R y el entorno de programación Rstudio. En adición a esto, se implementarán otras herramientas para facilitar la recolección de información dentro de las cuales destacamos el SelectorGadget.

R es un lenguaje relativamente joven y muy peculiar ya que, a diferencia de otros lenguajes de programación como Python y Java con los que el desarrollador puede llevar a cabo cualquier tipo de tarea, R está dedicado únicamente para la computación estadística y gráfica. Por lo que se decidió utilizar este lenguaje gracias a su facilidad de extensión y multitud de funciones para descargar y tratar los datos, el cual posibilita el acceso directo al lenguaje de programación, facilitando así su dinamismo y actualización.

Para la elaboración de los códigos, se utilizó un IDE (entorno integrado de desarrollo) conocido como Rstudio. El cual es un sistema de software para el diseño de aplicaciones que combina herramientas comunes para desarrolladores en una sola interfaz de usuario gráfica (GUI), es decir, permite a los desarrolladores ejecutar las tareas sin tener que pasar de una aplicación a otra.

Rstudio, al ser un IDE, permite llevar a cabo de manera sencilla tareas comunes, difíciles y tediosas de realizar como (editar de código fuente, administrar los archivos que se están usando, gestionar el entorno de trabajo, automatización de compilaciones locales, depurador, análisis del código a tiempo real, entre otras...).

Otro criterio que privó la selección de este lenguaje, fue que al ser un software gratuito que tiene facilidad para compartir códigos, crear herramientas para solucionar problemas comunes. Así mismo, se caracteriza por ser una alternativa amigable y con mucha información accesible en la red (librerías) que facilita el desarrollo del proyecto. Dentro de las principales librerías implementadas, se pueden destacar:

- DiagrammeR: Paquete para la creación de diagrama de flujos.
- kableExtra: Para construir tablas complejas comunes y manipular sus estilos.
- Mltools: Una colección de funciones auxiliares de aprendizaje automático, que ayudan en particular en la fase de análisis exploratorio de datos.
- Psych: Es un paquete que se utiliza para la creación de gráficos de correlación.
- Readr: Permite leer el fichero cvs.
- Robots.txt: Permite averiguar si las secciones de la página web son legalmente accesible. Si sale disallow es que son accesibles.
- Rvest: Está diseñado para trabajar junto a magrittr para expresar operaciones complejas de manera elegante y fácil de entender.
- Tidyverse: Es un conjunto de paquetes en R diseñados para ciencia de datos. Entre los paquetes que incluye, Se utilizarán:
 - Broom: Toma la salida desordenada y las convierte en tablas ordenadas.

- Dplyr: Ayuda a la manipulación de datos, permite mutar o crear nuevas variables, seleccionar, filtrar, sumarizar, acomodar o agrupar las variables.
 - Forcats: Es un paquete especial para lidiar con factores y datos categóricos.
 - Ggplot2: Se utiliza para la visualización de los datos mediante gráficos.
 - Patchwork: Diseñado para hacer la composición de gráficos.
 - Purrr: Este paquete facilita el trabajo con vectores y funciones.
 - Tibble: Es una transformación del data frame. Aprovecha lo bueno de los data frame y mejora algunas opciones que quedaron obsoletas.
 - Tidyr: Ayuda a transformar los datos para hacerlos más eficientes o tidy. Es decir, permite transformar las filas (para que cada una sea una observación), las columnas (para que cada una sea una variable) y las celdas (para que cada una sea un valor), lo que ayuda a la hora de modelar y graficar los datos.
 - Stargazer: Produce código LaTeX, código HTML y texto ASCII para tablas bien formateadas que contienen resultados de análisis de regresión.
 - Stringr: Ayuda a la preparación y limpieza de los datos.
- Xml2: Permite analizar documentos XML.

Asimismo, dentro de RStudio se utilizó una herramienta denominada R Markdown para llevar a cabo la elaboración del informe.

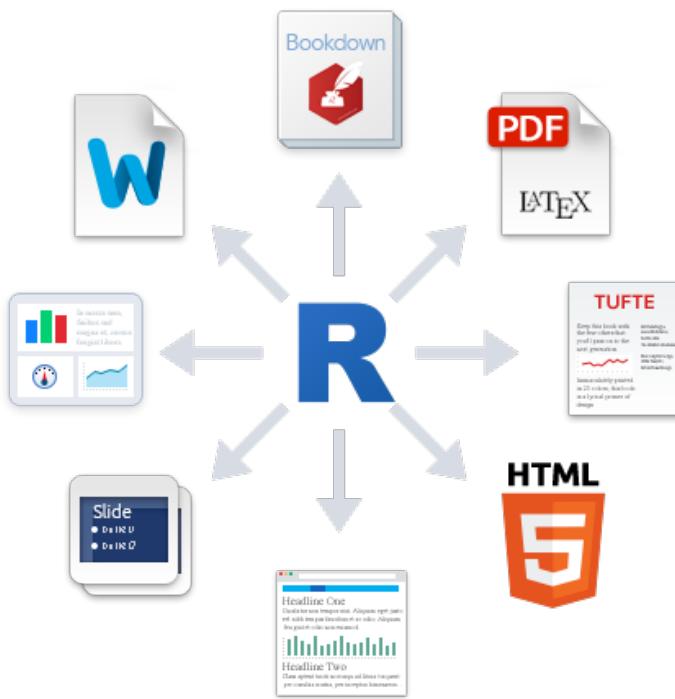
4.2. R Markdown

R Markdown es una herramienta ofimática diseñada con el fin de facilitar la creación de contenido y a su vez muchos tipos de salida (HTML, PDF, MS_Word, entre otros tipos expuestos en la figura 4.2.) a partir del archivo .Rmd. el cual permite combinar código en lenguaje R y contenido en formato de texto. Así mismo, permite:

- Automatizar tareas, manipular y analizar datos.
- Crear visualizaciones de datos de alta calidad.
- Crear dashboards para visualizar y analizar datos.
- Crear informes automáticos.
- Disponer de herramientas de análisis estadístico para ahondar en el conocimiento de los datos.

También se caracteriza por proporcionar informes rápidos y reproducibles de R que, a partir de los fragmentos incrustados de código R que son ejecutables con la sintaxis de Knitr, se puede actualizar el documento en cualquier momento volviendo a tejer los fragmentos de código.

Figura 4.2: Posible formatos estáticos y dinámicos de salida.



Fuente: ([Xie, 2022a](#))

4.3. ¿Por qué utilizar R markdown para escribir un trabajo de fin de grado?

Muchos de los trabajos que se realizan tanto en el ámbito laboral como estudiantil generan resultados numéricos, gráficos y tablas cuya integración en los programas mencionados anteriormente, se requiere en la mayoría de las ocasiones de las acciones de copiado y pegado en el portapapeles para insertarlos en el documento.

Al ser un trabajo muy artesanal, en muchas ocasiones ocurre que bien porque se han detectado errores o porque se quiere volver a recalcular con unos nuevos parámetros, se tiene que repetir todo el proceso. En cambio, R Markdown permite incluir elementos producidos por el lenguaje R, (código R, gráficos, tablas, etc) en el mismo documento evitando así que se tenga que recurrir al copiado y pegado al portapapeles para construir un documento de resultados.

Es por ello que suele ser una herramienta básica para el trabajo habitual de un profesional de la Estadística o de una persona que en su trabajo requiera la inclusión de análisis de datos o creación de informes o análisis periódicos (“reporting”). Ahora bien, para poder llevar a cabo un informe hay que tener en cuenta que es necesario instalar las siguientes herramientas: lenguaje R, la aplicación RStudio, la utilidad Pandoc y el sistema LaTeX (tinytex), que será el encargado de producir el documento pdf.

Capítulo 5

Obtención y descripción de la base de datos

5.1. Fuente de obtención de datos

Gracias a la gran cantidad de información accesible en internet, es posible llevar a cabo dicho estudio. Para ello, se han recolectado los datos necesarios de la [página web Europa-Camiones.com](#), la cual está editada por el grupo Vía Mobilis, referente en más de cuarenta países como experto en anuncios de materiales industriales y de vehículos pesados usados o también llamados de segunda mano en Internet.

Figura 5.1: Páginas web Grupo Vía Mobilis.



Fuente: ([Vía-Mobilis, 2022](#))

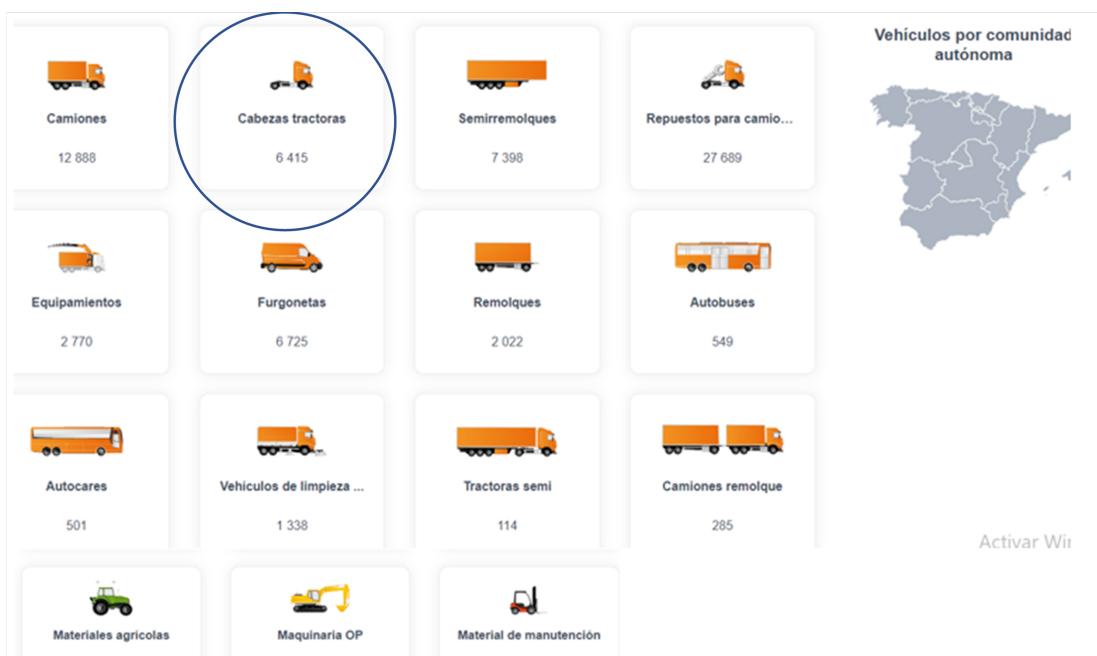
El objetivo de este grupo es ofrecer una amplia variedad de servicios adaptados a las necesidades de los profesionales de diferentes sectores (transporte, manutención, construcción, agricultura) a través de una red de 4 páginas web 100 % especializadas en la figura 5.1.

Si bien este grupo ofrece una gran cantidad de anuncios, para el desarrollo del estudio solo se hará uso de los datos obtenidos mediante la página web de Europa camiones, ya

que es la web que ofrece los anuncios correspondientes a las cabezas tractoras objeto de análisis.

Dicha página a fecha de este trabajo - junio 2022- tienen 209.029 anuncios publicados, por lo que no seríamos capaces de listarlos uno a uno. Sin embargo, la página oficial sugiere una división general por temas, indicando 4 grupos y cada grupo contiene de 1 a 15 subgrupos, con el fin de poder agrupar los datos contenidos en cada grupo. No obstante, en realidad la base de datos cuenta con una clasificación en cascada mucho más detallada la cual se puede apreciar en la figura 5.2 donde se ve el conjunto de datos, visto desde la página web:

Figura 5.2: Servicios especializados del grupo Vía Mobilis.



Fuente: ([Vía-Mobilis, 2022](#))

En la figura se puede apreciar la fuente específica utilizada en el trabajo correspondiente a las Cabezas tractoras las cuales, en junio 2022 tiene disponibles 6415 anuncios. Cuando hablamos de las cabezas tractoras se hace referencia al vehículo que está equipado de un acoplamiento o quinta rueda que le permite combinarse con un semirremolque que se engancha a la cabina donde se colocan todas las mercancías para su transporte. Si bien, está por sí sola no tienen la consideración de camiones al no estar integrado con el resto de la estructura del camión y pueden funcionar sin necesidad de cargarlo, es considerada una pieza de gran importancia en el sector del transporte y más concretamente del transporte de mercancías por carretera. Además, esta posibilidad de acople y desacople lo convierte en vehículos más flexibles y adaptables.

Marcas como Volvo, Renault, Mercedes o Scania tienen a la venta diferentes tipos de modelos. Sin embargo, ante los altos costos de los vehículos nuevos, se tiene la opción de acudir al mercado de segunda mano o mercado de ocasión donde la oferta es considerable

y depende de cada usuario tomar una decisión de acuerdo con sus necesidades, circunstancias personales y económicas que, serán las que determinen cuál será la alternativa más idónea. Por lo tanto, resulta fundamental la recolección de un conjunto de datos adecuado que nos ayude a la hora de la toma de decisiones.

Dentro de cada una de estas maquinarias podemos encontrar características técnicas que las diferencian entre sí, entre ellas, las más relevantes son:

- Modelo de cabeza tractora: existen diferentes modelos en el mercado, 4×2, 4×4, 6×2, 6×4, 6×6, 8×4... su diferencia fundamental es el tamaño que más convenga para desarrollar de la mejor manera su actividad.
- Precio: Suele ser decisiva a la hora de elegir una cabeza tractora. No obstante, es fundamental que no se convierta en el factor más determinante de todos, puesto que el factor más importante es que la compra pueda satisfacer las necesidades. Si se necesita una cabeza tractora muy potente para obtener las prestaciones necesarias para poder llevar a cabo el trabajo perfectamente, se tendrá que sacrificar, en parte, el precio de la maquinaria.
- Kilómetros acumulados: Las cabezas tractoras son vehículos muy fiables que acumulan una gran cantidad de kilómetros sin que su mecánica tenga porque resentirse. Sin embargo, cuando el odómetro de los vehículos acumula una gran cantidad de kilómetros recorridos, la maquinaria suele acusar de desgaste.
- Estado del vehículo: El gran inconveniente que presenta el mercado de ocasión es que, por lo general, no se conoce la procedencia del vendedor y resulta más complicado saber si es de fiar y si el vehículo se encuentra en buen estado. Una opción es confiar solo en empresas de venta con cierto bagaje, reputación y fiabilidad comprobada que garanticen que el vehículo vaya a tener un buen rendimiento en el futuro.
- Capacidad de carga: Cada fabricante varía los límites adaptando la capacidad de carga de las cabezas tractoras en función de del modelo ofertado.
- Suspensión: resorte/ neumática; neumática/ neumática; resorte/ resorte.
- País de localización: zona Euro.
- Medio de adquisición: Alquiler, leasing, compra - venta.
- Peso limpio: una de las características más importante a la hora de seleccionar la cabeza es tractora es su peso limpio (sin carga ni conductor en su interior).
- Las marcas conocidas que pueden tener las cabezas tractoras son: BMC, DAF, Iveco, MAN, Mercedes Benz, Renault, Scania, Volvo.

Figura 5.3: Marcas de vehículos ofertados en la páginas web.



Fuente: (Elaboración propia)

5.2. Recolección de la información

Como se ha comentado anteriormente, debido a la complejidad y gran cantidad de información disponible, en la actualidad el manejo y organización de los datos tiene mucho más valor puesto que, a la hora de realizar un análisis sirve como orientación y/o punto de partida para la toma de decisiones. Ahora bien, el proceso de obtención, recopilación y preparación de datos confiables y de calidad necesarios para el análisis puede ser una tarea muy difícil, por lo que es importante resaltar la globalización y el aporte del desarrollo tecnológico en materia de datos, que pone una gran cantidad de información al alcance de todo el público, representando una ganancia en tiempo y dinero en el proceso de obtención de los datos.

La problemática hoy en día no gira en torno al acceso a los datos, sino a su tratamiento e interpretación, cuyas disciplinas son la ciencia de datos y el Business Analytics respectivamente. Es por ello que, para el análisis o transcripción de los ingentes recursos disponibles resulta fundamental el uso de una aplicación tecnológica que permita llevar a cabo la recolección de datos específicos, precisos, fiables y ajustados a las necesidades de cada caso de forma automatizada.

Para ello, no existe una única solución, sino que existen numerosas aplicaciones y lenguajes de programación disponibles que permiten llevar a cabo este proceso. Para la consecución de los objetivos propuestos, se ha decidido utilizar para la obtención de la información de manera fiable y eficiente el WEB SCRAPING o raspado web, técnica utilizada mediante programas de software para extraer información de sitios web, la cual simula la navegación de un humano en la World Wide Web ya sea utilizando el protocolo HTTP manualmente, o incrustando un navegador en una aplicación.

5.3. Priorización de datos

Una vez conocido la cantidad de datos disponibles, se procede a priorizar los más confiables para su almacenamiento y centralizarlos con una plataforma de análisis. Esta plataforma hace posible una preparación de datos más rápida que permite crear dashboards personalizados para miles de vehículos.

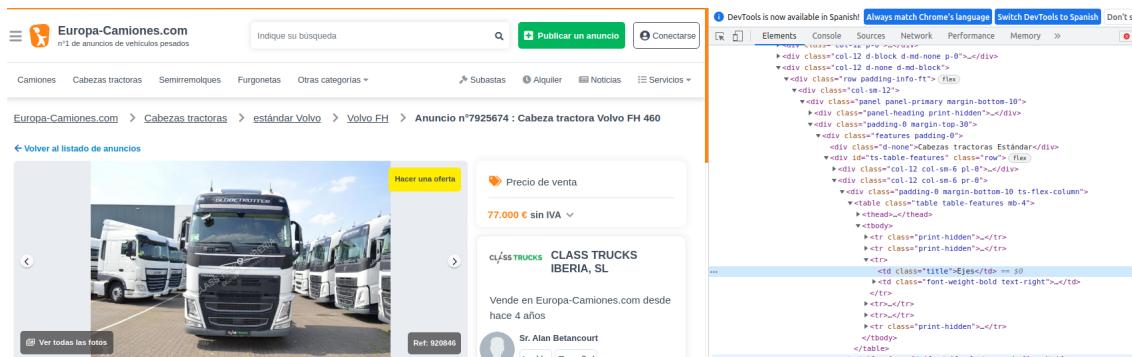
La primera fase de este trabajo consiste en analizar la depreciación en el mercado de las cabeceras tractoras de segunda mano. Para ello, se lleva a cabo la recolección de

información ofertadas en la [página web Europa-Camiones.com](#).

Sin embargo, tras la revisión de Europa-Camiones.com se destaca que utilizar el árbol de navegación es incómodo y puede ser confuso a la hora de buscar datos, debido a que el conjunto de datos ofrecidos no sigue un único patrón, es decir, que permite introducir información incompleta de cabezas tractoras, lo que genera complicaciones a la hora de recolectar los datos, ya que las tablas variaban de longitud. Es por ello que, para la obtención de las variables (Potencia, Kilometraje) se realizó en primer lugar la selección de varias celdas y posteriormente la limpieza de los datos, que de no realizarse generaría errores de selección y/o incluso no se seleccionaba la información aun encontrándose disponible.

Generalmente las páginas webs están escritas en un código de lenguaje específico, el cual es interpretado por los buscadores (Chrome, Edge...), y puede presentar distinto formato de código como el HTML (Hypertext Markup Language) y el CSS (Cascading Style Sheets). Al abrir una página web, no se ve ni el código ni el formato, si no una representación del mismo que, para poder observarlo, es necesario hacer click derecho en la página web + inspeccionar.

Figura 5.4: Ejemplo de visualización de código y formato de página web.



Fuente: (Elaboración propia)

Como se puede observar en la figura 5.4, esta presenta mucha información complicada e inentendible por lo que, para realizar la recopilación de los datos, son necesarias algunas herramientas y paquetes que faciliten el trabajo, dentro de la cual podemos destacar:

- Rvest: Ayuda en la extracción de información de las páginas web.
- Robots.txt
- SelectorGadget: Es una extensión de Chrome, que ayuda a analizar HTML y a rastrear la web para obtener la información que necesitamos. Al activar esta herramienta será importante seleccionar la información deseada en color verde y amarillo y la no deseada en color rojo, obteniendo un resultado como el que se visualiza el código y la información seleccionada como en la figura 5.5, lo que facilita al usuario el acceso a la información de interés.

Figura 5.5: Visualización herramienta Selector Gadget.

The screenshot shows the Selector Gadget interface with three main sections: 'Informacion principal' (main info), 'Características Renault T-Series' (Renault T-Series features), and 'Información complementaria Renault 480 T4X2 E6' (complementary Renault 480 T4X2 E6 info). The 'Informacion principal' section contains 'Ref. cliente : 10289' and 'Ref. web : 7758533'. The 'Características Renault T-Series' section shows 'Potencia 480 CV' and 'Kilometraje 923 000 km'. The 'Información complementaria Renault 480 T4X2 E6' section shows 'Caja de cambios' and 'Caja de cambios automática'. A cookie consent banner at the bottom states: 'Nuestro sitio utiliza cookies para garantizarle una experiencia de usuario óptima, ofrecerle servicios adaptados a sus intereses y ofrecerle publicidad relevante, tanto en nuestras páginas como en la de terceros. Al hacer clic en "Acepto", usted acepta el uso de estas cookies para estos fines.' Below the sections are buttons for 'print-sm-12:nth-child(2) .table-features:nth-child(1) .print', 'Clear (1)', 'Toggle Position', 'XPath', '?', and 'X'.

Fuente: ([Vía-Mobilis, 2022](#))

La recolección de la información se llevó a cabo de dos etapas:

1. La primera etapa consiste en la extracción de la información a partir de la página principal que muestra de forma resumida la información del vehículo. De ella se extrajo cuatro variables, Nombre del vehículo, Precio, País y Ciudad en el que se encuentra.

Figura 5.6: Visualización de la información que se encuentra en la página principal.



Fuente: ([Vía-Mobilis, 2022](#))

2. En la segunda etapa, para poder ampliar las características que se ofrecían de cada cabeza tractora en la página principal, se llevó a cabo una segunda forma de extracción de la información, que permitió conseguir datos interesantes y útiles para los posteriores análisis como (Marca, Potencia, Kilometraje, Precio, entre otros).

Figura 5.7: Visualización de la información característica de las cabezas tractoras.

Información principal		Características Volvo FMX	
Ref. cliente :	8402	Potencia	460 CV
Ref. web :	3070674	Kilometraje	428 094 km
Carrocería	Estándar	Ejes	6x4
Marca	Volvo	Ejes (Marca)	Volvo 6x4
Gama (marca : Volvo)	FMX	País (gama)	PORTUGAL
Estado	Ocasión	Energía	Diesel
Fecha 1ª matriculación	06/03/2013	Conducción	Volante a la izquierda
País	PORTUGAL	Normas Europeas	Volvo Euro 5
Constructor	Volvo	Información cabina FMX	

Fuente: (Vía-Mobilis, 2022)

Para obtener los links de cada vehículo y poder recolectar la información que presentaba en su interior, fue necesario llevar a cabo la inspección del código. Este proceso fue complejo debido a que el código en el que se encontraba la información era un formato CSS.

Figura 5.8: Visualización de código y formato CSS de la página web.

Fuente: (Elaboración propia)

Para poder llevar a cabo el WebScraping (extracción de los datos) en Rstudio, se implementaron una serie de funciones, dentro del cual se pueden destacar:

- `Read_html («url»)` Con esta función se crea un objeto que contiene todo el código o etiquetas HTML.
- `html_nodes(«objeto html», «etiqueta css»)`: Es usada para seleccionar partes del objeto que contiene todo el código html. El segundo parámetro es la clase CSS que está relacionada con la sección que se desea extraer.

- `html_name()`: Obtiene los atributos html.
- `html_text()`: Extrae el texto html.
- `html_attr()`: Regresa los atributos específicos html.
- `html_table()`: Convierte una tabla html en una estructura de datos en R.
- `str_match()`: Evalua si el valor con el que se compara sigue una estructura (pattern) específico.
- `str_trim()`: Coloca el texto en la forma deseada (sin espaciado) i.e. remove //n...
- `str_split()`: Se usa para separar string.
- `rbind`: Se emplea para no crear el dataframe por cada raspado de la página, sino que solo ir adjuntando los valores de la página por la que esté pasando, manteniendo el resto de los valores.
- `Sys.sleep (Time)`: Se utiliza para suspender la función de ejecutar el raspado de la información y con ello, no sobrecargar el servidor y evitar que te baneen.

Al culminar con la técnica de WebScraping, los datos se organizan en tablas de datos con formato ISO-8859-1 y fueron almacenados en forma de fichero tipo .csv, logrando recolectar una muestra de 6348 filas referidas a los productos en oferta (cabezas tractoras) y 8 columnas que indican las diferentes características que pueden llegar a presentar las cabezas tractoras (país, ciudad, marca, precio, kilometraje, potencia) que, a la fecha 15/06/2022 se encuentren disponibles en el mercado de ocasión ofrecidas en la página web. Estos datos serán pilar fundamental para poder llevar a cabo un buen análisis.

Una vez obtenida la muestra, se procedió con la eliminación de algunas celdas que no presentaban valor (NA), con el fin de evitar predicciones y/o análisis erróneos,

Tabla 5.1: Número de celdas de cada variable sin valor (NA).

	Marca	País	Potencia	Precio	km	Edad	Total Datos
Con NAs	16	0	1359	1564	212	117	6348
Quitando NAs	0	0	836	0	0	0	4625

Pudiendo observarse que el número de datos que se utilizarán para llevar a cabo el análisis se ha visto reducido en 1723, resultando un total de 4625 valores.

Algunas valores interesantes de las características técnicas a estudiar se encuentran en la siguiente tabla:

Tabla 5.2: Valores estadísticos de la base de datos bruta.

	n	Media	sd	mad	curtosis
Potencia	3779	460.56	56.55	29.65	6.20
Precio	4615	37997.56	24702.19	23795.73	3.74
Km	4615	624378.99	325111.10	287971.33	1.42
Edad	4615	8.80	6.44	4.45	5.30

En la tabla 5.2, podemos apreciar como el número de observaciones de la potencia es diferente al resto de valores por la cantidad de celdas con valores (NA). Además, se observa como todos presentan una kurtosis positiva, la cual indica que los datos muestran más valores atípicos extremos que una distribución normal. También se aprecia como era de esperar como las variables de precio y kilometraje presentan la desviación absoluta media (MAD) y desviación estándar (sd) más elevadas.

Así mismo, se realizó la creación de nuevas variables:

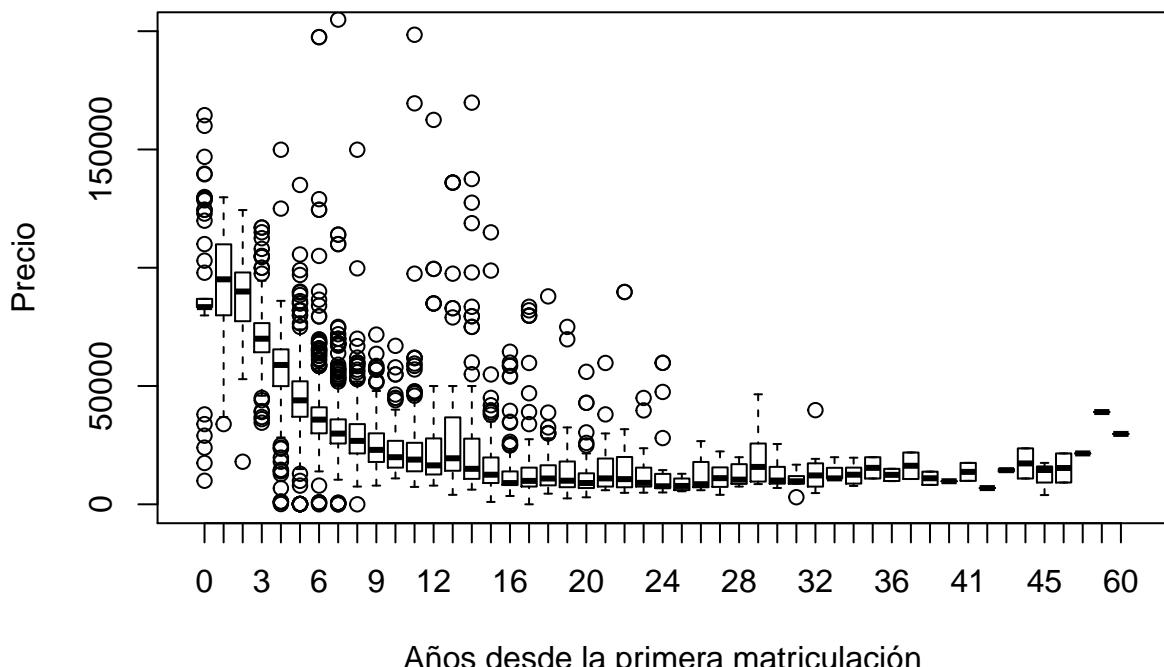
- years: En el que se calcula la edad de la cabeza tractora a partir de la fecha de la primera matriculación de la misma.
- years.category: En el que se separa por tramos de edades las cabezas tractoras ofrecidas, siguiendo el siguiente formato: $[0,5]$ = Nuevo, $[6,14]$ = Medio, $[16,28]$ = Viejo, $[28,\infty]$ = Muy viejo.

Capítulo 6

Limpieza de la base de datos

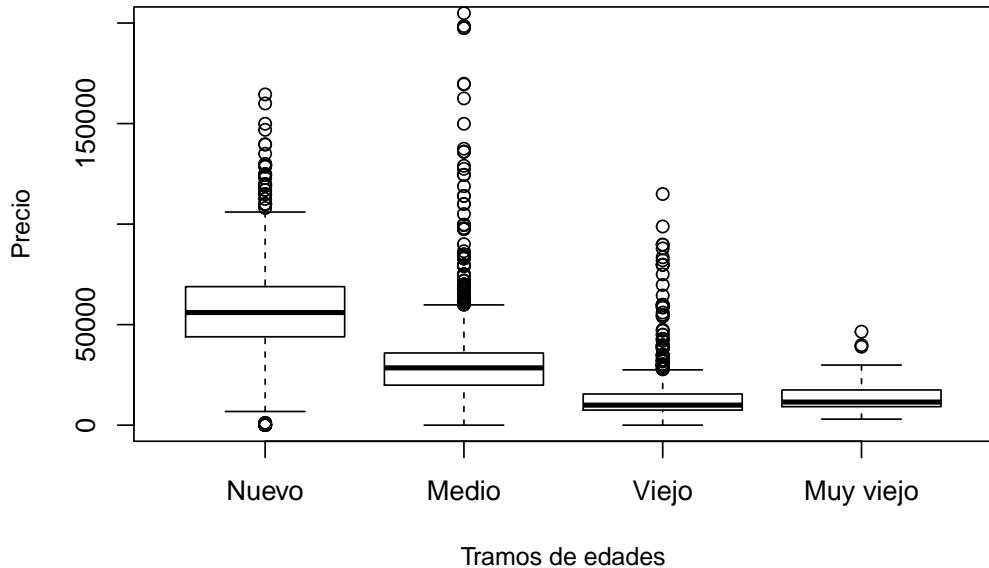
Una de las ramas de estudio más importantes de la estadística es el análisis de la calidad de los datos puesto que, en caso de presentar problemas (heterogeneidad, faltantes, etc.) pueden conducir a formular predicciones incorrectas y tomar decisiones erróneas con graves consecuencias. Entre los principales problemas que pueden presentar los datos, se encuentran los conocidos como valores atípicos (outliers) que “es una observación que se desvía mucho de otras observaciones y despierta sospechas de ser generada por un mecanismo diferente” ([Muñoz García et al., 2013](#)).

Figura 6.1: Representación gráfica de los datos con outliers.



Fuente: (Elaboración propia)

Figura 6.2: Representación gráfica de los datos con outliers.



Fuente: (Elaboración propia)

Basado en lo anteriormente expuesto, a partir de la base de datos bruta recolectada que como se aprecia en las figuras 6.1 y 6.2 presenta valores anómalos, se lleva a cabo la limpieza de registros enviados incorrectamente y de los datos que son sospechosos de ser erróneos, en el sentido que no siguen el mismo patrón que la mayoría. En primer lugar, se eliminan aquellos datos de cabezas tractoras que presenten “NA”, es decir, que no presenten información completa en (precio, kilómetros recorridos y año de primera matriculación). En segundo lugar, se procedió a realizar dos métodos para la detección de estos valores atípicos (Interquartile Range IQR y Median Absolute Deviation MAD) con el fin de eliminar y/o suavizar aquellos datos que puedan afectar y con ello producir un error en la extracción de la información.

Como primera alternativa, se utilizó la diferencia de rango intercuartílico, al ser una buena medida de variabilidad para distribuciones asimétricas y conjuntos de datos con valores atípicos. Este se calculó utilizando las siguientes formulas, con una constante $c=1.5$ para valores atípicos leves:

$$IQR = Q_3 - Q_1 \quad (6.1)$$

$$T_{min} = Q_1 - (c * IQR) \quad (6.2)$$

$$T_{max} = Q_3 + (c * IQR) \quad (6.3)$$

Como segunda alternativa, se utilizó la desviación absoluta alrededor de la mediana, debido a que como indica (Leys et al., 2013), la mediana al igual que la media, es una medida de tendencia central que apenas se ve afectada por la presencia de valores atípicos. El MAD se calculó utilizando la siguiente formula:

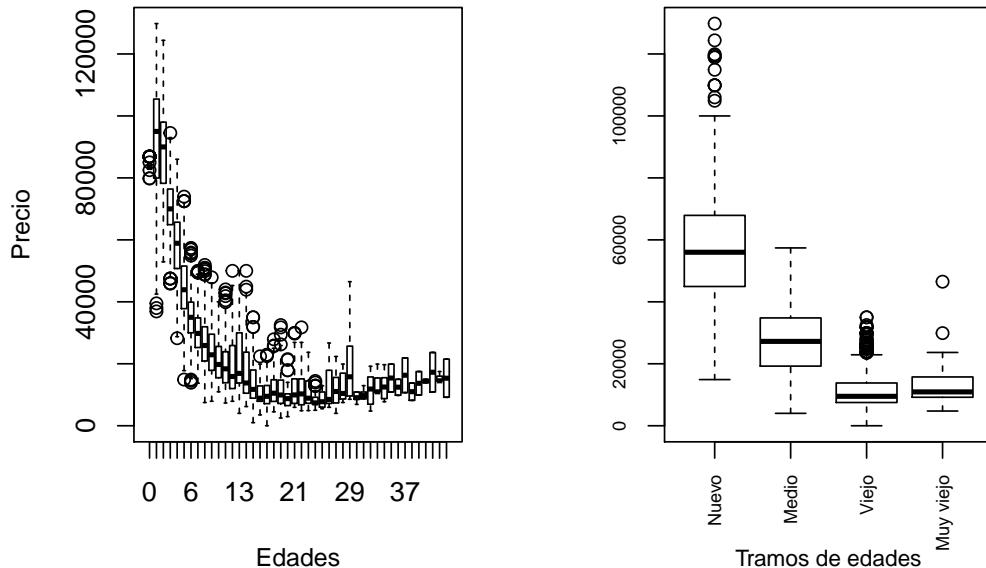
$$MAD = b * \text{median}(|x_i - \text{median}(x)|) \quad (6.4)$$

Con este valor se procedió a calcular el umbral conservador ($\alpha = 3$) para la detección de valores anómalos:

$$T_{min}, T_{max} = \text{median}(x) \pm (\alpha * MAD) \quad (6.5)$$

Una vez culminada la limpieza de datos con ambos métodos, se observaron los siguientes resultados del primer método (IQR) en la figura 6.3:

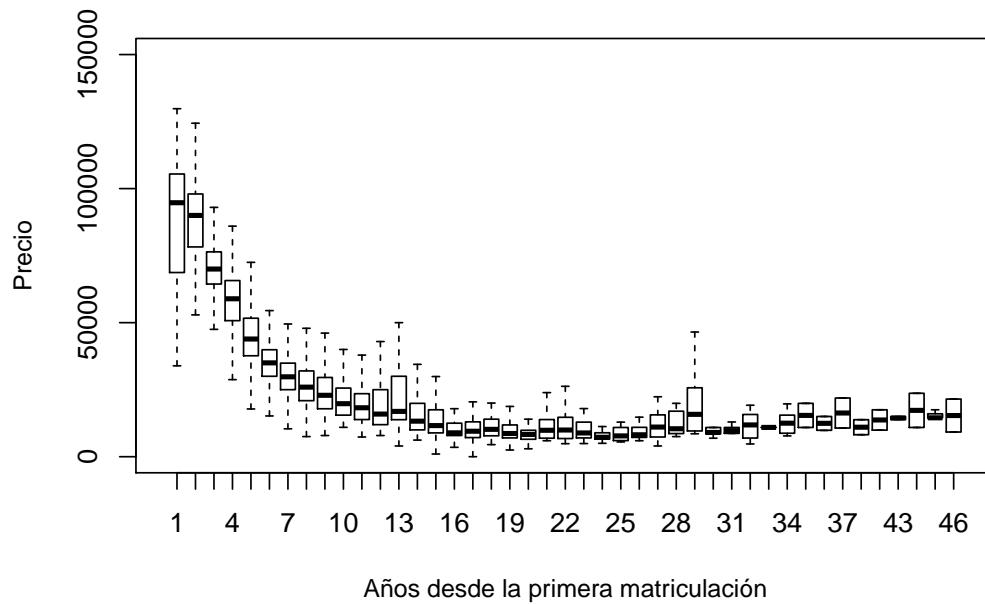
Figura 6.3: Representación gráfica de los datos posterior a la aplicación del método de diferencia de rango intercuartílico.



Fuente: (Elaboración propia)

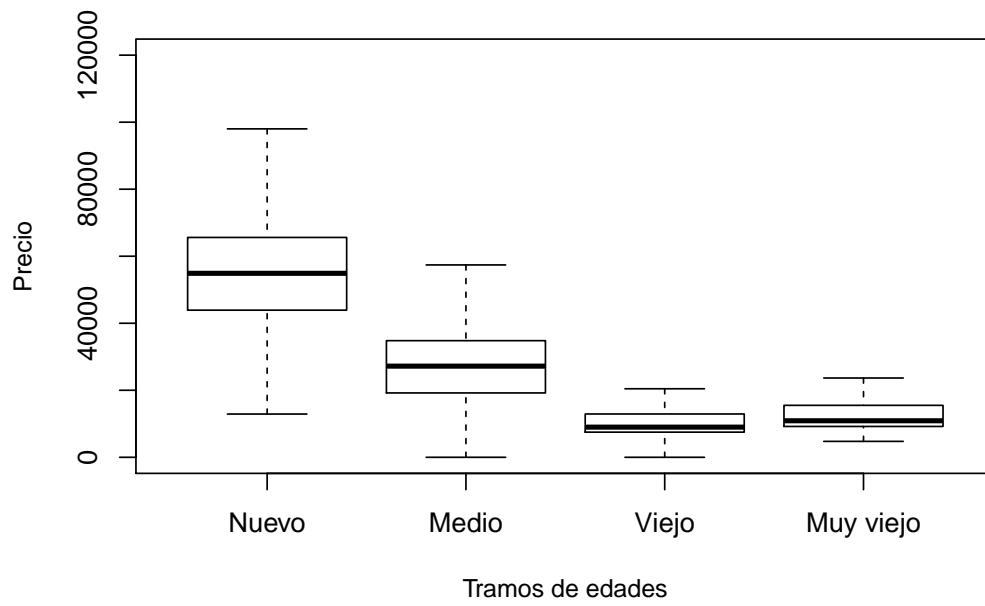
Se observa como en la figura 6.3 se reducen considerablemente el número de valores anómalos, produciendo una reducción en la amplitud del rango intercuartílico y de los límites de la caja de bigotes. A su vez, obtenidos los resultados del segundo método (MAD), representados en las figuras 6.4 y 6.5:

Figura 6.4: Representación gráfica de los datos sin outliers posterior a la aplicación del método MAD.



Fuente: (Elaboración propia)

Figura 6.5: Representación gráfica de los datos sin outliers posterior a la aplicación del método MAD.



Fuente: (Elaboración propia)

Se puede observar como en la figura 6.5 los tramos de edad nuevo y medio presentan distribución es simétrica, en cambio para los tramos de edad viejo y muy viejo, estos presentan una distribución asímetrica negativa.

Además se aprecia como en las figuras 6.4 y 6.5 con el método MAD se eliminan una mayor cantidad de observaciones atípicas en comparación al método IQR, produciendo

una reducción del total de la muestra a 4188 cabezas tractoras. Se ve como reduce la amplitud del rango intercuartílico y de los límites de la caja de bigotes. Es por ello que se utilizará el método MAD con el fin de reducir las posibilidades de conducir a decisiones erróneas y llevar a cabo el análisis de manera fiable y robusta.

Capítulo 7

Análisis exploratorio

Resulta impensable hoy en día tomar una decisión adecuada en cualquier ámbito partiendo del desconocimiento, por lo que se considera fundamental el valor y potencial de los datos, así como toda información disponible a la hora de tomar una decisión, puesto que el realizar una selección errónea y/o representación visual inadecuada de los datos puede dificultar o incluso imposibilitar el proceso de extracción de la información, y conllevar a una toma de decisión inadecuada.

Tabla 7.1: Valores estadísticos de las variables a estudiar posterior a la implementación del método MAD.

	n	Media	sd	mad	curtosis
Potencia	3406	457.82	52.61	29.65	5.55
Precio	4173	35117.79	20600.80	22387.26	0.19
Km	4173	653070.65	312219.13	278934.88	1.74
Edad	4173	8.80	6.09	4.45	4.37

Estos datos a analizar fueron procesados con el método MAD previamente con el fin de evitar valores erróneos que pudieran generar una malinterpretación de la información, pudiendo observar como se han reducido tanto las observaciones como los coeficientes de curtosis, MAD, y desviación estándar.

Finalmente, se estudiaran 4173 alternativas de cabeza tractoras, las cuales presentan características técnicas explicadas en las siguientes tablas y figuras.

Tabla 7.2: Número de observaciones marca.

Marca	Datos
Otro	61
Volvo	463
Mercedes	466
MAN	514
Renault	518
Scania	547
Iveco	614
DAF	995

En la tabla 7.2 podemos ver como dentro de las diferentes marcas de cabeza tractoras, la que presentan un mayor número de observaciones es DAF Trucks NV, a su vez, vemos como casi todas presentan un número muy similar de valores, a diferencia de la marca DAF y la marca Otros, que representan algunas marcas aisladas que pueden ser útiles para el análisis. Estas maquinarias pueden llegar a provenir de diversas partes de europa, como se representa en la tabla 7.3.

Tabla 7.3: Número de cabezas tractoras según el país en que se encuentra.

País	Datos
GRECIA	8
LUXEMBURGO	9
PORTUGAL	93
POLONIA	253
ITALIA	278
ALEMANIA	366
BÉLGICA	456
FRANCIA	621
ESPAÑA	861
HOLANDA	1237

Dentro de las observaciones expuestas en la tabla 7.3, vemos como a diferencia de la tabla 7.2 hay una gran variación entre la distribución de las observaciones, vemos como el mayor número de vehículos proviene de Holanda con 1237 y dentro de los países con menor número de cabezas tractoras encontramos a Grecia y a Luxemburgo.

Tabla 7.4: Número de cabezas tractoras por años que han transcurrido desde su primera matriculación.

Edad	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Datos	36	38	196	624	646	485	405	293	207	136	146	95	69	123

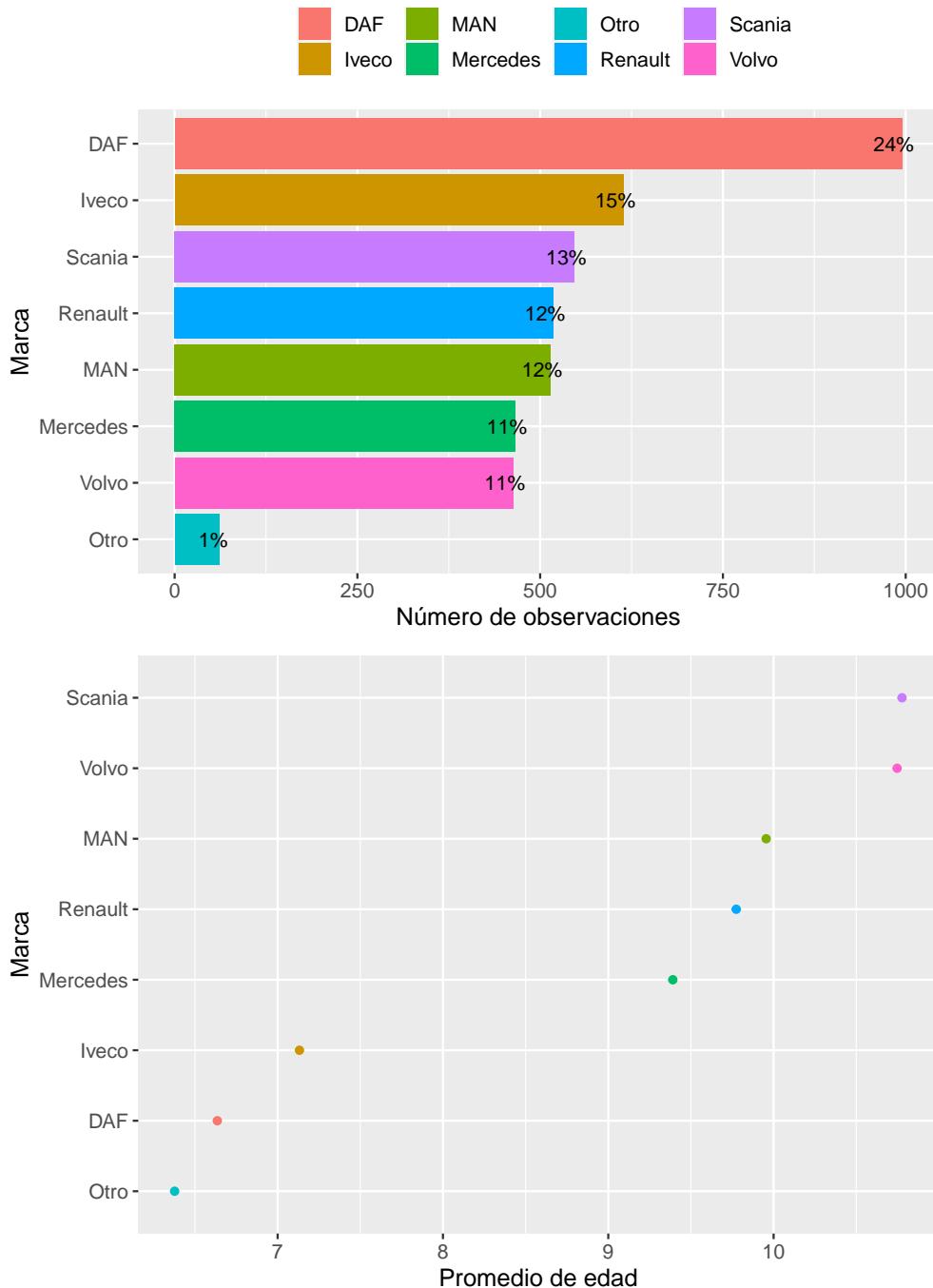
Edad	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Datos	98	94	95	50	46	39	43	35	30	32	25	10	16	7

Edad	29	30	31	32	33	34	35	36	37	39	41	43	44	45
Datos	7	6	6	9	2	5	2	4	2	2	2	2	1	3

En la tabla 7.4 podemos ver el número de observaciones según los años que presenten la cabeza tractoras. En esta, vemos como la mayor cantidad de datos se encuentran en las cabezas tractoras entre los 4 y 9 años desde su fecha de primera matriculación.

La correcta visualización de los datos permite detectar y comprender las tendencias, los valores atípicos y los patrones ocultos en grandes conjuntos de datos. Y, además de la tablas, existen muchos tipos de herramientas de visualización populares que permiten mostrar la información de manera eficaz como lo son los gráficos de barras para comparaciones, mapas para datos espaciales, gráficos de líneas para datos temporales y el diagrama de dispersión para comparar dos medidas, entre otros.

Figura 7.1: Datos característicos de las diferentes marcas de cabezas tractoras.

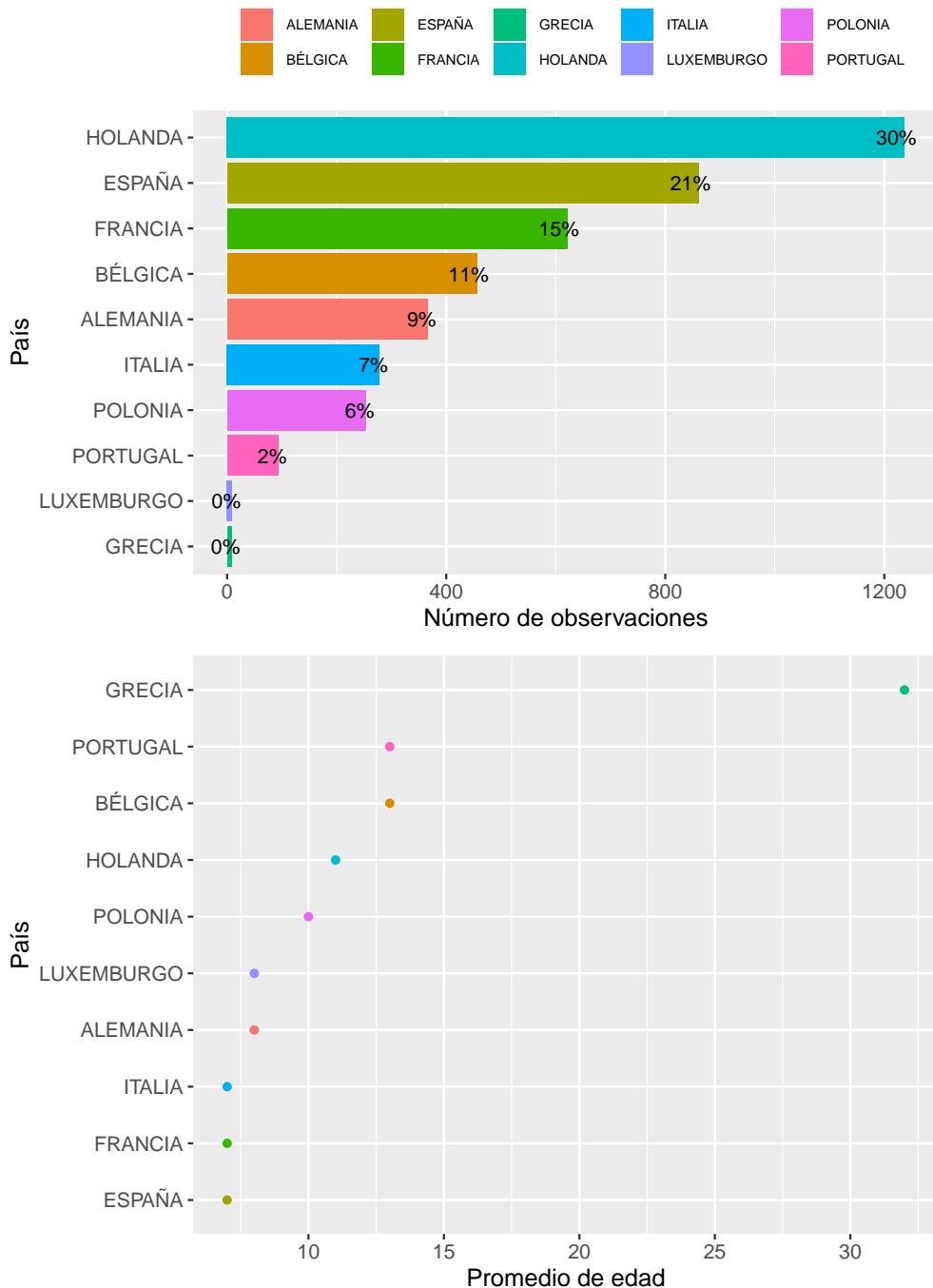


Fuente: (Elaboración propia)

En la figura 7.1 se representan las marcas de las cabezas tractoras respecto al número medio de edad en el que se ha realizado la matriculación y respecto al número de cabezas tractoras que se encuentran de cada marca. De la misma podemos destacar que el mayor número de cabezas tractoras representadas provienen del fabricante de vehículos neerlandés DAF Trucks NV, y que estos presentan una edad promedio de 6,6 años. Además, vemos como los vehículos más antiguos corresponden a las marcas suecas Scania y Volvo,

con una edad promedio próxima a los 11 años.

Figura 7.2: Datos característicos de las diferentes países que ofrecen las cabezas tractoras.

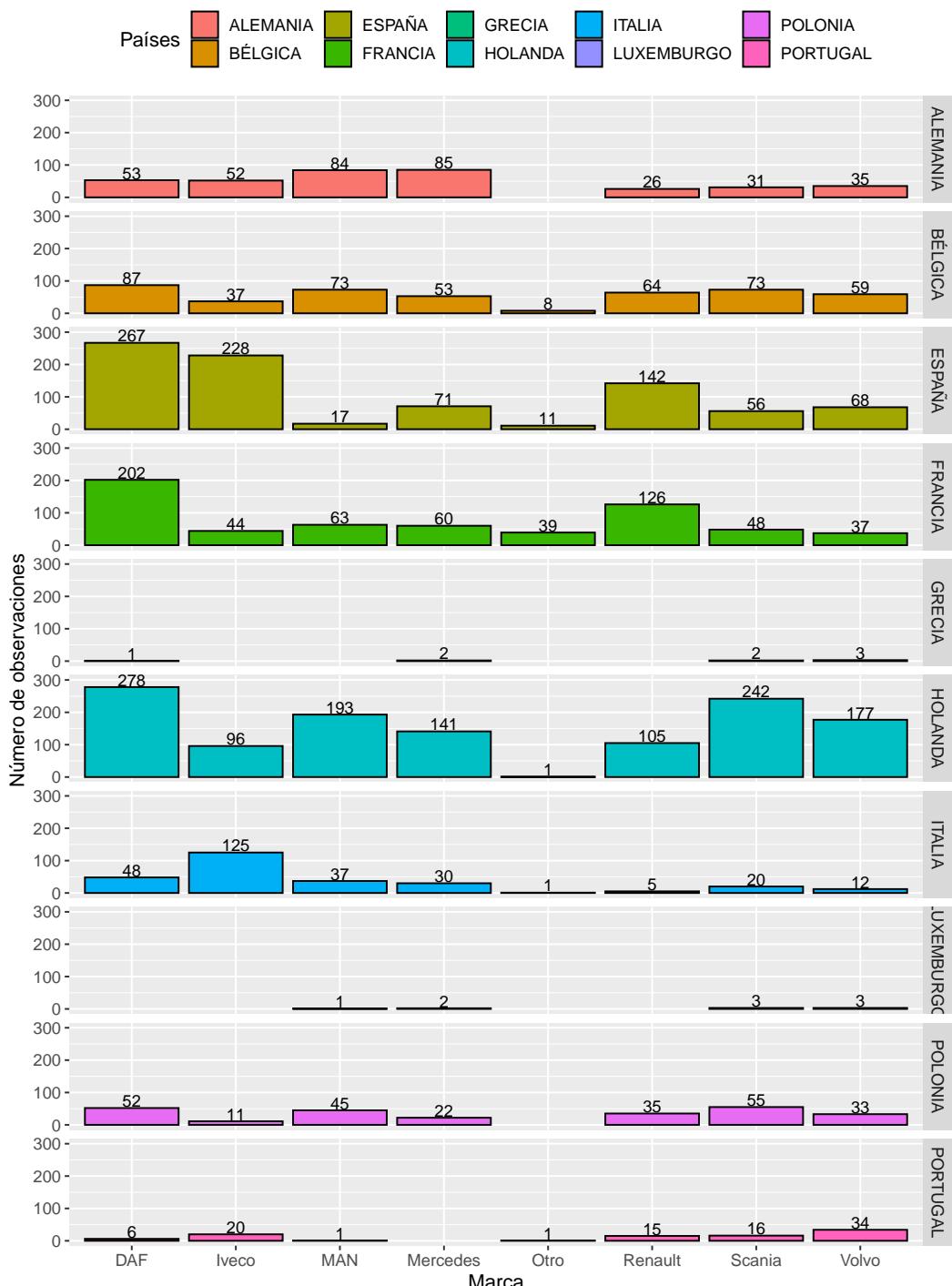


Fuente: (Elaboración propia)

En la figura 7.2 se representan los datos teniendo en cuenta los países en los cuales se encuentra ofertado el vehículo en función del número medio de edad en el que se ha realizado la matriculación y respecto al número de cabezas tractoras que se encuentran de

cada país. En esta vemos como Grecia presenta el menor número de vehículos ofertados y, además, presentan el mayor promedio de edad (32 años). Así mismo, como se comentó en la tabla 7.2, vemos como el mayor número de vehículos proviene de Holanda y que los países en los que la edad promedio de los vehículos ofertados más bajo corresponden a España, Francia y a Italia.

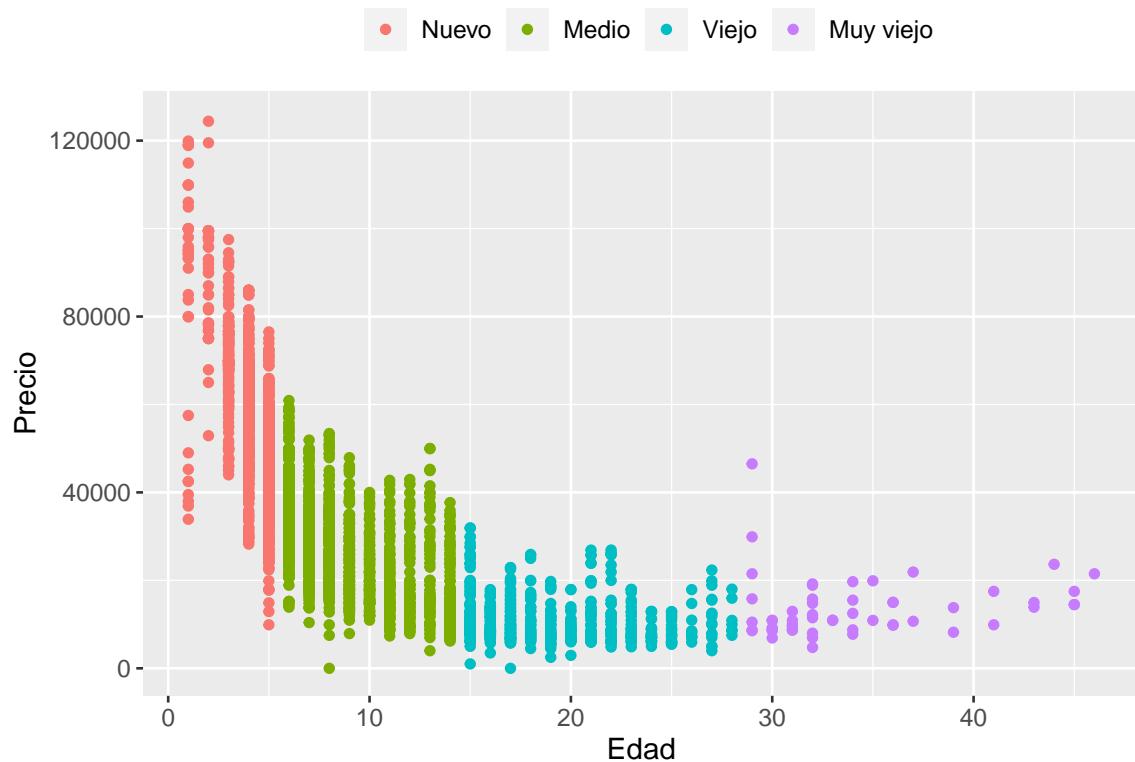
Figura 7.3: Número de las cabezas tractoras según el país de venta y marca.



Fuente: (Elaboración propia)

En la figura anterior, se puede apreciar a modo resumen de las figuras 7.2 y 7.3 una representación gráfica de entre las marcas de cabezas tractoras y el país donde se está vendiendo el vehículo.

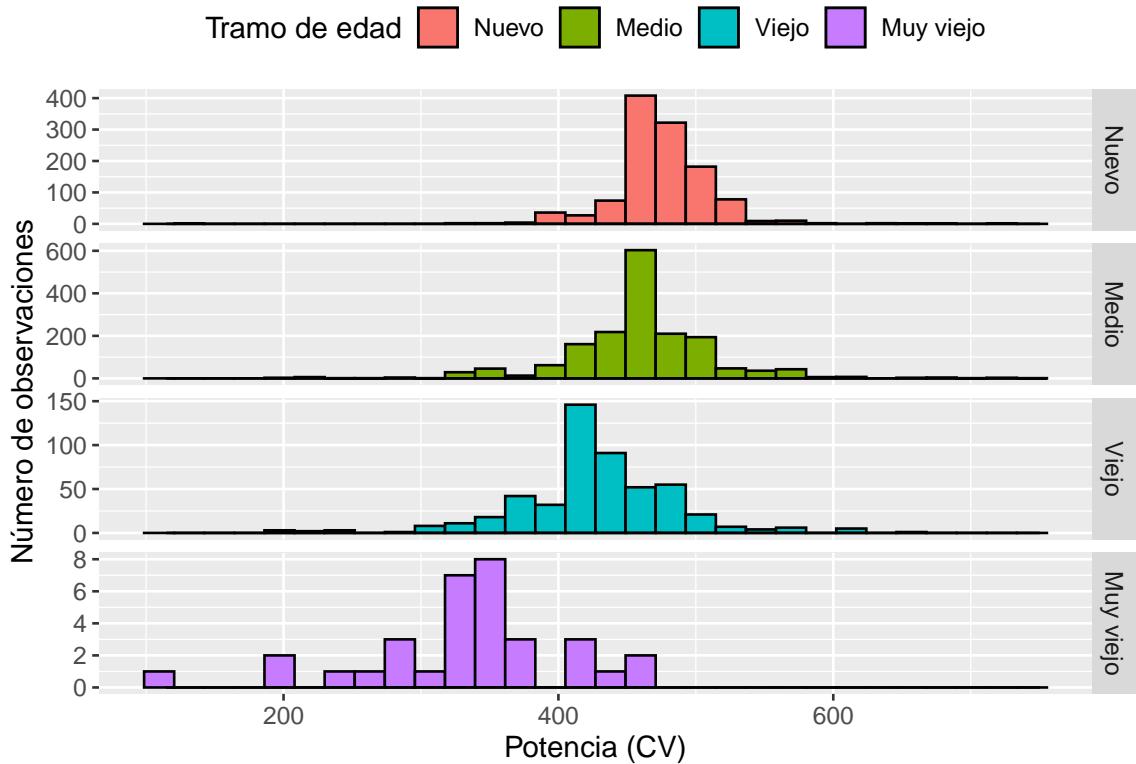
Figura 7.4: Gráfico de dispersión de los precios de las diferentes cabezas tractoras respecto al tramo de edad al que pertenece.



Fuente: (Elaboración propia)

A partir de la figura de dispersión representada anteriormente, se aprecia una tendencia negativa en el comportamiento del precio de venta con el paso del tiempo. Si bien hay unos matices diferenciadores en cuanto al comportamiento del precio de venta en los años iniciales y en los años ubicados hacia el final de su vida útil, se aprecia que, a nivel general, a medida que el vehículo sea más antiguo, el precio disminuye.

Figura 7.5: Representación de las potencias de los distintos grupos de cabezas tractoras.

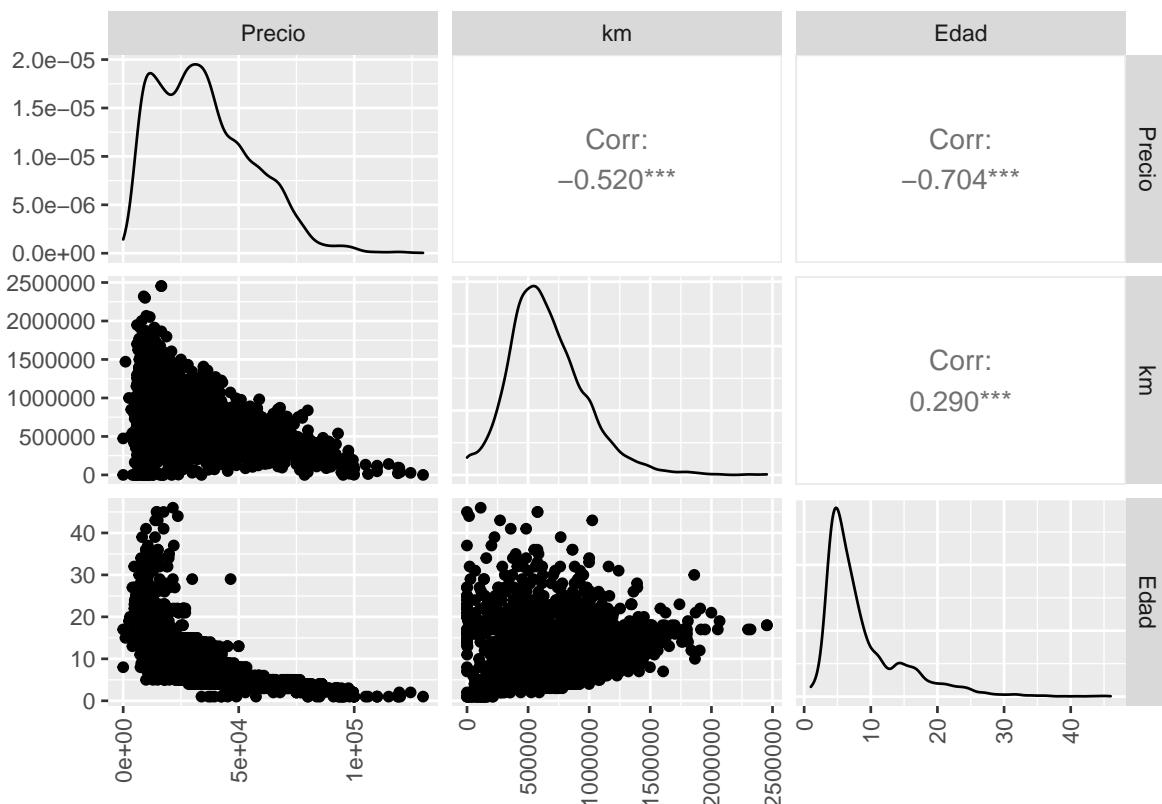


Fuente: (Elaboración propia)

En la siguiente figura 7.5 se puede observar las diferentes potencias de las cabezas tractoras que se encuentran ofertadas, estas se han separado según un rango de edad establecido anteriormente. Dentro de esta se ve como el mayor número de cabezas tractoras suelen presentar una potencia entre los 400 y los 600 CV y tienen una edad inferior a 15 años. Además, se evidencia como aquellos vehículos que presentan una edad superior tienden a tener una menor potencia, principalmente por los avances tecnológicos, que han traído una mayor potencia en la producción de estas cabezas tractoras.

Una vez explicado algunas características de la base de datos con la que se va a llevar a cabo el análisis, resulta fundamental, si presentan alguna relación la variable endógena (precio de venta) con las variables exógenas (edad del maquinaria, kilómetros recorridos).

Figura 7.6: Representación de la correlación respecto a la edad, kilometraje y precio.



Fuente: (Elaboración propia)

En la figura 7.6 se aprecia la relación entre edad y el kilometraje presentan una tendencia positiva, en donde los valores se encuentran un poco dispersos, si bien el coeficiente de correlación no es muy elevado, parece ser significativa.

Del mismo modo, en la figura 7.6 se observa como tanto el kilometraje como la edad respecto al precio presentan una tendencia decreciente (una relación inversa). Además, parece existir una mayor fuerza de relación entre la variable precio de venta y la edad desde que se produjo la primera matriculación de la cabeza tractora, al presentar un coeficiente de correlación más cercano a [-1,1], es por ello que en la representación se visualiza una menor dispersión entre las observaciones. Por lo que será importante llevar a cabo diversos modelos OLS de regresión para analizar el comportamiento del precio en función a las variaciones de edad.

Capítulo 8

Formulación de modelos OLS

Para llevar a cabo la modelización que puede llegar a tener la depreciación del valor de las cabezas tractoras en función de diferentes variables y evaluar cómo estas afectan tanto de forma independiente como conjunta a la depreciación de la maquinaria, se utilizarán diversos modelos OLS de regresión por el método de mínimos cuadrados ordinarios.

Estos modelo se basan en la relación de la variable exógena (precio de venta) con la variable o las variables endógenas (edad del maquinaria, kilómetros recorridos, marca, potencia, entre otros). Por tanto, no es un modelo de valoración, ya que no utilizamos para valorar un activo sino para obtener la tasa de depreciación.

Como indica ([Frías Domínguez, 2008](#)), un modelo de regresión lineal: es una técnica de modelado estadístico que consiste en generar un modelo de regresión para describir una variable dependiente, o respuesta continua Y como una función de una o varias variables independientes, o predictoras X . Este modelo permite comprender y predecir el comportamiento de sistemas complejos o a analizar datos experimentales, financieros y biológicos. Esta se describe de acuerdo con la siguiente ecuación:

$$Y = \beta_0 + \sum_{i=1}^n \beta_i X_i \quad (8.1)$$

Resulta fundamental llevar a cabo diferentes modelizaciones en busca de encontrar un modelo que explique mejor el comportamiento de estos activos, debido a que existe la posibilidad que el conjunto de datos no se llegue a explicar bien por una relación lineal, por esta razón, partiendo de la anterior ecuación, se llevaran a cabo unas variaciones para encontrar un método de regresión exponencial y potencial.

Un modelo de regresión exponencial es una técnica que busca encontrar la ecuación de la función exponencial $Y = \beta_0 * \beta_1^X$ que se ajuste mejor a un conjunto de datos, y que, al aplicar el logaritmo, se obtendría la siguiente ecuación:

$$\ln Y = \ln \beta_0 + \sum_{i=1}^n \beta_i X_i \quad (8.2)$$

Un modelo de regresión potencial es aquel en el que la función de ajuste es una función del tipo potencial del tipo $Y = \beta_0 * X^{\beta_1}$, el cual, según ([Frías Domínguez, 2008](#)), al linealizarlo, se obtendría la siguiente expresión:

$$\ln Y = \ln \beta_0 + \sum_{i=1}^n \beta_n * \ln X_n \quad (8.3)$$

En la siguiente tabla se puede observar un resumen de las diferentes ecuaciones empleadas para la modelización de la depreciación de las cabezas tractoras:

Tabla 8.1: Ecuaciones de las diferentes modelizaciones

Modelo	Ecuación de ajuste	Transformación de la ecuación	Tasa de depreciación
Lineal	$Y = \beta_0 + \sum_{i=1}^n \beta_n * X_n$	$Y = \beta_0 + \sum_{i=1}^n \beta_n * X_n$	$dr = \frac{\beta}{X}$
Exponencial	$\ln Y = \ln \beta_0 + \sum_{i=1}^n \beta_n X_n$	$Y = \beta_0 * \beta_1^X$	$dr = X^\beta$
Potencial	$\ln Y = \ln \beta_0 + \sum_{i=1}^n \beta_n * \ln X_n$	$Y = \beta_0 * X^{\beta_1}$	$dr = \beta$

De acuerdo con la Ley 27/2014, de 27 de noviembre, el artículo 12.1.a) del Impuesto sobre Sociedades (LIS) español, la maquinaria se amortizará durante su vida útil, y será admisible cualquier método que contemple las características técnico-económicas del activo, o derecho y, si no pudieran determinarse de forma fiable se adoptará el método lineal de amortización. Sin embargo, a diferencia de este, en el mercado secundario, se reconoce que estos activos tienen una vida más larga que la legal o la contable. ([BOE, 2014](#))

Como se ha ido comentando, en la depreciación de la maquinaria puede influir diversos factores como la marca, la potencia, el kilometraje, etc. Sin embargo, como se representó en la tabla 8.1, para evaluar la depreciación de la cabeza tractora se estableció un coeficiente de depreciación básico basado en la edad de los activos.

8.1. Bondad del ajuste del modelo

La bondad de ajuste de un modelo de regresión mide como las variables exógenas explican la variabilidad de la variable endógena. Las medidas más utilizadas para medir la calidad del ajuste son: error estándar de los residuos RSE , el test F y el utilizado en este trabajo, el coeficiente de determinación R^2 .

El coeficiente de determinación R^2 describe la proporción de variabilidad observada en la variable dependiente Y explicada por el modelo y relativa a la variabilidad total, es decir, valora si la regresión que estamos aplicando es representativa y puede predecir el valor de manera adecuada. Su coeficiente está acotado entre 0 y 1, de modo que cuanto más próximo sea el valor a 1, mejor será la bondad de ajuste del modelo de regresión a la distribución conjunta de las variables. Al ser ajustado, permite la comparación entre diferentes modelos, aunque el valor mínimo que tiene que alcanzar el coeficiente para comenzar a considerar que es representativo varía en función del que realiza el estudio.

Este valor se calcula mediante la ecuación (8.4), en donde SS_r es la medida de la variabilidad del modelo de regresión y la SS_t es la correspondiente a la medida de la variabilidad del valor sin considerar el efecto de la variable explicativa t .

$$R^2 = \frac{SS_R}{SS_t}, 0 \leq R^2 \leq 1 \quad (8.4)$$

8.2. Modelización del precio respecto a la edad

Después de haber aplicado la técnica de MAD para suprimir los valores atípicos, se ajustaron tres modelos de regresión lineal por mínimos cuadrados ordinarios del precio respecto a los años que han pasado desde la primera matriculación de cada cabeza tractora comentados con anterioridad. Como se representó en la tabla 7.1 en la página 24, se utilizarán 4173 datos para llevar a cabo el análisis.

En la siguiente tabla se representa de manera resumida los principales coeficientes estadísticos obtenidos de cada modelo según a los datos calculados en la página 57, en las tablas (C.1, C.2 y C.3), entre ellos, el coeficiente de ajuste, el coeficiente de depreciación, la desviación estándar de los residuos, los coeficientes β y debajo de estos, el error estándar.

Figura 8.1: Diferentes modelos de regresión del precio respecto a la edad.

<i>Dependent variable:</i>			
	price Modelo1 (1)	log(price) Modelo2 (2)	Modelo3 (3)
Lineal(years)	-2,382.975 *** (37.190)	-0.090 *** (0.001)	
Log(years)			-0.979 *** (0.010)
Constante	56,099.160 *** (398.104)	11.048 *** (0.013)	12.200 *** (0.021)
Coeff depre	-0.0425	0.9144	-0.9787
Observations	4,173	4,173	4,173
R ²	0.496	0.574	0.685
Residual Std. Error	14,626.130	0.470	0.404
F Statistic	4,105.633 ***	5,617.744 ***	9,060.815 ***

Note:

*p<0.1; **p<0.05; ***p<0.01

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Fuente: (Elaboración propia)

En la anterior tabla se puede ver como para todos los tipos de maquinaria y modelos muestran un nivel de significación de 1% “***”, es decir, presentan un nivel de 1% o inferior de significación. Lo que significa que hay evidencias estadísticas de que hay una diferencia que puede deberse en parte a la edad de la máquina.

Así mismo, proporciona información sobre los coeficientes β_1 , pudiendo ver como se esperaba, que todos los valores resultan negativos. Esto se debe a una tendencia negativa en el comportamiento del precio de venta con el paso del tiempo. También vemos que la tabla proporciona información sobre el coeficiente de R² de cada modelo, tal y como se definen en la ecuación 8.4. Como se ha comentado, R² explica que parte de la variabilidad de un factor puede ser causada o explicada por otro, en este caso, el (price) en función de la variable explicativa (years). Se aprecia como en los tres modelos propuestos, el coeficiente es superior o muy próximo a 0.5, esto implica que la variable explicativa “edad” como se comentó con anterioridad, es representativa para determinar el precio de este tipo

de activos, sin embargo hay otros factores que afectan al valor de los activos que no se han tenido en cuenta en el análisis. A su vez, vemos como tanto el modelo exponencial como potencial son superiores al modelo lineal, siendo el segundo el que mayor coeficiente presenta.

Además, se calculó el coeficiente de depreciación de mercado, tal y como se definen en la tabla 8.1 con el que se puede evaluar el desgaste de la maquinaria, a fin de comparar este valor con el coeficiente establecido por los organismos oficiales. En este se ve como las cabezas tractoras presentan en la regresión lineal un coeficiente de 4,25 %, viniendo a ser su vida útil aproximada de 23,5 años.

No obstante, si se analiza el modelo de regresión exponencial mediante un ejemplo de una adquisición de 1.000 euros, al tener un coeficiente de depreciación de 0,9144, en los cinco primeros años de vida útil presentarían un valor sobre el precio inicial del 63,93 %, en el décimo año este porcentaje desciende hasta el 40,87 %, en el vigésimo año se situaría en el 16,70 % y, finalmente, a los 50 años este activo tan solo tendría un 1,14 % del valor inicial.

Con el coeficiente hallado a través del modelo potencial, se sufrirían un deterioro mucho más acelerado en sus primeros años. Suponiendo el mismo ejemplo anterior, con el coeficiente de depreciación de 0,9787, al quinto año tendríamos un valor de 20,69 % y al decimo año, un valor de 10,5 %.

Destacar que para los tres modelos representados en la página 57, en las tablas (C.1, C.2 y C.3), el p-value modelo es significativo (2.2e-16) por lo que se puede aceptar que el modelo no es por azar.

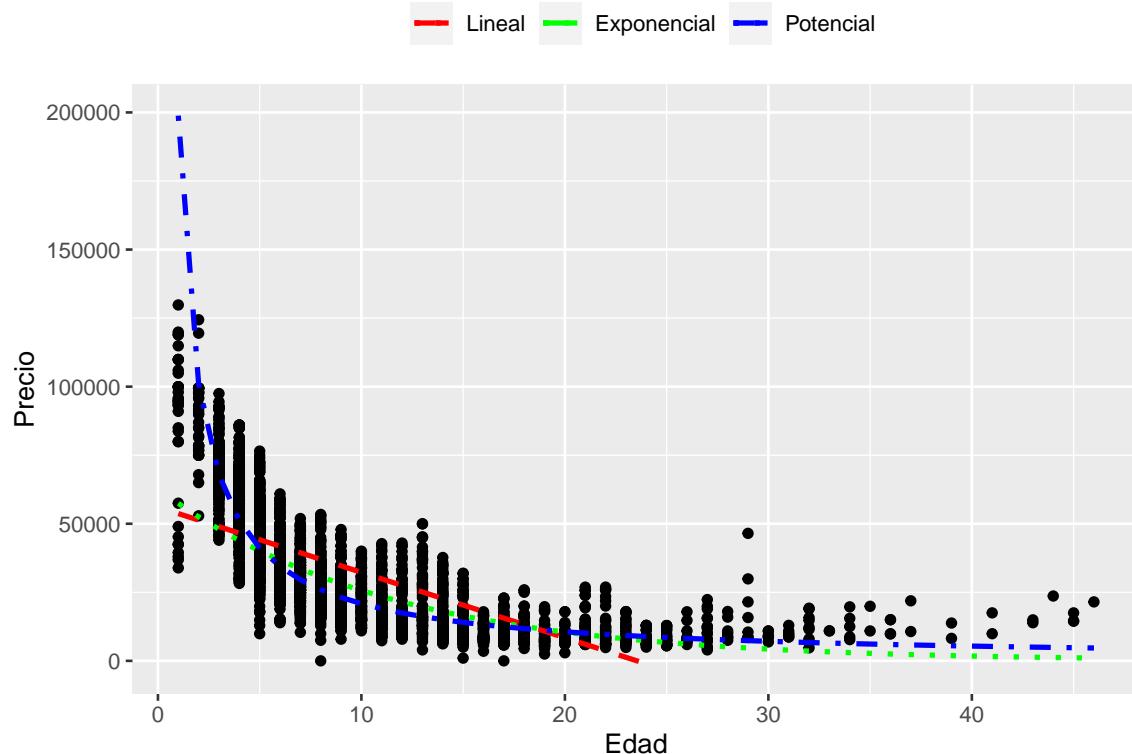
Figura 8.2: Intervalo de confianza de todos los modelos de los diferentes modelos de regresión lineal.

	2.5%	97.5%
Lineal- (Intercept)	55,318.660	56,879.650
Lineal - years	-2,455.887	-2,310.062
Exp - (Intercept)	11.023	11.073
Exp - years	-0.092	-0.087
Power - (Intercept)	12.158	12.242
Power - years	-0.999	-0.959

Fuente: (Elaboración propia)

También, se calculó el intervalo de confianza de los resultados previamente calculados, estos nos indican que con una confianza del 95 %, la pendiente se encuentra entre esos valores.

Figura 8.3: Gráfico de dispersión de los datos de las cabezas tractoras a analizar y su respectivos modelos de regresión.



Fuente: (Elaboración propia)

En la figura 8.3 se ve como en los primeros años de vida de una máquina, los valores caen rápidamente con los modelos exponencial y de potencia (depreciación acelerada), teniendo esta ultima inicialmente una pendiente más pronunciada. Posteriormente, sus curvas se suavizan (depreciación moderada), mientras que el modelo lineal tiene un descenso constante de los valores.

Capítulo 9

Modelos de regresión multivariable

Como se ha ido explicando a lo largo del informe, el objetivo de este trabajo de fin de grado es analizar el comportamiento de los valores y depreciación en el mercado de la maquinaria de ocasión, para esto resulta fundamental averiguar y explicar la relación entre una o varias variables con el valor de las cabezas tractoras. Después de encontrar en el anterior capítulo la relación entre el precio y la edad de la primera matriculación, en este capítulo se buscará encontrar la posible relación de multiples variables (kilometraje, brand) con respecto al precio de la maquinaria; para esto se llevarán a cabo diferentes modelos de regresión multivariable utilizando las ecuaciones anteriormente planteadas en la página 34, tabla 8.1. No se tendrán en cuenta la potencia, al no tener suficientes observaciones para llevar a cabo un modelo fiable.

Antes de llevar a cabo la modelización de los diferentes modelos, es fundamental aclarar que el coeficiente de determinación R^2 múltiple se encarga de describir la relación entre las variables predictoras y la variable respuesta. Al igual que el R^2 , este está acotado entre 0 y 1, donde 1 indica una relación lineal perfecta, mientras que un R^2 múltiple de 0 indica que no hay ninguna relación lineal.

Así mismo, también debemos tener presente la posibilidad de encontrar problemas de multicolinealidad, en donde las variables endógenas independientes en el modelo de regresión estén correlacionadas.

9.1. Modelización del precio respecto a la edad y el kilometraje

Como en el capítulo anterior, se llevaron a cabo tres modelos de regresión lineal por mínimos cuadrados ordinarios siguiendo las ecuaciones planteadas en la tabla 7.1 en la página 24. Esto con el fin de analizar el efecto y relación de dos variables endógenas (los años que han pasado desde la primera matriculación y el kilometraje) con respecto a la variable exógena (precio de venta).

Figura 9.1: Modelización del precio respecto a la edad y el kilometraje.

	<i>Dependent variable:</i>		
	price Modelo1 (1)	log(price) Modelo2 (2)	Modelo3 (3)
years	-3,274.508*** (60.313)	-0.098*** (0.002)	
km		-0.050*** (0.001)	-0.00000*** (0.00000)
years:km		0.002*** (0.0001)	0.00000*** (0.000)
log(years)			-0.570*** (0.078)
log(km)			0.100*** (0.017)
log(years):log(km)			-0.033*** (0.006)
Constant	82,769.700*** (769.233)	11.547*** (0.027)	10.929*** (0.211)
Observations	4,173	4,173	4,173
R ²	0.654	0.637	0.688
Residual Std. Error	12,129.700	0.433	0.402
F Statistic	2,621.689***	2,443.826***	3,059.024***

Note:

* p<0.1; ** p<0.05; *** p<0.01

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Fuente: (Elaboración propia)

En la anterior tabla se puede observar como para todos los tipos de maquinaria y modelos muestran un nivel de significación de 1% “***”, Sin embargo, al evaluar los coeficientes estadísticos obtenidos, la variable kilometraje parece presentar el principio parsimonia (el modelo nunca puede llegar a ser una descripción completamente precisa de la realidad; para describir la realidad, es preciso desarrollar un modelo tan complejo que éste sería de poca utilidad práctica) en donde parece no aportar mucha información y por lo tanto, no genera diferencias considerables al modelo anteriormente planteado, es por ello que no resultaría muy interesante integrar esta variable al modelo, teniendo en cuenta que al incluir la interacción entre las variables podría conllevar a aumentar la

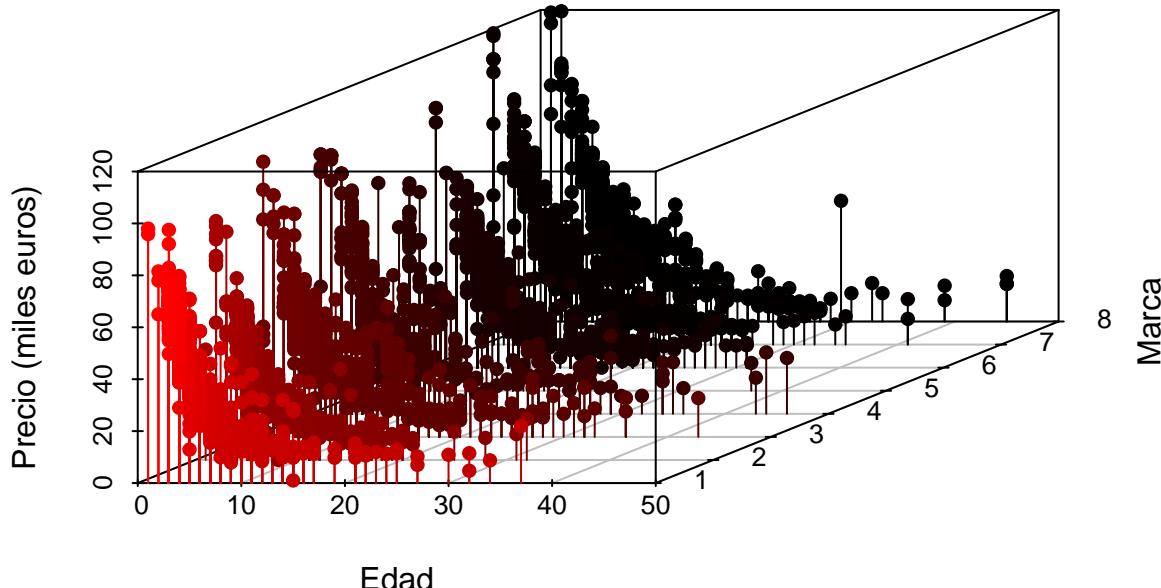
probabilidad de problemas de colinealidad y menor fiabilidad en el modelo.

Además, vemos como se reduce la precisión de las estimaciones, al aumentar el error estándar de los coeficientes de regresión.

9.2. Modelización del precio respecto a la edad y las marcas

Después de observar la presencia de principio parsimonia en el planteamiento anterior, se procedió a llevar cabo tres modelos de regresión lineal por mínimos cuadrados ordinarios siguiendo las ecuaciones planteadas en la tabla 7.1 de la página 24. Esto con el fin de analizar el efecto y relación de dos variables endógenas (los años que han pasado desde la primera matriculación y las marcas) con respecto a la variable exógena (precio de venta).

Figura 9.2: Regresión multivariadas del precio respecto a los años y la marca.



Fuente: (Elaboración propia)

En la figura 9.2 se representó el precio de la cabezas tractoras respecto a su edad y marca, tomando esta última valores del 1 al 8 correspondientes a las marcas DAF, Iveco, MAN, Mercedes, Otro, Renault, Scania, Volvo. Se observa como presentan un patrón de evolución general, teniendo todas las marcas una tendencia negativa.

Posteriormente, para llevar a cabo el análisis de regresión por factores, previamente se implementó el One-Hot-encoding a la marca, con el fin de convertir los datos categóricos

en numéricos. El método de codificación One-Hot nos permite etiquetar a qué clase pertenecen los datos, en donde se le asigna 0 a toda la dimensión, excepto para la clase a la que pertenecen los datos, que se le asigna 1 .

Tabla 9.1: One-Hot-encoding según la marca de cabeza tractora.

DAF	Iveco	MAN	Mercedes	Otro	Renault	Scania	Volvo
0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1
0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	0
0	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0

Según la codificación One-Hot, se estableció como variable referencia “DAF”, por lo que al tener n niveles (marcas) en los siguientes modelos aparecerán $n-1$ niveles, debido a que cuando todas las variables binarias tomen como valor 0, el modelo estaría prediciendo la variable de referencia “DAF”.

Figura 9.3: Diferentes modelos de regresión lineal multivariante.

	<i>Dependent variable:</i>		
	price Modelo1 (1)	log(price) Modelo2 (2)	Modelo3 (3)
years	-3,620.121*** (101.998)	-0.115*** (0.003)	
log(years)			-1.098*** (0.024)
Iveco	-12,045.770*** (1,373.850)	-0.135*** (0.044)	-0.030 (0.077)
MAN	-15,964.350*** (1,344.607)	-0.358*** (0.043)	-0.507*** (0.070)
Mercedes	-17,950.600*** (1,351.875)	-0.512*** (0.043)	-0.717*** (0.076)
Otro	-4,951.232* (2,935.286)	-0.137 (0.093)	-0.669*** (0.160)
Renault	-16,041.660*** (1,419.173)	-0.233*** (0.045)	-0.046 (0.075)
Scania	-5,835.191*** (1,400.732)	-0.097** (0.045)	-0.085 (0.072)
Volvo	-7,360.476*** (1,381.368)	-0.104** (0.044)	-0.002 (0.075)
years:Iveco	722.219*** (169.655)	-0.006 (0.005)	
years:MAN	1,641.344*** (135.545)	0.037*** (0.004)	
years:Mercedes	2,029.584*** (138.385)	0.065*** (0.004)	
years:Otro	1,373.660*** (360.253)	0.044*** (0.011)	
years:Renault	1,264.063*** (144.422)	0.005 (0.005)	
years:Scania	964.809*** (136.852)	0.018*** (0.004)	
years:Volvo	1,401.130*** (133.079)	0.030*** (0.004)	
log(years):Iveco			-0.054 (0.041)
log(years):MAN			0.239*** (0.035)
log(years):Mercedes			0.401*** (0.038)
log(years):Otro			0.453*** (0.091)
log(years):Renault			-0.050 (0.037)
log(years):Scania			0.086** (0.035)
log(years):Volvo			0.088** (0.036)
Constant	66,789.090*** (805.500)	11.272*** (0.026)	12.428*** (0.044)
Observations	4,173	4,173	4,173
R ²	0.554	0.631	0.720
Residual Std. Error	13,778.660	0.438	0.381
F Statistic	344.604***	473.807***	713.504***

Note:

* p<0.1; ** p<0.05; *** p<0.01

Fuente: (Elaboración propia)

Tras analizar la base de datos después de haberse aplicado la técnica de MAD para suprimir los valores atípicos, los modelos de regresión planteados ofrecen los resultados que se presentan en la figura 9.3. Como hemos observado en los modelos previos, si bien existen matrices diferenciadoras en cuanto al comportamiento del precio, los tres presentan una tendencia negativa en el comportamiento del precio de venta de la cabeza tractora con el paso del tiempo.

Se aprecia como en los tres modelos propuestos, el coeficiente R^2 ha incrementado su valor. Donde el primer modelo continúa presentando el menor coeficiente y, el tercer modelo (potencial) continua presentando el mayor coeficiente y por lo tanto, el que mejor se explica la variabilidad del efecto que presenta el precio de venta.

Se ve como para el modelo lineal y exponencial, casi todos los β presentan un nivel de significación de " *** " y " ** " salvo la marca "Otro" con un nivel de significación de "*" en modelo lineal y un nivel de significación superior a 10 % en el modelo exponencial; y la interacción entre la edad con la marca "Iveco" y con "Renault" que presentan un nivel de significación superior a 10 % en el modelo exponencial.

A diferencia de los dos primeros modelos, el potencial aunque es el que presenta mayor coeficiente de determinación, suele presentar niveles de significación inferiores, es decir, algunos de estos no presentan evidencias estadísticas de que afecten a la variable en cuestión.

Capítulo 10

Conclusiones

Para realizar el análisis de los valores y depreciación de cabezas tractoras en el mercado secundario de maquinaria, se ha perseguido una serie de objetivos parciales, necesarios para la extracción y recolección de los datos así como la gestión de los mismos. Para esto, fue fundamental el desarrollo de la técnica WebScraping, la cual permitió acceder a una mayor cantidad de datos.

La mayor disponibilidad de observaciones permitió demostrar que el comportamiento de la depreciación de la maquinaria de transporte y construcción concretamente en el mercado de tractores usados, muestran su patrón de depreciación según el mercado.

Posteriormente, se llevó a cabo la modelización del valor de las cabezas tractoras, en donde se encontró que la variable que mejor explicaba el comportamiento del precio era la edad de la primera matriculación. Una vez encontrada dicha variable, y establecido el modelo inicial, aún sabiendo que el modelo nunca puede llegar a ser una descripción completamente precisa de la realidad, se buscó obtener la relación de múltiples variables (kilometraje y marcas) con respecto al precio de la maquinaria; para aumentar la fiabilidad y robustez de modelo.

En el caso del análisis de la variable km, al comparar los tres modelos de regresión OLS, se observó que la variable kilometraje parece no generar diferencias considerables al modelo anteriormente planteado y presentar el principio parsimonia y que, el incluir dicha variable, podría conllevar a aumentar la probabilidad de problemas de colinealidad, menor fiabilidad en el modelo, así como la reducción de la precisión de las estimaciones y el incremento del error.

En cuanto al análisis de la variable Marca, al comparar los modelos de regresión lineal, exponencial y potencial, se observó que la variable Marca, parece presentar un patrón de evolución general por marca esperable, teniendo todas las marcas una tendencia negativa, siendo el modelo potencial la que mejor explica el comportamiento del precio de venta del activo, rechazando la hipótesis en la que se plantea la depreciación lineal como el modelo que mejor se ajusta a la realidad de la maquinaria de obra pública en España. Además, se vio que el coeficiente R² ha incrementado su valor, siendo el modelo potencial el que mejor explica la variabilidad del efecto que presenta el precio de venta.

Para finalizar, se ha visto como el desarrollo de este TFG ha permitido aumentar la comprensión de lenguaje R y así como en capacidades en el tratamiento de datos, la cual podría ser aplicable a cualquier caso de valoración de muebles e inmuebles. Resultaría

interesante disponer de una base de datos más amplia y actualizada, así como llevar a cabo la validacion cruzada de los modelos de regresión aplicada, ya que permitirían obtener una mayor robustez estadística a los diversos coeficientes de depreciación calculados para los activos.

Apéndice A

Apéndice A:

A.1. Código de las librerías utilizadas en R

```
rm(list=ls()) # clean
list.of.packages <- c("rvest", "magrittr", "tidyverse", "dplyr", "ggplot2", "tidy", "purrr", "tibble", "stringr", "forcats",
  "data.table", "robottxt", "readr", "xml2", "rmarkdown", "patchwork", "psych", "scatterplot3d", "broom", "DiagrammeR",
  "kableExtra", "knitr", "stargazer", "GGally", "data.table", "mltools")
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[, "Package"])]
if (length(new.packages))
  install.packages(new.packages)
library(rvest)
library(magrittr)
library(tidyverse)
library(dplyr)
library(ggplot2)
library(tidy)
library(purrr)
library(tibble)
library(stringr)
library(forcats)
library(data.table)
library(robottxt) #evaluar la accesibilidad de la pag
library(readr)
library(xml2) # analizar documentos XML
library(rmarkdown)
library(patchwork) #para combinar graficos
library(psych)
library(scatterplot3d)
library(broom)
library(DiagrammeR)
library(kableExtra)
library(knitr)
library(stargazer)
library(GGally)
library(data.table)
library(mltools)
rm(list.of.packages, new.packages)
```

A.2. Creación del diagrama

```
DiagrammeR::grViz("digraph {
  graph [layout = dot, rankdir = TB]

  node [shape = rectangle
    fontname = Helvetica
    penwidth = 2.0]
  rec1 [label = 'Step 1. Data gathering']
  rec2 [label = 'Step 2. Outlier treatment (MAD)']
  rec3 [label = 'Step 3. Clean database']
  rec4 [label = 'Step 4. OLS models']
  rec5 [label = 'Step 5. Best fit model']

  node [shape = rectangle penwidth = 0.5]
  rec4_1 [label = 'Linear model']
  rec4_2 [label = 'Exponential model']
  rec4_3 [label = 'Power model']")
```

```

# edge definitions with the node IDs # edge [arrowhead = diamond]
#rec1 -> rec2; rec2 -> rec3; rec3 -> rec4; rec4 -> rec4_1; rec4 -> rec4_2; rec4 -> rec4_3; rec4_1 -> rec5; rec4_2 -> rec5; rec4_3 -> rec5;
rec1 -> rec2 -> rec3 -> rec4; rec4 -> rec4_1 -> rec5; rec4 -> rec4_2 -> rec5; rec4 -> rec4_3 -> rec5
#si pongo en el edge [label = ''], me aparece un nombre en la linea, no en el cuadrado

graph [nodesep = 0.1]
}#, height = 500

```

A.3. Código del WebScraping en R

```

{
  get_marca = function(gen_link){
    gen_page = read_html(gen_link)
    brand = gen_page %>% html_nodes("#ts-table-features .text-right span , .under u , .pl-0 .mb-4:nth-child(1) tr:nth-child(2) .text-right ,
    .pl-0 .mb-4:nth-child(1) tr:nth-child(1) .text-right") %>% html_text()
    date_pattern = "[a-zA-Z]+"
    brand <- str_match(brand , date_pattern)
    brand = brand[4]
    brand <- str_trim(brand, side="both")
    return(brand)
  }

  get_pot = function(gen_link){
    gen_page = read_html(gen_link)
    power = gen_page %>% html_nodes(".mb-4:nth-child(1) .print-hidden:nth-child(4) .text-right , .mb-4:nth-child(1) .print-hidden:nth-child(3)
    .text-right , .mb-4:nth-child(1) .print-hidden:nth-child(2) .text-right , .mb-4:nth-child(1) .print-hidden:nth-child(1)
    .text-right") %>% html_text() %>% paste(" ", .)
    date_pattern = "[0-9]+[-.]CV"
    power <- str_match(power , date_pattern)
    power <- power %>% na.omit()
    power = power[1]
    power = as.numeric(gsub('\\D+', '' , power))
    #%% str_split("CV") %% map_chr(1) #no lo borro, pero no lo pongo ahora por probar por si hay errores cuando no ponen la potencia
    return(power) #si en potencia no sale CV, es KM #puedo poner una de si no hay CV = NA
  }

  get_price = function(gen_link){
    gen_page = read_html(gen_link)
    price = gen_page %>% html_nodes(".color-primary.align-middle") %>% html_text()
    price= as.numeric(gsub('\\D+' , '' , price))
    price = price[1]
    return(price) #si el precio no sale numerico, saldra NA que corresponde a precio bajo pedido
  }

  get_km = function(gen_link){
    gen_page = read_html(gen_link)
    km = gen_page %>% html_nodes(".mb-4:nth-child(1) .print-hidden:nth-child(4) .text-right , .mb-4:nth-child(1) .print-hidden:nth-child(3)
    .text-right , .mb-4:nth-child(1) .print-hidden:nth-child(2) .text-right , .mb-4:nth-child(1) .print-hidden:nth-child(1) .text-right")
    %>% html_text() %>% paste(" ", .)
    date_pattern = "([0-9]{1,3})[-.]" + km"
    km <- str_match(km , date_pattern)
    km <- km %>% na.omit() #eliminate each row that has at least one NA (outliers)
    km = km[1]
    km = as.numeric(gsub('\\D+' , '' , km))
    return(km) #si no tiene km = NA,
  }

  get_year = function(gen_link){
    gen_page = read_html(gen_link)
    year = gen_page %>% html_nodes("#ts-table-features .text-right") %>% html_text() %>% paste("", , collapse="")
    #year <- str_replace(year, "/", 3)
    #year = year[1]
    date_pattern = "[0-9]{4}"
    year<- str_match(year, date_pattern)
    if (length(year)>22){
      year <- year %>% na.omit() #eliminate each row that has at least one NA (outliers)
      year = year[,2]
    } else {
      year <- year
    }
    return(year)
  }

  #-----Code-----
  datos = data.frame()
  datos1 = data.frame()
  datos2 = data.frame()
  dato = data.frame()

  #ask if is allowed to download the information --> binary answer if the answer is false you are fuck
  link <- "https://www.europa-camiones.com/cabezas-tractoras-usadas/1-31/anuncios-cabezas-tractoras.html?"
  paths_allowed(paths = c(link))
  page = read_html(link)
  #page_result=55
  #for loop
  for(page_result in seq(from = 309, to = 316, by = 1)){
    link = paste0("https://www.europa-camiones.com/cabezas-tractoras-usadas/1-31/anuncios-cabezas-tractoras.html?p=",
    page_result) #page by default has space between strings that you are trying to concatenate
    page = read_html(link)
  }
}

```

```

city<-0
country <-0
brand<-0
km<-0
name<-0
price<-0
year<-0
power<-0
gen_link<-0

#assign the class,
name = page %>% html_nodes(".card-title") %>% html_text() %>% str_trim()
country = page %>% html_nodes(".meta-location") %>% html_text() %>% str_split("-") %>% map_chr(1) %>% str_trim(side="both")
{
  city = page %>% html_nodes(".meta-location") %>% html_text() %>% str_split("-") %>% map_chr(2)
  city <- str_trim(city,side="both")
  date_pattern = "([\\a-zA-Z\\u00d1\\u00d1\\.\\-]{1,24}) #([\\a-zA-Z\\-\\.]{1,24})"
  #24 es un numero random, es para que incluya Rio de Loba- Viseu
  city <- str_match(city, date_pattern)
  city <- city[,2]
  city <- str_trim(city,side="both")
}

#-----
#obtain the links of each vehicle to search more information
gen_link <- page %>% html_nodes(".card-body") %>% html_elements("[href]") %>% html_attr("href")
date_pattern = "(/([a-zA-Z0-9\\-.]+)+/([a-zA-Z]+).html"
gen_link <- str_match(gen_link , date_pattern)
gen_link <- gen_link %>% na.omit()
gen_link = gen_link[,1]
gen_link <- gen_link %>% paste("https://www.europa-camiones.com", , sep="")#put the link you get after www-europa...
gen_link

#-----
#with the links, it will go to each one, execute the brand function and the result will put it in a vector,
#we put the USERNAMES = FALSE so that the link does not appear in the vector
brand = sapply(gen_link, FUN = get_marca, USE.NAMES = FALSE)# to avoid the url -->FALSE
price = sapply(gen_link, FUN = get_price, USE.NAMES = FALSE)# to avoid the url -->FALSE
km = sapply(gen_link, FUN = get_km, USE.NAMES = FALSE)# to avoid the url -->FALSE
#year <- c(seq(1:21))
year = sapply(gen_link, FUN = get_year, USE.NAMES = FALSE)# to avoid the url -->FALSE
year = year[2,]

#power -
{
  power = sapply(gen_link, FUN = get_pot, USE.NAMES = FALSE)
  #date_pattern = "Pot/(0-9){3}[-.]CV"
  #power <- str_match(power , date_pattern)
  #power = power[,1]
  #power = as.numeric(gsub('\\D+', '', power))
}
#view(power)
#-----

#datos = rbind(datos, data.frame(name, brand, country, city, power, price, km, year, stringAsFactors=FALSE))
#rbind para no crearlo siempre, solo ir adjuntando
datos1 = rbind(datos1, data.frame(name, brand, country, city, power, price, km, year, gen_link, stringAsFactors=FALSE))
#rbind para no crearlo siempre, solo ir adjuntando
#datos = rbind(datos1, data.frame(name, country, city, gen_link))
print(paste("Page: ", page_result)) #tracking problems
#tStart<-Sys.time()
Sys.sleep(time=runif(1, min=14, max=24))
#tFin<- Sys.time() - tStart
}

sapply(datos,class)
datos1 = datos1[,-9]
datos2 = rbind(datos,datos1)
datos = datos2

write.csv(datos, file= 'WS.csv')    #exporting data to a CSV file

#-----NOTES-----
#html_text we retrieve all the information
#html_nodes go to the source code of the web page and look for the node that we want, and remove it from us
#str_trim() put the text in the correct form, i.e. remove //n...
#we use str_split to cut and use the first part with map_chr(1) instead of using gsub

```

Apéndice B

Apéndice B: Códigos implementados

B.1. Código del Fichero: capítulo05.Rmd

```
options(readr.show_col_types = FALSE) #to quiet the full column specification
## Importing data from a CVS file
data_WebScraping <- suppressMessages(read_csv("WebScraping.csv",
                                             locale = locale(encoding = "ISO-8859-1",
                                                             asciiify = TRUE), na = "NA"))
data_WebScraping <- select(data_WebScraping, -1, -stringAsFactors) # to display on the screen

datos <- data_WebScraping # to modify
data_WebScraping <-%>% rename ("power (CV)" = colnames(data_WebScraping)[5], "price (€)" = colnames(data_WebScraping)[6])
#view(data_WebScraping)

#----- CAPITULO_05 -----
#resume the number of NA
resume = data.frame(t(colSums(is.na(datos)))) #to obtain for each column the number of empty data (NA)
resume = cbind(resume, dim(datos)[1]) #we can see the total number of data (row)

datos <- datos[!is.na(datos$price), ] #omit rows with NA in specific column of data frame
datos <- datos[!is.na(datos$year), ]
datos <- datos[!is.na(datos$km), ]

resume1 = data.frame(t(colSums(is.na(datos))))
resume1 = cbind(resume1, dim(datos)[1])
resume = rbind(resume, resume1)

rownames(resume) <- c("Con NAs", "Quitando NAs")
colnames(resume)[9] <- "Total Datos"
view(resume) #cuadro resumen de quitar NA

#create variable (years and years.category)
date <- Sys.Date() ## [1] "2022-07-11"
date <- as.numeric(format(date, 'Y'))
datos <- datos %>% mutate(years = date-year) #creamos la columna con los años que tiene el camion funcionando

datos$years.category <- cut(datos$years, breaks = c(-1, 3, 15, 30, Inf), labels = c("New", "Medium", "Old", "Very old"))
# Convertir var numerica "years" en categorica usando puntos de corte

datos <- datos %>% filter(years>-1 & years<100)

sapply(datos,class) #we can see the different number of classes
summary(datos)
resume04<- rbind(describe(datos$power), describe(datos$price), describe(datos$km), describe(datos$years))
#similar as summary but it show the result as a dataframe
rownames(resume04) <- c("Power", "Price", "Km", "Year" )
view(resume04)

#-----#
list_marca <- datos[2]
datos <- datos %>% mutate(list_marca = ifelse(brand == "Est", "Otro", brand))
datos <- datos [, c ("name", "list_marca", "country", "city", "power", "price", "km", "year", "years", "years.category")]
colnames(datos)[2] <- "brand"

rm(date, resume1, list_marca) #elimino variable no utilizadas

##-----#
```

B.2. Código del Fichero: capítulo06.Rmd

```

#----- CAPITULO_06 -----
#-----DETECTION OF OUTLIERS-----
### https://universeofdatascience.com/how-to-remove-outliers-from-data-in-r/
### https://www.adictosaltrabajo.com/2019/11/28/deteccion-y-reemplazo-de-outliers-con-r/
#https://aakinshin.net/posts/harrell-davis-double-mad-outlier-detector/
#https://www.reneshbedre.com/blog/find-outliers.html
#https://universeofdatascience.com/how-to-remove-outliers-from-data-in-r/

#-----#
#dim(datos) ## la dimension es de [1] filas=4615 y columnas=2
AGE <- unique(datos$years) # obtenemos los AGEs por los que partir el dataframe

#get threshold values for outliers
list_quantiles <- tapply(datos$price, datos$years, quantile)
Q1s <- sapply(1:length(AGE), function(i) list_quantiles[[i]][2])
Q3s <- sapply(1:length(AGE), function(i) list_quantiles[[i]][4])
IQRs <- tapply(datos$price, datos$years, IQR)
minis <- sapply(1:length(AGE), function(i) list_quantiles[[i]][1])
maxis <- sapply(1:length(AGE), function(i) list_quantiles[[i]][5])
Mes <- sapply(1:length(AGE), function(i) list_quantiles[[i]][3]) #median
MAD <- tapply(datos$price, datos$years, mad)

#Intercuartil
Lowers <- Q1s - 1.5*IQRs #valor atípico leve dista --> q < Q1 - 1,5 · IQR o bien q > Q3 + 1,5 · IQR
Uppers <- Q3s + 1.5*IQRs

#Hampel filter -->MAD
lower_bound <- Mes - 3 * MAD
Upper_bound <- Mes + 3 * MAD

#for to collect the data that pass the filter
list_year <- split(datos, datos$years)
data_no_outlier <- NULL
data_no_outlier_Mad <- NULL
for (i in 1:length(AGE)){
  out <- subset(list_year[[i]], list_year[[i]]$price > Lower[i] & list_year[[i]]$price < Uppers[i]) #Intercuartile range
  data_no_outlier <- rbind(data_no_outlier, out)
  #data_no_outlier <- rbind(out)
  outMad <- subset(list_year[[i]], list_year[[i]]$price > lower_bound[i] & list_year[[i]]$price < Upper_bound[i]) #Hampel filter -->MAD
  data_no_outlier_Mad<- rbind(data_no_outlier_Mad, outMad)
}
#dim(data_no_outlier) ## [1] 4297   2
#dim(data_no_outlier_Mad) ## [1] 4188   2

#-----
#elimino
rm(AGE, i, IQRs, lower_bound, Lower, MAD, maxis, Minis, Q1s, Q3s, Upper_bound, Uppers, out, outMad)
#-----

#grafico
par(mfrow = c(1,1))
#boxplot(data_no_outlier$price~data_no_outlier$years)
boxplot(data_no_outlier_Mad$price~data_no_outlier_Mad$years,outline=FALSE)
boxplot(data_no_outlier_Mad$price~data_no_outlier_Mad$years.category,outline=FALSE)
##-
list_marca <- data_no_outlier_Mad[2]
data_no_outlier_Mad <- data_no_outlier_Mad %>% mutate(list_marca = ifelse(brand == "Est", "Otro", brand))
data_no_outlier_Mad <-
  data_no_outlier_Mad [, c ("name", "list_marca", "country", "city", "power", "price", "km", "year", "years", "years.category")]
colnames(data_no_outlier_Mad)[2] <- "brand"

```

B.3. Código del Fichero: capítulo07.Rmd

```

#----- filtro datos segun una condicion y almaceno valores en listas -----
#agrupa por paises sin hacer filtro
df_brand <- data_no_outlier_Mad %>% group_by(brand) %>%
  summarise( n=n(), mean_year = mean(year), mean_years= mean(years), mean_price = ceiling(mean(km)), median_price=median(km),
             variance_price = var(km), sd_price = sd(km)) %>% arrange(n)
df_country <- data_no_outlier_Mad %>% group_by(country) %>%
  summarise( n=n(), mean_year = ceiling(mean(year)), mean_years= ceiling(mean(years)), mean_price = ceiling(mean(km)),
             median_price=median(km), variance_price = var(km), sd_price = sd(km)) %>% arrange(n)

#calcula las variables, elimina aquellas que no cumplan condiciones propuestas, como el no tener mas de dos valores por marca, por pais,
# uno agrupa por paises y otro por marcas y dice cuanto hay para cada uno, ademas saca la media de km y la suma total
df_country <- data_no_outlier_Mad %>% group_by(country) %>%
  summarise( n=n(), mean_year = ceiling(mean(year)), mean_years= ceiling(mean(years)), mean_price = ceiling(mean(km)),
             median_price=median(km), variance_price = var(km), sd_price = sd(km)) %>% filter(n>2) %>% arrange(n)
df_brand <- data_no_outlier_Mad %>% group_by(brand) %>%
  summarise( n=n(), mean_year = mean(year), mean_years= mean(years), mean_price = ceiling(mean(km)), median_price=median(km),
             variance_price = var(km), sd_price = sd(km)) %>% filter(n>2) %>% arrange(n)

#agrupa en un vector aquellos paises que tengan mas de dos cabezas tractoras y en otro aquellas marcas que tengan mas de dos cabezas tractoras
countries <- as.vector(data.frame(t(df_country[,1])))
brands <- as.vector(data.frame(t(df_brand[,1])))

#en data_no_outlier_Mad solo mantendremos los valores que cumplen las dos condiciones anteriormente planteadas... y
#dim(data_no_outlier_Mad) # 4188 10
data_no_outlier_Mad <- filter(data_no_outlier_Mad, country %in% countries, brand %in% brands)
#dim(data_no_outlier_Mad) # 4173 10

#summary as a dataframe of power, country, years and brand
resume07<- rbind(describe(data_no_outlier_Mad$power), describe(data_no_outlier_Mad$price), describe(data_no_outlier_Mad$km),
describe(data_no_outlier_Mad$years))
rownames(resume07) <- c("Power", "Price", "Km", "Year" )

#separa en una lista segun power, country, years y brand
list_power <- split(data_no_outlier_Mad, data_no_outlier_Mad$power)
list_country <- split(data_no_outlier_Mad, data_no_outlier_Mad$country)
list_years <- split(data_no_outlier_Mad, data_no_outlier_Mad$years)
list_brand <- split(data_no_outlier_Mad, data_no_outlier_Mad$brand)

#del cuadro que esta, para el analisis, sacamos un resumen de cada variable y la almacenamos en la lista_resumen:
df_year <- data_no_outlier_Mad %>% group_by(years) %>% summarise( n=n(), mean_price = ceiling(mean(price)))
years = df_year[,1:2]
brands = df_brand[,1:2]
countries = df_country[,1:2]
years.categories <- data_no_outlier_Mad %>% group_by(years.category) %>% summarise( n=n()) %>% arrange(n)
lista_resumen <- list(years, brands, countries, years.categories) #accedo a este lista_resumen[[2]]
rm(yearss, countries, years.categories, brands)
#----- plots
#plots
##Bar graph
par(mfrow = c(2,1))
ggplot(df_country, aes(x = reorder(country, n), y = n, fill = country)) + geom_col()
+ geom_text(aes(label = n), hjust = +0.7, colour = "black") + coord_flip()
+ theme(legend.position = "top", text = element_text (size = 12), legend.text = element_text (size = 7))
+ labs(fill = element_blank(), x = "País", y = "Número de observaciones") #numbers of cabezas tractoras for every country
ggplot(df_country, aes(x = reorder(country, mean_years), y = mean_years, colour = country)) + geom_point()
+ coord_flip() + theme(legend.position='none', text = element_text (size = 12))
+ labs(x="País", y = "Promedio de edad") #edad promedio de cabezas tractoras for every country

data_no_outlier_Mad %>%
  ggplot() +
  aes(x = brand, fill= country) +
  geom_bar(color="black") +
  facet_grid(country ~ .) +
  geom_text(aes(label = ..count..), stat = "count", vjust = -0.1, colour = "black") +
  theme(legend.position = "top") +
  labs(fill = "Paises", x="Marca", y = "Número de observaciones")

ggplot(data_no_outlier_Mad, aes(x = years, y = price, colour = years.category)) +
  geom_point() +
  theme(legend.position = "top") +#scatter plots y grafico de dispersion
  labs(colour = element_blank(), x="Edad", y = "Precio")

ggplot(data_no_outlier_Mad, aes(x = power, fill = years.category)) +
  geom_histogram(colour = "black") +
  geom_density() +
  facet_grid(years.category ~ ., scales = "free") +    ###histograma
  theme(legend.position = "top")+
  labs(fill = "Tramo de edad", x="Potencia (CV)", y = "Número de observaciones")

#borro cosas que no uso
rm(df_brand, df_country, df_year, resume, resume04, resume07)

#https://www.statology.org/multiple-linear-regression-r/
sal <- data_no_outlier_Mad[,c(6,7,9)]
# library(GGally)
colnames(sal) <- c("Precio", "km", "Edad")
# pairs(sal[-1], pch = 18, col = "steelblue") #pairs() function to create a scatterplot of every possible pair of variables:
ggpairs(sal) #create a similar plot that contains the actual linear correlation coefficients for each pair of variables

```

B.4. Código del Fichero: capítulo08.Rmd

```

#-----CAPITULO_08-----
#-----modelos de regresion lineal una variable, los comentados son el calculo manual sin funcion-----
resume08<- rbind(describe(data_no_outlier_Mad$years), describe(data_no_outlier_Mad$price), describe(data_no_outlier_Mad$km))
#similar as summary but it show the result as a dataframe
rownames(resume08) <- c("Year", "Price", "Km")

### Modelo 1
#-----
#lineal
#calculo manual, este da resultados con mas decimales
# b_lineal = cov(data_no_outlier_Mad$years, data_no_outlier_Mad$price) / (resume07 [4,4])^2
# a_lineal = resume07 [2,3] - (resume07 [4,3] * b_lineal)
# r_2_lineal = (cov(data_no_outlier_Mad$years, data_no_outlier_Mad$price) / ((resume07 [4,4]) * (resume07 [2,4])))^2
# r_2_lineal = (cor(data_no_outlier_Mad$years, data_no_outlier_Mad$price))^2 #da igual, pero es otra forma
# coef_depre_lineal = b_lineal / a_lineal
# data_no_outlier_Mad <- data_no_outlier_Mad%>% mutate(y_lineal = a_lineal + b_lineal * data_no_outlier_Mad$years) #Creamos la Y
#calculo con funcion lm
reg_years_price_lineal = lm(price ~ years, data = data_no_outlier_Mad) #nos da el a y el b
data_no_outlier_Mad <- data_no_outlier_Mad %>% mutate(y_lineal = fitted(reg_years_price_lineal)) #fitted nos da la Y del modelo
summary(reg_years_price_lineal)

#residuals(reg_years_price_lineal) #saco los valores residuales
lineal_years_price_coeff_beta <- reg_years_price_lineal$coefficients
lineal_years_price_coeff_depre <- round(lineal_years_price_coeff_beta[2]/lineal_years_price_coeff_beta[1], digits = 4)
lineal_years_price_coeff_R <- summary(reg_years_price_lineal)$r.square
#lineal_years_price_coeff_R <- summary(reg_years_price_lineal)$adj.r.square #para el R ajustado
#-----

# #exp
#manual
# b_exp = cov(data_no_outlier_Mad$years, log(data_no_outlier_Mad$price)) / (resume07 [4,4])^2
# a_exp = mean(log(data_no_outlier_Mad$price)) - (resume07 [4,3] * b_exp)
# a_exp = exp(a_exp)
# r_2_exp = (cov(data_no_outlier_Mad$years, log(data_no_outlier_Mad$price)) / (sd(log(data_no_outlier_Mad$price)) * (resume07 [4,4])))^2
# r_2_exp = (cor(data_no_outlier_Mad$years, log(data_no_outlier_Mad$price)))^2 #no se pq me da diferente a lo que da en el excel
# coef_depre_exp = exp(b_exp)

#con formula
reg_years_price_exp = lm(log(price) ~ years, data = data_no_outlier_Mad) #nos da el a y el b
data_no_outlier_Mad <- data_no_outlier_Mad %>% mutate(y_exp = exp(fitted(reg_years_price_exp))) #fitted nos da la Y del modelo
summary(reg_years_price_exp)

#residuals(reg_years_price_exp) #saco los valores residuales
exp_years_price_coeff_beta <- reg_years_price_exp$coefficients
exp_years_price_coeff_depre <- round(exp(exp_years_price_coeff_beta[2]), digits = 4)
exp_years_price_coeff_R <- summary(reg_years_price_exp)$r.square
#exp_years_price_coeff_R <- summary(reg_years_price_exp)$adj.r.square #para el R ajustado
#-----

#power
#manual
# b_power = cov(log(data_no_outlier_Mad$years), log(data_no_outlier_Mad$price)) / (sd(log(data_no_outlier_Mad$years)))^2
# a_power = mean(log(data_no_outlier_Mad$price)) - (mean(log(data_no_outlier_Mad$years)) * b_power)
# a_power = exp(a_power)
# r_2_power = (cov(log(data_no_outlier_Mad$years), log(data_no_outlier_Mad$price)))
# / (sd(log(data_no_outlier_Mad$years) * sd(log(data_no_outlier_Mad$price))))^2
# r_2_power = (cor(log(data_no_outlier_Mad$years), log(data_no_outlier_Mad$price)))^2
# coef_depre_power = b_power

#con formula
reg_years_price_power = lm(log(price) ~ log(years), data = data_no_outlier_Mad) #nos da el a y el b
data_no_outlier_Mad <- data_no_outlier_Mad %>% mutate(y_power = exp(fitted(reg_years_price_power))) #fitted nos da la Y
summary(reg_years_price_power)

#residuals(reg_years_price_power) #saco los valores residuales
power_years_price_coeff_beta <- reg_years_price_power$coefficients
power_years_price_coeff_depre <- round(power_years_price_coeff_beta[2], digits = 4)
power_years_price_coeff_R <- summary(reg_years_price_power)$r.square
#power_years_price_coeff_R <- summary(reg_years_price_power)$adj.r.square #para el R ajustado
#-----

#include models with stargazer
#stargazer(reg_years_price_lineal, reg_years_price_exp, reg_years_price_power, type="text", df=FALSE) #coef de a faltaria hacer el exp(a)
#https://www.jakeruss.com/cheatsheets/stargazer/
#https://rdrr.io/cran/stargazer/man/stargazer_stat_code_list.html #stargazer: list of statistic codes

my_models <- list()
my_models[[ 'reg_years_price_lineal']] <- lm(price ~ years, data = data_no_outlier_Mad)
my_models[[ 'reg_years_price_exp']] <- lm(log(price) ~ years, data = data_no_outlier_Mad)
my_models[[ 'reg_years_price_power']] <- lm(log(price) ~ log(years), data = data_no_outlier_Mad)
stargazer::stargazer(my_models, type = "text", column.labels = c("Model1", "Model2", "Model3"),
                      covariate.labels=c("Lineal(years)", "Log(years)", "Constante"), single.row = TRUE, df=FALSE,
                      omit.stat=c("LL", "adj.rsq"), no.space=FALSE, add.lines = list(c("Coef depre",
                      lineal_years_price_coeff_depre, exp_years_price_coeff_depre, power_years_price_coeff_depre)),
                      notes = c("-----",
                      "*** Significant at the 1 percent level.", "** Significant at the 5 percent level.", "* Significant at the 10 percent level."),
                      out="Modelos_regresion/Modelos_regresion.html")
knitr:::include_url("file:///home/macosal/Escritorio/MemoriaTFE/MemoriaMACM/Modelos_regresion/Modelos_regresion.html")
#keep.stat="n", solo deja observaciones
#star.cutoffs = c(0.05, 0.01, 0.001) # star.cutoffs = NULL by default
#df=FALSE hace que no salga el 4171 de valores que usa en el modelo del df
#title="Tabla 1. Diferentes modelos OLS",
#digits = 1 numeros de decimales de la tabla salvo los agregados

```

```

#-----#
#saco intervalo de confianza de todos los modelos
confi <- list()
confi[['Lineal1']] <- c((confint(reg_years_price_lineal))[1,1], (confint(reg_years_price_lineal))[1,2])
confi[['Lineal2']] <- c((confint(reg_years_price_lineal))[2,1], (confint(reg_years_price_lineal))[2,2])
confi[['Exp1']] <- c((confint(reg_years_price_exp))[1,1], (confint(reg_years_price_exp))[1,2])
confi[['Exp2']] <- c((confint(reg_years_price_exp))[2,1], (confint(reg_years_price_exp))[2,2])
confi[['Power1']] <- c((confint(reg_years_price_power))[1,1], (confint(reg_years_price_power))[1,2])
confi[['Power2']] <- c((confint(reg_years_price_power))[2,1], (confint(reg_years_price_power))[2,2])

confi<- t(as.data.frame(confi))
colnames(confi) <- c("2.5 %", "97.5 %")
rownames(confi) <- c("Lineal - (Intercept)", "Lineal - years",
                     "Exp - (Intercept)", "Exp - years",
                     "Power - (Intercept)", "Power - years")
stargazer::stargazer(confi, type = "text", align = TRUE, df=FALSE, out="Modelos_regresion/confi.html")
knitr::include_url("file:///home/macosmai/Escritorio/MemoriaTFE/MemoriaMACM/Modelos_regresion/confi.html")
# flip = TRUE,
# tidy(reg_years_price_lineal) #covariate.labels=c("", "2.5", "97.5","2.5", "97.5","2.5", "97.5),
# tidy(summary(reg_years_price_lineal))
# augment(reg_years_price_lineal)
# glance(reg_years_price_lineal)

#-----#
#tendencia que realmente sigue
ggplot( data = data_no_outlier_Mad,
        mapping = aes( x= years,
                      y = price))++
  geom_point()+
  geom_smooth()

#-----#
ggplot(data_no_outlier_Mad, aes(years)) + # basic graphical object
  geom_point(aes(y = price)) +
  geom_line(aes(y=y_lineal, colour="Lineal"), linetype=2, size=1) + # first layer
  geom_line(aes(y=y_exp, colour="Exponencial"), linetype=3, size=1) + # second layer
  geom_line(aes(y=y_power, colour="Potencial"), linetype=4, size=1)+ # second layer
  scale_color_manual("", 
                     breaks = c("Lineal", "Exponencial", "Potencial"),
                     values = c("red", "green", "blue"))+
  labs(title = "TITULO")+
  labs( x="Edad", y = "Precio")+
  theme(plot.title = element_text(hjust = 0.5), legend.position = "top")
#-----#
#other options
# <!-- plot2<- -->
# <!-- {plot(data_no_outlier_Mad$years, data_no_outlier_Mad$price) #, main = "TITULO" -->
# <!-- legend("topright", legend = c("lineal", "power", "semilog"), lty= c(2, 3, 4), lwd = 1, col = c("red", "blue", "green")) -->
# <!-- lines(data_no_outlier_Mad$years, data_no_outlier_Mad$y_lineal , col = "red", linetype=2) -->
# <!-- lines(data_no_outlier_Mad$years, data_no_outlier_Mad$y_exp , col = "blue", linetype=3) -->
# <!-- lines(data_no_outlier_Mad$years, data_no_outlier_Mad$y_power , col = "green", linetype=4)} -->

# <!-- coeff<- list() -->
# <!-- coeff[["Modelo1"]]<- reg_years_price_lineal$coefficients -->
# <!-- coeff[["Modelo2"]]<- reg_years_price_exp$coefficients -->
# <!-- coeff[["Modelo3"]]<- reg_years_price_power$coefficients -->
# <!-- coeff[["Modelo4"]]<- t(as.data.frame(c(lineal_years_price_coeff_depre, summary(reg_years_price_lineal)$r.square ))) -->
# <!-- coeff[["Modelo5"]]<- t(as.data.frame(c(exp_years_price_coeff_depre, summary(reg_years_price_exp)$r.square ))) -->
# <!-- coeff[["Modelo6"]]<- t(as.data.frame(c(power_years_price_coeff_depre, summary(reg_years_price_power)$r.square ))) -->
# <!-- dat <- rbind(coeff[["Modelo4"]], coeff[["Modelo5"]], coeff[["Modelo6"]]) -->
# <!-- colnames(dat) <- c("coeff_depre", "r.square") -->
# <!-- rownames(dat) <- c("lineal", "exp", "power") -->
# <!-- coeff <- t(as.data.frame(coeff[1:3])) -->
# <!-- colnames(coeff) <- c("B0", "B1") -->
# <!-- coeff <- cbind(dat,coeff) -->
# <!-- stargazer::stargazer(coeff, type = "text", align = TRUE, df=FALSE, style = "qje", out="coeff.html") -->

```

B.5. Código del Fichero: capítulo09.Rmd

```

#----- CAPITULO_09 -----
#----- modelos de regresion lineal multivariable-----

#relation between price, years and km
my_models <- list()
my_models[['reg_years_price_lineal']] <- lm(price ~ years * km, data = data_no_outlier_Mad)
my_models[['reg_years_price_exp']] <- lm(log(price) ~ years * km, data = data_no_outlier_Mad)
my_models[['reg_years_price_power']] <- lm(log(price) ~ log(years) * log(km), data = data_no_outlier_Mad)
stargazer::stargazer(my_models, type = "text", column.labels = c("Modelo1", "Modelo2", "Modelo3"), single.row = TRUE,
df=FALSE, omit.stat=c("LL","adj.rsq"), no.space=FALSE, notes = c("-----",
"*** Significant at the 1 percent level.", "** Significant at the 5 percent level.", "* Significant at the 10 percent level."),
out="Modelos_Regresion/Modelos_Regresion_multii1.html")

#column.labels = c("Modelo1", "Modelo2", "Modelo3"), covariate.labels=c("Lineal(years)", "Log(years)", "Constante"),
knitr::include_url("file:///home/macosmai/Escritorio/MemoriaTFE/MemoriaMACM/Modelos_Regresion/Modelos_Regresion_multii1.html")

# -----
levels <- c("DAF", "Iveco", "MAN", "Mercedes", "Otro", "Renault", "Scania", "Volvo")
data_no_outlier_Mad$brand_factor <- match(data_no_outlier_Mad$brand, levels)

attach(data_no_outlier_Mad)
scatterplot3d(x=years, y=brand_factor, z=price, pch=16, cex.lab=1, zlim=c(0, 120000, 60000),
highlight.3d=TRUE, type="h", xlab='Edad', ylab='Marca', zlab='Precio')

# -----
# library(data.table)
# library(mitoools)
#solo las 8 marcas (para mostrar el resultado del one-hot encoding. --> no para todas las muestras)
sal <- brands[1]
sal$brand <- as.factor(sal$brand)
onehot <- one_hot(as.data.table(sal))
colnames(onehot) <- c("DAF", "Iveco", "MAN", "Mercedes", "Otro", "Renault", "Scania", "Volvo")
# -----
#para las 8 marcas, considerando todos los datos
sal <- data_no_outlier_Mad[2]
sal$brand <- as.factor(sal$brand)
onehot <- one_hot(as.data.table(sal))
colnames(onehot) <- c("DAF", "Iveco", "MAN", "Mercedes", "Otro", "Renault", "Scania", "Volvo")

#no me deja descargar dummyVars
# dummy <- dummyVars(" - . ", data=sal)
# onehot <- data.frame(predict(dummy, onehot = data))

sal <- cbind(data_no_outlier_Mad[,c(1,6,7,9)], onehot)
#onehot-encoding: si tienes n niveles (marcas) te aparecen n-1, cuando no aparece DAF
#es porque es la marca de referencia, si todas las binarias son cero, el modelo predice para DAF
#broom::tidy(model1) #información estadística de los modelos.

#relation between price, years, brand and km, esto vendría a ser lo que hago despues con el onehot-encoding
my_models <- list()
sal <- cbind(data_no_outlier_Mad[,c(1,6,7,9)], onehot) #sin DAF, ya que si todas las binarias son cero, el modelo predice para DAF
my_models[['reg_years_price_lineal']] <- lm(price ~ years *(Iveco + MAN + Mercedes + Otro + Renault + Scania + Volvo), sal[,c(-1)])
my_models[['reg_years_price_exp']] <- lm(log(price) ~ years *(Iveco + MAN + Mercedes + Otro + Renault + Scania + Volvo), sal[,c(-1)])
my_models[['reg_years_price_power']] <- lm(log(price) ~ log(years) * (Iveco + MAN + Mercedes + Otro + Renault + Scania + Volvo), sal[,c(-1)])
stargazer::stargazer(my_models, type = "text", column.labels = c("Modelo1", "Modelo2", "Modelo3"), single.row = TRUE,
df=FALSE, omit.stat=c("LL","adj.rsq"), no.space=FALSE, out="Modelos_Regresion/Modelos_Regresion_multii5.html")
knitr::include_url("file:///home/macosmai/Escritorio/MemoriaTFE/MemoriaMACM/Modelos_Regresion/Modelos_Regresion_multii5.html")

# -----
#https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html
library(corrplot)
library(psych)
correlation <- round(cor(sal[,c(-1)]),3)
corrplot(correlation, method = 'number')
# corrplot.mixed(correlation)
# corrplot(correlation, method = 'circle', type = 'lower')
corrplot(correlation, method = 'pie', type = 'lower')
pairs.panels(sal[,c(-1)], pch=21)
round(cor(sal[,c(-1)]),3)

# -----
#diversas pruebas
# ``{r include= FALSE}
# #relation between price, years, brand and km, esto vendría a ser lo que hago despues con el onehot-encoding
# my_models <- list()
#sal <- cbind(data_no_outlier_Mad[,c(1,6,7,9)], onehot) #sin DAF, ya que si todas las binarias son cero, el modelo predice para DAF
# my_models[['reg_years_price_lineal']] <- lm(price ~ years * km * brand, data = data_no_outlier_Mad)
# my_models[['reg_years_price_exp']] <- lm(log(price) ~ years * km * brand, data = data_no_outlier_Mad)
# my_models[['reg_years_price_power']] <- lm(log(price) ~ log(years) * log(km) * brand, data = data_no_outlier_Mad)
# stargazer::stargazer(my_models, type = "text", single.row = TRUE, df=FALSE, omit.stat=c("LL","adj.rsq"), no.space=FALSE,
# out="Modelos_Regresion/Modelos_Regresion_multii4.html")
# knitr::include_url("file:///home/macosmai/Escritorio/MemoriaTFE/MemoriaMACM/Modelos_Regresion/Modelos_Regresion_multii4.html")
# ```

# pruebas de graficos 3d

# scatterplot3d(x=brand_factor, y=years, z=price, pch=16, cex.lab=1,
# highlight.3d=TRUE, type="h", xlab='brands', ylab='years', zlab='Price (\u20ac)')
#
# attach(data_no_outlier_Mad)
# scatterplot3d(x=brand_factor, y=price, z=years, pch=16, cex.lab=1,
# highlight.3d=TRUE, type="h", xlab='brands', ylab='Price (\u20ac)', zlab='years')
# ## years-km-price

```

```

# ````{r eval=TRUE,echo=FALSE}
# attach(data_no_outlier_Mad)
# scatterplot3d(x=years, y=km, z=price, pch=16, cex.lab=1,
#                 highlight.3d=TRUE, type="h", xlab='AGE',
#                 ylab='Distancia (km)', zlab='Price (€)')
#
# price-years-km
# attach(data_no_outlier_Mad)
# scatterplot3d(x=price, y=years, z=km, pch=16, cex.lab=1,
#                 highlight.3d=TRUE, type="h", xlab='AGE',
#                 ylab='Distancia (km)', zlab='Price (€)')
#
# years-power-price
# attach(data_no_outlier_Mad)
# scatterplot3d(x=years, y=power, z=price, pch=16, cex.lab=1,
#                 highlight.3d=TRUE, type="h", xlab='AGE',
#                 ylab='Distancia (km)', zlab='Price (€)')
#
# ## years-years.category-price
# attach(data_no_outlier_Mad)
# scatterplot3d(x=years, y=years.category, z=price, pch=16, cex.lab=1,
#                 highlight.3d=TRUE, type="h", xlab='AGE',
#                 ylab='Distancia (km)', zlab='Price (€)')
#
# years.category-price-power
# attach(data_no_outlier_Mad)
# scatterplot3d(x=years.category, y=price, z=power, pch=16, cex.lab=1,
#                 highlight.3d=TRUE, type="h", xlab='AGE category',
#                 ylab='Price (€)', zlab='power (CV)')
#
#
# pruebas de regresión multivariante
# ````{r include=FALSE}
# # relation between price and km
# ## #lineal
# reg_years_price_lineal_3d = lm(price ~ km, data = data_no_outlier_Mad)
# ###exp
# reg_years_price_exp_3d = lm(log(price) ~ km, data = data_no_outlier_Mad)
# ###power
# reg_years_price_exp_3d = lm(log(price) ~ log(km), data = data_no_outlier_Mad)
# ...
#
# ````{r include=FALSE}
# # relation between price and brand
# ## #lineal
# reg_years_price_lineal_3d = lm(price ~ brand, data = data_no_outlier_Mad)
# ###exp
# reg_years_price_exp_3d = lm(log(price) ~ brand, data = data_no_outlier_Mad)
# ...
#
# ````{r include=FALSE}
# #relation between price, years and km
# ###lineal
# reg_years_price_lineal_3d = lm(price ~ years + km, data = data_no_outlier_Mad)
# ###exp
# reg_years_price_lineal_3d = lm(log(price) ~ years + km, data = data_no_outlier_Mad)
# ###power
# reg_years_price_lineal_3d = lm(log(price) ~ log(years) + km, data = data_no_outlier_Mad)
# ###power
# reg_years_price_lineal_3d = lm(log(price) ~ log(years) + log(km), data = data_no_outlier_Mad)
# ...
#
# ````{r include=FALSE}
# #relation between price, years and brand
# ###lineal
# reg_years_price_lineal_3d = lm(price ~ years + brand, data = data_no_outlier_Mad)
# ###exp
# reg_years_price_lineal_3d = lm(log(price) ~ years + brand, data = data_no_outlier_Mad)
# ###power
# reg_years_price_lineal_3d = lm(log(price) ~ log(years) + brand, data = data_no_outlier_Mad)
# ...
#
# ````{r include=FALSE}
# ###lineal
# reg_years_price_lineal_3d = lm(price ~ years + km + brand, data = data_no_outlier_Mad)
# ###exp
# reg_years_price_lineal_3d = lm(log(price) ~ years + km + brand, data = data_no_outlier_Mad)
# ###power
# reg_years_price_lineal_3d = lm(log(price) ~ log(years) + km + brand, data = data_no_outlier_Mad)
# ###power2
# reg_years_price_lineal_3d = lm(log(price) ~ log(years) + log(km) + brand, data = data_no_outlier_Mad)
# ...

```

Apéndice C

Apéndice C: Modelos de regresión OLS

C.1. Modelos del precio respecto a la edad

Tabla C.1: Modelo lineal del precio respecto a la edad.

```
##  
## Call:  
## lm(formula = price ~ years, data = data_no_outlier_Mad)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -37034 -10684  -3035   8342  76084  
##  
## Coefficients:  
##                 Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 56099.16     398.10 140.92 <2e-16 ***  
## years       -2382.97     37.19 -64.08 <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 14630 on 4171 degrees of freedom  
## Multiple R-squared:  0.4961, Adjusted R-squared:  0.4959  
## F-statistic: 4106 on 1 and 4171 DF,  p-value: < 2.2e-16
```

Tabla C.2: Modelo exponencial del precio respecto a la edad.

```
##  
## Call:  
## lm(formula = log(price) ~ years, data = data_no_outlier_Mad)  
##  
## Residuals:  
##      Min       1Q    Median       3Q      Max  
## -10.3316 -0.2584   0.0101   0.2820   3.0458  
##  
## Coefficients:  
##                 Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 11.047701   0.012784 864.16 <2e-16 ***  
## years       -0.089514   0.001194 -74.95 <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.4697 on 4171 degrees of freedom  
## Multiple R-squared:  0.5739, Adjusted R-squared:  0.5738  
## F-statistic:  5618 on 1 and 4171 DF,  p-value: < 2.2e-16
```

Tabla C.3: Modelo potencial del precio respecto a la edad.

```
##  
## Call:  
## lm(formula = log(price) ~ log(years), data = data_no_outlier_Mad)  
##  
## Residuals:  
##      Min       1Q    Median       3Q      Max  
## -10.1651 -0.1910   0.0203   0.2106   1.8425  
##  
## Coefficients:  
##                 Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 12.20022   0.02133 572.10 <2e-16 ***  
## log(years) -0.97868   0.01028 -95.19 <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.404 on 4171 degrees of freedom  
## Multiple R-squared:  0.6848, Adjusted R-squared:  0.6847  
## F-statistic:  9061 on 1 and 4171 DF,  p-value: < 2.2e-16
```

C.2. Modelos del precio respecto a la edad, km y la marca

Tabla C.4: Modelo lineal del precio respecto a la edad, kilometraje y marca.

```
##
## Call:
## lm(formula = price ~ years * km * (Iveco + MAN + Mercedes + Otro +
##     Renault + Scania + Volvo), data = sal[, c(-1)])
##
## Residuals:
##    Min      1Q Median      3Q     Max
## -47613   -6706   -745   6383  75212
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               9.861e+04  1.447e+03  68.170 < 2e-16 ***
## years                  -4.380e+03  1.528e+02 -28.667 < 2e-16 ***
## km                     -6.608e-02  2.552e-03 -25.895 < 2e-16 ***
## Iveco                  -1.315e+04  2.684e+03 -4.898 1.00e-06 ***
## MAN                     -1.697e+04  2.480e+03 -6.844 8.84e-12 ***
## Mercedes                -1.243e+04  2.341e+03 -5.308 1.17e-07 ***
## Otro                     -3.304e+04  4.438e+03 -7.445 1.18e-13 ***
## Renault                 -2.427e+04  2.668e+03 -9.097 < 2e-16 ***
## Scania                  -7.353e+03  2.677e+03 -2.747 0.006047 **
## Volvo                   -1.036e+04  2.510e+03 -4.129 3.71e-05 ***
## years:km                 3.154e-03  2.150e-04 14.673 < 2e-16 ***
## years:Iveco              -4.069e+02  2.940e+02 -1.384 0.166447
## years:MAN                 9.686e+02  2.143e+02  4.519 6.40e-06 ***
## years:Mercedes            1.875e+03  2.016e+02  9.301 < 2e-16 ***
## years:Otro                 2.686e+03  3.612e+02  7.436 1.25e-13 ***
## years:Renault              1.205e+03  2.353e+02  5.123 3.15e-07 ***
## years:Scania               5.046e+02  2.159e+02  2.337 0.019466 *
## years:Volvo                1.176e+03  1.956e+02  6.012 2.00e-09 ***
## km:Iveco                  2.540e-03  4.775e-03  0.532 0.594757
## km:MAN                     7.754e-03  4.285e-03  1.809 0.070453 .
## km:Mercedes                2.773e-03  3.923e-03  0.707 0.479610
## km:Otro                     5.776e-02  1.308e-02  4.416 1.03e-05 ***
## km:Renault                 2.193e-02  4.556e-03  4.813 1.54e-06 ***
## km:Scania                  1.999e-02  4.072e-03  4.910 9.47e-07 ***
## km:Volvo                   1.856e-02  4.032e-03  4.604 4.26e-06 ***
## years:km:Iveco             8.548e-04  3.960e-04  2.159 0.030923 *
## years:km:MAN                -2.709e-04  3.161e-04 -0.857 0.391522
## years:km:Mercedes           -1.130e-03  3.138e-04 -3.603 0.000319 ***
## years:km:Otro                -4.581e-03  9.446e-04 -4.849 1.28e-06 ***
## years:km:Renault             -1.196e-03  3.497e-04 -3.421 0.000631 ***
```

```

## years:km:Scania -9.790e-04 2.988e-04 -3.277 0.001058 **
## years:km:Volvo -1.195e-03 2.823e-04 -4.232 2.36e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10650 on 4141 degrees of freedom
## Multiple R-squared: 0.7347, Adjusted R-squared: 0.7327
## F-statistic: 370 on 31 and 4141 DF, p-value: < 2.2e-16

```

Tabla C.5: Modelo exponencial del precio respecto a la edad, kilometraje y marca.

```

##
## Call:
## lm(formula = log(price) ~ years * km * (Iveco + MAN + Mercedes +
##     Otro + Renault + Scania + Volvo), data = sal[, c(-1)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.2152 -0.1677  0.0219  0.1916  2.4600
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.189e+01  5.329e-02 223.172 < 2e-16 ***
## years                  -1.136e-01  5.628e-03 -20.192 < 2e-16 ***
## km                     -1.333e-06  9.400e-08 -14.178 < 2e-16 ***
## Iveco                  -1.156e-02  9.887e-02 -0.117 0.906951
## MAN                    -3.334e-01  9.135e-02 -3.650 0.000266 ***
## Mercedes               -3.534e-01  8.624e-02 -4.098 4.24e-05 ***
## Otro                   -7.179e-01  1.635e-01 -4.391 1.16e-05 ***
## Renault                -3.703e-01  9.827e-02 -3.768 0.000167 ***
## Scania                 -2.865e-01  9.861e-02 -2.906 0.003684 **
## Volvo                  -4.236e-01  9.246e-02 -4.582 4.75e-06 ***
## years:km                4.462e-08  7.918e-09  5.635 1.87e-08 ***
## years:Iveco             -3.829e-02  1.083e-02 -3.536 0.000411 ***
## years:MAN                9.271e-03  7.896e-03  1.174 0.240410
## years:Mercedes           5.101e-02  7.426e-03  6.869 7.42e-12 ***
## years:Otro                6.061e-02  1.331e-02  4.555 5.39e-06 ***
## years:Renault             -3.800e-03  8.669e-03 -0.438 0.661170
## years:Scania              5.536e-03  7.952e-03  0.696 0.486375
## years:Volvo                2.123e-02  7.204e-03  2.947 0.003222 **
## km:Iveco                 -3.011e-07  1.759e-07 -1.712 0.086981 .
## km:MAN                    1.073e-07  1.579e-07  0.679 0.496866
## km:Mercedes               4.642e-08  1.445e-07  0.321 0.748052
## km:Otro                   1.341e-06  4.818e-07  2.783 0.005402 **
## km:Renault                4.198e-07  1.678e-07  2.501 0.012416 *
## km:Scania                 6.819e-07  1.500e-07  4.546 5.62e-06 ***

```

```

## km:Volvo          8.454e-07  1.485e-07  5.692 1.34e-08 ***
## years:km:Iveco   4.098e-08  1.459e-08  2.810 0.004984 **
## years:km:MAN     1.014e-08  1.164e-08  0.870 0.384101
## years:km:Mercedes -1.662e-08  1.156e-08 -1.437 0.150673
## years:km:Otro     -1.022e-07  3.480e-08 -2.936 0.003344 **
## years:km:Renault  -1.631e-08  1.288e-08 -1.266 0.205650
## years:km:Scania   -2.147e-08  1.100e-08 -1.951 0.051091 .
## years:km:Volvo    -2.956e-08  1.040e-08 -2.842 0.004499 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3923 on 4141 degrees of freedom
## Multiple R-squared:  0.7049, Adjusted R-squared:  0.7026
## F-statistic:  319 on 31 and 4141 DF,  p-value: < 2.2e-16

```

Tabla C.6: Modelo potencial del precio respecto a la edad, kilometraje y marca.

```

##
## Call:
## lm(formula = log(price) ~ log(years) * log(km) * (Iveco + MAN +
##   Mercedes + Otro + Renault + Scania + Volvo), data = sal[,,
##   c(-1)])
##
## Residuals:
##      Min        1Q     Median       3Q      Max
## -10.0157 -0.1691  0.0039  0.1870  1.7875
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                13.917101  0.909197 15.307 < 2e-16 ***
## log(years)                 -1.071049  0.381478 -2.808 0.005014 **
## log(km)                    -0.119515  0.070343 -1.699 0.089391 .
## Iveco                      1.945373  1.750326  1.111 0.266446
## MAN                        -2.577976  0.965547 -2.670 0.007616 **
## Mercedes                   -1.086143  1.098508 -0.989 0.322846
## Otro                       -3.768780  1.017495 -3.704 0.000215 ***
## Renault                     -2.507108  1.326675 -1.890 0.058859 .
## Scania                      -5.652403  1.131524 -4.995 6.11e-07 ***
## Volvo                      -5.078110  1.153674 -4.402 1.10e-05 ***
## log(years):log(km)          0.002038  0.029442  0.069 0.944825
## log(years):Iveco            -0.972918  0.713744 -1.363 0.172919
## log(years):MAN              0.387179  0.409423  0.946 0.344372
## log(years):Mercedes         1.253922  0.473524  2.648 0.008126 **
## log(years):Otro              0.870904  0.414863  2.099 0.035854 *
## log(years):Renault           -0.254488  0.529383 -0.481 0.630736
## log(years):Scania            1.549029  0.463049  3.345 0.000829 ***

```

```

## log(years):Volvo          1.056617  0.443036  2.385 0.017126 *
## log(km):Iveco           -0.152481  0.135402 -1.126 0.260172
## log(km):MAN              0.166551  0.074962  2.222 0.026349 *
## log(km):Mercedes         0.033436  0.085844  0.390 0.696924
## log(km):Otro              0.274701  0.081525  3.370 0.000760 ***
## log(km):Renault           0.195697  0.102902  1.902 0.057271 .
## log(km):Scania            0.436362  0.087574  4.983 6.52e-07 ***
## log(km):Volvo             0.398119  0.089628  4.442 9.15e-06 ***
## log(years):log(km):Iveco  0.070020  0.055111  1.271 0.203971
## log(years):log(km):MAN    -0.016557  0.031686 -0.523 0.601313
## log(years):log(km):Mercedes -0.067409  0.036914 -1.826 0.067911 .
## log(years):log(km):Otro   -0.052961  0.033443 -1.584 0.113355
## log(years):log(km):Renault 0.010206  0.041053  0.249 0.803685
## log(years):log(km):Scania  -0.118517  0.035800 -3.311 0.000939 ***
## log(years):log(km):Volvo   -0.080532  0.034473 -2.336 0.019532 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3698 on 4141 degrees of freedom
## Multiple R-squared:  0.7378, Adjusted R-squared:  0.7358
## F-statistic: 375.8 on 31 and 4141 DF,  p-value: < 2.2e-16

```

Apéndice D

Apéndice D: Objetivos de desarrollo sostenible

ANEXO

OBJETIVOS DE DESARROLLO SOSTENIBLE



Reflexión sobre la relación del TFG con los ODS en general y con el/los ODS más relacionados.

El Pacto Mundial de las Naciones Unidas desarrollada en el año 2000, es una iniciativa que pretende crear un mundo sostenible y sensibilizar a empresas para que estas “alineen sus estrategias y operaciones con Diez Principios Universales, sobre derechos humanos, normas laborales, medioambiente y lucha contra la corrupción”.(Agut and Del Pilar, 2015).

Los objetivos de este plan entraron en vigor en el año 2015, con la aprobación de la Agenda 2030, la cual estaría compuesta por 17 objetivos globales (Objetivos de Desarrollo Sostenible (*ODS*)) que permiten evaluar el impacto que presentan sus actividades en el ecosistema para con este asegurar el bienestar del planeta y mejorar la calidad de vida de las personas al tratar.

Figura D.1: Objetivos de Desarrollo Sostenible.



Fuente: ([NACIONES-UNIDAS, 2022](#))

Para el cumplimiento de dichos objetivos propuestos por la ONU, es necesario la involucración tanto del gobierno, el sector privado como la sociedad civil y la sociedad en general. Es por ello que se ha decidido aprovechar la realización del presente Trabajo de Fin de Grado para reflexionar la importancia de este en la consecución de los ODS.

El tema seleccionado, tiene varios objetivos de desarrollo sostenible ligados a su creación, principalmente, por la influencia que presenta en el futuro de las empresas los factores sociales y medioambientales. Dentro de estos objetivos ligados, inicialmente se puede destacar el objetivo número 1: Fin de la pobreza y el objetivo número 8: Trabajo decente y crecimiento económico.

En cuanto al objetivo nº 8, al estar transfiriendo maquinarias, se generarán unas ventas que enriquecerán a los propietarios y el dinero generado con el negocio, se podrá reinvertir, produciendo mayor riqueza fiscal y recuperación económica en el sector. Así mismo, se considera que está relacionado con el objetivo nº 1 al fomentar el crecimiento y formalización de PYMES permitiéndoles llevar a cabo sus actividades económicas gracias a la compra de estos productos a un precio considerablemente menor, logrando así un incremento de la actividad económica, una mayor oferta de empleo y con ello la reducción del desempleo.

ANEXO I. RELACIÓN DEL TRABAJO CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE DE LA AGENDA 2030

Anexo al Trabajo de Fin de Grado y Trabajo de Fin de Máster: Relación del trabajo con los Objetivos de Desarrollo Sostenible de la agenda 2030.

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza.	x			
ODS 2. Hambre cero.				x
ODS 3. Salud y bienestar.				x
ODS 4. Educación de calidad.				x
ODS 5. Igualdad de género.				x
ODS 6. Agua limpia y saneamiento.				x
ODS 7. Energía asequible y no contaminante.				x
ODS 8. Trabajo decente y crecimiento económico.	x			
ODS 9. Industria, innovación e infraestructuras.				x
ODS 10. Reducción de las desigualdades.			x	
ODS 11. Ciudades y comunidades sostenibles.				x
ODS 12. Producción y consumo responsables.				x
ODS 13. Acción por el clima.				x
ODS 14. Vida submarina.				x
ODS 15. Vida de ecosistemas terrestres.				x
ODS 16. Paz, justicia e instituciones sólidas.				x
ODS 17. Alianzas para lograr objetivos.				x

Descripción de la alineación del TFG/TFM con los ODS con un grado de relación más alto.

***Utilice tantas páginas como sea necesario.

Anexo al Trabajo de Fin de Grado y Trabajo de Fin de Máster: Relación del trabajo con los Objetivos de Desarrollo Sostenible de la agenda 2030. (Numere la página)

El tema seleccionado, tiene varios objetivos de desarrollo sostenible ligados a su creación, principalmente, por la influencia que presenta en el futuro de las empresas los factores sociales y medioambientales. Dentro de estos objetivos ligados, inicialmente se puede destacar el objetivo número 1: Fin de la pobreza y el objetivo número 8: Trabajo decente y crecimiento económico.

En cuanto al objetivo nº 8, al estar transfiriendo maquinarias, se generarán unas ventas que enriquecerán a los propietarios y el dinero generado con el negocio, se podrá reinvertir, produciendo mayor riqueza fiscal y recuperación económico en el sector. Así mismo, se considera que está relacionado con el objetivo nº 1 al fomentar el crecimiento y formalización de PYMES permitiéndoles llevar a cabo sus actividades económicas gracias a la compra de estos productos a un precio considerablemente menor, logrando así un incremento de la actividad económica, una mayor oferta de empleo y con ello la reducción del desempleo.

Bibliografía

- Agut, M. and Del Pilar, M. (2015). Objetivos de desarrollo sostenible (ods, 2015-2030) y agenda de desarrollo post 2015 a partir de los objetivos de desarrollo del milenio (2000-2015). *Valencia: Universidad de Valencia*.
- Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., and Iannone, R. (2022). *rmarkdown: Dynamic Documents for R*. R package version 2.17; Disponible en: <https://CRAN.R-project.org/package=rmarkdown>.
- Amat Rodrigo, J. (2016). Estadística y machine learning con r. Web - accedido el 20-09-2022; Disponible en: <https://github.com/JoaquinAmatRodrigo/Estadistica-con-R>.
- Aparicio, J., Martínez, M., and Morales, J. (2004). Modelos lineales aplicados en r. *Dto. Estadística, Matemáticas e Informática*.
- BOE (2014). Ley 27/2014, de 27 de noviembre, del impuesto sobre sociedades. Disponible en: [BOE-A-2014-12328](https://www.boe.es/eli/es/l/2014/11/27/27/con), <https://www.boe.es/eli/es/l/2014/11/27/27/con>.
- Chang, W. (2018). *R graphics cookbook: practical recipes for visualizing data*. O'Reilly Media.
- Frías Domínguez, M. D. (2008). Estadística. tema 2. modelos de regresión. Web - accedido el 20-09-2022; Disponible en: https://ocw.unican.es/pluginfile.php/1156/course_section/1396/tema_02.pdf.
- Gareth, J., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.
- Gil, C. (2018). R para profesionales de los datos: una introducción. Disponible en: https://datanalytics.com/libro_r/_main.pdf.
- Gil-Terrón Rodríguez, F. J. (2020). *Diseño y programación de un tablero de indicadores económicos*. PhD thesis, Universitat Politècnica de València.
- Hernández Barajas, F. and Mazo Lopera, M. (2020). *Ánalysis de regresión con R*. Universidad de Colombia. Disponible en: https://fhernanb.github.io/libro_regresion/.
- Hlavac, M. (2022). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2.3; Disponible en: <https://CRAN.R-project.org/package=stargazer>.

- León Merán, J. L. d. et al. (2019). Desafíos de la nueva economía desde la perspectiva del flujo circular de la renta. Disponible en: <https://uvadoc.uva.es/bitstream/handle/10324/36083/TFG-O-1619.pdf?sequence=1&isAllowed=y>.
- Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology*, 49(4):764–766.
- Luque-Calvo, P. L. (2017). *Escribir un Trabajo Fin de Estudios con R Markdown*. Disponible en: <http://destio.us.es/calvo>.
- Molina, G. and Rodrigo, M. F. (2009). El modelo de regresión lineal. Web - accedido el 20-09-2022; Disponible en: http://ocw.uv.es/ciencias-de-la-salud/pruebas-1/1-3/t_09nuevo.pdf.
- Muñoz García, J. A., Amón Uribe, I., et al. (2013). Técnicas para detección de outliers multivariantes.
- NACIONES-UNIDAS (2022). Objetivos de desarrollo sostenible. Web - accedido el 02-10-2022; Disponible en: <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>.
- Nava Rosillón, M. A. (2009). Análisis financiero: una herramienta clave para una gestión financiera eficiente. *Revista venezolana de Gerencia*, 14(48):606–628. Disponible en: http://ve.scielo.org/scielo.php?script=sci_arttext&pid=s1315-99842009000400009.
- Ortí Lahoz, A. et al. (1974). Los mercados de ocasión. una aproximación al proceso de formación de sus precios.
- Pedersen, T. L. (2022). *patchwork: The Composer of Plots*. <https://patchwork.data-imaginist.com>, <https://github.com/thomasp85/patchwork>.
- Peiró Biosca, A. (2015). *Modelización de curvas de depreciación en maquinaria usada*. PhD thesis, Universitat Politècnica de València.
- Peng, R. D. (2016). *R programming for data science*. Leanpub Victoria, BC, Canada.
- Piquerias, V. Y. (2015). *Coste, producción y mantenimiento de maquinaria para construcción*. Editorial de la Universidad Politécnica de Valencia.
- Postiguallo, D., Blasco, A., and Ribal, F. J. (2015). Modeling the depreciation rate of construction machinery. an ordinary least-squares approach and quantile regression.
- Postiguallo, D., Ribal, J., and Blasco, A. (2018). Case study: model for economic lifetime of public construction machinery with ordinary least squares regression and cross validation techniques. *Finance, Markets and Valuation*, 4(1):41–57.
- R-Core-Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponible en: <https://www.R-project.org/>.
- Reboyras Soleto, F. (2020). *Utilización de la base de datos Eurostat. Elaboración de informes automáticos con R*. PhD thesis, Universitat Politècnica de València.

- RStudio-Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA. Disponible en: <http://www.rstudio.com/>.
- Sanchez, G. (2013). Handling and processing strings in r. *Trowchez Editions, Berkeley*.
- Vía-Mobilis (2022). Europa-camiones. Web - accedido el 20-09-2022; Disponible en: <https://www.europa-camiones.com/>.
- Wickham, H. (2016a). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Disponible en: <https://ggplot2.tidyverse.org>.
- Wickham, H. (2016b). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Disponible en: <https://ggplot2.tidyverse.org>.
- Wickham, H. (2019). *Advanced r*. CRC press.
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., and Dunnington, D. (2022a). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. R package version 3.3.6; Disponible en: <https://CRAN.R-project.org/package=ggplot2>.
- Wickham, H., François, R., Henry, L., and Müller, K. (2022b). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.10; Disponible en: <https://CRAN.R-project.org/package=dplyr>.
- Wickham, H. and Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc.".
- Xie, Y. (2014). knitr: A comprehensive tool for reproducible research in R. In Stodden, V., Leisch, F., and Peng, R. D., editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC. Disponible en: <http://www.crcpress.com/product/isbn/9781466561595>.
- Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. Disponible en: <https://yihui.org/knitr/>.
- Xie, Y. (2016). *bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman and Hall/CRC, Boca Raton, Florida. Disponible en: <https://bookdown.org/yihui/bookdown>.
- Xie, Y. (2022a). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.29; Disponible en: <https://CRAN.R-project.org/package=bookdown>.
- Xie, Y. (2022b). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.40; Disponible en: <https://yihui.org/knitr/>.
- Xie, Y., Allaire, J., and Grolemund, G. (2018). *R Markdown: The Definitive Guide*. Chapman and Hall/CRC, Boca Raton, Florida. Disponible en: <https://bookdown.org/yihui/rmarkdown>.
- Xie, Y., Dervieux, C., and Riederer, E. (2020). *R Markdown Cookbook*. Chapman and Hall/CRC, Boca Raton, Florida. Disponible en: <https://bookdown.org/yihui/rmarkdown-cookbook>.

Bibliografía

Yang, J., Rahardja, S., and Fränti, P. (2019). Outlier detection: How to threshold outlier scores? Disponible en: <https://doi.org/10.1145/3371425.3371427>.

Zhu, H. (2021). *kableExtra: Construct Complex Table with kable and Pipe Syntax*. R package version 1.3.4; Disponible en: <https://CRAN.R-project.org/package=kableExtra>.