



FACULTAD DE ESTUDIOS ESTADÍSTICOS

GRADO EN ESTADÍSTICA APLICADA

Curso 2024/2025

Trabajo de Fin de Grado

**TÍTULO: ANÁLISIS GEOGRÁFICO Y TEMPORAL
DE LAS MATRICULACIONES DE TURISMOS EN
ESPAÑA**

Alumno: Gonzalo Flores Sevilla

Tutor: Fernando Pérez Contreras

Junio de 2025



UNIVERSIDAD COMPLUTENSE
MADRID

ÍNDICE

Resumen.....	5
1. INTRODUCCION.....	7
1.1-Contexto y Relevancia del Estudio.....	7
1.2-Objetivos del Estudio	8
1.3-Estructura del trabajo	9
1.4-Contribuciones Esperadas.....	10
2 Marco teórico.....	10
2.1-Análisis de datos de Matriculaciones de turismos.....	10
3. Metodología.....	10
3.1-Análisis Clúster: Clustering Jerárquico.....	11
3.2-Análisis Univariante de Series Temporales (modelo SARIMA).....	12
3.3-Análisis Multivariante de Series Temporales (Modelo VAR).....	14
4.1-Estructura y variables.....	16
4.2-Procesamiento de los datos:.....	17
4.2.1-Análisis de datos faltantes	17
4.2.2-Análisis de outliers	17
4.3-Análisis exploratorio de datos	18
5 Soluciones.....	20
5.1 Análisis Clúster.....	20
5.1.1 introducción al análisis.....	20
5.1.2-Resultados globales (junio 2020 – diciembre 2024)	21
5.1.3-Análisis por tipo de tecnología (año 2024).....	23
5.1.4-Turismos con combustibles fósiles (2024).....	25
5.1.5-Turismos con combustibles renovables (2024).....	26
5.1.6-Turismos con combustibles eléctricos (2024)	27
5.2 Análisis Univariante.....	29
5.2.1-Visualización de las series temporales	30
5.2.2-Análisis de la tendencia	31
5.2.3-Análisis de la homocedasticidad	33
5.2.4-Análisis de la estacionariedad	35

5.2.5-Análisis de la estacionalidad	38
5.2.6-Ajuste del modelo ganador	39
5.2.7-Evaluación de los modelos ajustados y diagnóstico de los residuos	41
5.2.8-Pronósticos para 2025	43
5.3 Análisis Multivariante.....	45
5.3.1-Análisis de la homocedasticidad	45
5.3.2-Proceso de desestacionalización.....	46
5.3.3-Análisis de estacionariedad	47
5.3.4-Test de causalidad de Granger.....	49
5.3.5-Ajuste del modelo VAR	50
5.3.6-Análisis Impulso-Respuesta.....	53
5.3.7-Descomposición de la Varianza.....	55
6. Conclusiones	57
7. Bibliografía.....	60
8. Anexo	61

RESUMEN.

En este Trabajo de Fin de Grado se analiza la evolución de las matriculaciones de los turismos nuevos entre el 20 de junio de 2020 hasta diciembre de 2024.

Los datos inicialmente incluían información desde 2015, pero finalmente, el análisis parte de junio de 2020 porque es cuando se declaró oficialmente el fin del estado de emergencia por COVID-19 y este periodo presenta una caída drástica de las matriculaciones, que afecta a las estimaciones y predicciones en el uso de series temporales, por ello se considera a partir de esa fecha.

El análisis se realiza sobre los turismos matriculados en España del tipo diésel, gasolina, gasolina híbridos, gasolina híbridos enchufables, diésel híbridos y sobre los eléctricos matriculados en España. Estos datos han sido obtenidos a partir de los microdatos provistos por la Dirección General de Tráfico, los cuales primero se depuran y clasifican por tipo de combustible y provincia.

Una vez depurados y transformados, se realiza un análisis clúster por tipo de vehículo y provincia, los cuales se agrupan en función de algunas variables demográficas extraídas del Instituto Nacional de Estadística, sobre las que se realizan también transformaciones para aplicarlas en el análisis.

A parte de este análisis geográfico, posteriormente, se aplica series temporales con el modelo *Seasonal Autorregressive Integreted Average* por cuatro tipos de combustión seleccionados.

Y para finalizar, un análisis multivariante en el que se usan los modelos vectoriales autoregresivos para estudiar posibles relaciones entre series.

Las conclusiones obtenidas por cada uno de los análisis permiten entender cómo ha evolucionado y como está evolucionado el parque automovilístico en España. Además, el uso de series temporales ofrecerá una proyección a futuro de las matriculaciones de los turismos nuevos por tipo de combustión.

Abstract.

This Final Degree Project analyzes the evolution of new passenger car registrations between June 20, 2020 and December 2024.

The data initially included information from 2015, but finally, the analysis starts from June 2020 because that is when the end of the state of emergency was officially declared by COVID-19 and this period presents a drastic drop in registrations, which affects the estimates and predictions in the use of time series, so it is considered from that date.

The analysis is carried out on diesel, gasoline, gasoline hybrids, gasoline plug-in hybrids, gasoline hybrids, diesel hybrids and electric cars registered in Spain. These data have been obtained from the microdata provided by the Dirección General de Tráfico, which are first cleaned and classified by fuel type and province.

Once cleaned and transformed, a cluster analysis is performed by type of vehicle and province, which are grouped according to some demographic variables extracted from the Instituto Nacional de Estadística, on which transformations are also carried out to apply them in the analysis.

In addition to this geographical analysis, time series are then applied with the Seasonal Autoregressive Integrated Average model for four selected types of combustion.

And finally, a multivariate analysis in which vector autoregressive models are used to study possible relationships between series.

The conclusions obtained from each of the analyses allow us to understand how the Spanish car fleet has evolved and how is evolving. In addition, the use of time series will provide a future projection of new car registrations by type of combustion.

1. INTRODUCCION.

En España, el parque automovilístico estos últimos años ha experimentado considerables cambios en el mercado de las nuevas matriculaciones. A parte de ir aumentando año tras año en cuanto volumen, también han llegado hace relativamente poco los vehículos eléctricos e híbridos que actualmente están en pleno auge y cada vez más se ven por las calles españolas.

Es por eso por lo que, realizar el análisis de las matriculaciones ofrece oportunidades para entender y estudiar cómo ha evolucionado cada tipo de combustible durante estos años. Además, hoy en día el mercado está evolucionando hacia una movilidad más sostenible con más turismos eléctricos e híbridos.

La pandemia COVID-19 provocó un fuerte impacto en el mercado automovilístico, ocasionando caídas abruptas en el volumen de las matriculaciones. Por eso, aunque se explicará más adelante, el análisis que se realiza en este trabajo será desde el momento en el que se declaró el fin del estado de emergencia el 20 de junio de 2020 para evitar distorsiones en el uso de series temporales (Crisis sanitaria COVID-19, 2020).

Del mismo modo, a partir de este año es cuando empieza a ser notable la presencia de vehículos más sostenibles.

Este estudio comienza con la obtención de los microdatos de la Dirección General de Tráfico (DGT, 2024) de todas las matriculaciones diarias del parque automovilístico español desde enero de 2015 hasta finales de 2024. Sin embargo, se filtrará desde un inicio por los turismos matriculados nuevos y a medida que vaya avanzando este trabajo se añaden más variables. Todos los análisis se realizan mediante el software estadístico R.

El objetivo final es comprender qué ha ocurrido durante estos años en España con los turismos, en que zonas geográficas se están comprando más turismos renovables (eléctricos e híbridos) y qué proyecciones habrá para 2025 (52 semanas).

1.1-Contexto y Relevancia del Estudio

Uno de los retos medioambientales principales de la mayoría de los países es la transición hacia un parque automovilístico más sostenible aumentando el número de matriculaciones de los eléctricos e híbridos y tratando de disminuir progresivamente los turismos de combustión fósil. Este objetivo está vinculado con varios Objetivos de Desarrollo Sostenible (ODS) definidos por la Organización de las Naciones Unidas (ONU, 2015) en la Agenda de 2030.

1. ODS 7: Energía asequible y no contaminante: Un parque automovilístico de turismos eléctricos será sostenible si va acompañada del uso de energías

renovables. Promover los coches eléctricos ayuda a reducir la demanda de combustibles fósiles.

2. ODS 9: Industria, Innovación e infraestructura: Fomenta la transición hacia vehículos eléctricos y la mejora de infraestructuras, como puntos de recarga para este tipo de transporte.
3. ODS 11: Ciudades y comunidades sostenibles: Los coches eléctricos, al no emitir gases contaminantes, contribuyen a reducir la contaminación del aire, mejorando la calidad de vida en las ciudades.
4. ODS 12: Producción y consumo responsables: Incentiva la movilidad eléctrica, que está en línea con la adopción de tecnologías más sostenibles.
5. ODS 13: Acción por el clima: Los vehículos eléctricos (si usan energía limpia) ayudan a reducir las emisiones de CO₂ del transporte, uno de los mayores contribuyentes del cambio climático.

Aunque los objetivos (ODS) no son exclusivamente para los coches eléctricos, su desarrollo y adopción son muy importantes. Ya que favorecen el uso de energías limpias (ODS 7), necesitan infraestructuras modernas como puntos de recarga (ODS 9), mejoran la calidad del aire en las ciudades (ODS 11), fomentan un consumo más responsable (ODS 12) y contribuyen directamente contra el cambio climático (ODS 13). Por lo tanto, los turismos eléctricos son una transición clave.

El plan España Auto 2030 es una de las iniciativas más relevantes e importantes promovida por el Gobierno y respaldada por la Asociación Española de Fabricantes de Automóviles y Camiones (ANFAC). El objetivo de este plan está centrado en el vehículo eléctrico y consiste en alcanzar metas como las 100.000 matriculaciones de turismos eléctricos en 2025. Se ayudará mediante incentivos y mejoras en las infraestructuras de carga de este tipo de vehículos (Soria, 2024).

Dicho esto, con el análisis de las matriculaciones se podrá averiguar el ritmo y la dirección hacia esta transformación más sostenible. Además, con la aplicación de las series temporales se podrá analizar las proyecciones a futuro y confirmar este proceso de transformación.

1.2-Objetivos del Estudio

El objetivo principal de esta investigación es analizar en profundidad la evolución de las matriculaciones de turismos nuevos en España después del periodo COVID-19 por tipo de combustible.

Con mapas geográficos se pretende identificar qué zonas y provincias tienen mayor o menor volumen de matriculaciones de turismos renovables, indicando si se dirigen hacia un cambio más sostenible o si de lo contrario optan por las tecnologías más tradicionales, como los de gasolina o los de diésel. Para ello, se incorporan tres

variables clave: la renta neta media por hogar, las matriculaciones por cada 1000 hogares y la proporción de población rural.

Por otro lado, se busca determinar qué modelos estadísticos se ajustan mejor a la evolución de cada serie, con el fin de generar predicciones que ayuden a anticipar la evolución del mercado en los próximos años.

Por último, se analiza si existe algún tipo de relación entre las matriculaciones de distinto tipo de combustible, es decir, si el crecimiento o descenso de un tipo puede estar afectando directamente al comportamiento de las demás.

1.3-Estructura del trabajo

La estructura de este trabajo se organiza en diferentes apartados que abarcan desde el marco teórico hasta las conclusiones.

En el **marco teórico**, se realiza un análisis temporal de las matriculaciones por tipo de combustible.

En la **metodología** aplicada se describen detalladamente todas las técnicas estadísticas empleadas por cada tipo de análisis.

En el **procesamiento de los datos** se analiza la estructura y las variables de los datos. Incluidos el procesamiento y análisis exploratorio.

El **análisis clúster** se usará para analizar en profundidad zonas geográficas con características similares en el número de matriculaciones por tipo de combustión en función de tres variables: la renta neta media por hogar, el número de matriculaciones cada mil hogares y un índice de ruralidad.

El **análisis univariante** corresponde al uso de series, donde se identificarán tendencias y estacionalidades. Se implantará una función de búsqueda exhaustiva del mejor modelo en base a las predicciones para realizar pronósticos a futuro con el menor error posible.

El **análisis multivariante** es el último objetivo del estudio, donde se aplicará métodos de series temporales (modelos vectoriales autorregresivos) para estudiar posibles relaciones entre tipos de combustión en función de aumentos o disminuciones del número de las matriculaciones.

Finalmente, el apartado de conclusiones expone los principales resultados obtenidos a lo largo del estudio y plantea recomendaciones orientadas a futuras investigaciones que apliquen o complementen los análisis realizados.

1.4-Contribuciones Esperadas.

Este trabajo busca aportar el conocimiento y la información sobre la evolución de las matriculaciones desde una perspectiva geográfica y temporal.

Los resultados podrían ser de utilidad para empresas del sector o responsables de políticas públicas al ofrecer una comparación muy detallada y una visión del mercado español por provincia y tipo de combustible.

Por otro lado, se espera que las proyecciones sirvan para analizar en qué dirección se dirige el parque automovilístico y para anticipar tendencias.

En definitiva, este estudio pretende ofrecer una visión más amplia sobre la transformación del parque automovilístico español hacia la sostenibilidad, aportando herramientas que faciliten la toma de decisiones.

2 MARCO TEÓRICO.

Este apartado se enfoca en estudiar las tendencias y patrones por tipo de combustible y provincia a lo largo del tiempo, aplicando técnicas estadísticas y modelos de series temporales para extraer información útil.

2.1-Análisis de datos de Matriculaciones de turismos.

Los datos de las matriculaciones de turismos nuevos incluyen información específica sobre el tipo de vehículo, aunque este estudio se centrará únicamente en su tipo de combustible (diésel, gasolina, eléctricos e híbridos), su uso (particular o de uso público) y su zona geográfica.

Las series de datos al ser datos ordenados cronológicamente permiten analizar la evolución que han tenido las matriculaciones a lo largo del tiempo. Identificando estacionalidades, tendencias, variaciones extremas y además ofrece la posibilidad de realizar proyecciones al capturar la estructura de la serie.

3. METODOLOGÍA.

Para el análisis de las matriculaciones en España, se utilizarán tres técnicas estadísticas. En primer lugar, clúster, después modelos Seasonal Autoregressive Integrated Moving Average (SARIMA) y por último los modelos vectoriales autorregresivos (VAR). El *clustering* permitirá identificar provincias con comportamientos similares en cuanto las matriculaciones.

Los modelos SARIMA, de carácter univariante, se usarán para analizar tendencias, capturar patrones estacionales y realizar predicciones por cada tipo de combustible.

Por último, los modelos VAR, de tipo multivariante, servirán para analizar y capturar relaciones entre los distintos tipos de turismos.

3.1-Análisis Clúster: Clustering Jerárquico

Siguiendo Québec Centre for Biodiversity Science (2023) y Olah, Antal y Kónya (2023), y tomando como base sus explicaciones, se han desarrollado los modelos de análisis de clúster jerárquico aplicados en este trabajo, así como los métodos de validación mediante los índices de Silhouette, Dunn y Calinski-Harabasz.

El *Clustering* jerárquico por conglomerados es una técnica que permite agrupar las provincias en base a su similitud teniendo en cuenta una serie de variables.

Mediante la creación de un dendograma se facilita la interpretación de las relaciones entre las provincias. Esta técnica es una herramienta adecuada para análisis exploratorios.

La técnica que se emplea es por conglomerados y consiste en agrupar cada provincia inicialmente en un único clúster. Progresivamente aquellas provincias que presenten características similares se irán reagrupando en el mismo clúster. La distancia entre provincias se mide con la distancia euclidiana y el criterio de selección que se utiliza es el de Ward.D2, que minimiza la varianza intra-clúster en cada iteración.

Ward.D2, se basa en minimizar el aumento de la distancia euclidiana al cuadrado que se produce al unir dos conglomerados y esto favorece a la creación de agrupaciones de provincias lo más homogéneas posibles. Para este análisis, se estandariza las variables porque esto favorece a la creación de clústeres de la manera más efectiva.

El procedimiento se aplica sobre tres variables socioeconómicas:

Renta neta media por hogar de cada provincia, un índice de matriculaciones que cuantifica el número de matriculaciones por cada mil hogares en cada provincia y por último un índice que indica el porcentaje de ruralidad por provincia también.

El número óptimo de clústeres se selecciona a través del método del codo y del dendograma. Con el primer método se observa el punto de inflexión en la reducción de la varianza intra-clúster, siendo el óptimo el que menor varianza produzca.

Se evalúa la similitud interna con el índice de Silhouette, la variabilidad dentro de cada grupo con el índice de Dunn y la validez de los clústeres establecidos con el índice Calinski-Harabasz.

También, se valida estadísticamente la existencia de diferencias significativas entre los grupos sobre cada variable con contrastes de normalidad (Shapiro-Wilk) y con la

prueba de Levene para verificar la igualdad de varianzas. Se comprueba la normalidad y la igualdad de varianzas para poder realizar un análisis de la varianza (ANOVA).

En el caso de no cumplirse los supuestos, se aplica una prueba no paramétrica (Kruskal-Wallis) para asegurar las diferencias entre grupos todo bajo un nivel de significación del 5%.

Esta metodología, se aplica sobre dos enfoques diferentes. El primero sobre todos los tipos de combustión entre junio de 2020 y diciembre de 2024. Y el segundo sobre diferentes grupos de tipos de combustión haciendo foco en el año 2024.

Finalmente, los resultados se representan en mapas geográficos y gráficos de barras por clúster para facilitar la interpretación de las agrupaciones con la renta neta media por hogar, el índice de matriculaciones y la proporción media de población rural.

3.2-Análisis Univariante de Series Temporales (modelo SARIMA)

Para la elaboración de este apartado se han seguido principalmente los apuntes de series temporales (Álvarez Liébana, 2024). Y para una explicación más amplia sobre series temporales, puede consultarse Hyndman y Athanasopoulos (2021).

El modelo SARIMA es una extensión del modelo ARIMA diseñada para capturar patrones no estacionales (p,d,q) y estacionales (P,D,Q) en series temporales. Su notación es SARIMA(p,d,q)(P,D,Q)[S] y S es la componente estacional.

- (p, d, q) representa la parte no estacional:
 - p es el orden autorregresivo (AR por sus siglas en inglés)
 - d es el número de diferenciaciones necesarios para corregir la estacionariedad de la serie en términos de tendencia.
 - q es el orden del componen media móvil (MA por sus siglas en inglés)

- (P, D, Q) representa la parte estacional:
 - P es el orden autorregresivo estacional (AR)
 - D es el número de diferenciaciones necesarios para corregir la estacionariedad de la serie en términos de estacionalidad.
 - Q es el orden del componen media móvil estacional (MA)

El análisis univariante de series temporales es una herramienta muy importante para evaluar tendencias, estacionalidades y entender el comportamiento que ha tenido cada serie a lo largo del tiempo

En este trabajo, el análisis se centra en las matriculaciones de turismos nuevos gasolina, diésel, gasolina híbridos y eléctricos al ser los tipos de combustión con más matriculaciones y con más interés en este estudio. Se estudia individualmente la evolución de cada una de las series y se construye un modelo de predicción

Se inicia con la creación de una serie temporal por cada tipo de combustión con frecuencia mensual. Con la visualización de las series se detecta una caída drástica en el periodo COVID-19.

El principal motivo es para mejorar la capacidad explicativa y la calidad de las estimaciones y de las predicciones. Al considerar cada serie desde esa fecha, con los datos mensuales se obtienen muy pocas observaciones, concretamente cincuenta y cinco, por lo que se decide reestructurar la serie en frecuencia semanal para contar con más observaciones.

Las tendencias en series temporales indican la dirección (positiva o negativa) en la que la serie se dirige. Se evalúa si la tendencia es significativa o no con una regresión lineal y la presencia de una tendencia significativa indica indicios de no estacionariedad en la serie (supuesto que se debe cumplir).

Después de estimar el modelo de regresión lineal para analizar la tendencia, se evalúa si la varianza de los errores del modelo es constante (homocedasticidad). La homocedasticidad es un requisito que se debe cumplir para garantizar la fiabilidad de las predicciones y de las estimaciones y se usa la prueba de Breusch-Pagan para detectar si la varianza de los residuos es constante o no. En el caso de rechazar la hipótesis nula de homocedasticidad, se aplica transformaciones para estabilizar la varianza.

La estacionalidad es el patrón que se repite a lo largo de la serie temporal y se identifica esa componente estacional mediante la visualización de gráficos.

La estacionariedad es una propiedad o concepto en series temporales que garantiza que las propiedades estadísticas (media, varianza y autocorrelación) se mantengan constantes a lo largo de la serie. Esto permite generar predicciones y estimaciones fiables

Esta propiedad se observa con gráficos de autocorrelación (ACF por sus siglas en inglés) y gráficos de autocorrelación parcial (PACF por sus siglas en inglés). ACF mide la correlación entre el valor actual de la serie con los valores pasados y no hay estacionariedad si las barras verticales de este gráfico están por encima de las bandas horizontales y disminuyen progresivamente. PACF mide la correlación entre una observación y sus rezagos, pero limando el efecto de los rezagos intermedios. En este tipo de gráficos la no estacionariedad existe si la barra del primer rezago está por encima de las bandas horizontales y muy próxima del uno.

Estadísticamente se evalúa con la prueba de Dickey-Fuller aumentado (ADF) la presencia o no de una raíz unitaria bajo la hipótesis nula de no estacionariedad.

La presencia de la raíz unitaria quiere decir que el valor actual de la serie depende del valor anterior (autocorrelación). Cuando la serie no es estacionaria se tiene que aplicar diferenciaciones en términos de tendencia o de estacionalidad para conseguir la estacionariedad en la serie.

Para la selección del modelo, se elaboró una función de R que se puede consultar en el Anexo. Esta función lo que hace es comparar 1024 modelos con diferentes combinaciones de los parámetros no estacionales (p,d,q) y estacionales (P,D,Q) donde q, p, Q y P van desde cero a tres y los términos que hacen referencia a la diferenciación (d, D) van desde cero a uno.

En cada combinación se ajusta un modelo en el conjunto de entrenamiento y se evalúa sobre el de prueba (52 semanas) calculando las métricas error cuadrático medio (MSE por sus siglas en inglés) y error absoluto medio (MAE por sus siglas en inglés).

El modelo que se selecciona es aquel que consigue mejores errores de predicción y además ayuda a saber si es necesario o no aplicar diferenciaciones. Este proceso de búsqueda exhaustiva tiene un inconveniente y es el tiempo de ejecución, ya que por serie se ha necesitado un día para encontrar el modelo óptimo.

Con el modelo óptimo ajustado, se evalúa los errores de predicción del modelo sobre el conjunto de entrenamiento y de prueba con gráficos *boxplot* y con métricas de precisión como MAE, raíz del error cuadrático medio (RMSE por sus siglas en inglés) y error porcentual absoluto medio (MAPE por sus siglas en inglés). Además, se realiza un diagnóstico de los residuos para comprobar que son ruido blanco, es decir, que tienen media constante en 0, varianza constante y no tienen autocorrelación.

3.3-Análisis Multivariante de Series Temporales (Modelo VAR)

Siguiendo los tutoriales de Kotzé (2019) y Lee (2025), en este apartado se desarrolla el modelo VAR aplicado a series temporales multivariantes.

El modelo VAR es una herramienta que se utiliza para analizar series temporales conjuntamente.

Sirve para analizar si por ejemplo el aumento o la disminución de las matriculaciones de un tipo de turismo influyen en las matriculaciones de otro tipo de combustión. Funciona modelando cada serie temporal como una función lineal de sus propios valores pasados de la serie y además de los valores pasados de las otras variables (tipos de combustión) para capturar posibles relaciones o no entre series.

Para este trabajo, se usó el modelo VAR con las series mensuales de las matriculaciones de los turismos nuevos diésel, gasolina, gasolina híbridos y eléctricos.

En este análisis además se usa procedimientos aplicados en las series temporales univariantes para asegurar que el modelo estimado es el óptimo en base al criterio de información escogido.

Las matriculaciones diarias se agruparon en mensuales para mejor interpretación de los resultados a partir del fin del estado de emergencia por COVID-19 y así evitar posibles distorsiones. Para verificar la homocedasticidad y la estacionariedad se

aplicó la prueba de Breusch-Pagan, gráficos ACF y PACF y pruebas estadísticas como el test ADF.

Para los modelos VAR se requiere que las series sean estacionarias en cuanto a tendencia, además, se desestacionaliza cada una de las series para eliminar las variaciones de las matriculaciones mensuales y para tener estimaciones más precisas. Porque al desestacionalizar, se aprecia mejor la tendencia positiva o negativa.

Se selecciona el número óptimo de rezagos mediante cuatro criterios de información: Akaike Information Criterion(AIC), Hannan-Quinn information Criterion (HQ), Schwarz Criterion (SC) y final prediction error (FPE). Estos criterios eligen el número óptimo de esta forma:

- AIC: Suele elegir modelos que son más grandes, es decir, con más meses hacia atrás.
- HQ y BIC: Son todo lo contrario, eligen modelos más simples.
- FPE: Se enfoca en predecir bien el futuro, desde el punto de vista de AIC ya que suele elegir también modelos menos simples.

Aunque algunos de estos criterios se usan para evaluar la capacidad predictiva del modelo, en este caso, el objetivo es analizar las relaciones entre tipo de vehículo y no realizar predicciones precisas. Por eso se prioriza un buen ajuste del modelo a los datos seleccionando el modelo según estos criterios.

Y tras la estimación del modelo se realiza cuatro pruebas de diagnóstico para verificar que se cumple los supuestos requeridos por este tipo de modelo de series temporales. Las pruebas son:

- Prueba de Portmanteau: Para verificar la ausencia de autocorrelación.
- Test de Jarque-Bera: Para comprobar que los residuos del modelo siguen una distribución normal.
- Test de ARCH: Para descartar la presencia de heterocedasticidad (varianza de los errores no constante)
- Estabilidad del sistema: Para comprobar que el módulo de las raíces del polinomio del modelo VAR estimado son menores que uno.

Después de comprobar que se han cumplido todos estos supuestos, se aplican algunos análisis para extraer información:

- Test de Granger: Sirve para comprobar si los valores pasados de una serie ayudan a predecir la de otra. Evalúa si las matriculaciones de un tipo de turismo anticipan las de otro.
- Análisis de impulso-Respuesta (IRF por sus siglas en inglés): Sirve para evaluar qué ocurre en las matriculaciones de cada tipo de combustión cuando se produce un shock (aumento) inesperado en una de ellas. De esta forma se identifican relaciones entre las variables o no.

- Descomposición de la varianza (FEVD por sus siglas en inglés): Sirve para evaluar cómo influyen las predicciones de las matriculaciones a futuro de un tipo de tecnología por el resto de las tecnologías y por sí misma.

4. Estudio y depuración de la base

La base de datos que se utiliza para esta investigación se construye a partir de los microdatos de las matriculaciones diarias de cualquier tipo de vehículo que ofrece la DGT.

El portal estadístico de la DGT (DGT, 2024) es una plataforma en línea destinada a ofrecer acceso público a una amplia gama de datos y estadísticas relacionadas con el tráfico y la seguridad vial. Este portal incluye información sobre accidentes de tráfico, infracciones, matriculaciones de vehículos, permisos de conducción y otros datos relevantes que pueden ser útiles para investigadores, profesionales del sector y el público en general.

Concretamente se han descargado 120 ficheros mensuales con más de doce millones de registros correspondientes a matriculaciones diarias desde 2015 a 2024, los cuales se han organizado, leído y depurado con el software estadístico R.

Esas doce millones de observaciones (matriculaciones diarias) posteriormente se redujeron aproximadamente a diez millones y medio de datos al filtrar por matriculaciones de turismos nuevos.

4.1-Estructura y variables

Los datos de origen están en formato texto con columnado fijo y contenían 69 variables de las que se seleccionaron durante el proceso de depuración las indicadas abajo. Con el lenguaje de programación R se procesó la lectura iterativa de los 120 ficheros en un bucle.

Cada fila representa una única matriculación y cada columna corresponde a una variable. Las variables extraídas son:

1. *Fecha_Matriculación*: fecha de matriculación del vehículo.
2. *Vehículo*: Se ha filtrado únicamente por turismos para esta investigación (categoría 40).
3. *Tipo_Propulsión*: Se ha filtrado por las categorías 0, 1 y 2. Haciendo referencia a turismos de gasolina, diésel y eléctricos.
4. *Provincia*: Código de la provincia donde está domiciliado el vehículo. Para este trabajo se reemplaza el código por el nombre de la provincia correspondiente.

5. *Nuevo_Usado*: Indica si el vehículo es nuevo o usado (N, U) al momento de la matriculación. Para este estudio, se filtró por turismos nuevos.
6. *Servicio*: Código del servicio del vehículo con varias categorías. Se ha filtrado para el análisis clúster por las categorías B00, B06, B07, B09, B17, B18, B19 y B21 las cuales hacen referencia al uso particular del turismo.
7. *Categoría_Elétrico*: Las categorías son PHEV (eléctrico enchufable), REEV (eléctrico de autonomía extendida), HEV (eléctrico híbrido) y BEV (eléctrico de batería).

Además, se han creado nuevas variables a partir de las mencionadas.

1. *Matriculaciones*: Corresponde al número total de matriculaciones. Dependiendo del objetivo en este estudio, se calcula mediante la agrupación por fecha, tipo de vehículo y provincia.
2. *Tipo_Vehículo*: Se crea esta variable a partir de la combinación de las variables *Tipo_Propulsión* y *Categoría_Elétrico*, con el fin de clasificar los turismos en seis categorías: gasolina, gasolina híbrido, gasolina híbrido enchufable, diésel, diésel híbrido y eléctricos.

Dependiendo de cada parte del análisis, los datos se agregan o por semanas, meses o años.

4.2-Procesamiento de los datos:

4.2.1-Análisis de datos faltantes

Se ha verificado que la base de datos no contiene ningún dato faltante por lo que directamente se puede proceder al procesamiento sin la necesidad de realizar técnicas de imputación o alguna limpieza adicional en los datos.

4.2.2-Análisis de outliers

No se ha realizado ningún análisis de valores atípicos ya que las variables utilizadas no presentan desviaciones que perjudiquen al conjunto de datos.

Sin embargo, en análisis clúster, se observa cómo la Comunidad Madrid presenta estadísticas muy superiores respecto al resto de provincias en todos los enfoques que se han realizado.

Debido a su alta renta neta media por hogar, a su alto número de matriculaciones cada mil hogares y su bajo índice de ruralidad, su comportamiento podría considerarse atípico en términos relativos.

4.3-Análisis exploratorio de datos

Esta investigación se inicia con un análisis exploratorio ya que es una fase fundamental para una comprensión inicial de la base de datos.

Primero se realiza una descripción general del número de matriculaciones por tipo de combustión incluyendo para cada tipo sus estadísticas descriptivas en la tabla 1.

En esta base de datos, las variables son categóricas por lo que se representa la evolución de cada combustión en un gráfico de líneas (ilustración 1). Gracias a este primer gráfico se identifica y se compara por primera vez la evolución de los diferentes tipos de combustión en los últimos años en España.

Además, se añade un gráfico (ilustración 3) de barras apiladas en el que se muestra el peso que tiene cada tecnología en el mercado con el paso de los años para complementar al gráfico de líneas.

	Acumulado Anual				% Crecimiento			
	2021	2022	2023	2024	2021	2022	2023	2024
Total Matriculaciones	843723	789121	908540	977689	-1,4%	-6,5%	15,1%	7,6%
Gasolina	388115	343562	388426	381335	-8,5%	-11,5%	13,1%	-1,8%
Diesel	186370	146182	120522	107052	-28,7%	-21,6%	-17,6%	-11,2%
Total Veh Comb Fósil	574485	489744	508948	488387	-16,2%	-14,8%	3,9%	-4,0%
Diesel_Hibrido	28988	23001	29652	34948	43,0%	-20,7%	28,9%	17,9%
Electrico	24214	31595	54880	58951	31,6%	30,5%	73,7%	7,4%
Gasolina_Hibrido	181198	206244	262286	348334	61,8%	13,8%	27,2%	32,8%
Gasolina_Hibrido_Enchufable	34838	38537	52774	47069	83,2%	10,6%	36,9%	-10,8%
Total Veh Comb Renov	269238	299377	399592	489302	105,5%	11,2%	33,5%	22,5%

Tabla 1: Resumen de matriculaciones por tipo de vehículo

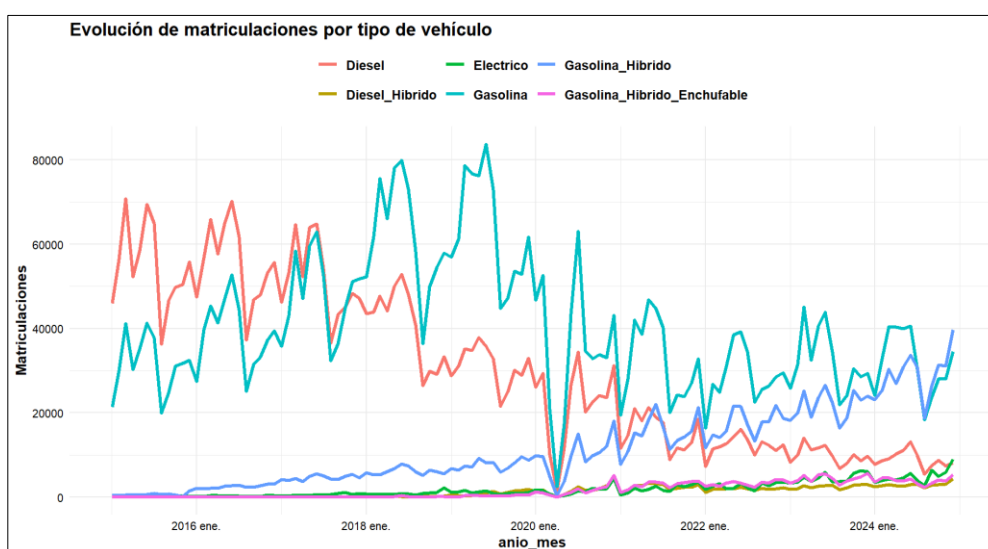


Ilustración 1: Evolución por tipo de vehículo

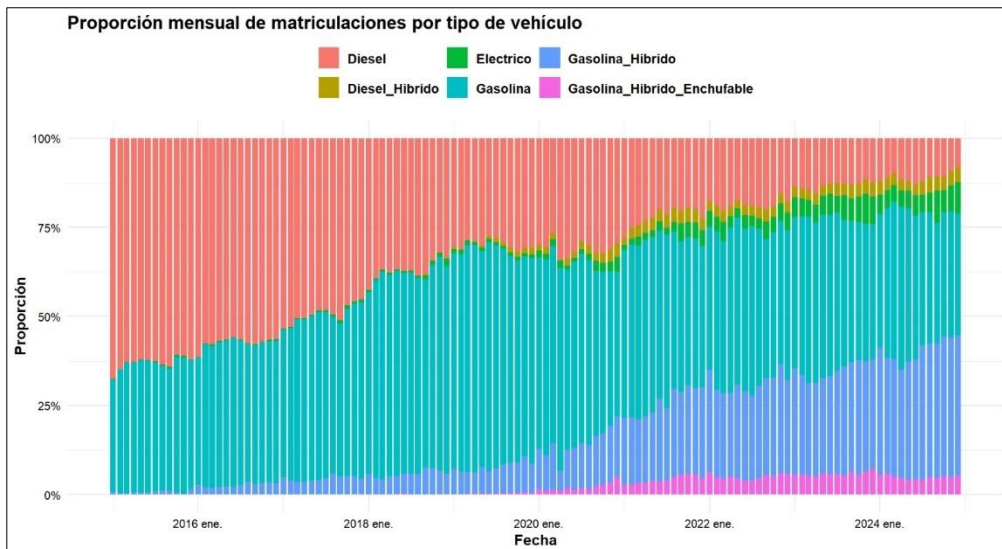


Ilustración 2: Gráfico de barras apiladas

Las ilustraciones 1 y 2 representan la evolución del número de matriculaciones mensual entre enero de 2015 y diciembre de 2024 distinguiendo entre los tipos de propulsión de vehículos matriculados. Podemos observar varios aspectos interesantes en los últimos 10 años en este mercado:

De la ilustración 1 se destaca:

- Un claro comportamiento estacional anual. Se ve cómo recurrentemente en junio/julio aparecen altos niveles de matriculación mientras que en agosto/septiembre decrecen drásticamente. Esto va a ser muy relevante, pues el análisis de series temporales con factor estacional tiene ciertas particularidades.
- Vemos también como los niveles de matriculación posteriores al COVID-19 son inferiores a los previos y que a partir de ahí es cuando aparecen los vehículos más sostenibles.

En la ilustración 2, al inicio de la serie se observa como el volumen de las matriculaciones de vehículos diésel estaban por encima de las matriculaciones de gasolina. Sin embargo, los turismos gasolina a partir de 2017 se alternan con los diésel, ya que estos comienzan a descender reduciendo el número de matriculaciones hasta el final de la serie.

Por otro lado, el tipo de turismo que más crecimiento muestra es el de gasolina híbridos. Superando a los eléctricos, diésel híbridos y gasolina híbridos enchufables. A finales de 2024 alcanzan casi el mismo volumen de matriculaciones que los de gasolina.

Esto se puede deber principalmente a la normativa aprobada por la Unión Europea que exige reducir en un 55% las emisiones de CO₂ de turismos nuevos para 2030

(respecto 2021), como primer paso antes de prohibir la venta de turismos que no sean de emisiones cero a partir de 2050 (Acciones de la UE, 2022).

Esta medida afecta directamente a los turismos de combustión tradicional y afectará a la población que tenga un coche con este tipo de tecnología.

Esto refleja una de las opciones con mayor popularidad hoy en día entre los españoles sin dar el paso a la categoría de los 100% eléctricos. Los cuales también presentan una tendencia positiva con el paso de los años.

En los inicios de 2020 se observa la caída de las matriculaciones provocada por el COVID-19. Este suceso a la hora de modelizar series temporales distorsiona y perjudica las estimaciones y las predicciones. Por eso, en este estudio, se analiza la evolución de las matriculaciones después del periodo COVID-19, es decir, cuando se declara el fin del estado de emergencia el 20 de junio de 2020.

5 SOLUCIONES

5.1 Análisis Clúster

5.1.1 introducción al análisis.

Este apartado tiene como objetivo aplicar *clustering* para analizar y comprender que zonas geográficas son similares entre sí en función del número de matriculaciones y de tres variables socioeconómicas:

- Renta neta media por hogar de cada provincia (INE, 2022a)
- Índice de matriculaciones por cada mil hogares, que surge a partir de dividir el número de matriculaciones entre el número de hogares por provincias (INE, 2021)
- Proporción de población rural por provincias a partir de los datos de Padrón Continuo (INE, 2022b), considerando población rural aquella que reside en municipios de menos de diez mil habitantes.

Las dos primeras variables se han considerado en términos de hogares ya que mayoritariamente la decisión de compra de turismos para uso particular se suele decidir a nivel hogar y el nivel de renta puede jugar un papel importante. La proporción de población rural se usa para añadir un componente geográfico en la clasificación.

Este estudio permite analizar cuáles son las provincias con características similares.

La variable *servicio* que se menciona en la sección 4.1 se filtra por las categorías que identifican que los turismos son de uso particular. Las cuales son B00, B06, B07, B09, B17, B18, B19 y B21.

Se plantea dos enfoques diferentes. El primero con todas las matriculaciones entre 2020 junio y 2024 diciembre y con todos los tipos de turismos agrupados. De esta manera se evalúa e identifica qué provincias tienen un mayor volumen de matriculaciones.

En segundo lugar, se analiza un enfoque parecido, pero solo para 2024, que se divide en tres análisis: uno para los turismos fósiles (diésel y gasolina agrupados), otro con todos los turismos sostenibles (eléctricos y todos los tipos de híbridos) y el último únicamente con los turismos eléctricos.

Se analizan conjuntamente estos tres análisis para obtener una perspectiva de cuáles son las zonas geográficas con más matriculaciones en las tecnologías tradicionales y cuáles en las tecnologías más modernas en función de las agrupaciones creadas a partir de las variables.

5.1.2-Resultados globales (junio 2020 – diciembre 2024)

Se comienza con un primer análisis por conglomerados considerando todas las matriculaciones de los turismos nuevos entre junio 2020 y diciembre 2024.

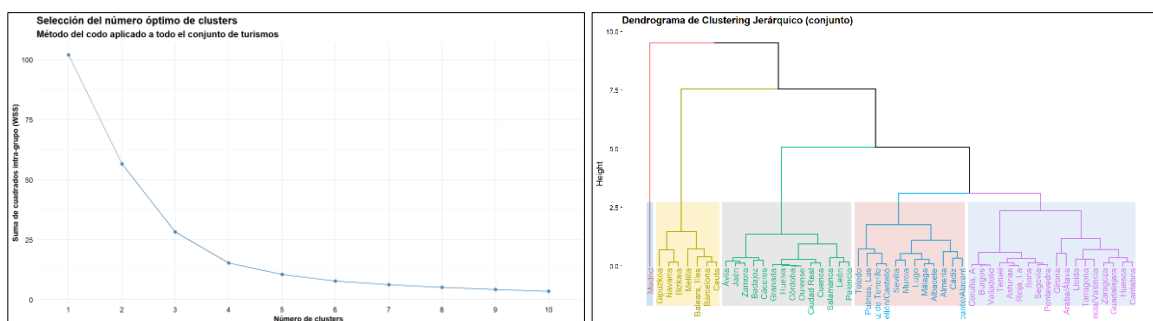


Ilustración 3: Gráfico del codo y dendrograma

En la ilustración 3, con el gráfico del codo se observa como a partir del quinto punto la varianza intra-cluster no se reduce apenas. Por ello, la selección óptima del número de agrupaciones es cinco ($k=5$). En el dendrograma, ilustración 3, se visualizan los cinco grupos de provincias con características socioeconómicas similares.

La creación de estos cinco grupos se aprecia con más claridad mediante la creación de un mapa geográfico asignando un color a cada provincia en función de su agrupación. Además, se añade tres gráficos de barras con las estadísticas socioeconómicas por clúster para que se facilite la interpretación.

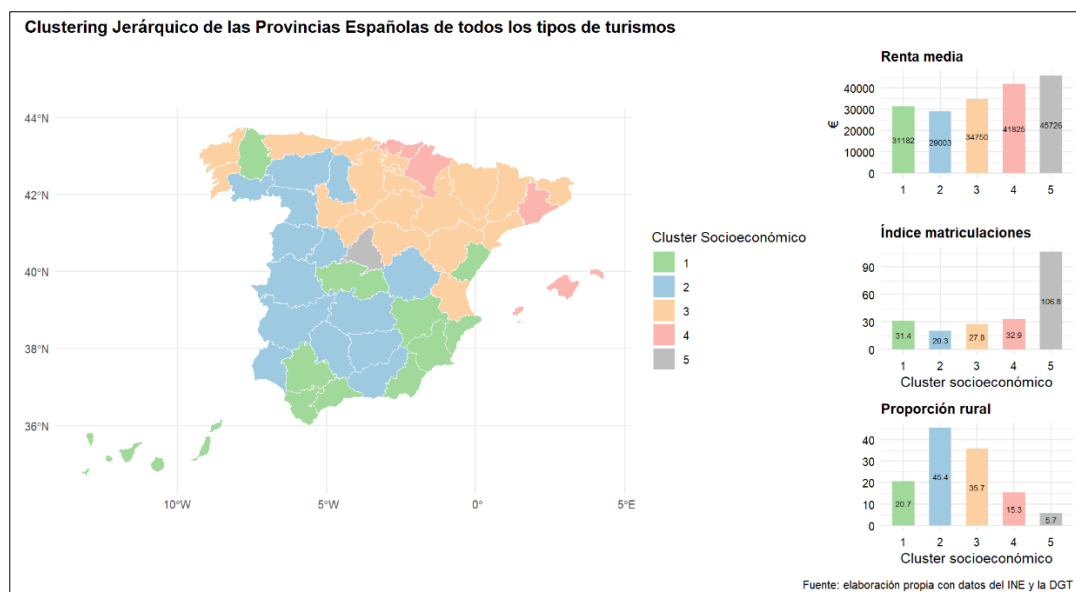


Ilustración 4: Mapa geográfico de todos los tipos de turismo

El clúster 5 se compone únicamente de la Comunidad de Madrid posicionándose como la provincia con mayor renta neta media, un mayor índice de matriculaciones y menor proporción rural.

En el extremo opuesto a Madrid, estaría el clúster 2 agrupando provincias del interior peninsular con un alto porcentaje de población rural (45%), y la menor renta neta media por hogar. Esto podría explicar el bajo número de matriculaciones por cada mil hogares y, por lo tanto, un menor acceso a la renovación de turismos.

Los clústeres 1 y 3 comparten estadísticas similares como el número de matriculaciones y la renta media. Sin embargo, las proporciones de población rural hace que se agrupen en diferentes clústeres. Por ejemplo, en ambos se aprecian grandes provincias como Sevilla, Málaga, Alicante, Valencia, Tarragona...

Lo significativo del clúster 4, principalmente Barcelona y el País Vasco, es que representa una renta media alta similar a la del clúster 5 y también una baja proporción de población rural pero no supera ni se acerca al índice de matriculaciones de Madrid

En conclusión, a nivel geográfico, Madrid es un caso especialmente llamativo ya que lidera en el índice de matriculaciones, lo cual puede deberse por factores como la alta intensidad urbana y un mayor nivel de renta por hogar que provoca una propensión a una renovación continua del parque automovilístico. Por otro lado, las provincias del interior y noroeste peninsular presentan una menor renovación de automovilística, en línea con su perfil más rural y económico bajo.

Métrica	Valor	Criterio esperado
Silhouette	NaN	Cercano a 1 (alta cohesión y separación)
Dunn	0.080	Mayor es mejor (mejor separación)
Davies-Bouldin	0.622	Menor es mejor (menor dispersión intra-cluster)
Calinski-Harabasz	64.1	Mayor es mejor (alta varianza entre clusters)
ANOVA (renta)	< 0.001	p < 0.05 (diferencias significativas)
Kruskal-Wallis (matriculaciones)	< 0.001	p < 0.05 (diferencias significativas)
Kruskal-Wallis (ruralidad)	< 0.001	p < 0.05 (diferencias significativas)

Tabla 2: resultados globales

De acuerdo con la Tabla 2, se aplican diferentes métricas de validación para asegurar que las agrupaciones creadas son estadísticamente correctas y no son formadas aleatoriamente. El índice de Silhouette no se puede calcular ya que hay un clúster con una sola observación, como es en el caso de Madrid.

El bajo resultado del índice de Davies-Bouldin indica que la dispersión de las provincias dentro de cada clúster es baja. Por otro lado, los resultados de Dunn y Calinski-Hasbez indican que hay buena separación entre grupos y que la cohesión interna es aceptable.

Con las pruebas ANOVA y Kruskal-Wallis se confirma que las diferencias entre clústeres son significativas estadísticamente.

En conclusión, todas las pruebas que se realizan indican que las agrupaciones se han creado correctamente con estadísticas socioeconómicas bastante similares.

5.1.3-Análisis por tipo de tecnología (año 2024)

En este apartado se vuelve a realizar el análisis clúster, pero por tipos de combustible y solo para las matriculaciones del año 2024. Incluyendo de nuevo las variables socioeconómicas para la creación de agrupaciones.

Para ello, se hacen 3 análisis individuales: fósiles (gasolina y diésel), renovables (híbridos y eléctricos) y eléctricos.

Lo interesante de este enfoque es que permite identificar qué zonas geográficas con una renta neta media, un índice de matriculaciones y una proporción de ruralidad similar se encuentran en una situación o un proceso hacia un parque móvil más sostenible o de si lo contrario se encuentran en un punto en el que siguen predominando las tecnologías más tradicionales.

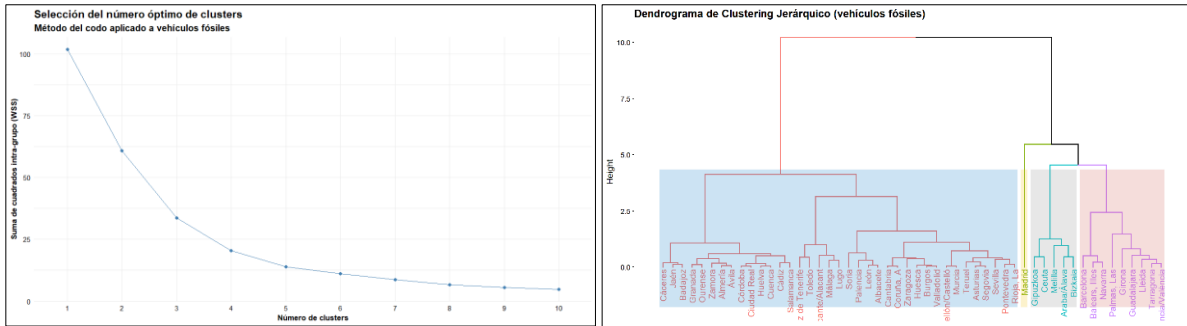


Ilustración 5: Gráfico del codo y dendrograma

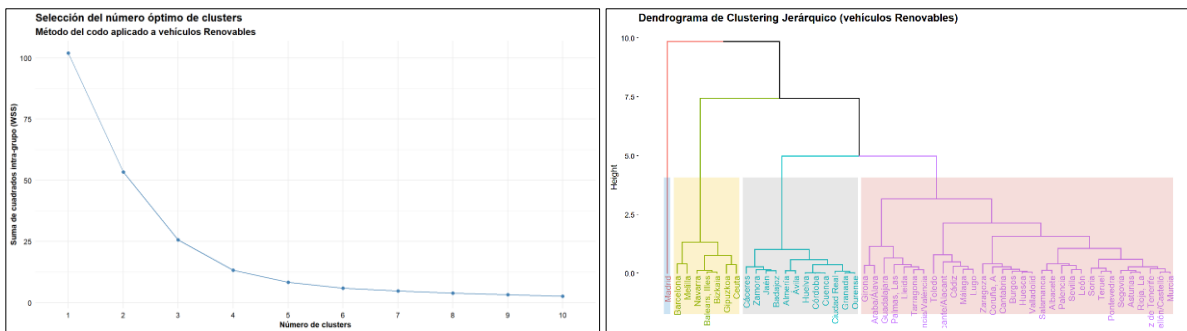


Ilustración 6: Gráfico del codo y dendrograma

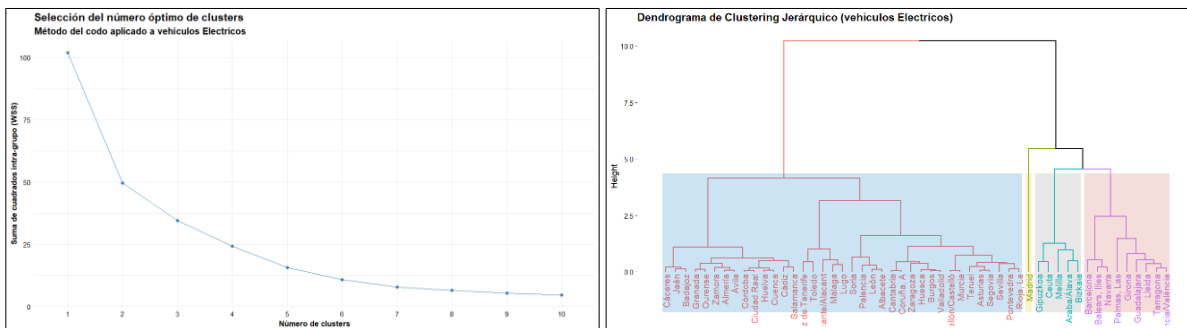


Ilustración 7: Gráfico del codo y Dendrograma

Con las ilustraciones 5, 6 y 7 se vuelve a seleccionar el número óptimo de agrupaciones para cada uno de los tres análisis con el mismo procedimiento explicado en el apartado 5.1.2.

Los análisis de codo en este caso sugieren una selección de cuatro clústeres ($k = 4$) por cada tipo de combustible.

Esta comparativa es mucho más visual con los mapas territoriales, donde las diferencias geográficas se representan de forma más intuitiva.

5.1.4-Turismos con combustibles fósiles (2024)

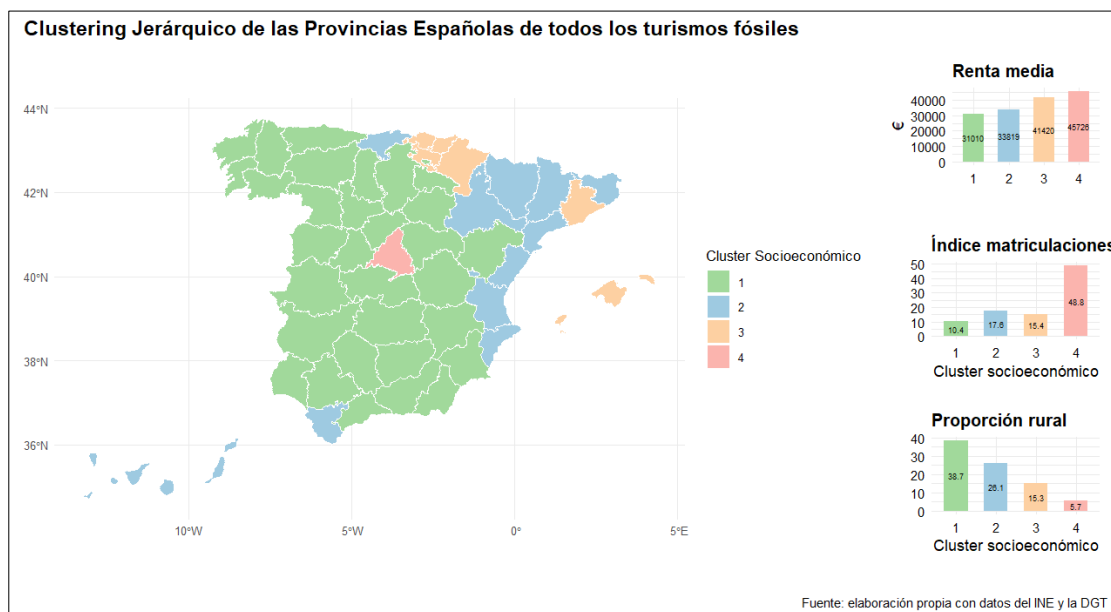


Ilustración 8: Mapa geográfico de todos los turismos fósiles

En la ilustración 8, Madrid vuelve a destacar excepcionalmente por encima del resto de las provincias en España por su elevado índice de matriculaciones, renta neta media y su baja proporción rural.

El clúster 1 formado prácticamente por todas las provincias del interior, sur y del norte peninsular son las que tienen un menor número de matriculaciones. Podría estar explicado por su baja población urbana y una menor renta media.

Vuelve a estar Barcelona y el País Vasco en el mismo clúster (3) debido a sus similitudes. Y aunque en niveles socioeconómicos son similares a los de Madrid, vuelve a estar diferenciado por el número de matriculaciones.

Curiosamente el clúster 2, agrupa la mayor parte de las provincias del noreste español junto con las Islas Canarias.

5.1.5-Turismos con combustibles renovables (2024)

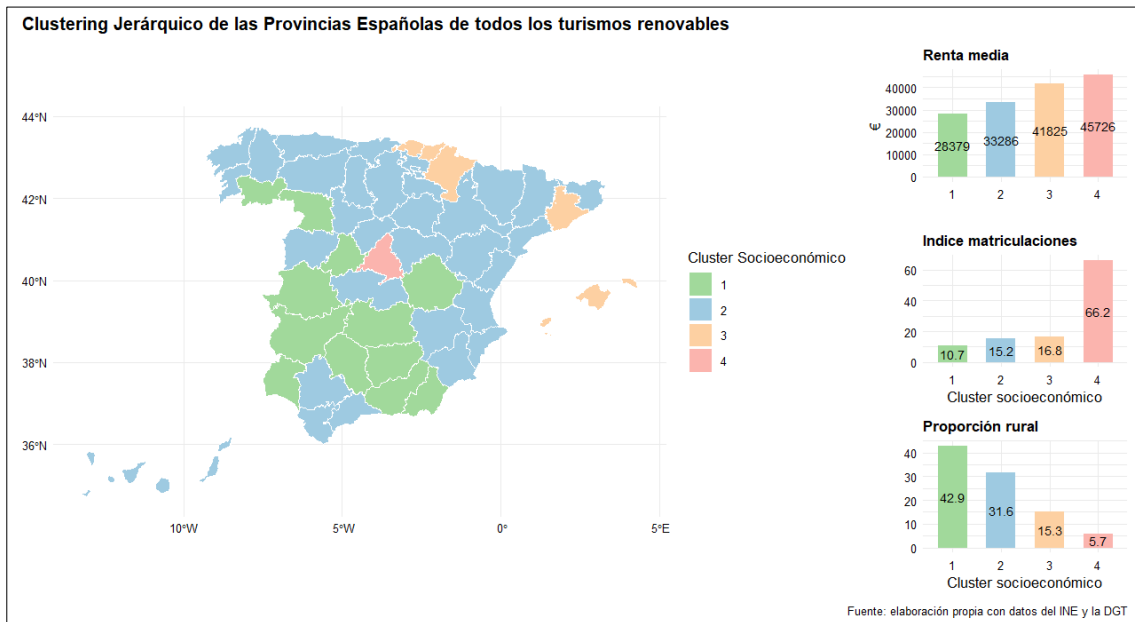


Ilustración 9: mapa geográfico de todos los turismos renovables

En la ilustración 7, cuando se analiza el caso de los turismos renovables (eléctricos e híbridos agrupados), si se compara con el mapa de los combustibles fósiles hay dos clústeres similares.

- Madrid (clúster 4) pero con un índice de matriculaciones de renovables cuatro veces superior con el resto de España.
- Barcelona, País Vasco, Navarra y Baleares (clúster 3) a pesar de tener una renta media alta y poca ruralidad, están todavía muy lejos en cuanto las matriculaciones que presenta Madrid.

La desigualdad territorial se observa sobre todo en el clúster 1, con provincias más del sur peninsular. Las cuales corresponden a la España profunda con menor urbanidad.

5.1.6-Turismos con combustibles eléctricos (2024)

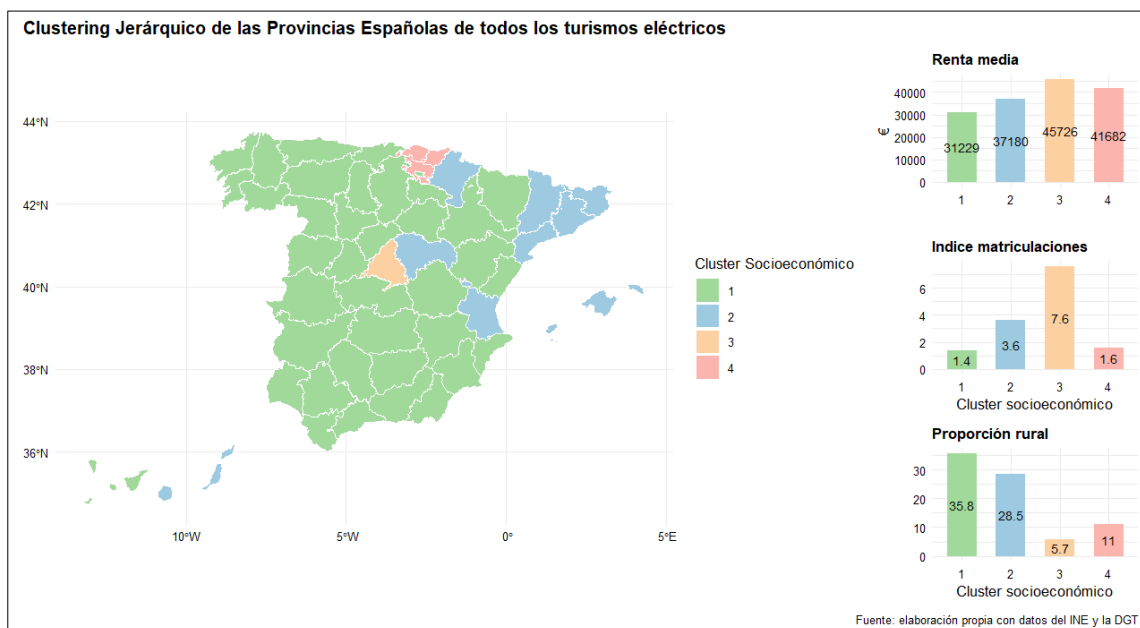


Ilustración 10: Mapa geográfico de los eléctricos

En la ilustración 10 se hace foco solamente en los turismos 100% eléctricos matriculados en 2024.

Una vez más Madrid (clúster 3) vuelve a destacar por encima del resto de provincias con casi el doble de matriculaciones. Pero se observa que el clúster 2 es el segundo grupo con mayor índice de matriculaciones localizado principalmente en el noreste peninsular: Cataluña, Valencia, Navarra y las Islas Baleares.

Para contrastar esta clara diferencia provincial, se consulta el mapa oficial de puntos de recarga por provincia (REVE), citado en el artículo de Soria (2025), el cual ha sido elaborado por el Ministerio para la Transición Ecológica. En el cual se resalta que Cataluña y Madrid son las comunidades con más puntos de recarga a nivel nacional.

Por otro lado, prácticamente el resto de las provincias de España que se agrupan en el clúster 1 se caracterizan principalmente por su bajo número de matriculaciones por cada mil hogares. Esto es un claro indicio de que en estas zonas geográficas este tipo de vehículo sostenible no tiene gran peso en el mercado. Esto puede deberse por la falta de puntos de recarga.

Las provincias del País Vasco muestran un comportamiento que no es habitual al tener la segunda renta neta media por hogar más alta y un índice de urbanidad muy alto ya que la cantidad de matriculaciones que tienen cada mil hogares es muy similar a la del clúster 1.

En definitiva, se confirma que el proceso hacia una movilidad más sostenible en 2024 no se está cumpliendo en todo el territorio o si se está haciendo, es lentamente.

En gran medida se debe a las diferencias socioeconómicas comentadas y la no fácil accesibilidad a puntos de recarga y por ello, es de gran interés que se tengan en cuenta estos resultados y estas diferencias si se busca un mercado con mayor volumen en las matriculaciones de turismos eléctricos en un futuro cercano.

Métrica	Valor	Criterio esperado
Silhouette	NaN	Cercano a 1 (alta cohesión y separación)
Dunn	0.152	Mayor es mejor (mejor separación)
Davies-Bouldin	0.663	Menor es mejor (menor dispersión intra-cluster)
Calinski-Harabasz	63.7	Mayor es mejor (alta varianza entre clusters)
ANOVA (renta)	< 0.001	p < 0.05 (diferencias significativas)
Kruskal-Wallis (matriculaciones)	< 0.001	p < 0.05 (diferencias significativas)
ANOVA (ruralidad)	0.0015	p < 0.05 (diferencias significativas)

Tabla 3: resultados turismos fósiles

Métrica	Valor	Criterio esperado
Silhouette	NaN	Cercano a 1 (alta cohesión y separación)
Dunn	0.118	Mayor es mejor (mejor separación)
Davies-Bouldin	0.462	Menor es mejor (menor dispersión intra-cluster)
Calinski-Harabasz	106.6	Mayor es mejor (alta varianza entre clusters)
Kruskal-Wallis (renta)	< 0.001	p < 0.05 (diferencias significativas)
Kruskal-Wallis (matriculaciones)	< 0.001	p < 0.05 (diferencias significativas)
ANOVA (ruralidad)	0.0047	p < 0.05 (diferencias significativas)

Tabla 4: resultados turismos renovables

Métrica	Valor	Criterio esperado
Silhouette	NaN	Cercano a 1 (alta cohesión y separación)
Dunn	0.256	Mayor es mejor (mejor separación)
Davies-Bouldin	0.542	Menor es mejor (menor dispersión intra-cluster)
Calinski-Harabasz	51.1	Mayor es mejor (alta varianza entre clusters)
ANOVA (renta)	< 0.001	p < 0.05 (diferencias significativas)
Kruskal-Wallis (matriculaciones)	< 0.001	p < 0.05 (diferencias significativas)
ANOVA (ruralidad)	0.0117	p < 0.05 (diferencias significativas)

Tabla 5: resultados turismos eléctricos

Para este enfoque se vuelve a comprobar que las agrupaciones son válidas con los mismos índices y pruebas para los tres casos (Tablas 3, 4 y 5). Los resultados que se obtienen mediante los índices indican que cada uno de los clústeres están formados por provincias con poca diferencia entre ellas. Reafirmando que las agrupaciones no han sido creadas aleatoriamente.

Por otro lado, las pruebas ANOVA y Kruskal-Wallis para cada variable presentan diferencias significativas entre clústeres.

En conclusión, todas estas pruebas garantizan que las agrupaciones de los clústeres se han creado eficazmente con características socioeconómicas parecidas.

5.2 Análisis Univariante

En esta sección se analiza la evolución de cuatro tipos de combustión mediante la aplicación de series temporales desde la semana 26 de 2020 hasta la semana 52 de 2024 con frecuencia semanal.

El análisis se centra a partir de la semana 26 de 2020 (fecha en la que finalizó el estado de alarma en España) porque la serie de matriculaciones de gasolina muestra un cambio estructural claro debido a la pandemia. Antes del COVID-19, la tendencia era creciente y después se estabiliza (ilustración 11). Modelar todo el periodo con modelo SARIMA, sería incorrecto ya que se mezclarían dos comportamientos distintos afectando así a las estimaciones y predicciones. Puesto que todos los tipos de turismos se vieron afectados de una forma u otra, se inicia desde ese periodo para todas las series.

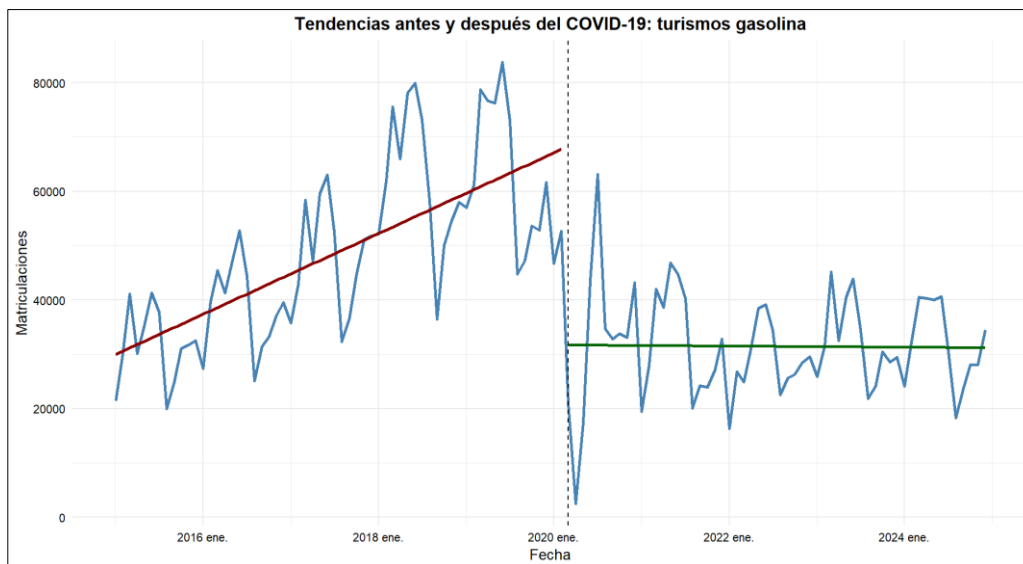


Ilustración 11: Tendencias antes y después del COVID-19 para turismos gasolina

Se centra en el análisis de la tendencia, homocedasticidad, estacionalidad, estacionariedad, en la búsqueda del mejor modelo predictivo para cada serie y por último una proyección para 2025 (52 semanas).

5.2.1-Visualización de las series temporales

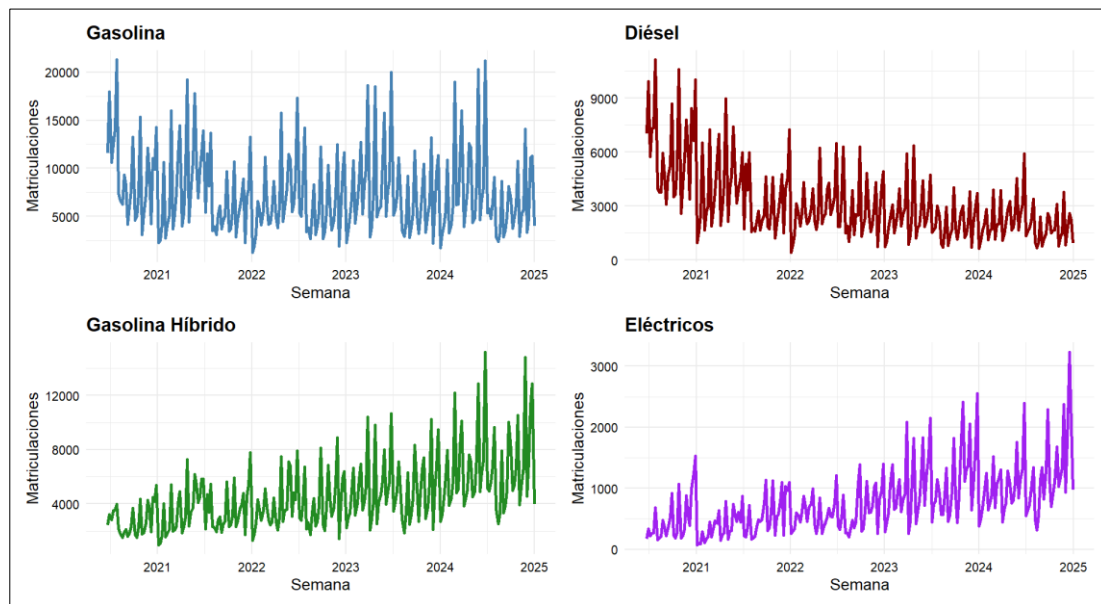


Ilustración 12: Serie temporal semanal por tipo de vehículo

La ilustración 12 representa la evolución de los turismos gasolina, diésel, gasolina híbridos y los eléctricos. En ella, se aprecia cómo cada serie tiene patrones y tendencias diferentes.

En la serie de la gasolina se aprecia una estacionalidad con patrones marcados y sin una tendencia clara, aunque, parece que es más negativa que positiva. Aun así, se analizará más adelante. De todos modos, son los que más matriculaciones tienen y han tenido a lo largo del periodo analizado.

Por otro lado, los vehículos diésel aparte de tener una estacionalidad igual que los de gasolina, presenta una disminución progresiva en el número de matriculaciones. En consonancia con el futuro abandono de este tipo de vehículos.

En cuanto a las tecnologías más modernas, actualmente se parecía con claridad su tendencia positiva y su aumento progresivo con el paso de los años.

En conclusión, todo apunta a una desaparición lenta y prolongada de las matriculaciones de los tipos de turismos más tradicionales y un aumento en los eléctricos y en los híbridos.

5.2.2-Análisis de la tendencia

Se evalúa la existencia o no de una tendencia significativa para cada serie mediante la creación de una regresión lineal considerando como variable explicativa la variable semana. Y se obtienen los resultados que muestran la Tabla 6:

Tecnología	Pendiente	Error Std.	t value	p-value	R ²
Gasolina	-0.9218	0.5744	-1.605	0.1099	0.0108
Diésel	-2.2035	0.2329	-9.463	<2e-16	0.2759
Híbrido	2.778	0.2917	9.524	<2e-16	0.2785
Eléctrico	0.6233	0.0587	10.621	<2e-16	0.3243

Tabla 6: Resultados regresión lineal

A partir de la Tabla 6 se usa un nivel de significación del 5% para contrastar la hipótesis nula (no existe relación significativa entre la semana y el número de matriculaciones). Y se obtiene que, para los vehículos de gasolina, no se puede rechazar H0 al tener un p-valor superior a 0,05 y se concluye que para estos turismos la tendencia no es significativa, reafirmando la suposición que se hizo en la visualización de cada una de las series temporales.

El p-valor de los turismos diésel es inferior al nivel de significación y además tiene un coeficiente negativo, por lo que se concluye una vez más que este tipo de turismos sigue una tendencia negativa y significativa.

En cuanto los de gasolina híbridos y los eléctricos, se concluye que las tendencias son significativas y que además son positivas.

A continuación, se repite el gráfico de líneas anterior añadiendo a cada serie una línea de tendencia para tener una mejor visualización de la misma.

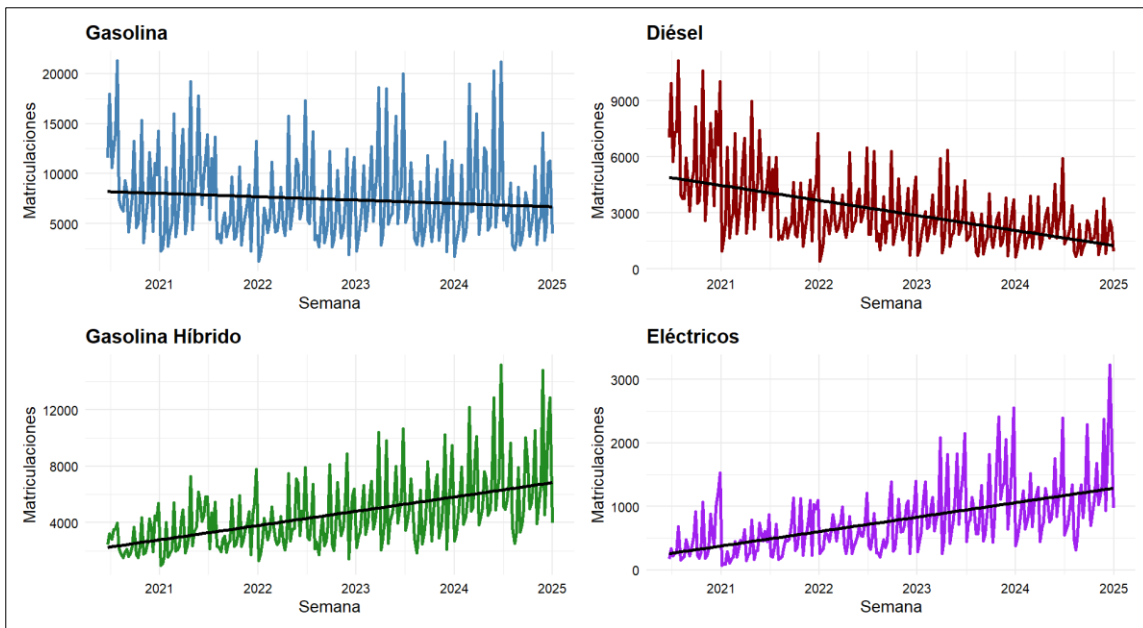


Ilustración 13: Serie temporal por tipo de vehículo con línea de tendencia

En la ilustración 13 se aprecia mejor la tendencia para todas las series menos para la gasolina que parece algo negativa.

Que la tendencia determinística sea significativa presenta un fuerte indicio de no estacionariedad en sentido estricto en la serie. Esto significa, que, en media, las matriculaciones no son constantes a lo largo de la serie y esto es un supuesto que se debe de cumplir para garantizar que las estimaciones y las predicciones sean precisas.

La estacionariedad se refiere a la ausencia de tendencia estocástica en una serie temporal. Sin embargo, la presencia de una tendencia determinística no necesariamente implica que la serie no sea estacionaria, siempre que esta tendencia pueda ser modelada y eliminada, esto es, la serie puede ser *trend-stationary*. Cuando realicemos el test ADF de raíz unitaria podremos determinar si las componentes determinísticas son *trend* (tendencia), *drift* (constante o deriva) o *none* (ninguna).

El análisis del test de raíz unitaria determinará si una tendencia observada es determinista (y se puede ajustar en el modelo) o es estocástica fruto de una raíz unitaria. Al eliminar la tendencia determinística y/o la raíz unitaria, se consigue eliminar la dependencia de cada valor de la serie con los del pasado. Y de esta manera se podría ajustar un modelo SARIMA,

5.2.3-Análisis de la homocedasticidad

Se comprueba si los residuos de los modelos de regresión creados para analizar la tendencia tienen varianza constante (Homocedasticidad) o no (Heterocedasticidad), con la prueba de Breusch-Pagan. Se rechaza o se acepta la hipótesis nula (H_0 =varianza de los errores constante) bajo un nivel de significación del 5%.

Tecnología	BP Statistic	gl	p-value
Gasolina	0.058848	1	0.8083
Diésel	27.425	1	1.633e-07
Gasolina Híbrido	22.28	1	2.356e-06
Eléctrico	15.71	1	7.383e-05

Tabla 7: Resultados prueba de homocedasticidad

De acuerdo con la Tabla 7, en el caso de los turismos gasolina, el p-valor es superior al nivel de confianza, por lo que no se rechaza la hipótesis nula de que la varianza de los residuos es constante a lo largo de la serie.

Sin embargo, para el resto de los tipos se rechaza la H_0 , por lo que se concluye los residuos de cada uno de estos modelos presentan heterocedasticidad.

Para las series que se rechaza la hipótesis nula, se tiene que aplicar una transformación sobre la variable de las matriculaciones para estabilizar la varianza ya que la varianza constante de los errores es uno de los supuestos principales de los modelos de regresión lineal.

Primero se opta por aplicar una transformación logarítmica y se vuelve a evaluar la homocedasticidad.

Tecnología	BP Statistic	gl	p-value
Diésel (log)	0.42897	1	0.5125
Gasolina Híbrido (log)	0.01724	1	0.8955
Eléctrico (log)	8.1339	1	0.00434

Tabla 8: Resultados prueba de homocedasticidad para el modelo de regresión de la variable transformada (log)

En la tabla 8, los nuevos resultados que se obtienen de la prueba de Breusch-Pagan mediante la transformación logarítmica indican que sí se ha corregido la heterocedasticidad, consiguiendo estabilizar la varianza de los errores para los turismos diésel y gasolina híbridos.

Sin embargo, la transformación logarítmica en los eléctricos no ha corregido este supuesto ya que el p-valor es inferior a 0,05. Por lo que es necesario aplicar otro tipo de transformación.

Tecnología	BP Statistic	gl	p-value
Eléctrico (Box-Cox)	2.1872	1	0.1392

Tabla 9: Resultados prueba de homocedasticidad para el modelo de regresión la variable transformada (Box-Cox)

Se opta por aplicar una transformación con Box-Cox y se obtiene un valor de lambda óptimo de 0,14 en la Tabla 9. Se aplica la transformación con este parámetro y se evalúa de nuevo la homocedasticidad.

Esta vez, se confirma que la varianza de los residuos sí es consonante para los eléctricos porque el p-valor es mayor a 0,05. Por lo que para esta serie se concluye que es necesaria una transformación mediante Box-Cox.

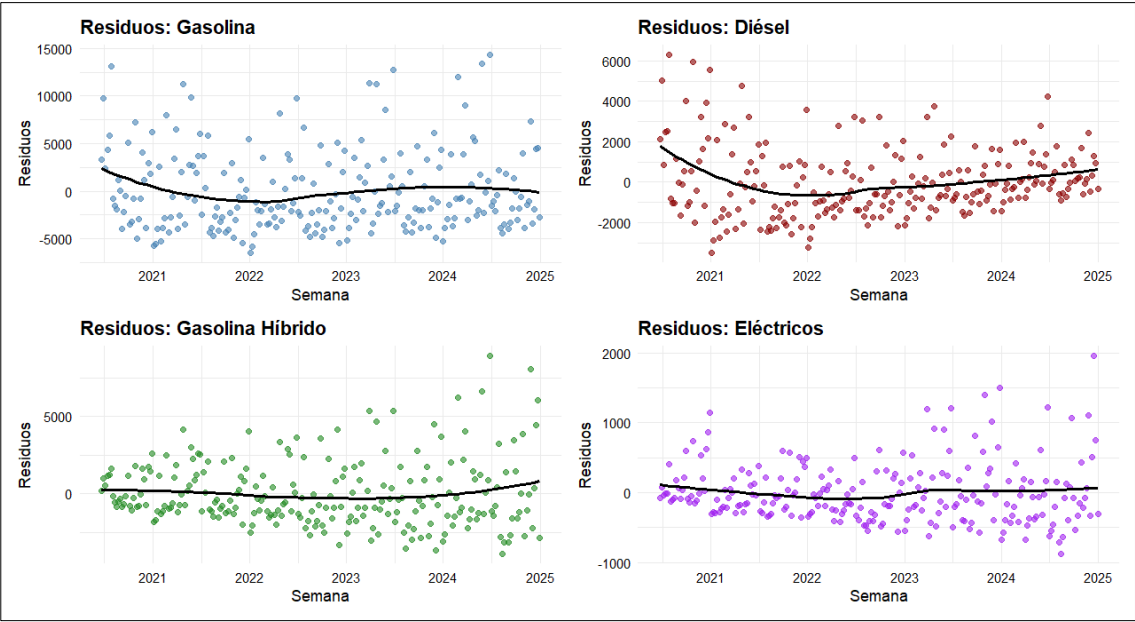


Ilustración 14: Gráfico residuos variables sin transformar

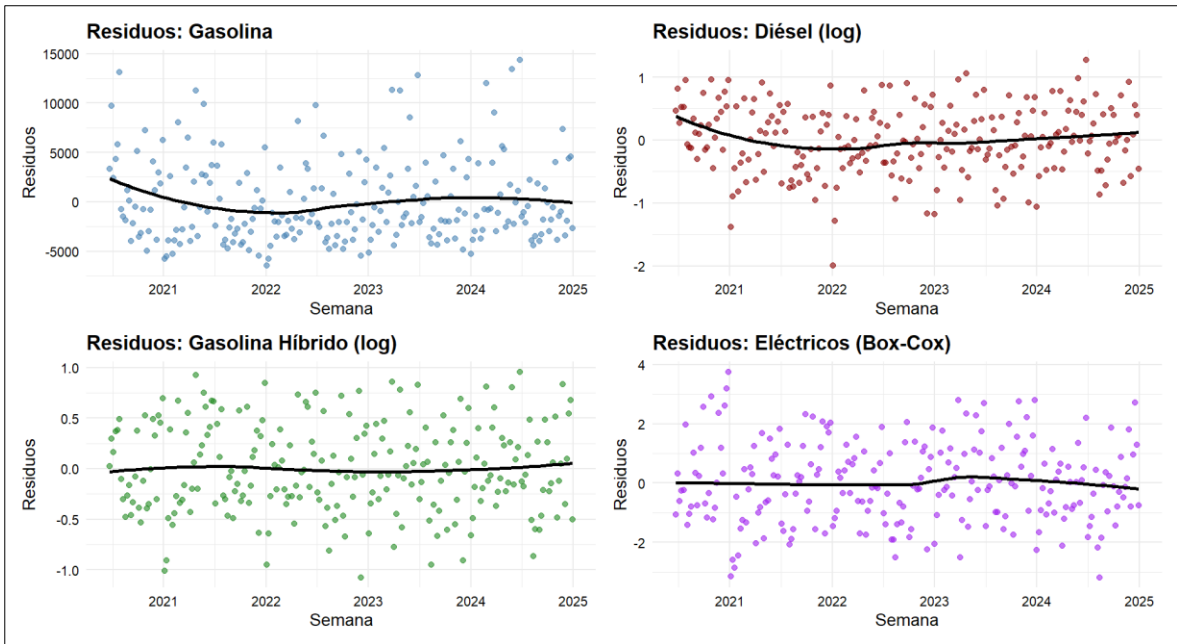


Ilustración 15: Gráfico Residuos con variables transformadas

Tras averiguar las transformaciones oportunas, la ilustración 14 y 15 muestran la dispersión de los residuos de cada modelo en el antes y después de aplicarlas transformaciones para confirmar visualmente que la varianza de los errores se ha corregido a lo largo de la serie. Se aprecia como la dispersión en la ilustración 14 no está en torno al cero y después sí (ilustración 15), concluyendo que se acepta el supuesto de homocedasticidad.

5.2.4-Análisis de la estacionariedad

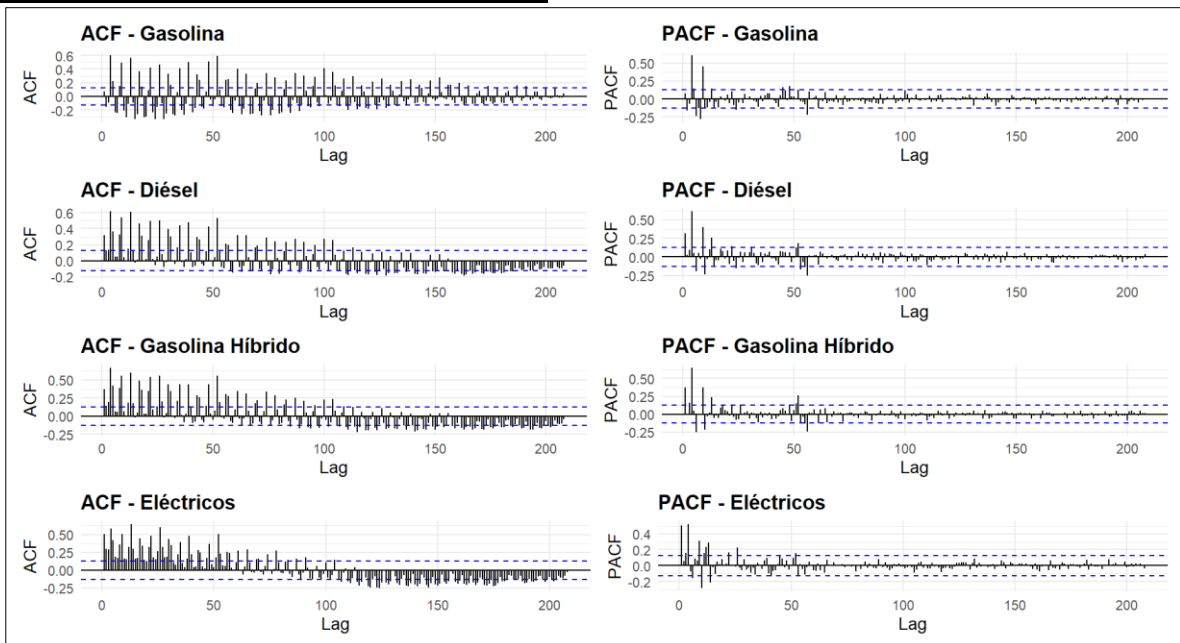


Ilustración 16: Gráficos ACF y PACF

En la ilustración 16 se muestran los gráficos de autocorrelación (ACF) y de autocorrelación parcial (PACF) de las cuatro series. No se observa la posible existencia de una raíz unitaria puesto que, en la PACF, en ninguno de los casos, el primer retardo es igual a 1, pero sí hay autocorrelación entre los rezagos que es lo que trataremos de modelar con el SARIMA.

Con los ACF se aprecia como las barras verticales decrecen lentamente hasta estar por debajo de las bandas horizontales de confianza. Esto indica correlación entre rezagos.

Aunque los correlogramas no indiquen la presencia de raíz unitaria, se aplica pruebas estadísticas como las del test ADF para contrastar la presencia o no de raíz unitaria.

Tecnología	Especificación	ADF / Valor crítico	Tendencia significativa	Constante significativa	phi2 / Valor crítico	phi3 / Valor crítico	phi1 / Valor crítico
Diésel	Tendencia	-3.202 / -3.43	Sí	Sí	3.971 / 4.75	5.228 / 6.49	—
Diésel	Constante	-1.941 / -2.88	—	No	2.592 / 4.63	—	—
Diésel	Ninguna	-1.264 / -1.95	—	—	—	—	—
Gasolina	Tendencia	-3.295 / -3.43	No	Sí	3.640 / 4.75	5.447 / 6.49	—
Gasolina	Constante	-3.300 / -2.88	—	Sí	—	—	5.458 / 4.63
Híbrido gasolina	Tendencia	-5.239 / -3.43	Sí	Sí	9.462 / 4.75	13.745 / 6.49	—
Eléctrico	Tendencia	-2.800 / -3.43	Sí	Sí	3.532 / 4.75	3.929 / 6.49	—
Eléctrico	Constante	-0.905 / -2.88	—	No	—	—	1.741 / 4.63
Eléctrico	Ninguna	1.510 / -1.95	—	—	—	—	—

Tabla 10: Resultados de las pruebas ADF

Se contrasta la hipótesis nula (H_0 = existencia de raíz unitaria) usando un nivel de significación del 5% con tres especificaciones: con tendencia (*trend*), con constante (*drift*) y sin estructura (*none*) fijándonos en la Tabla 10.

Para los diésel y los eléctricos, para el test ADF con tendencia y constante no se rechaza la hipótesis nula de raíz unitaria al 5% de significación, ya que el estadístico tau3 (ADF) es mayor que el valor crítico, por lo que se concluye que no son estacionarias. Los estadísticos phi2 y phi3 indican que tanto la constante como la tendencia no son necesarios en el modelo, sin embargo, ambos coeficientes son estadísticamente significativos. Por lo tanto, este modelo se descarta para ambas series y se prueba con la especificación de constante. El modelo con constante tampoco logra rechazar H_0 , y con el modelo sin constante ni tendencia tampoco se pudo rechazar. En ninguna de las tres especificaciones se rechaza, por tanto, la serie no es estacionaria y es necesaria la diferenciación.

En el caso de los gasolina, la serie es estacionaria bajo la especificación con constante, además el estadístico phi1 justifica la necesidad de mantener la constante (se rechaza H_0 de que hay una raíz unitaria en un modelo sin constante). Como la constante es significativa y se rechaza H_0 , este es el mejor modelo y no es necesaria aplicar la diferenciación.

Para los gasolina híbridos, la serie es estacionaria bajo el modelo con tendencia y con constante. Se rechaza H_0 al ser el estadístico ADF menor que el valor crítico (τ_3), y además tanto la tendencia como la constante son necesarias en el modelo. Aunque los estadísticos ϕ_2 y ϕ_3 no necesariamente justifiquen incluir constante y tendencia en el modelo, en el momento en el que el test ADF ya rechace raíz unitaria, implica que también se rechace ϕ_2 y ϕ_3 . Por lo tanto, este sería el mejor modelo y no es necesario aplicar ninguna diferenciación.

Para el caso de los diésel y los eléctricos se aplica una diferenciación y se vuelve a realizar la prueba ADF sobre los modelos *trend*, *drift* y *none*.

Tecnología	Especificación	ADF / Valor crítico	Tendencia significativa?	Constante significativa?	ϕ_2 / Valor crítico	ϕ_3 / Valor crítico	ϕ_1 / Valor crítico
Diésel	Tendencia	-6.121 / -3.43	No	No	12.554 / 4.75	18.822 / 6.49	—
Diésel	Constante	-6.128 / -2.88	—	No	—	—	18.786 / 4.63
Diésel	Ninguna	-6.047 / -1.95	—	—	—	—	—
Eléctrico	Tendencia	-9.447 / -3.43	No	No	29.761 / 4.75	44.640 / 6.49	—
Eléctrico	Constante	-9.428 / -2.88	—	No	—	—	44.445 / 4.63
Eléctrico	Ninguna	-9.283 / -1.95	—	—	—	—	—

Tabla 11: Resultados de las pruebas ADF con las series diésel y eléctrico diferenciadas

Tras aplicar la diferenciación en la serie de diésel y eléctricos se observa que al aplicar de nuevo el test ADF (Tabla 11) con tendencia y constante, ahora sí se rechaza la hipótesis nula de no estacionariedad, indicando que la serie es estacionaria bajo esta especificación. No obstante, ni la tendencia ni la constante son significativas por lo que se descarta este modelo. Al aplicar el test con constante, H_0 también se rechaza, pero la constante tampoco es significativa. Por tanto, se utiliza finalmente el modelo sin constante y ni tendencia ya que también se rechaza H_0 . Se concluye que las series diferenciadas de diésel y eléctricos son estacionarias y no requieren transformaciones adicionales.

5.2.5-Análisis de la estacionalidad

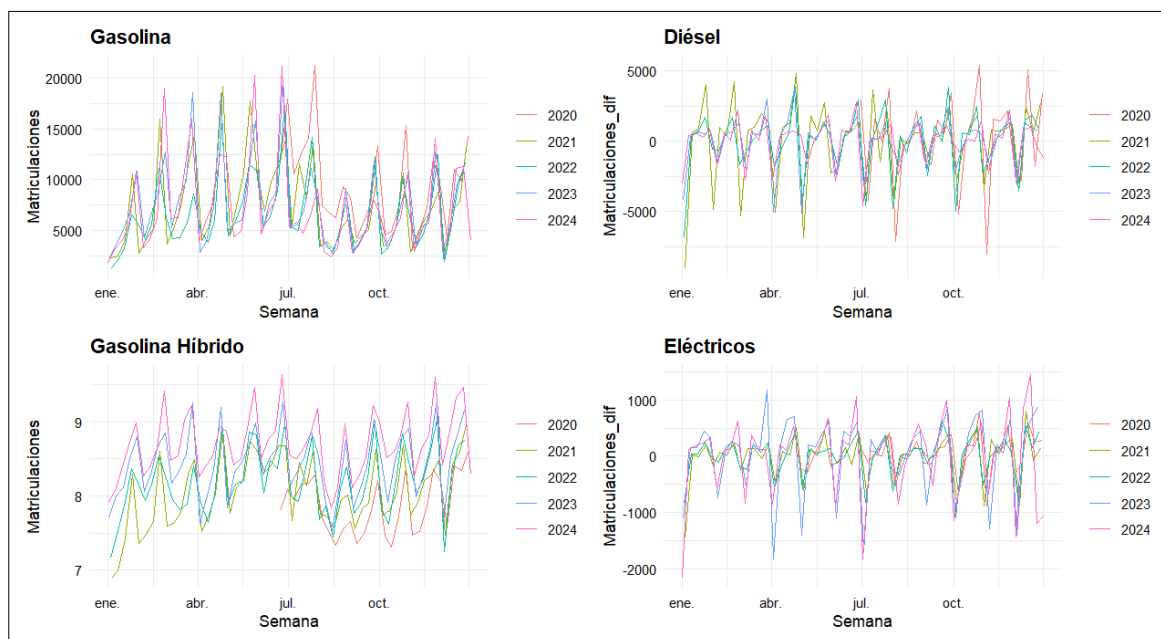


Ilustración 17: Gráfico estacionalidad

La mejor forma de concluir o no la presencia del componente estacional es mediante gráficos (Ilustración 17). Se observa como para cada línea de cada año se superponen o siguen un patrón similar. Esto indica la presencia de una estacionalidad anual, en este caso 52 semanas.

Además, con este gráfico se puede apreciar la tendencia creciente, decreciente o sostenida por tipo de vehículo. Ya que cada línea hace referencia a un año.

Por ejemplo, para los turismos de gasolina se aprecia como las líneas se superponen indicando que el número de las matriculaciones ni aumenta ni disminuye, sino que es constante (sin tendencia significativa).

Sin embargo, para los gasolina híbridos y los eléctricos, las líneas según van avanzado los años están por encima del resto, indicando un claro aumento de la demanda de este tipo de turismos (tendencia positiva). Para los diésel ocurre exactamente lo contrario (tendencia negativa).

5.2.6-Ajuste del modelo ganador

Tecnología	Modelo SARIMA	d	D
Gasolina	(2,1,0)(1,1,1)[52]	1	1
Diésel	(1,1,2)(0,0,3)[52]	1	0
Gasolina Híbrido	(3,1,0)(0,1,0)[52]	1	1
Eléctrico	(0,1,0)(3,0,3)[52]	1	0

Tabla 12: Mejores modelos predictivos por serie

A la hora de seleccionar el mejor modelo, se busca el que mejor predice, y, por tanto, no tienen porque coincidir con el modelo que mejor ajusta. Esto significa que podría ocurrir que en el análisis exploratorio no se evidencie una raíz unitaria, pero que luego el modelo elegido sea uno en el que se diferencia la serie.

Esta estrategia consiste en evaluar concretamente 1024 combinaciones en los cuales se alterna los valores de los parámetros (p , q , P , Q) y de las diferenciaciones (d , D). Se selecciona el modelo óptimo en función de su capacidad predictiva, aquellos que minimicen MSE o MAE.

Para llevar a cabo este proceso, se ha dividido cada serie en dos subconjuntos: uno de prueba (junio 2020-diciembre 2023) y un conjunto de prueba compuesto por las 52 semanas de 2024. El modelo se ajusta con los datos del conjunto de entrenamiento y su rendimiento se evalúa en el de prueba (2024).

Aunque este enfoque no es el más preciso, ya que la forma más robusta sería analizar las predicciones a un solo paso y reestimar el modelo de forma iterativa a mediada que se añaden observaciones, se ha optado por este procedimiento porque implica un coste computacional menor en comparación con la forma iterativa. Aun así, con el método realizado, se ha requerido poco más de 24 horas en encontrar el mejor modelo debido al gran volumen de datos y combinaciones.

Los modelos SARIMA seleccionados mediante la búsqueda exhaustiva bajo las mejores predicciones sobre el conjunto de prueba muestran modelos diferentes en la Tabla 12.

Esta búsqueda establece la existencia de una raíz unitaria ya que para encontrar cada uno de los mejores modelos se ha necesitado una diferenciación en la componente permanente. Además, para las series gasolina y gasolina híbrido, la función determina que también es necesaria una diferenciación estacional.

Cada componente del modelo se interpreta así:

- El parámetro autorregresivo no estacional (p) indica cuántas semanas anteriores influyen sobre el valor actual. Por ejemplo, $p=3$ en los híbridos indica una dependencia reciente (tres semanas).
- El parámetro media móvil (q) evalúa cómo influye los errores pasados sobre el valor actual de la serie. Por ejemplo, $q=2$ en los diésel indica que los errores de las dos últimas semanas afectan al valor actual de la serie.
- Los términos P y Q indican la influencia de las observaciones y de los errores de la misma semana del año anterior.

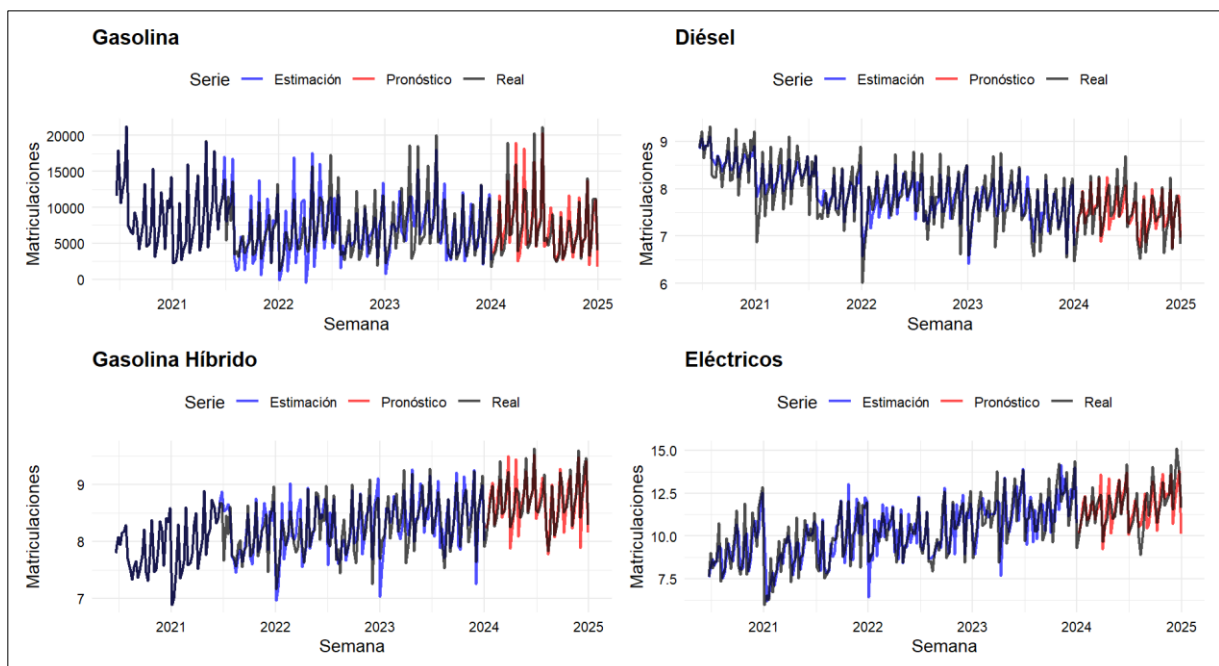


Ilustración 18: Gráfico de las estimaciones y predicciones por tipo de turismo

Generalmente se observa como las estimaciones y las predicciones de cada uno de los modelos ajustados capturan con bastante precisión las series originales.

El conjunto original se dividió en dos subconjuntos, en el de entrenamiento y en el de prueba (compuesto de las 52 últimas semanas de la serie) que es donde se evalúa.

En el caso de los gasolina, el modelo al inicio de la serie entre 2020 y mediados de 2021 casi no genera error. A medida que avanza la serie algunas variaciones no se capturan, pero nada preocupante ya que en general es bastante preciso.

El mejor modelo ajustado para los diésel es bastante adecuado. Captura prácticamente toda la estructura de la serie original y, además, la tendencia proyectada para 2024 es afín a la dirección que venía siguiendo en los años anteriores.

Los gasolina híbrido tienen un modelo que captura el inicio de una serie igual de bien que los gasolina, donde las estimaciones y las predicciones son bastante estables siguiendo la misma tendencia. Por lo que la diferenciación estacional ha favorecido a la hora de capturar los patrones de la serie.

Finalmente, el mejor modelo ajustado de los eléctricos comparte prácticamente las mismas características del resto de modelos en cuanto a estimaciones y predicciones.

5.2.7-Evaluación de los modelos ajustados y diagnóstico de los residuos

En este apartado se analizan las métricas de error y los residuos de cada modelo SARIMA para comprobar y afirmar su precisión y validez.

El error de predicción es la diferencia entre el valor real y el valor estimado. El mejor modelo en cada caso se elige para la muestra de prueba en base a los criterios MAE y MSE.

El objetivo consiste en verificar que los modelos ajustados para cada serie predicen con fiabilidad y que los errores no presentan patrones que se repiten con el tiempo. Cabe destacar que la selección del modelo se basa en el menor error en la muestra de prueba, no en la de entrenamiento.

Tipo	Conjunto	MAE	RMSE	MAPE
Gasolina	Entrenamiento	1370.2863445	2115.0396378	23.109032
Gasolina	Prueba	1336.427698	2018.1120872	18.485203
Diésel	Entrenamiento	0.2006687	0.2642318	2.582317
Diésel	Prueba	0.1778177	0.2497481	2.356406
Híbrido	Entrenamiento	0.1455659	0.2271829	1.779641
Híbrido	Prueba	0.1639433	0.2233624	1.874659
Eléctrico	Entrenamiento	0.7958624	1.0635796	8.195712
Eléctrico	Prueba	0.6106564	0.8049575	5.166538

Tabla 13: Métricas de error de predicción por serie temporal

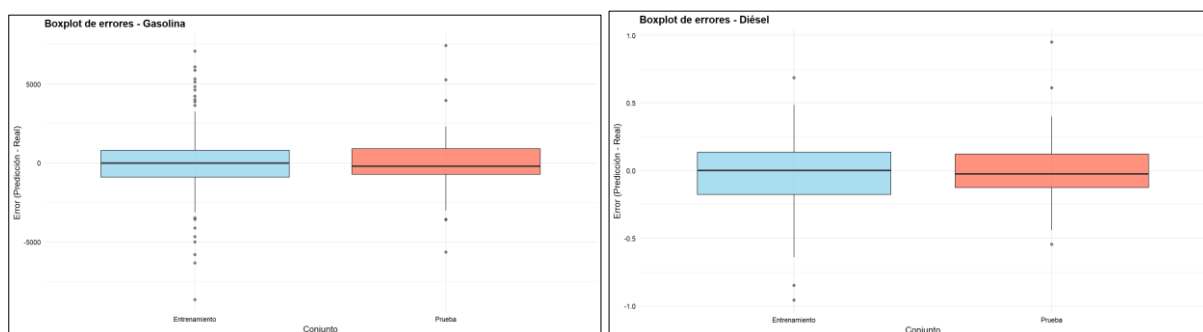


Ilustración 19: BoxPlot de errores de predicción en la muestra de prueba

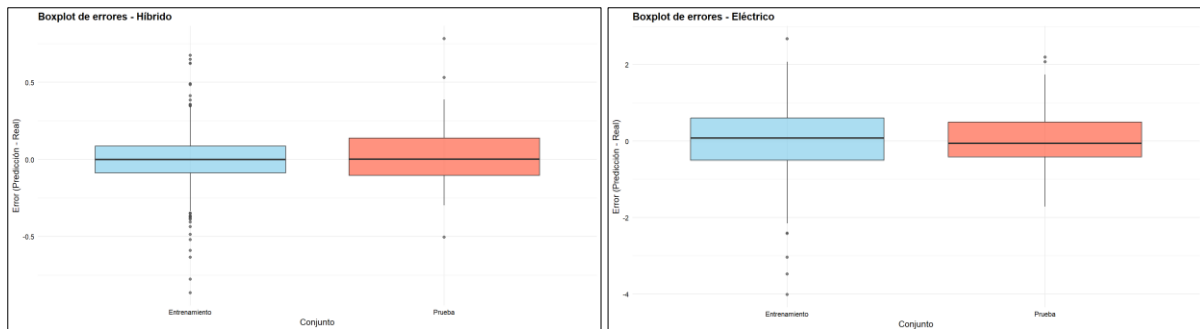


Ilustración 20: Boxplot de errores de predicción en la muestra de prueba

Se observa que no se puede comparar las métricas de error, como el MAE y el RMSE de cada uno de los modelos SARIMA ya que están en escalas diferentes debido a las transformaciones que se hicieron para estabilizar la varianza de los errores.

En cambio, el MAPE es una métrica que se utiliza para evaluar la precisión del modelo de predicción expresando el error promedio en forma de porcentaje respecto al valor real.

Los errores muestran errores porcentuales bajos en los de gasolina híbridos (1,87%) y diésel (2,36%), moderados para eléctricos (5,17%) y más elevados en el modelo de gasolina (18,48%). Para casi todos los modelos el error es inferior en el conjunto de prueba, indicando buena capacidad predictiva sobre datos nuevos (muestra de prueba).

Si analizamos el ajuste del modelo en la muestra de entrenamiento, estos resultados se refuerzan visualmente mediante los *boxplots* de errores donde la mediana se sitúa cerca de 0 en todas las tecnologías. Que la mediana de los errores este centrada en el 0, indica que no hay sesgo.

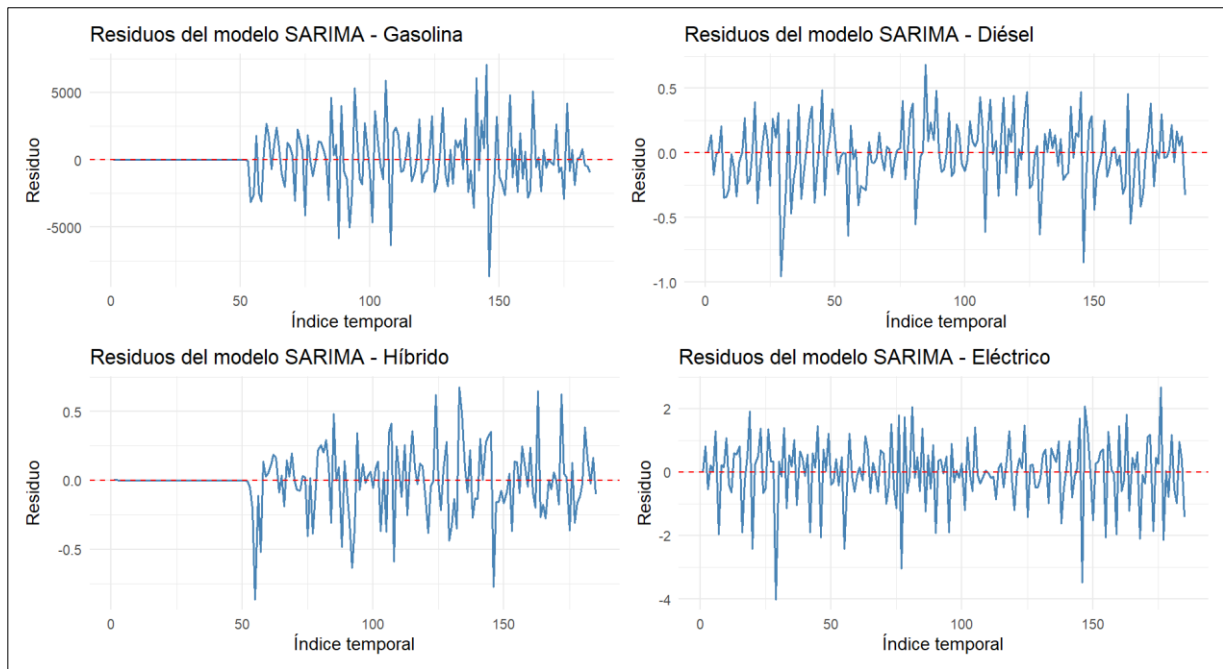


Ilustración 21: Gráfico de los residuos en el conjunto de entrenamiento

Se finaliza este apartado con un gráfico de líneas de los residuos del modelo para el conjunto de entrenamiento, ya que estos residuos permiten evaluar si el modelo ha capturado bien la estructura interna de la serie. Si se observa, como en este caso, que la serie de los errores oscila en torno al cero sin patrones a lo largo del eje, indica que los residuos son ruido blanco y que los modelos han capturado adecuadamente la estructura de cada serie.

5.2.8-Pronósticos para 2025

Para finalizar el análisis de series temporales, se realiza un pronóstico para 2025 con el objetivo de evaluar e interpretar la posible evolución de cada tipo de turismo. Analizando la tendencia y averiguar si va a seguir la misma dirección como los años anteriores.

Para ello, una vez encontrado el modelo óptimo según la capacidad predictiva para 2024, se vuelve a ajustar utilizando todos los datos disponibles hasta diciembre de este año.

Este reentrenamiento permite realizar predicciones más fiables al contar con los datos más recientes. Por lo tanto, a partir del modelo actualizado, se realizan las proyecciones para 2025.

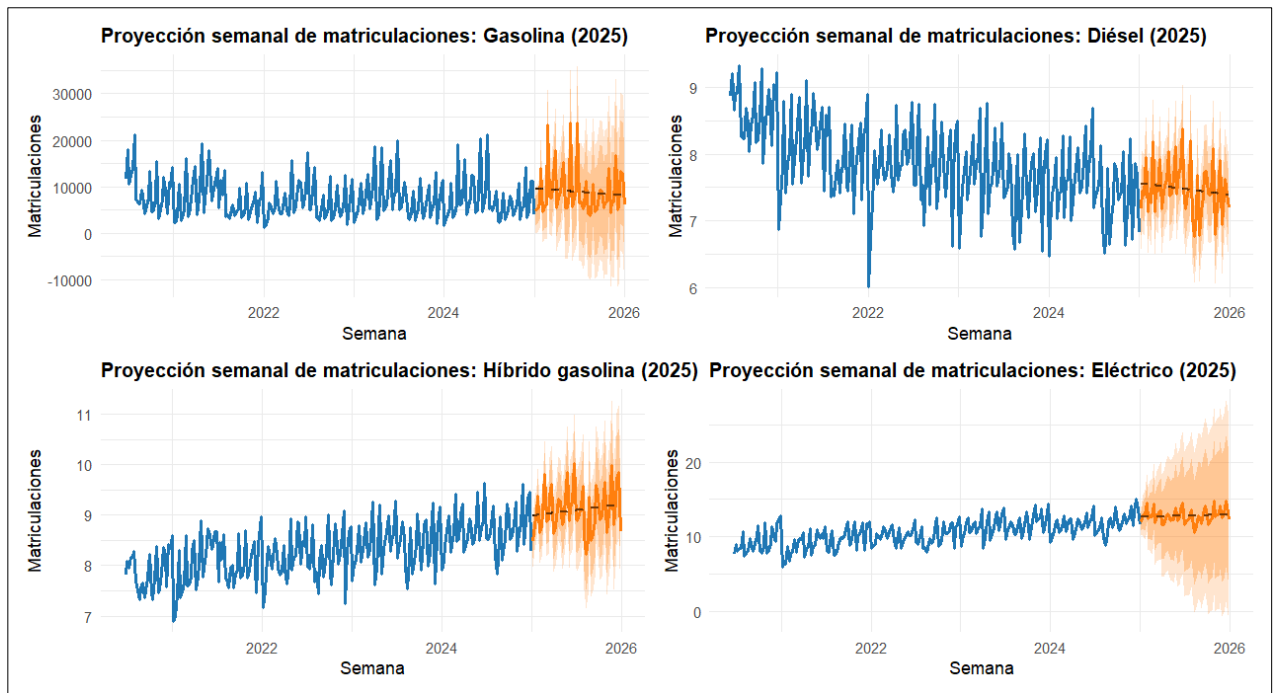


Ilustración 22: Gráfico de las proyecciones para 2025 por tipo de turismo

En primer lugar, se aprecia que la proyección generada por los modelos de los gasolina y de los diésel (Ilustración 22) siguen una tendencia negativa, aunque no muy pronunciada. Esta tendencia sigue la misma dirección que los años anteriores indicando que estas tecnologías tienden a ir desapareciendo con el paso de los años.

Los turismos 100% eléctricos siguen una tendencia positiva aunque no muy pronunciada. Esta proyección ascendente puede indicar que, con el paso del tiempo, este tipo de turismos irán ganando peso en el mercado. Sin embargo, se aprecia cierta preferencia por los gasolina híbridos.

Esta evolución podría deberse por los incentivos y ayudas a la hora de comprar un coche de estas características y también por políticas medioambientales más estrictas que prohíban el acceso a zonas con turismos que no sean cero emisiones, que ya se están viendo en grandes ciudades en España.

En cambio, para los gasolina híbridos, las proyecciones generadas por el modelo indican una 2025 con un mayor volumen de matriculaciones con una tendencia claramente positiva. Es más que evidente que este tipo de turismos se va a ver cada vez más por las calles debido a las ventajas que ofrece.

Las bandas sombreadas en color naranja representan los intervalos de confianza de las predicciones realizadas por cada uno de los modelos. Las zonas más oscuras hacen referencia al intervalo del 80% y la más clara a la del 95%. Estas zonas indican el rango del cual es probable que se sitúen las matriculaciones reales en 2025. Cuanto más anchas son estas bandas, mayor será la variabilidad.

En conclusión, los resultados confirman una evolución en España hacia tecnologías más modernas y sostenibles. Donde el peso de las matriculaciones de los diésel y gasolina seguirán siendo altas pero que, a medida que avancen los años, irán desapareciendo.

En cuanto los eléctricos y los híbridos, se espera todo lo contrario, un aumento progresivo pero lento.

5.3 Análisis Multivariante

En esta sección se analiza series temporales conjuntamente considerando un nivel de agregación mensual.

Con el objetivo de averiguar si el aumento o disminución de las matriculaciones de un tipo de tecnología influyen en las de la otra. Y, también se verifica cuánto influye cada tecnología en la variabilidad de otro tipo de turismo.

Se centra en el análisis de homocedasticidad, en la desestacionalización, estacionariedad y en la estimación del mejor modelo VAR, esto es el modelo que mejor ajusta (no el que mejor predice).

A parte se incluye interpretaciones de IRF (análisis-impulso-respuesta) y de la descomposición de varianza (FEVD).

5.3.1-Analisis de la homocedasticidad

Igual que en el análisis univariante, se comprueba que la varianza de los errores de los modelos es homocedástica para un nivel de desagregación mensual.

Este análisis se realiza sobre las series originales con agregación mensual, con un fin exploratorio, para comprobar la homocedasticidad de cada serie y aplicar transformaciones si es necesario. Para ello, se ajusta un modelo de regresión lineal simple sobre la tendencia temporal en cada serie. Y posteriormente se aplica el test de Breusch-Pagan a los residuos de cada modelo, ya que este test se usa para detectar heterocedasticidad en el contexto de modelos de regresión.

Tecnología	BP Statistic	gl	p-value
Gasolina	3.3419	1	0.06754
Diésel	9.0028	1	0.002696
Gasolina Híbrido	2.4409	1	0.1182
Eléctrico	3.4185	1	0.06447

Tabla 14: Resultados prueba de homocedasticidad

En la Tabla 14, se observa que sólo se rechaza la hipótesis nula de homocedasticidad para los diésel. Para el resto de las series, se concluye que la varianza de los errores es constante a lo largo de la serie.

Aunque no todas las series sean heterocedásticas, se aplican transformaciones logarítmicas en todas las series incluido en las que son homocedásticas para tener todas en la misma escala y facilitar así las interpretaciones.

Tecnología	BP Statistic	gl	p-value
Gasolina	1.7708	1	0.1833
Diésel	2.6832	1	0.1014
Gasolina Híbrido	3.713	1	0.05399
Eléctrico	3.5443	1	0.05975

Tabla 15: Resultados prueba de homocedasticidad para la variable transformada (log)

En la Tabla 15, se muestra que tras aplicar las transformaciones se concluye que han sido efectivas porque para cada una de ellas no se rechaza H_0 al ser el p-valor mayor que 0,05.

5.3.2-Proceso de desestacionalización

Los modelos tipo VAR suponen que las series no tienen estacionalidad. Por ello, es necesario desestacionalizar cada una de las series transformadas antes de ajustar el modelo. Porque si no se hace, puede confundir las relaciones entre cada uno de los tipos de turismo.

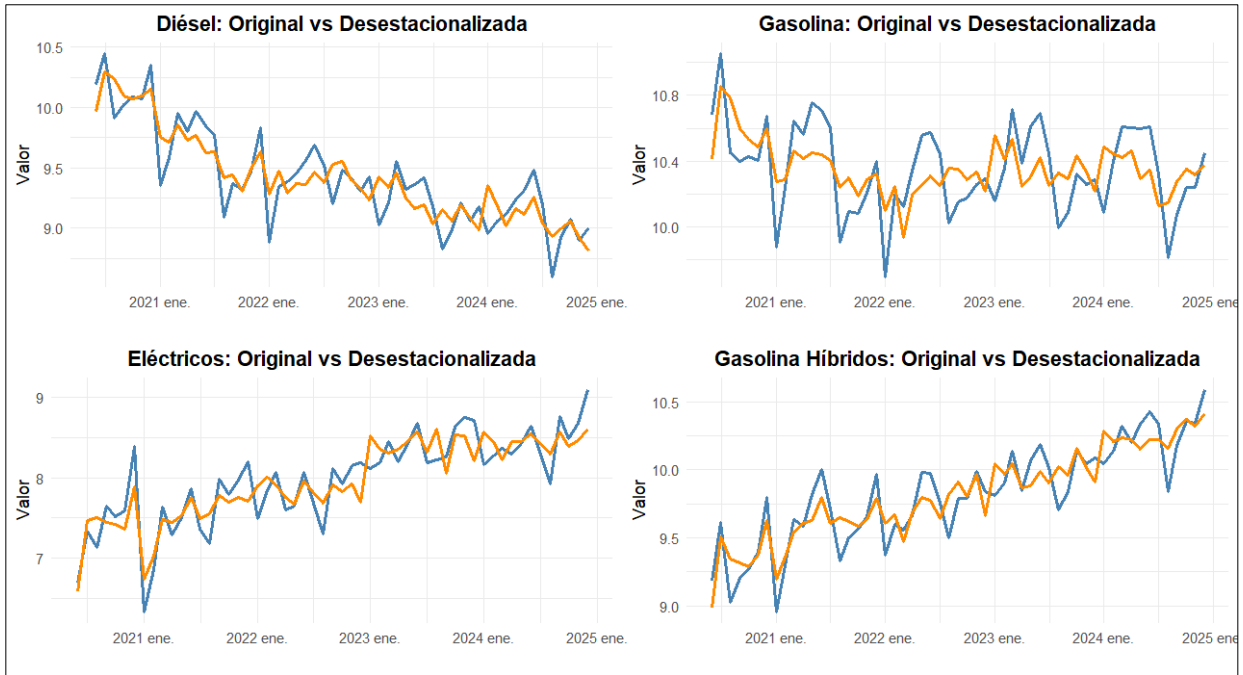


Ilustración 23: Series transformadas desestacionalizadas

Al desestacionalizar cada una de las series se observa las tendencias que siguen de forma más precisa.

5.3.3-Análisis de estacionariedad

En este apartado se repite el mismo proceso que se aplica en series temporales univariantes para detectar o no la presencia de una raíz unitaria y por consecuencia la no estacionariedad de las series.

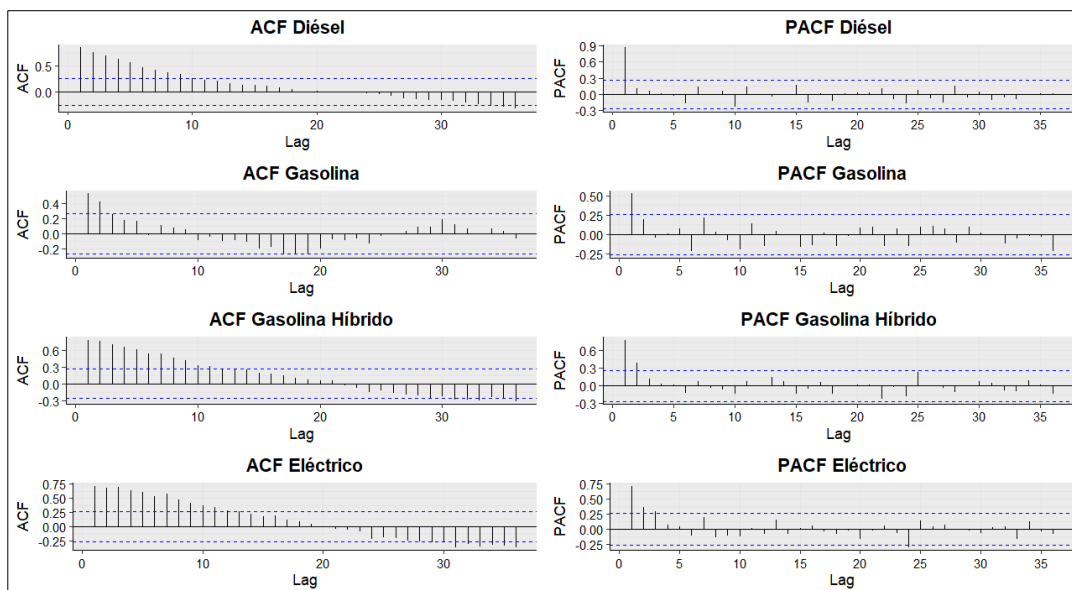


Ilustración 24: Gráficos ACF y PACF de las series mensuales

Para todos los gráficos ACF, salvo para los gasolina, las barras verticales están por encima de las bandas horizontales y van disminuyendo progresivamente. Esto significa autocorrelación.

En los gráficos de la derecha, menos para los gasolina, se observa como el primer rezago es significativo, llegando a casi 0,9 para diésel y a 0,75 para eléctricos, lo que indicaría la persistencia de estas series. Esta persistencia no implica necesariamente raíz unitaria. El test ADF es posible que la detecte en diésel y eléctricos, pero menos probable en gasolina y gasolina híbrido ya que presentan menor autocorrelación inicial.

Se comprueba estos resultados con la prueba ADF calculada sobre los modelos *trend*, *drift* y *none* aplicado sobre la serie transformada y desestacionalizada.

Tecnología	Especificación	Estadístico ADF / Valor crítico	¿Tendencia significativa?	¿Constante significativa?	phi2 / Valor crítico	phi3 / Valor crítico	phi1 / Valor crítico
Diésel	Tendencia	-3.899 / -3.45	Sí	Sí	5.879 / 4.88	7.606 / 6.49	—
Gasolina	Tendencia	-3.227 / -3.45	No	Sí	3.626 / 4.88	5.421 / 6.49	—
Gasolina	Constante	-3.028 / -2.89	—	Sí	—	—	4.602 / 4.71
Gasolina híbrido	Tendencia	-4.859 / -3.45	Sí	Sí	9.116 / 4.88	12.435 / 6.49	—
Eléctrico	Tendencia	-3.069 / -3.45	Sí	Sí	3.546 / 4.88	4.712 / 6.49	—
Eléctrico	Constante	-1.092 / -2.89	—	No	—	—	1.815 / 4.71
Eléctrico	Ninguna	1.510 / -1.95	—	—	—	—	—

Tabla 16: Resultados de las pruebas ADF

Si el valor del estadístico es menor que el calor crítico (Tabla 16), se rechaza la hipótesis nula (H_0 =No es estacionaria, es decir, presenta una media que cambia con el tiempo (raíz unitaria)).

Para los diésel y los gasolina híbridos, el test ADF con tendencia y constante se rechaza la hipótesis nula de raíz unitaria al 5% de significación, ya que el estadístico tau3 (ADF) es menor que el valor crítico, por lo que se concluye que son estacionarias. Los estadísticos phi2 y phi3 indican que tanto la constante como la tendencia son necesarios en el modelo y además ambos coeficientes son estadísticamente significativos. Por lo que no es necesario la diferenciación y el mejor modelo es el que incluye constante y tendencia.

En el caso de los gasolina, no se puede rechazar H_0 en el modelo con tendencia y constante por que el estadístico tau3 es mayor que el valor crítico a un nivel de significación del 5%. Además, los estadísticos phi2 y phi3 indican que ni la tendencia

ni la constante son necesarios en el modelo. La tendencia no es significativa pero la constante sí, aun así, se descarta este modelo y se prueba a hacer el ADF con constante solamente. Para esta segunda especificación, se rechaza la hipótesis nula de raíz unitaria y se concluye que es estacionaria y el estadístico phi1 indica que la constante no es necesaria. Finalmente, se elige el modelo con constante y no es necesaria la diferenciación.

En el caso de los eléctricos, no se rechaza en todos los casos que no es estacionaria por lo que es necesaria la diferenciación.

Para el caso de los eléctricos se aplica una diferenciación y se vuelve a realizar la prueba ADF sobre los modelos *trend*, *drift* y *none*.

Tecnología	Especificación	Estadístico ADF / Valor crítico	Tendencia significativa?	Constante significativa?	phi2 / Valor crítico	phi3 / Valor crítico	phi1 / Valor crítico
Eléctrico	Tendencia	-8.188 / -3.45	No	No	22.369 / 4.88	33.524 / 6.49	—
Eléctrico	Constante	-8.254 / -2.89	—	No	—	—	34.094 / 4.71
Eléctrico	Ninguna	-7.944 / -1.95	—	—	—	—	—

Tabla 17: Resultados de las pruebas ADF para eléctricos diferenciada

Tras aplicar la diferenciación en la serie de eléctricos se observa que al aplicar de nuevo el test ADF (Tabla 17) con tendencia y constante, ahora sí se rechaza la hipótesis nula de no estacionariedad, indicando que la serie es estacionaria bajo esta especificación. Aunque los estadísticos phi2 y phi3 superan los valores críticos, como ni la tendencia ni la constante son significativas, no se justifica mantener ambos términos en el modelo. La razón es que como tau3 es distinto de cero, eso ya nos lleva a rechazar phi2 y phi3, aunque la constante y tendencia fuesen ambas cero. Por ese motivo, se prueba con la especificación con constante, donde también se rechaza H0, pero de nuevo la constante no es significativa. Finalmente se evalúa el modelo sin constante y tendencia y el test vuelve a rechazar H0. Por lo que, se concluye que la serie diferenciada de los eléctricos es estacionaria, y que el modelo más adecuado es el que no incluye ni constante y ni tendencia.

5.3.4-Test de causalidad de Granger

Se aplica este test para analizar si las matriculaciones pasadas de un tipo de turismo ayuda a predecir las de otro. Permite identificar relaciones de dependencia temporal entre tipos de vehículo, mostrando si una serie “Granger-causa” (significa que una variable X ayuda a predecir a otra variable Y) a otra en términos predictivos.

Resultados del Test de Granger			
Causa	Dependiente	p-valor	Significativo
Gasolina	Diésel	0.02554	✓
Híbrido	Diésel	5.845e-06	✓
Eléctrico	Diésel	0.05687	.
Diésel	Gasolina	0.802	✗
Híbrido	Gasolina	0.1557	✗
Eléctrico	Gasolina	0.2691	✗
Diésel	Híbrido	6.542e-06	✓
Gasolina	Híbrido	0.0007308	✓
Eléctrico	Híbrido	0.0002031	✓
Diésel	Eléctrico	0.818	✗
Gasolina	Eléctrico	0.6774	✗
Híbrido	Eléctrico	0.496	✗

Tabla 18: Resultados test de Granger

De acuerdo con la Tabla 18, el test de causalidad de Granger muestra varias relaciones unidireccionales significativas. Por ejemplo, se detecta causalidad de gasolina hacia diésel con un p-valor inferior a 0,05. Esto sugiere que cambios en las matriculaciones de gasolina, provocan cambios en las matriculaciones de diésel, esto puede deberse a la competencia entre tecnologías tradicionales. También se observan efectos significativos de los gasolina híbridos a los diésel. Por otro lado, las matriculaciones pasadas de los diésel, gasolina y eléctricos causan a los híbridos en el sentido de Granger.

Como se ha encontrado una relación bidireccional entre gasolina híbridos y diésel, es decir, ambos tipos de turismos se influyen mutuamente en el tiempo, se justifica el uso de un modelo VAR, ya que permite analizar todas las interacciones entre todas las tecnologías para rezagos específicos.

5.3.5-Ajuste del modelo VAR

Primero con los criterios estadísticos se selecciona el número óptimo de rezagos que se debe incluir en el modelo mediante la evaluación de varios posibles rezagos.

Criterio	Rezago óptimo
AIC (Akaike)	2
HQ (Hannan-Quinn)	2
SC (Schwarz/BIC)	1
FPE (Error de predicción final)	2

Tabla 19: Selección del modelo con los criterios de información

En este caso, todos los criterios de la Tabla 19 indican que el mejor modelo solo incluye 2 rezagos salvo bajo el criterio BIC.

En otras palabras, esto significa que el modelo solo necesita saber qué pasó en los dos meses anteriores para poder explicar y tender el comportamiento de las series.

Antes de interpretar los parámetros del modelo, se debe verificar que el modelo estimado cumple cuatro supuestos: ausencia de autocorrelación, distribución normal de los residuos, homocedasticidad y la estabilidad del modelo.

Test de diagnóstico	p-valor / criterio	Resultado
Autocorrelación (Portmanteau)	0.3458	No se rechaza H0: sin autocorrelación
Normalidad (JB Test)	0.4642	No se rechaza H0: normalidad multivariada
Asimetría (Skewness)	0.4389	No se rechaza H0: sin asimetría significativa
Curtosis (Kurtosis)	0.416	No se rechaza H0: sin curtosis excesiva
Heterocedasticidad (ARCH)	0.8282	No se rechaza H0: homocedasticidad
Estabilidad del modelo (raíces < 1)	Cumple (todas < 1)	El modelo es estable

Tabla 20: Comprobación de los 4 supuestos

En todas las pruebas de la Tabla 20 se cumple con los supuestos mencionados. Esto asegura que el modelo ajustado sea válido y que ofrezcan interpretaciones fiables.

Después de verificar que el mejor modelo VAR (VAR (2)) no incumple los requisitos se obtiene los siguientes parámetros:

	diesel.l1	gasolina.l1	híbrido.l1	eléctrico.l1	diesel.l2	gasolina.l2	híbrido.l2	eléctrico.l2	constante
diesel_des	0.269	0.006	-0.205	-0.070	-0.087	0.548°	-0.539*	-0.019	9.241**
gasolina_des	-0.491*	0.682*	-0.102	-0.048	0.107	0.307	-0.250	-0.046	7.180**
hibrido_des	-0.124	0.121	0.347	-0.156°	-0.255	0.048	0.248	-0.079	5.840*
electrico_des	-0.185	0.349	-0.680	-0.566**	-0.386	0.246	0.142	-0.369*	4.560

Tabla 21: Estimación de los parámetros del VAR (2)

Cuando se estima el modelo, se construye una ecuación por cada tipo de tecnología. Y cada ecuación representa el comportamiento de una tecnología en función de sus dos meses anteriores y el del resto de tecnologías más un error y una constante.

Por ejemplo, la ecuación de los diésel, explica las matriculaciones que ha tenido en el mes actual en función de lo que pasó en los meses anteriores con los diésel, gasolina, gasolina híbridos y los eléctricos.

$$y_t^D = \omega_D + \alpha_1 y_{t-1}^D + \alpha_2 y_{t-2}^D + \delta_1 y_{t-1}^G + \delta_2 y_{t-2}^G + \beta_1 y_{t-1}^{GH} + \beta_2 y_{t-2}^{GH} + \gamma_1 y_{t-1}^E + \gamma_2 y_{t-2}^E + \varepsilon_t^D$$

Y para interpretar si algún tipo de turismo ha tenido un efecto significativo, se aprecia con la tabla si el parámetro es significativo.

Los valores con un ° significan que son significativos un nivel de significación del 10%, con un * al 5% y con ** al 1% (muy significativo).

En el caso del diésel, no se han detectado efectos significativos propios. Sin embargo, el rezago correspondiente a los gasolina híbridos sí lo es en el segundo rezago. Esto significa que un aumento de las matriculaciones de los gasolina híbridos hace dos meses, está asociado con una disminución en las matriculaciones de los diésel actuales.

Por otro lado, un aumento en las matriculaciones de los eléctricos en los meses anteriores está asociado con una disminución en las matriculaciones actuales. Este comportamiento sugiere que la demanda se ajusta rápidamente tras una subida puntual.

Los gasolina muestran que, si suben un mes, suelen seguir subiendo en el mes siguiente. Además, el diésel tiene un efecto negativo. Cuando las matriculaciones de los diésel suben en el mes anterior, las matriculaciones actuales de los gasolina tienden a disminuir.

Si relacionamos estos resultados con los del test de Granger (Tabla 18), para los diésel se observa un efecto negativo y significativo de los gasolina. Esta relación es bidireccional ya que el test de Granger lo indica.

En el caso de los gasolina, el VAR muestra que los diésel tienen un efecto negativo y significativo sobre los gasolina, pero el test de Granger ha detectado lo contrario, reflejando una competencia entre ambas. Esto no supone una contradicción, ya que

el test de Granger analiza relaciones dos a dos, mientras que el VAR estima un modelo multivariante analizando todas las relaciones conjuntamente.

Los gasolina híbridos no reciben efectos significativos en el VAR, pero el test de Granger muestra que las matriculaciones de este tipo de turismos están influenciadas por las demás, reforzando la idea de que su evolución depende del mercado.

Sin embargo, los eléctricos no reciben influencia significativa de otras tecnologías ni en el VAR ni en el test de Granger.

5.3.6-Análisis Impulso-Respuesta

Este análisis sirve para comprobar las relaciones entre las tecnologías cuando se produce un impulso (aumento) en una variable. Por ejemplo, si se produce un aumento repentino en los eléctricos, se puede comprobar qué pasa con los diésel en los meses siguientes.

La línea negra es la estimación y las líneas rojas son los intervalos de confianza al 95%. Y si el intervalo contiene al 0, el efecto no es significativo.

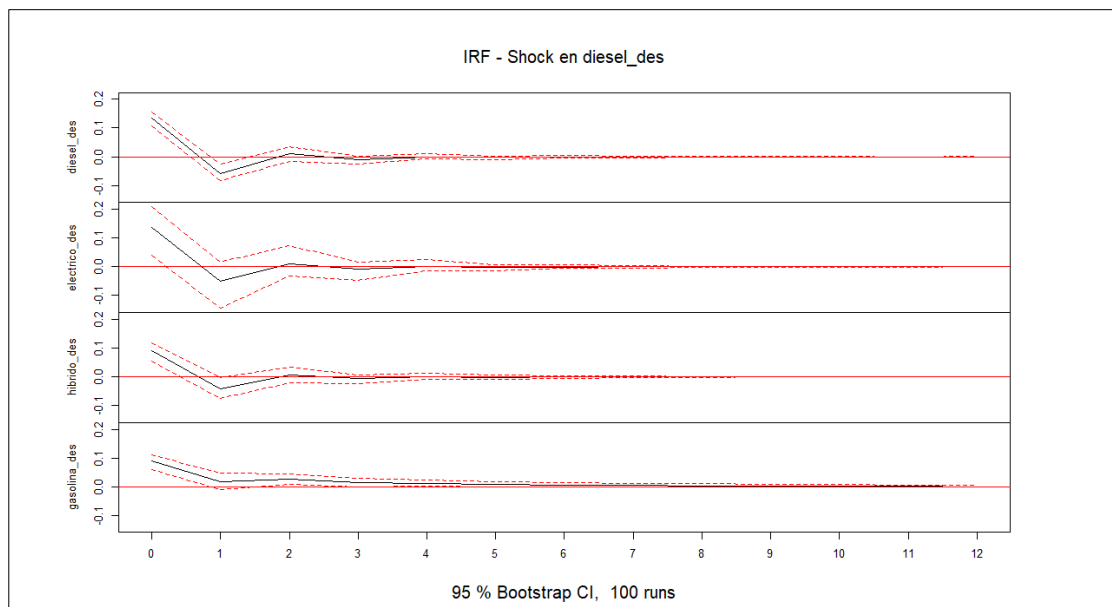


Ilustración 25: IRF de los diésel

Cuando se produce una subida inesperada de las matriculaciones de los diésel, se observa como para los eléctricos se produce una caída significativa. Para el resto de tipos no se producen efectos significativos ante una subida inesperada.

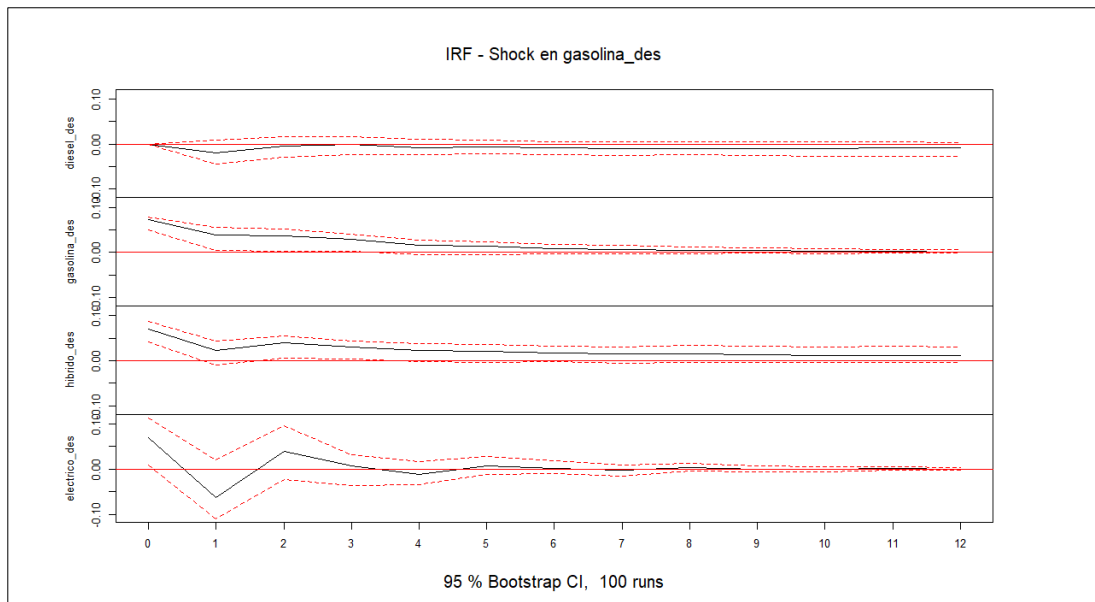


Ilustración 26: IRF de los gasolina

En el caso de los gasolina, un aumento inesperado solo produce un efecto significativo sobre ellos mismos. Ya que el intervalo los primeros 2-3 meses no incluye al 0 pero por bastante poco.

Esto quiere decir, que, ante un aumento, es muy probable que en los próximos meses también aumente. Pero ese aumento ira disminuyendo con el paso del tiempo.

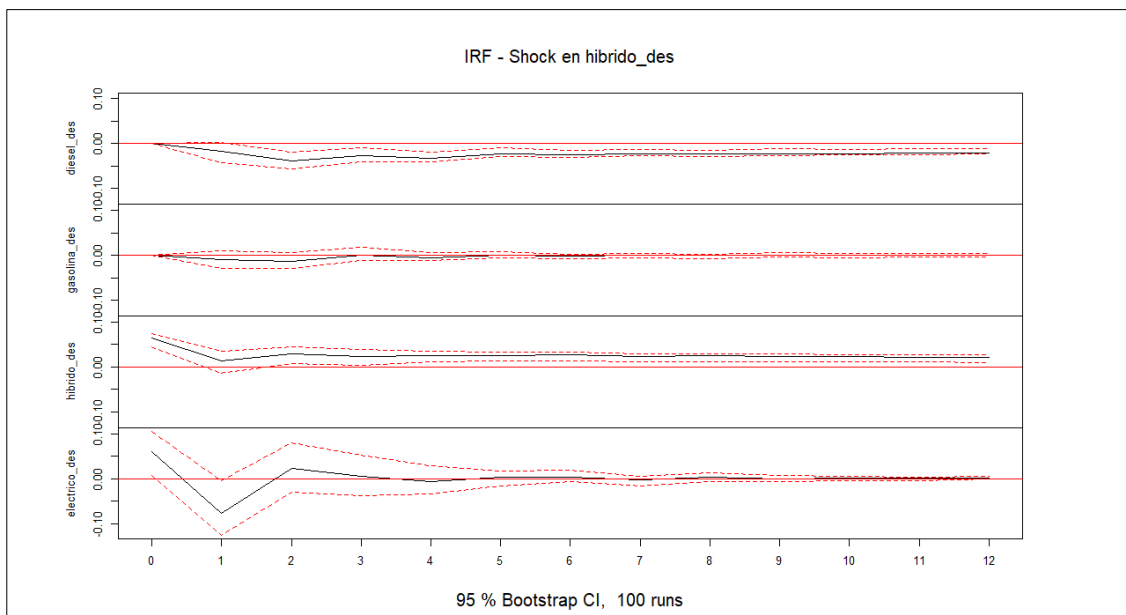


Ilustración 27: IRF de los gasolina hibridos

Para los gasolina híbridos se observa como un aumento inesperado de sus matriculaciones tiene un efecto negativo y significativo en los diésel. Un aumento de la demanda de los gasolina híbridos parece reducir de forma sostenida la de los diésel. Y respecto a sus propias matriculaciones pasa lo mismo, pero de forma positiva. Esto refleja una sustitución entre tecnologías.

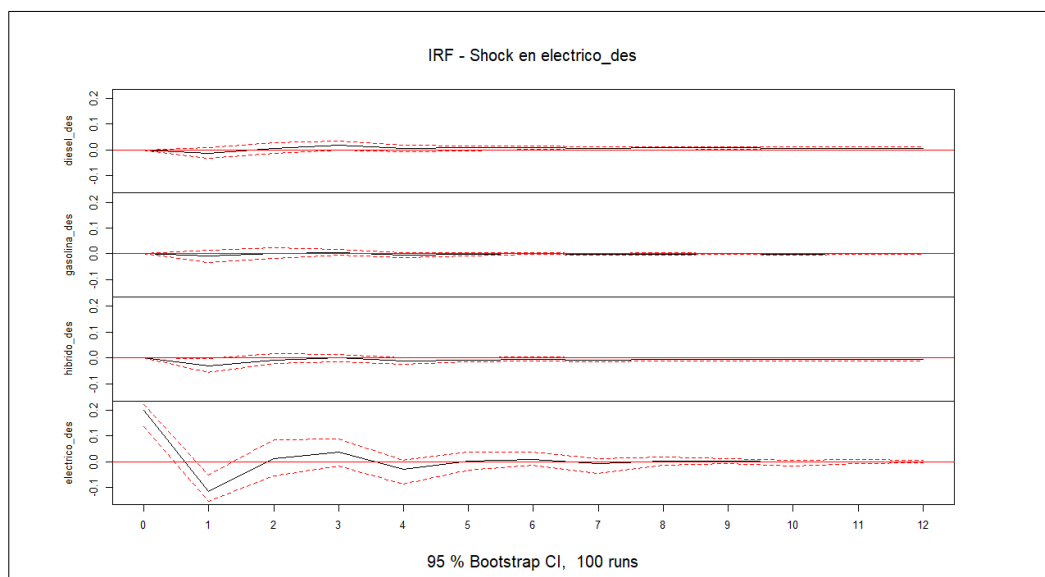


Ilustración 28: IRF de los eléctricos

Por último, el aumento inesperado en los eléctricos de un mes produce un efecto significativo en los eléctricos al mes siguiente reduciendo el número de matriculaciones.

5.3.7-Descomposición de la Varianza

Este análisis ayuda a entender comportamientos de cada una de las variables en el futuro de estas.

A modo de ejemplo: si las matriculaciones de los eléctricos aumentan mucho en un futuro, la descomposición de la varianza dice cuánto de ese cambio se debe a cambios en los diésel, gasolina y gasolina híbridos y cuánto se debe por los eléctricos.

La diferencia clave con IRF es que los IRF te dicen qué ocurre con los gasolina si de repente las matriculaciones los diésel crecen.

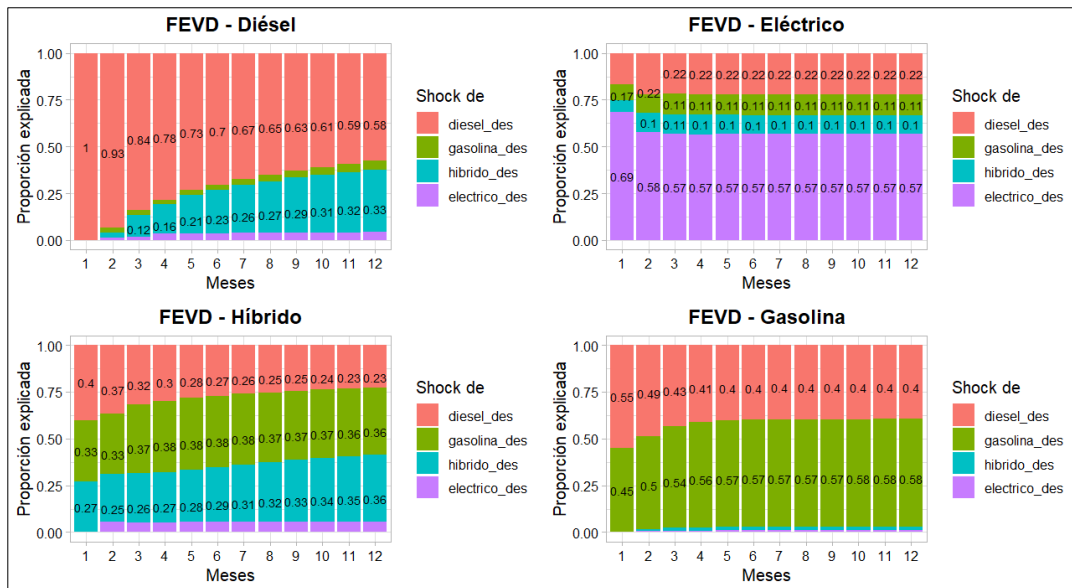


Ilustración 29: Descomposición de la varianza

La varianza futura de las matriculaciones diésel se debe casi al 100% por incrementos o disminuciones de los diésel. Aunque con el tiempo los gasolina híbridos explican hasta un 32% de la variabilidad.

En el caso de los eléctricos, los diésel apenas producen efecto sobre los eléctricos. Es decir, si las matriculaciones de los diésel cambian, las de los eléctricos no tienen por qué hacerlo. Pero mayoritariamente (alrededor del 60%) se debe a las propias variaciones de los eléctricos.

En los gasolina híbridos, sus matriculaciones futuras dependen bastante de lo que suceda con las de los gasolina, los diésel y sus propias matriculaciones.

La variabilidad de las matriculaciones en un futuro de los gasolina tiene mucho que ver con los aumentos y disminuciones de los gasolina. Y, en segundo lugar de las variaciones de los diésel.

Dado que los diésel influyen mucho en las variaciones del resto de tecnologías, esto sugiere que siguen teniendo un peso importante en el mercado, afectando a las tecnologías más modernas y a las tradicionales.

6. CONCLUSIONES

El análisis de la evolución de las matriculaciones por tipo de turismo durante estos años aplicando clúster jerárquico (datos anuales), un análisis univariante con modelos SARIMA en cada serie temporal (datos semanales) y por último el análisis multivariante analizando series conjuntas con el modelo VAR (datos mensuales) ofrece resultados bastante interesantes.

En primer lugar, la utilización de mapas geográficos con cada uno de los análisis ha permitido identificar qué zonas geográficas y provincias se comportan de manera similar mediante la inclusión de tres variables. Al usar la renta neta media por hogar, un índice de matriculaciones cada mil hogares y la proporción de población rural ayuda interpretar aún mejor los resultados.

En el caso de las tecnologías más tradicionales, como los vehículos diésel y gasolina se observa que, a nivel nacional, el volumen de matriculaciones no varía mucho, excepto en Madrid, que para los tres casos ha superado al resto de provincias con mayoría absoluta.

Cuando se estudia el comportamiento de las tecnologías más modernas en 2024, se observa algunas diferencias, donde las provincias con mayor renta y mayor índice de urbanidad tienen el mayor número de matriculaciones de este tipo de turismos cada mil hogares.

Cuando el análisis se enfoca únicamente en los turismos 100% eléctricos de nuevo para 2024, se aprecia la existencia de dos agrupaciones con mayor índice de matriculaciones, como es en el caso de Madrid y por otro lado las provincias de Cataluña, Valencia, Navarra y las Islas Baleares. Según El Ministerio de Transición Ecológica, citado en el artículo de Soria (2025), estas provincias son las que tienen el mayor punto de recargas a nivel nacional. En el resto del territorio español, las matriculaciones de estos turismos apenas llegaron a las 2 cada mil hogares en 2024.

La aplicación de modelos SARIMA ha ayudado a identificar tendencias y estacionalidades en cada una de las series.

Los turismos de gasolina y los diésel siguen una tendencia negativa y progresiva, en consonancia con las políticas medio ambientales que buscan un parque automovilístico más sostenible. Y, por otro lado, los eléctricos y los gasolina híbridos siguen una dirección positiva y aditiva.

Los mejores modelos seleccionados para cada una de las series mediante búsqueda exhaustiva han ofrecido proyecciones para 52 semanas de 2025. Se espera, según las predicciones de cada modelo, que las tendencias que seguían en los años anteriores sigan así, con una transición lenta hacia las tecnologías renovables. Aunque con gran peso aún en volumen de los diésel y los gasolina, sobre todo en las zonas geográficas con mayor índice de ruralidad y menor renta neta media como mostró el análisis clúster.

Para complementar el análisis de series temporales, el enfoque multivariante con la aplicación de modelos VAR ha servido para entender las posibles relaciones entre las diferentes tecnologías.

Las conclusiones que se obtienen de este análisis son, por un lado, que ante un aumento repentino de las matriculaciones de una tecnología, esto afecta negativamente en las matriculaciones propias al mes siguiente estabilizando ese aumento. Y, por otro lado, la variabilidad de las matriculaciones en un futuro de los diésel, gasolina, gasolina híbridos y de los eléctricos, se debe principalmente a los aumentos y disminuciones de los turismos tradicionales, como son los diésel. Aunque los gasolina híbridos también tienen un papel importante en la explicación de la variabilidad de los diésel con el paso del tiempo.

Por lo que, aunque esta tecnología siga una tendencia negativa, se reafirma que será lenta ya que, para proyecciones a futuro, afectará a la variabilidad del volumen de matriculaciones de los gasolina, de los gasolina híbridos y de los turismos 100% eléctricos. Veremos en los próximos años si el ritmo crece al nivel necesario para alcanzar los objetivos de sostenibilidad fijados.

Este estudio se centra en una comprensión profunda de la evolución de las matriculaciones en España, pero podría extenderse de distintas formas.

Para futuras investigaciones, con relación al *clustering*, sería interesante un estudio a nivel europeo, analizando qué zonas geográficas tienen características similares en cuanto a las matriculaciones de cada tipo de turismo.

Los datos de las matriculaciones vienen en formato diario, lo cual dificulta los análisis y que al tener tantos datos en esta frecuencia (aproximadamente 4 millones de datos si contamos desde el fin del estado de emergencia), se introduce mucho ruido y mucha variabilidad escondiendo los patrones importantes. Por otro lado, los datos diarios suelen tener varias estacionalidades a la vez: ciclos semanales (de 7 días, o 5 si contamos solo días laborales), también puede haber patrones anuales (365 días) y mensuales. Estos patrones pueden modelarse con variables *dummy*, pero su identificación es complicada.

Por todo esto, se decide agrupar las series por semanas o meses. Estas agregaciones consiguen suavizar las variaciones bruscas del día a día, lo cual hace más fácil interpretar la tendencia real. También se consigue reducir el impacto de valores atípicos, como los días festivos que afectan al número de matriculaciones.

En cuanto al modelo multivariante VAR, se probó inicialmente a modelarlo con datos semanales, pero la componente estacional no se lograba capturar, afectando este análisis. Por eso se optó por usar una agregación mensual, la cual era más estable y compatible con los requisitos del modelo. Sin embargo, para los modelos SARIMA, sí ha funcionado correctamente con la estacionalidad semanal ($S=52$) usando paquetes estadísticos de R.

Por último, es importante destacar y aclarar que se ha usado dos enfoques diferentes según el objetivo en este trabajo. Para el modelo SARIMA, se ha empleado un procedimiento de búsqueda exhaustiva que selecciona el mejor modelo según su capacidad predictiva. En este caso, el análisis previo que se hizo sobre la homocedasticidad, tendencia, estacionariedad y estacionalidad ha sido exploratorio sin influir en la selección del modelo final.

En cambio, para el modelo VAR el enfoque es distinto. En lugar de buscar el modelo que mejor predice, se ha querido determinar cómo se relacionan las matriculaciones entre los distintos tipos de turismo. Para este análisis, se ajusta el mejor modelo, y para ello las series deben ser estacionarias, ya que esto no se corrige automáticamente.

Las herramientas y métodos desarrollados pueden ser adaptados y utilizados en otros contextos para obtener información detallada y precisa de la situación actual y de la posible situación a futuro de las matriculaciones de turismos nuevos en España para plantear posibles mejoras hacia un parque móvil más sostenible.

7. BIBLIOGRAFÍA

- Acciones de la UE para reducir las emisiones de aviones y barcos.* (2022). Obtenido de <https://www.europarl.europa.eu/topics/es/article/20220610STO32720/acciones-de-la-ue-para-reducir-las-emisiones-de-aviones-y-barcos>
- Álvarez Liébana. (2024). *Apuntes de clase sobre análisis de series temporales.* Obtenido de <https://javieralvarezliebana.es/docencia/time-series/diapos/#/title-slide>
- Athanasopoulos, Hyndman. (2021). *Forecasting: Principles and Practice.* Obtenido de <https://otexts.com/fpp3/arima-ets.html>
- Clustering.* (2020). Obtenido de <https://r.qcbs.ca/workshop09/book-en/clustering.html>
- Crisis sanitaria COVID-19: Normativa e información útil.* (2020). Obtenido de https://administracion.gob.es/pag_Home/atencionCiudadana/Crisis-sanitaria-COVID-19.html
- Dirección General de Tráfico (DGT). (2024). *Microdatos de Matriculaciones de Vehículos (diarios).* Obtenido de <https://www.dgt.es/menusecundario/dgt-en-cifras/dgt-en-cifras-resultados/dgt-en-cifras-detalle/Microdatos-de-Matriculaciones-de-Vehiculos-diarios/>
- Feng Jiang, Jie Liu , Chunzhi Yi. (2023). *Comparative Analysis of the Clustering Quality in Self-Organizing Maps for Human Posture Classification.* Obtenido de <https://pmc.ncbi.nlm.nih.gov/articles/PMC10538130/#sec6-sensors-23-07925>
- Instituto Nacional de Estadística (INE). (2021). *Hogares según su tamaño por tamaño del municipio.* Obtenido de <https://www.ine.es/jaxi/Datos.htm?tpx=59538>
- Instituto Nacional de Estadística (INE). (2022a). *Atlas de distribución de renta de los hogares.* Obtenido de <https://www.ine.es/jaxiT3/Tabla.htm?t=53689>
- Instituto Nacional de Estadística (INE). (2022b). *Estadística del Padrón Continuo.* Obtenido de <https://www.ine.es/jaxi/Tabla.htm?path=/t20/e245/p04/provi/I0/&file=0tam002.px&L=0>
- Kotzé, Kevin. (2019). *Tutorial: Vector Autoregression Models.* Obtenido de https://kevinkotze.github.io/ts-7-tut/?utm_source
- Lee, Sarah. (2025). *Complete Guide to Vector Autoregression: Methods for Forecasting.* Obtenido de <https://www.numberanalytics.com/blog/complete-guide-vector-autoregression/>
- Organización de las Naciones Unidas. (2015). *Objetivos de Desarrollo Sostenible.* Obtenido de <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>
- Soria, Sara. (13 de 12 de 2024). *El reto de Anfac es alcanzar una cuota del 10% en coches eléctricos en 2025.* Obtenido de <https://www.coches.net/noticias/plan-espana-auto-2030-anfac>
- Soria, Sara. (10 de 01 de 2025). *¿Cuántos puntos de recarga eléctrica hay en España?* Obtenido de <https://www.coches.net/noticias/cuantos-puntos-de-recarga-electrica-hay-en-espana>

8. ANEXO

Función para la búsqueda exhaustiva del mejor modelo:

```
for (p in 0:max_order) {
  for (q in 0:max_order) {
    for (P in 0:max_order) {
      for (Q in 0:max_order) {
        for (d in d_vals) {
          for (D in D_vals) {

            combinaciones_totales <- combinaciones_totales + 1
            cat("Probando modelo: p =", p, "q =", q, "| P =", P, "Q =", Q, "| d =", d, "D =", D, "\n")

            fit <- tryCatch({
              Arima(train_data,
                order = c(p, d, q),
                seasonal = list(order = c(P, D, Q), period = 52),
                include.mean = TRUE)
            }, error = function(e) NULL)

            if (!is.null(fit)) {
              combinaciones_validas <- combinaciones_validas + 1

              pred <- forecast(fit, h = length(test_data))$mean
              mse <- mean((test_data - pred)^2, na.rm = TRUE)
              mae <- mean(abs(test_data - pred), na.rm = TRUE)

              # Guardar en log
              result_row <- data.frame(p, d, q, P, D, Q, MSE = mse, MAE = mae)
              write_csv(result_row, log_file, append = TRUE)

              if (mse < best_mse || mae < best_mae) {
                best_mse <- mse
                best_mae <- mae
                best_model <- fit
                best_params <- list(p=p, d=d, q=q, P=P, D=D, Q=Q)
              }

              # Limpieza de memoria
              rm(fit, pred)
              gc()
            }
          }
        }
      }
    }
  }
}
```

Declaración Responsable sobre Autoría y Uso Ético de Herramientas de Inteligencia Artificial (IA)

Yo,

FLORES SEVILLA, GONZALO

declaro de manera responsable que el/la presente:

- Trabajo de Fin de Grado (TFG)
- Trabajo de Fin de Máster (TFM)
- Tesis Doctoral

Titulado/a

ANÁLISIS GEOGRÁFICO Y TEMPORAL DE LAS MATRICULACIONES DE TURISMOS EN ESPAÑA

es el resultado de mi trabajo intelectual personal y creativo, y ha sido elaborado de acuerdo con los principios éticos y las normas de integridad vigentes en la comunidad académica y, más específicamente, en la Universidad Complutense de Madrid.

Soy, pues, autor del material aquí incluido y, cuando no ha sido así y he tomado el material de otra fuente, lo he citado o bien he declarado su procedencia de forma clara -incluidas, en su caso, herramientas de inteligencia artificial-. Las ideas y aportaciones principales incluidas en este trabajo, y que acreditan la adquisición de competencias, son mías y no proceden de otras fuentes o han sido reescritas usando material de otras fuentes.

Asimismo, aseguro que los datos y recursos utilizados son legítimos, verificables y han sido obtenidos de fuentes confiables y autorizadas. Además, he tomado medidas para garantizar la confidencialidad y privacidad de los datos utilizados, evitando cualquier tipo de sesgo o discriminación injusta en el tratamiento de la información.

En Madrid a 04/06/2025



FIRMA: