

Project Break I: EDA

Bootcamp Data Science

Autor: Justo Barco Cerro

Contents

Contexto	3
Descripción de las variables.....	3
Proceso.....	6
ANÁLISIS UNIVARIANTE.....	7
Categorías.....	7
MARCA	7
PROVINCIAS.....	8
PROPULSION.....	9
CATELECT (Categoría eléctrica).....	9
Numéricas.....	10
TARA	10
ANÁLISIS BIVARIANTE	11
Categoría – Categoría.....	11
CHI-2	11
Categoría – numérica	12
Mannwhitneyu.....	12
Anova.....	13
TARA-RENTING.....	15
Numérica – Numérica	16
TARA-PESO_MAX	16
CILINDRADA-POTENCIA.....	17
POTENCIA-KW.....	18
ANÁLISIS GENERAL	20
Evolución de matriculaciones.....	20
Evolución de matriculaciones ECO-CERO.....	21
Distribución total de matriculaciones ECO-CERO.....	23
Distribución por provincia de los vehículos	21

Contexto

El fichero de datos en los que se basa el EDA, se refiere al parque automovilístico completo de los coches matriculados en España hasta finales del año 2023.

Según la DGT, el objetivo de este fichero, es el análisis del número de vehículos, su composición y características técnicas. Este archivo puede descargarse a través de la siguiente URL:

<https://www.dgt.es/menusecundario/dgt-en-cifras/dgt-en-cifras-resultados/dgt-en-cifras-detalle/Microdatos-de-parque-de-vehiculos-anual/>

Para entender el archivo, es posible descargarse también un documento dónde se explica el detalle de la información contenida en el fichero (data/parque_consolidado_2023_ORIGINAL.txt).

Objetivo

El objetivo de este proyecto es analizar el parque de vehículos de España, centrándonos en vehículos considerados como ecológicos (etiqueda CERO y ECO) para ver si realmente se está tendiendo a un parque cada vez más ecológico.

Descripción de las variables

A continuación, se muestran las variables contenidas en el fichero original, así como la acción a realizar para cada una de ellas:

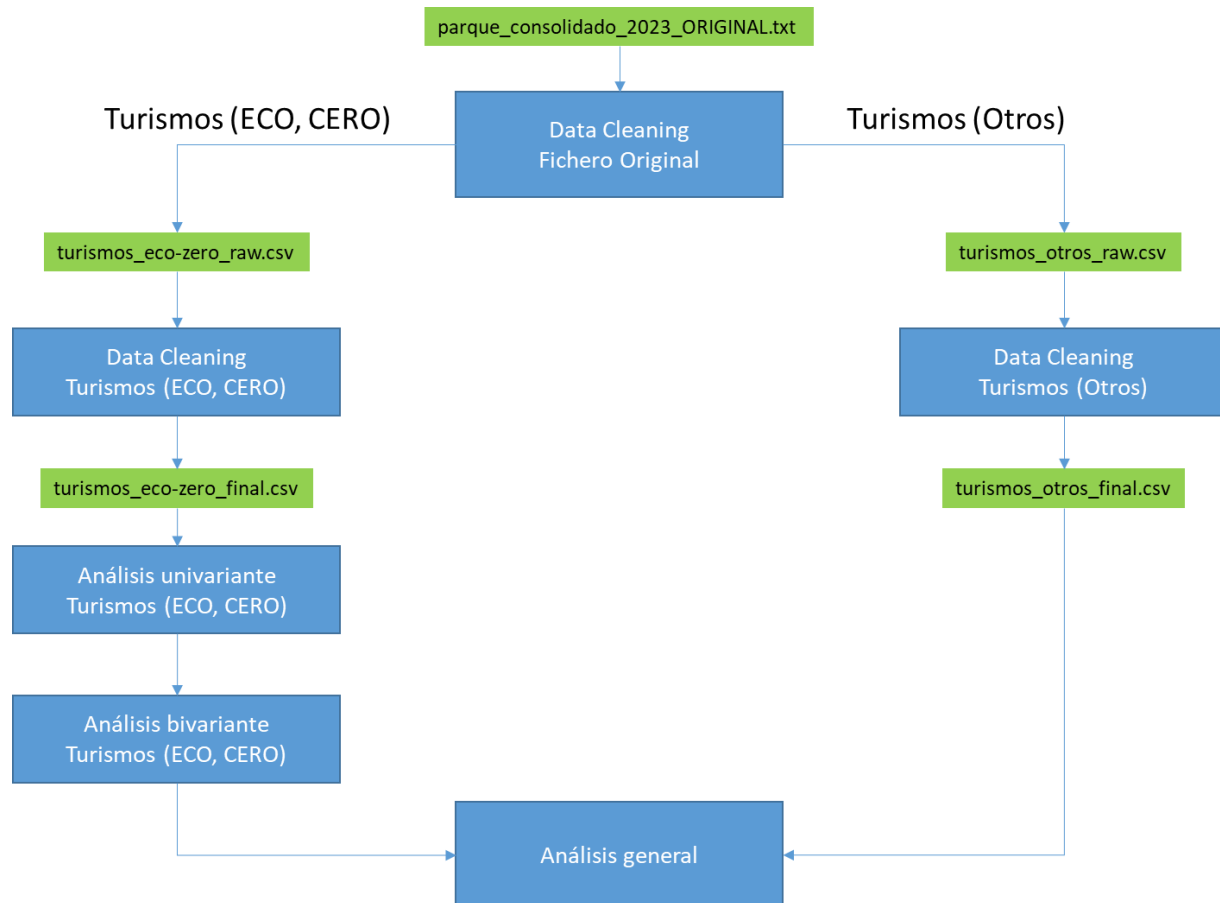
Nombre de campo	ACCION	Descripción de contenido
PROVINCIA	Mantener	Código de la provincia donde está domiciliado el vehículo.
MUNICIPIO	Eliminar	Código INE del municipio donde está domiciliado el vehículo.
FABRICANTE	Eliminar	Nombre del fabricante del vehículo completado.
MARCA	Mantener	Marca del fabricante del vehículo.
MODELO	Mantener	Denominación comercial del vehículo.
TIPO	Eliminar	Tipo homologado de vehículo.
VARIANTE	Eliminar	Si procede, identificación de la variante de dicho tipo.
VERSION	Eliminar	Si procede, identificación de la versión de dicho tipo.
PROVINCIA_MATR	Mantener	Código de la provincia de matriculación del vehículo.
FECHA_MATR	Mantener	Fecha de matriculación del vehículo.
FECHA_PRIM_MATR	Mantener	Fecha de la primera matriculación del vehículo.
CLASE_MATR	Eliminar	Código de clase de matrícula.
PROCEDENCIA	Eliminar	Código de la procedencia del vehículo.
NUEVO_USADO	Eliminar	Indicador de vehículo nuevo (N) o usado (U) al momento de la matriculación FECHA_MATR.

TIPO_TITULAR	Eliminar	Indicador del tipo de titular del vehículo: Persona Física o Jurídica.
NUM_TITULARES	Eliminar	Número de titulares del vehículo.
SUBTIPO_DGT	Eliminar	Código del subtipo de vehículo.
TIPO_DGT	Mantener	Tipo de vehículo: motocicleta, ciclomotor, turismo, camión, furgoneta, autobús, remolque/semirremolque, tractor industrial u 'otros vehículos'.
CAT_EURO	Mantener	Código de la categoría del vehículo.
CLAS_CONSTRUCCION	Eliminar	Código de clasificación del vehículo por criterio de construcción.
CLAS_UTILIZACION	Eliminar	Código de clasificación del vehículo por criterio de utilización.
SERVICIO	Eliminar	Código de servicio del vehículo.
RENTING	Mantener	Indicador de Renting (S/N).
TARA	Mantener	Tara del vehículo. Unidad: Kg. La carga útil se calculará como PESO_MAX - TARA.
PESO_MAX	Mantener	Masa máxima en carga admisible del vehículo en circulación (MMA, Masa Máxima Autorizada). Unidad: Kg.
MOM	Eliminar	Masa en orden de marcha. Unidad: Kg.
MMTA	Eliminar	Masa máxima en carga técnicamente admisible. Unidad: Kg.
CILINDRADA	Mantener	Cilindrada del vehículo. Unidad: cm3.
POTENCIA	Mantener	Potencia fiscal. Unidad: CVF.
KW	Mantener	Potencia del motor. Unidad: Kw.
PROPULSION	Mantener	Código del tipo de propulsión.
CATELECT	Mantener	Código de categoría eléctrica.
CONSUMO	Mantener	Consumo de energía eléctrica WLTP. Unidad: Wh/Km.
AUTONOMIA	Mantener	Autonomía eléctrica. Unidad: Km.
ALIMENTACION	Mantener	Código del tipo de alimentación.
TIPO_DISTINTIVO	Mantener	Código del distintivo ambiental.
EMISIONES_EURO	Mantener	Nivel de emisiones del motor que aparece en la homologación de tipo.
EMISIONES_CO2	Mantener	Emisiones específicas de CO2 (combinadas), valor WLTP. Unidad: g/Km.
CARROCERIA	Eliminar	Código de la carrocería del vehículo.
DISTANCIA_EJES	Eliminar	Distancia entre ejes indicada en la homologación de tipo. Unidad: mm.
EJE_ANTERIOR	Eliminar	Vía máxima del eje delantero, indicada en la homologación de tipo. Unidad: mm.
EJE_POSTERIOR	Eliminar	Vía máxima del eje trasero, indicada en la homologación de tipo. Unidad: mm.
PLAZAS	Eliminar	Número de plazas del vehículo. Para un vehículo de carga, este campo indicará el número de plazas máximo permitido cuando el vehículo está habilitado para carga de mercancías (ej: asientos posteriores abatidos).

PLAZAS_MAX	Eliminar	Número de plazas máximo del vehículo. Para un vehículo de carga, este campo indicará el número de plazas máximo permitido cuando el vehículo está habilitado para transporte de personas (ej: asientos posteriores desplegados).
PLAZAS_PIE	Eliminar	Número de plazas de pie para el caso de M2 y M3, de acuerdo con lo especificado en el Reglamento CEPE/ONU 36, 52, 107 ó respecto a la Directiva 2001/85/CE.

Proceso

En el siguiente diagrama se muestra el proceso que se ha utilizado para poder realizar el EDA:

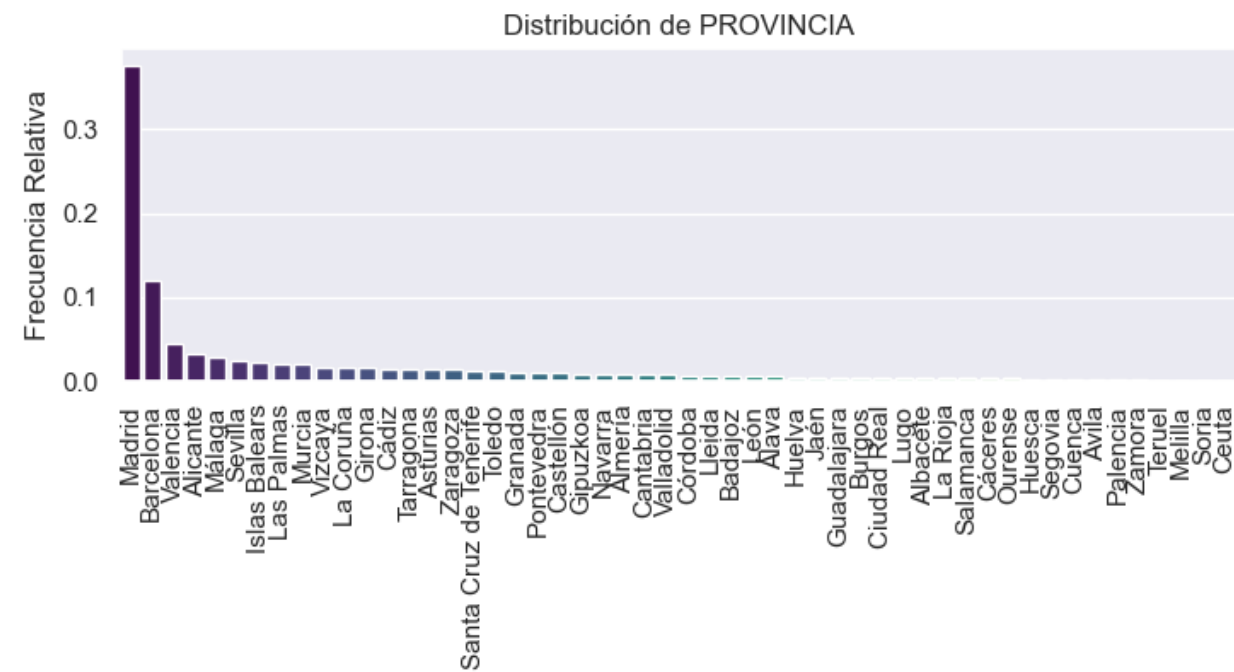


1. A partir del fichero original de la DGT, se han eliminado las columnas identificadas en la exploración inicial de variables y se han generado 2 archivos diferentes. Se ha optado por esta solución dada la gran cantidad de registros que contiene el fichero.
 - a. Turismos con distintivo ECO y CERO
 - b. Otros turismos con distintivo C, B o sin distintivo
2. Proceso de limpieza de datos
 - a. Para ambos archivos, se han llevado a cabo las siguientes acciones:
 - i. Identificar las columnas con valores nulos
 - ii. Transformar a tipo "date" las columnas FECHA_MATR y FECHA_PRIM_MATR
 - iii. Poner el nombre de los campos de provincia en lugar del identificador
 - iv. Para las columnas con identificadores que tienen una tabla maestra asociada, se han sustituido los ids por los nombres, para facilitar la comprensión de los datos.

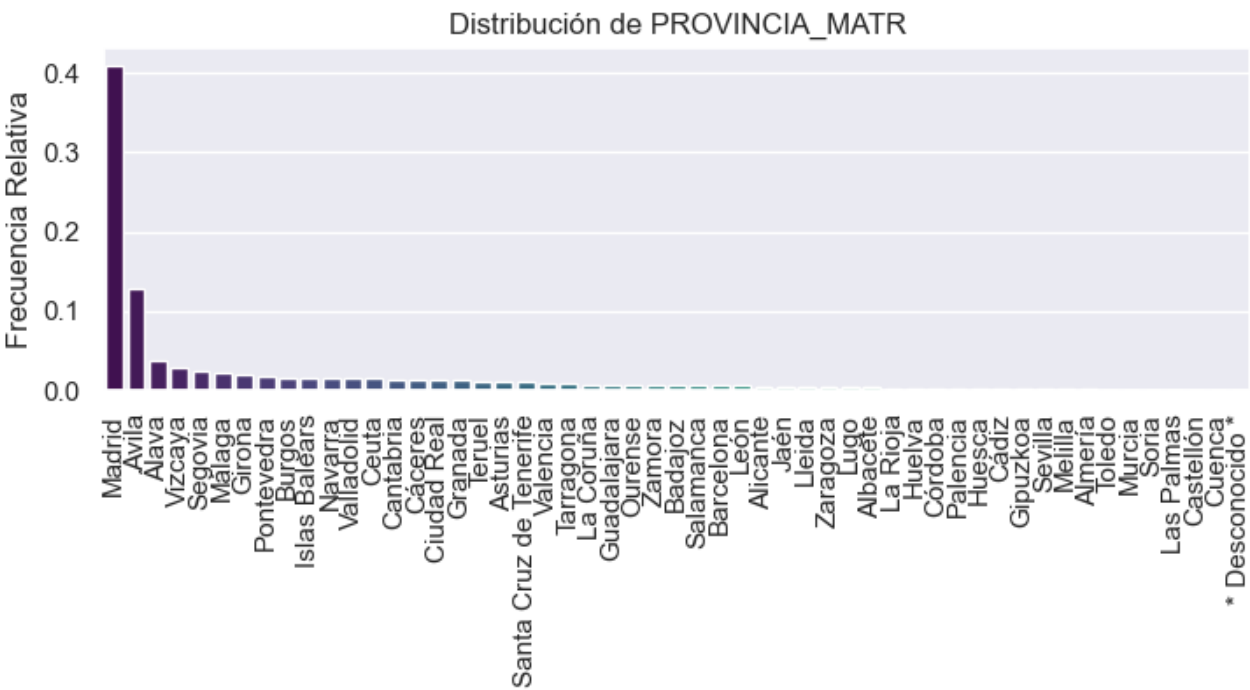
[illegible]

PROVINCIAS

Las provincias dónde hay más coches domiciliados son Madrid, Barcelona, Alicante y Malaga.

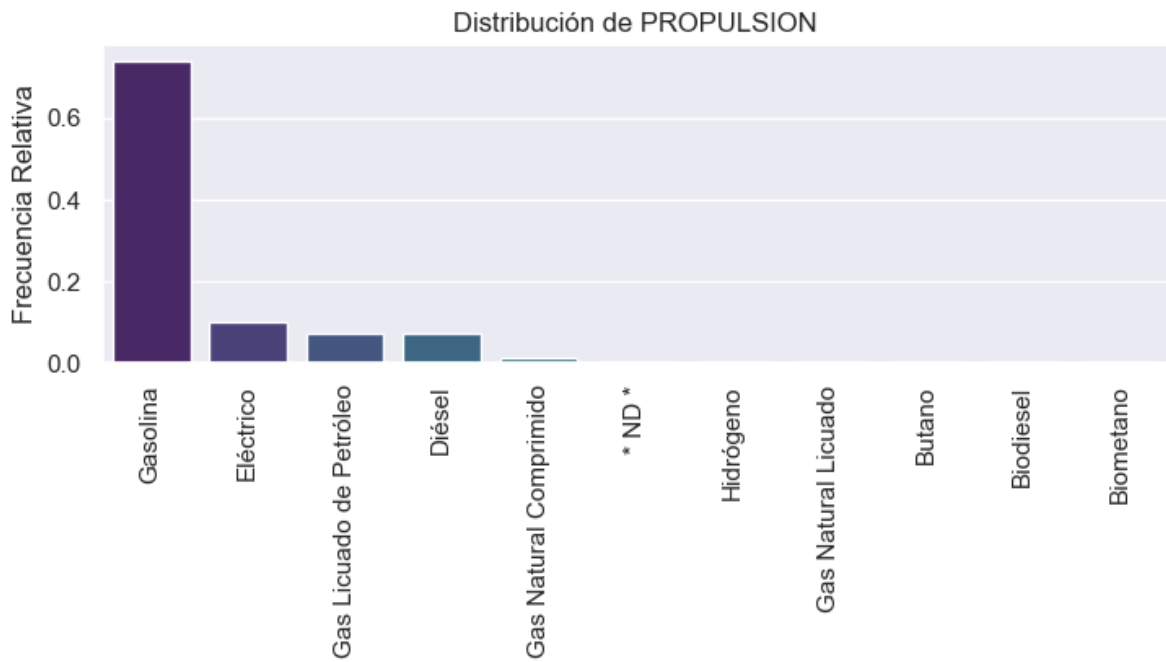


Las provincias con mayor número de matriculación de coches son Madrid, Avila, Alava, que difiere con la provincia dónde está domiciliado el vehículo. Es un dato interesante a analizar ya que puede haber condiciones de impuestos de matriculación especiales que promocionen a estas provincias.



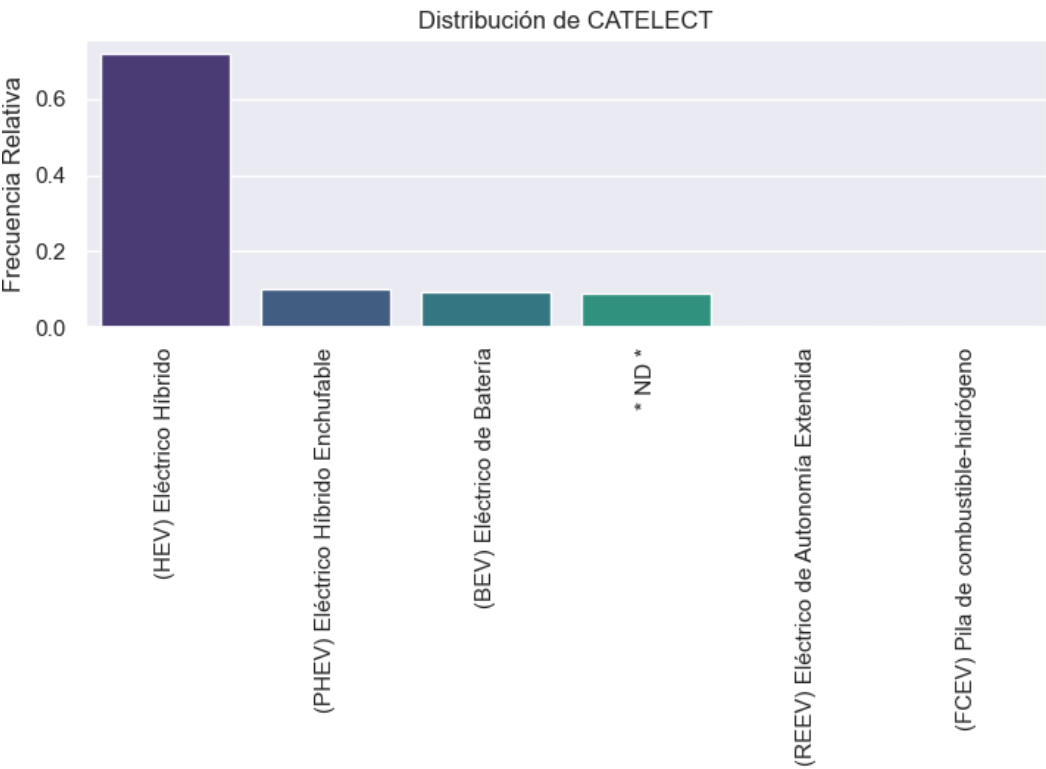
PROPULSION

La gran mayoría de los coches ECO o CERO utilizan la Gasolina.



CATELECT (Categoría eléctrica)

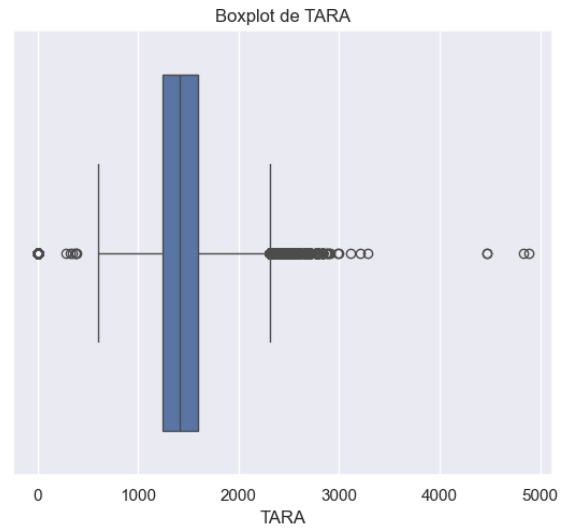
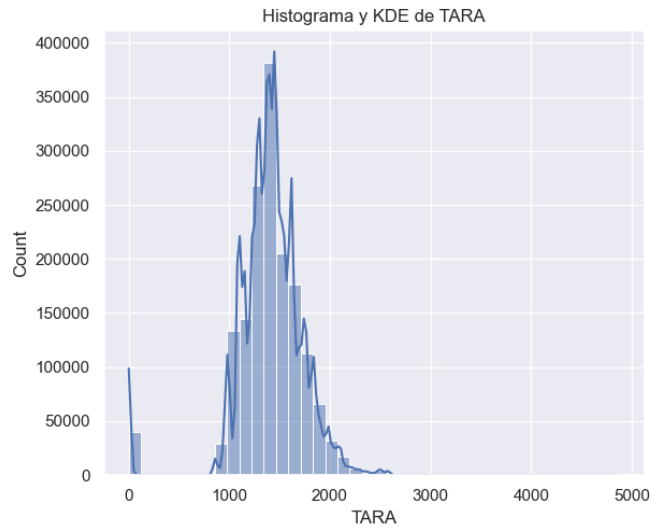
La gran mayoría de coches son Eléctricos híbridos



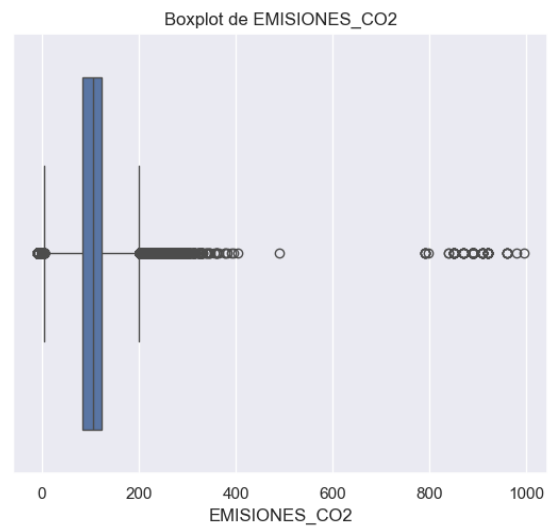
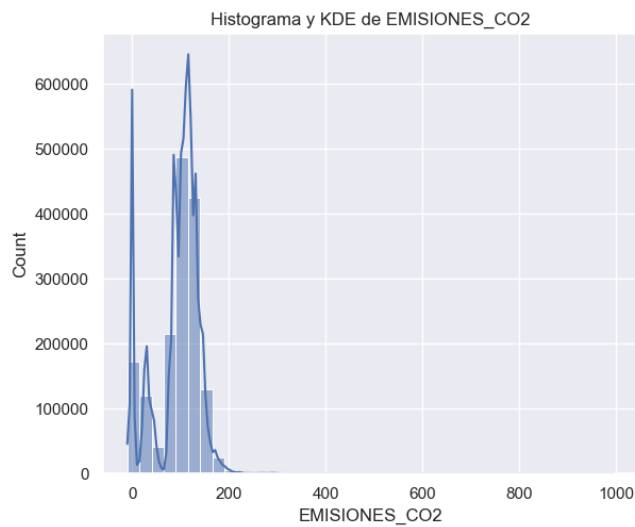
Numéricas

Se pinta un histograma y diagrama de caja para cada una de las variables numéricas para identificar datos relevantes. A continuación, se describen las que han destacado ya que su forma se podría asemejar a una campana de Gauss

TARA



EMISIONES_CO2



ANÁLISIS BIVARIANTE

Categórica – Categórica

CHI-2

Para la búsqueda de relaciones entre variables categóricas, utilizamos el método Chi-2. La siguiente tabla muestra el valor P de cada uno de los pares de columnas categóricas, cuyo resultado sea menor 0.05 y mayor que 0.

	CAT1	CAT2	RESULTADO
3	PROVINCIA	CAT_EURO	1.960031e-03
27	PROVINCIA_MATR	CAT_EURO	2.394080e-21
34	CAT_EURO	RENTING	4.656948e-04
35	CAT_EURO	PROPULSION	3.907724e-27
36	CAT_EURO	CATELECT	2.007429e-109
37	CAT_EURO	ALIMENTACION	3.955950e-32
38	CAT_EURO	TIPO_DISTINTIVO	1.229670e-03

Todas las relaciones están relacionadas con la columna CAT_EURO. Prácticamente todos los valores de esta columna son M1.



“La categoría M1 se tratan de los vehículos de motor fabricados para el transporte de personas y su equipaje. Para nosotros son los turismos identificados por la DGT, aunque hay otros valores que pueden considerarse que no se han categorizado correctamente”, según se dice en: <https://www.serviciositv.es/blog/informacion-itv/tipos-de-vehiculos-clases-y-categorias-de-vehiculos>. Tiene sentido ya que hemos hecho un filtro previo de tipo de vehículo como “Turismo” que es el que corresponde con M1.

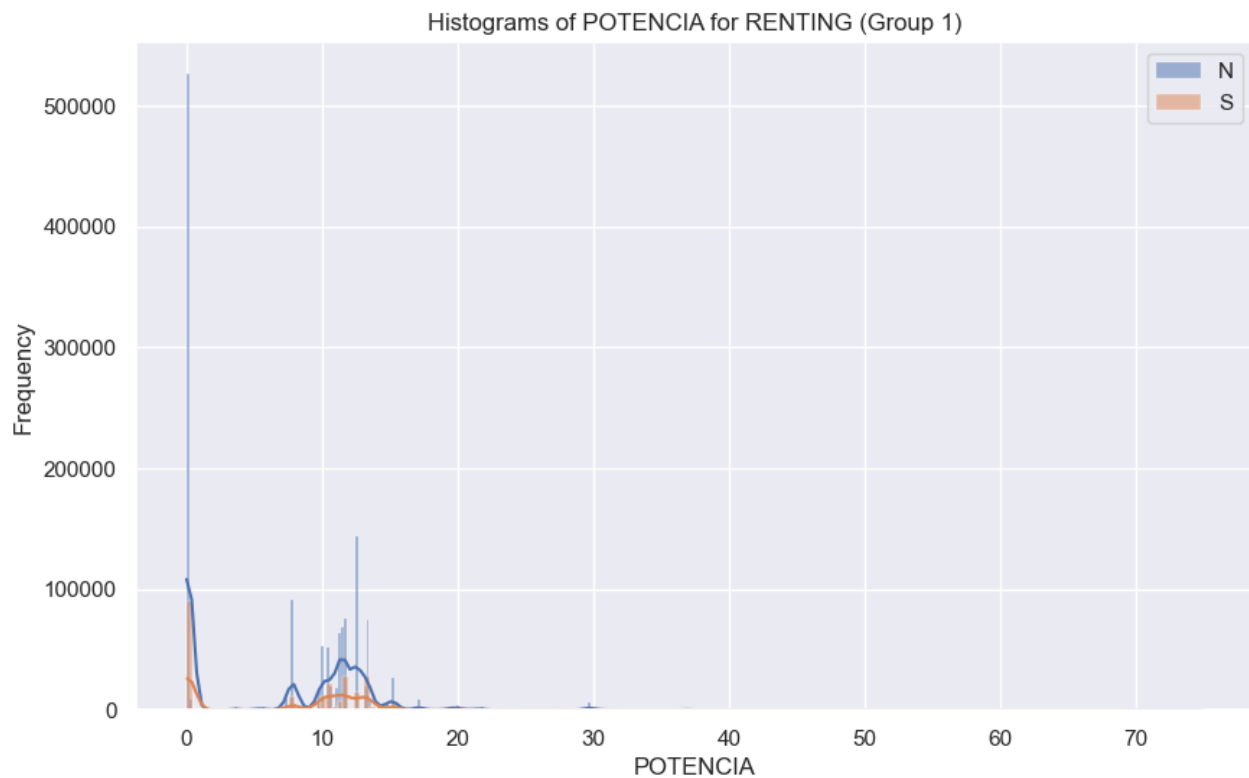
Catégorica – numérica

Mannwhitneyu

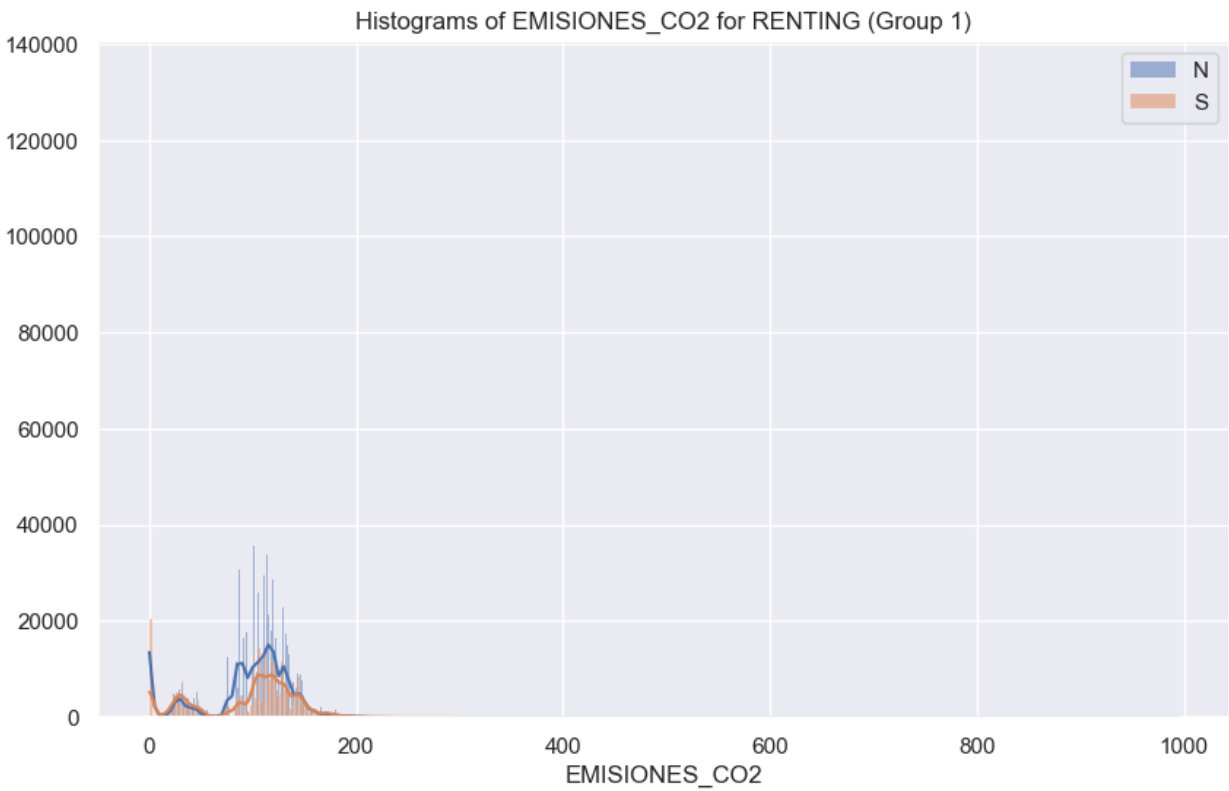
Para relacionar las variables categóricas binarias con numéricas, utilizamos el método “Mannwhitneyu”. La siguiente tabla muestra el valor P de cada uno de los pares de columnas categóricas, cuyo resultado sea menor 0.05 y mayor que 0.

	CAT	NUM	RESULTADO
5	RENTING	POTENCIA	1.435956e-116
9	RENTING	EMISIONES_CO2	9.806057e-207

RENTING-POTENCIA



RENTING-EMISIONES_CO2

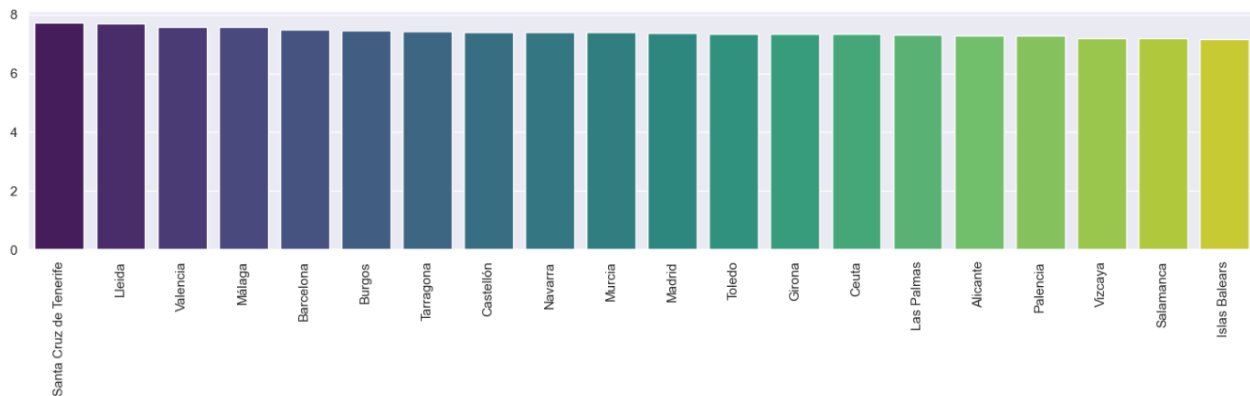


Anova

Para relacionar las variables categóricas no binarias con numéricas, utilizamos el método “Anova”. La siguiente tabla muestra el valor P de cada uno de los pares de columnas categóricas, cuyo resultado sea menor 0.05 y mayor que 0.

	CAT	NUM	RESULTADO
3	PROVINCIA	POTENCIA	1.050886e-184
32	CAT_EURO	TARA	1.450153e-108
33	CAT_EURO	PESO_MAX	1.179219e-261
34	CAT_EURO	CILINDRADA	1.170702e-60
35	CAT_EURO	POTENCIA	6.865338e-28
36	CAT_EURO	KW	4.655740e-141
37	CAT_EURO	CONSUMO	3.877564e-10
38	CAT_EURO	AUTONOMIA	1.786628e-02
39	CAT_EURO	EMISIONES_CO2	3.436816e-46

La potencia es muy similar a todas las provincias



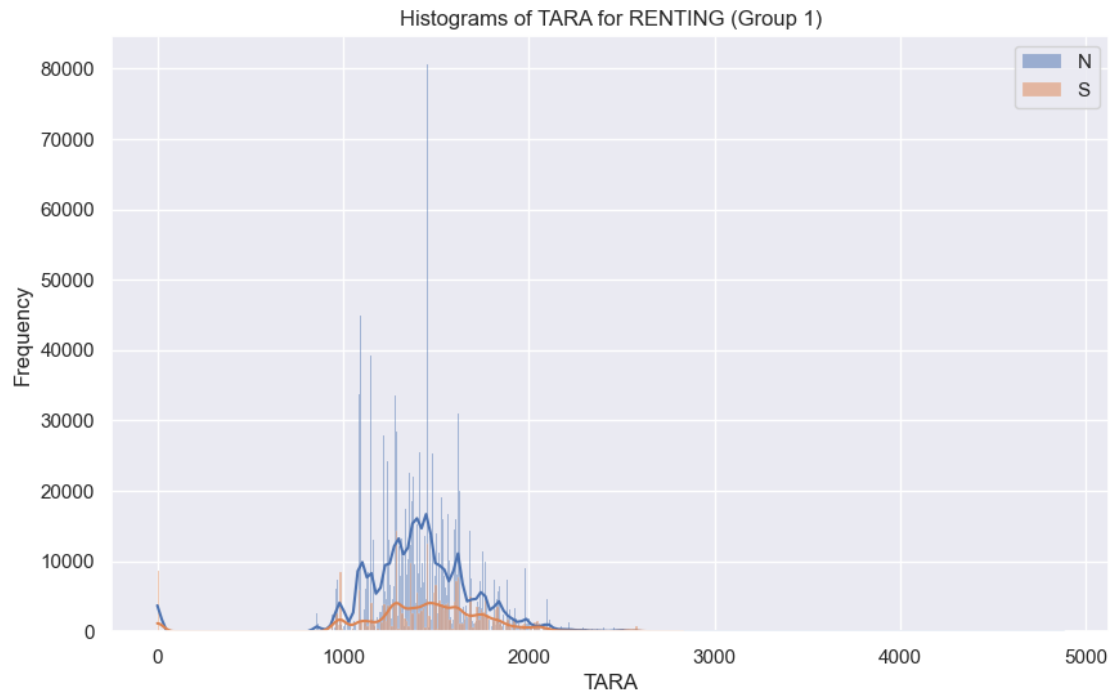
El resto de las relaciones están relacionadas con la columna CAT_EURO. Prácticamente todos los valores de esta columna son M1.



“La categoría M1 se tratan de los vehículos de motor fabricados para el transporte de personas y su equipaje. Para nosotros son los turismos identificados por la DGT, aunque hay otros valores que pueden considerarse que no se han categorizado correctamente”, según se dice en: <https://www.serviciositv.es/blog/informacion-itv/tipos-de-vehiculos-clases-y-categorias-de-vehiculos>. Tiene sentido ya que hemos hecho un filtro previo de tipo de vehículo como “Turismo” que es el que corresponde con M1.

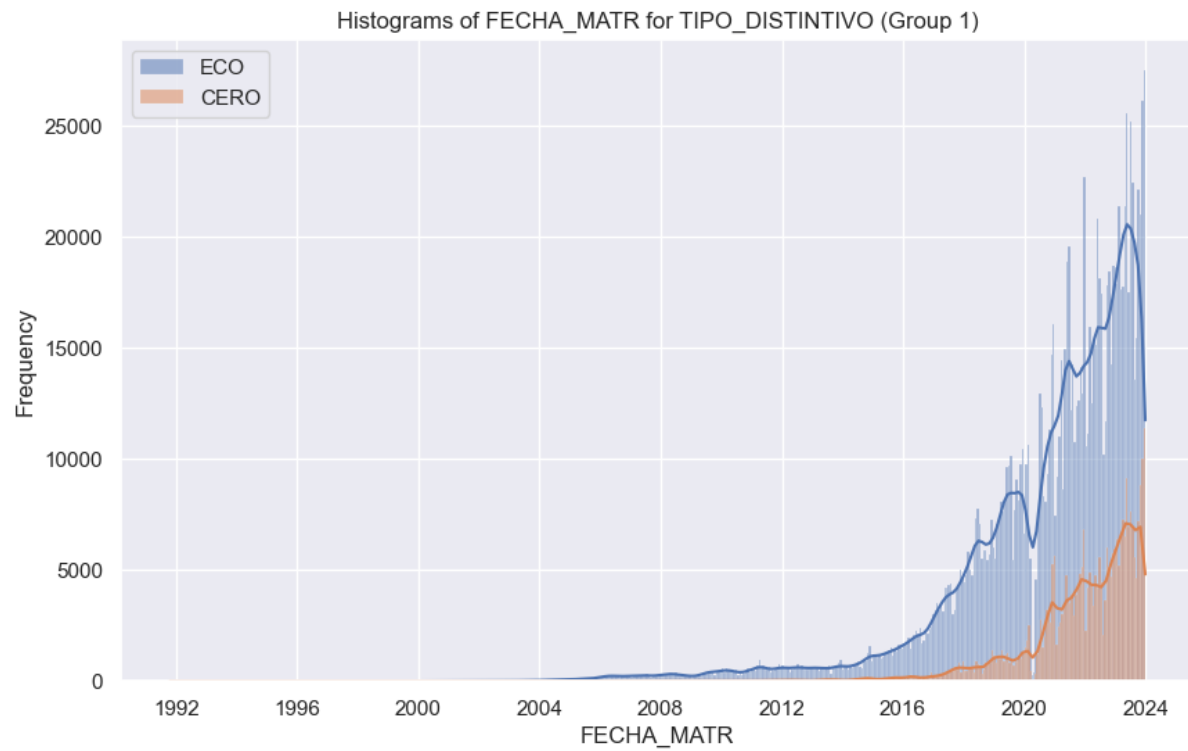
Caben destacar algunas gráficas de comparativa de columnas numéricas dada su forma y pueden ser interesantes para analizar:

TARA-RENTING



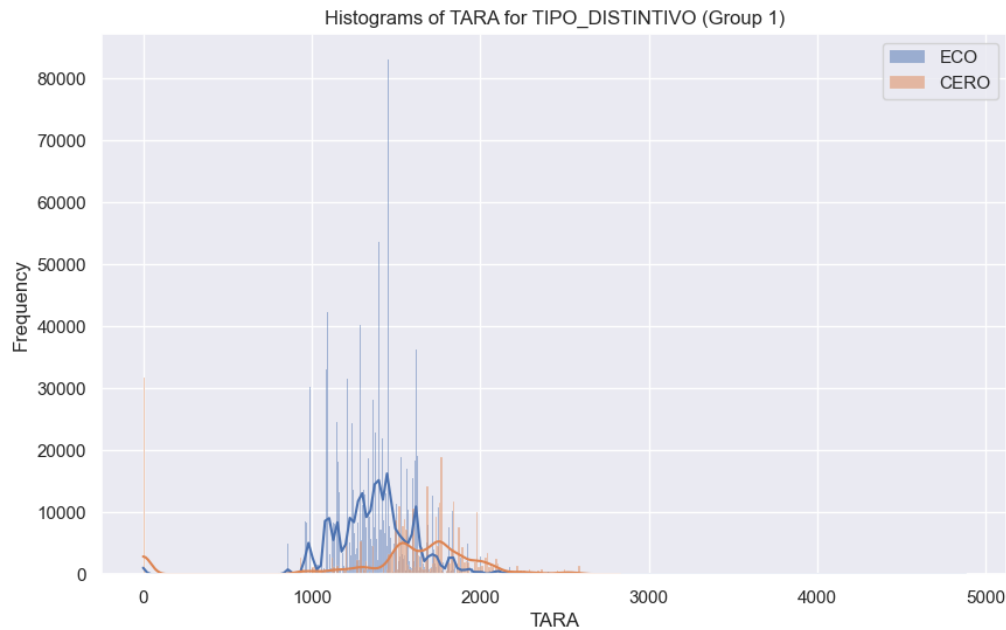
FECHA_MATR – DISTINTIVO

La evolución de las matriculaciones por distintivo muestra el aumento de matriculaciones de coches ECO-CERO y además es superior en el caso de ECO frente a ECO



DISTINTIVO – TARA

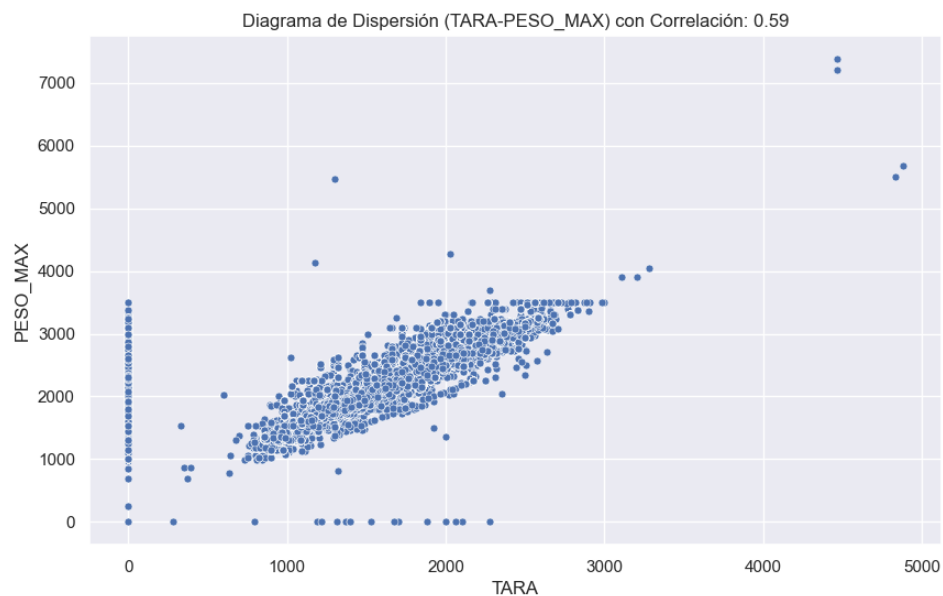
Se puede observar que hay mayor número de vehículos ECO y que el valor máximo de la tara es mayor para vehículos de etiqueta CERO



Numérica – Numérica

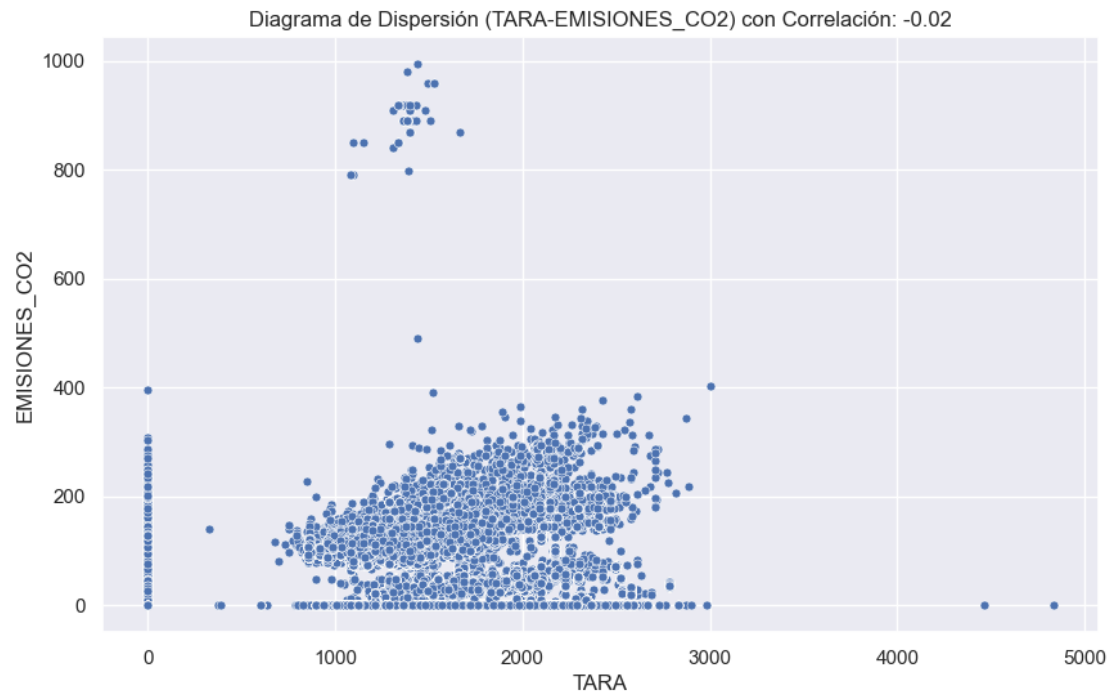
TARA-PESO_MAX

El peso máximo de un vehículo es la tara más la carga que puede alojar, es por eso que en el gráfico se puede observar una relación directa entre ambas.



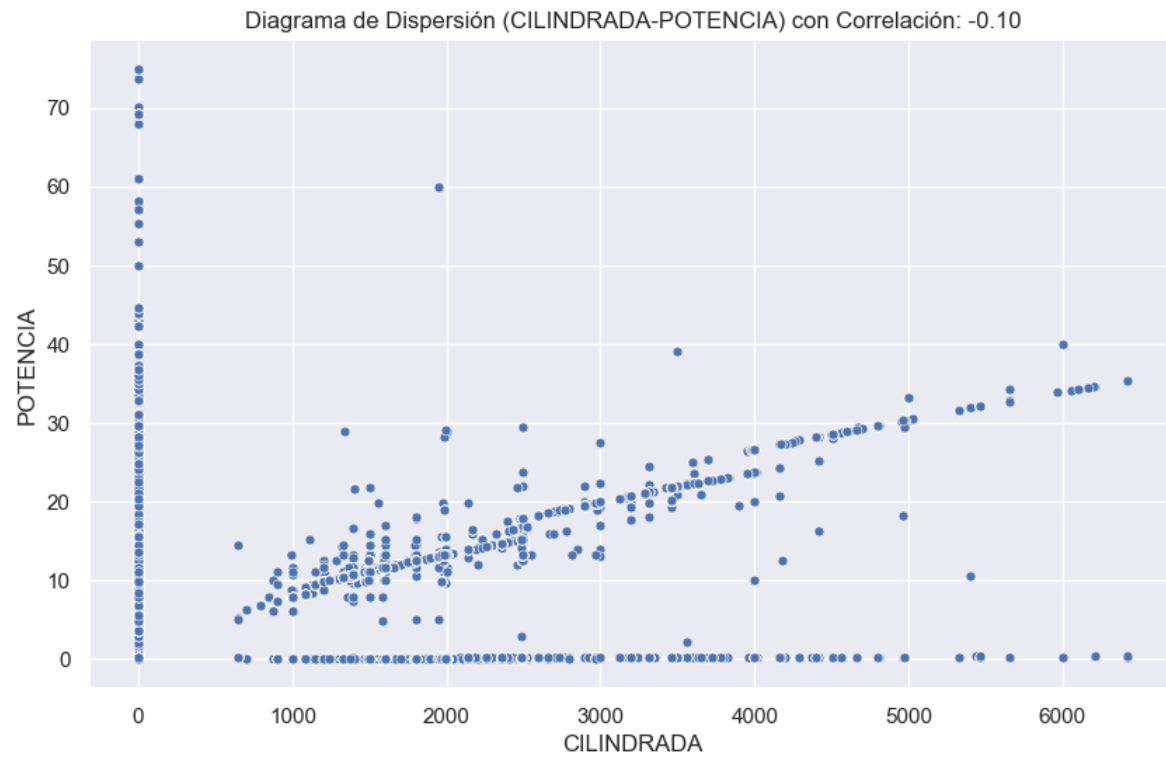
TARA-EMISIONES_CO2

Aunque no es una relación lineal, parece que, a mayor tara de un vehículo, más emisiones puede emitir



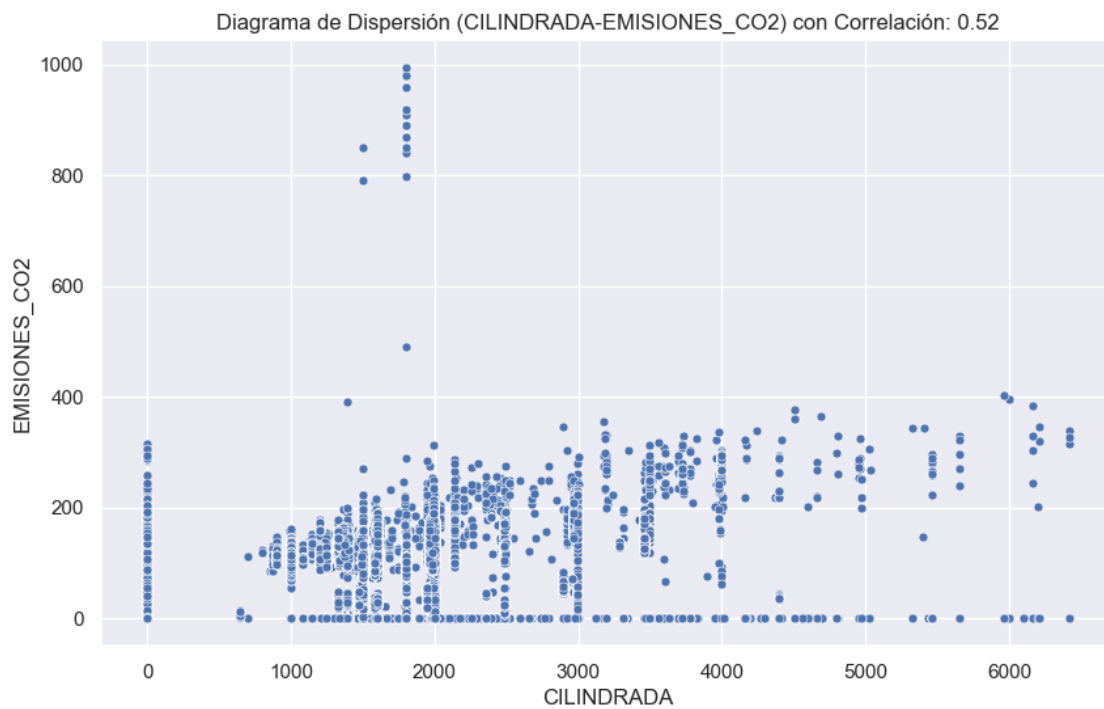
CILINDRADA-POTENCIA

Obviamente, la potencia y la cilindrada de un vehículo están íntimamente relacionados



CILINDRADA-EMISIONES_CO2

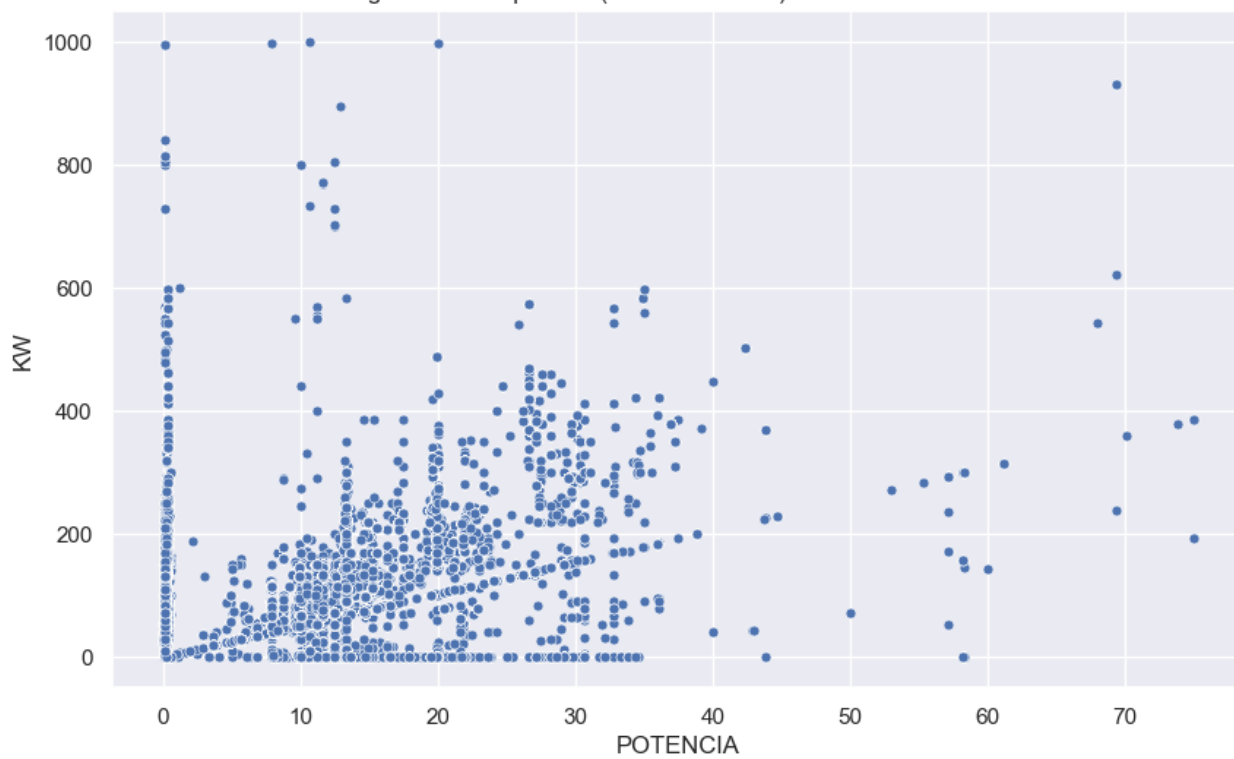
Parece observarse también que, a mayor cilindrada, hay una mayor cantidad de emisiones de CO2.



POTENCIA-KW

Aunque hay dispersión, se puede identificar una línea de tendencia que pudiera relacionar la potencia con los KW.

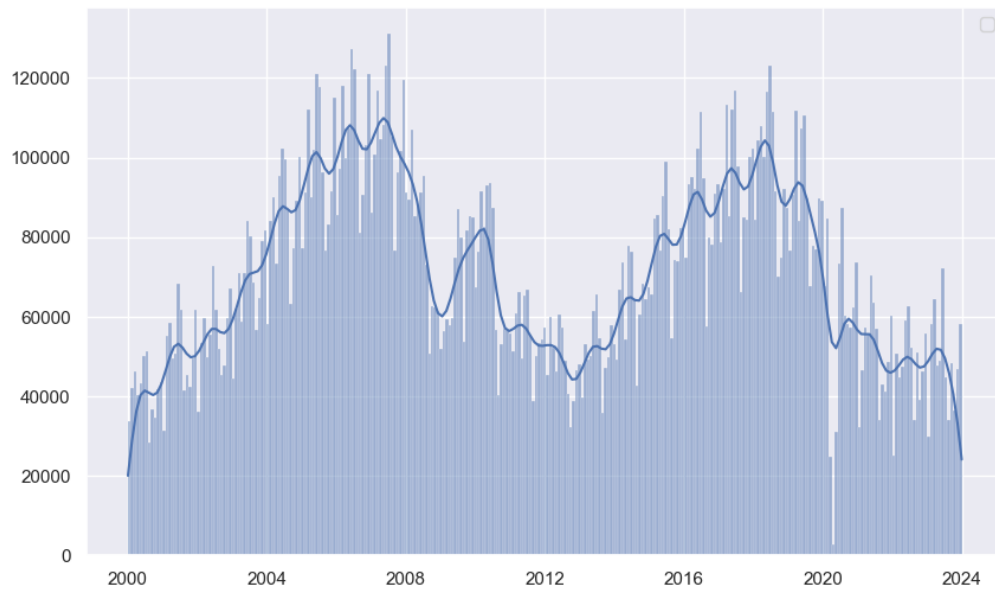
Diagrama de Dispersión (POTENCIA-KW) con Correlación: 0.31



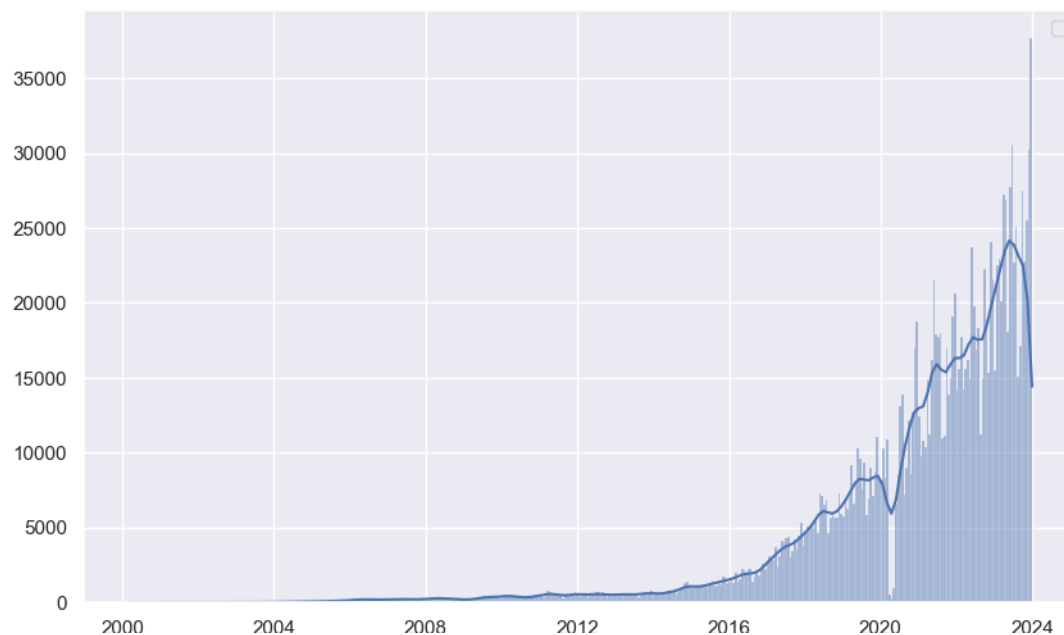
ANÁLISIS GENERAL

Evolución de matriculaciones

Se realiza una comparativa entre el fichero ECO-CERO y el fichero con otros distintivos para ver la evolución de las matriculaciones. Se puede observar que la matriculación de ECO-CERO está aumentando en los últimos años, mientras que el resto de vehículos, se mantiene en línea o quizás está experimentando una bajada después de la pandemia.



Distintivo C,B o sin distintivo



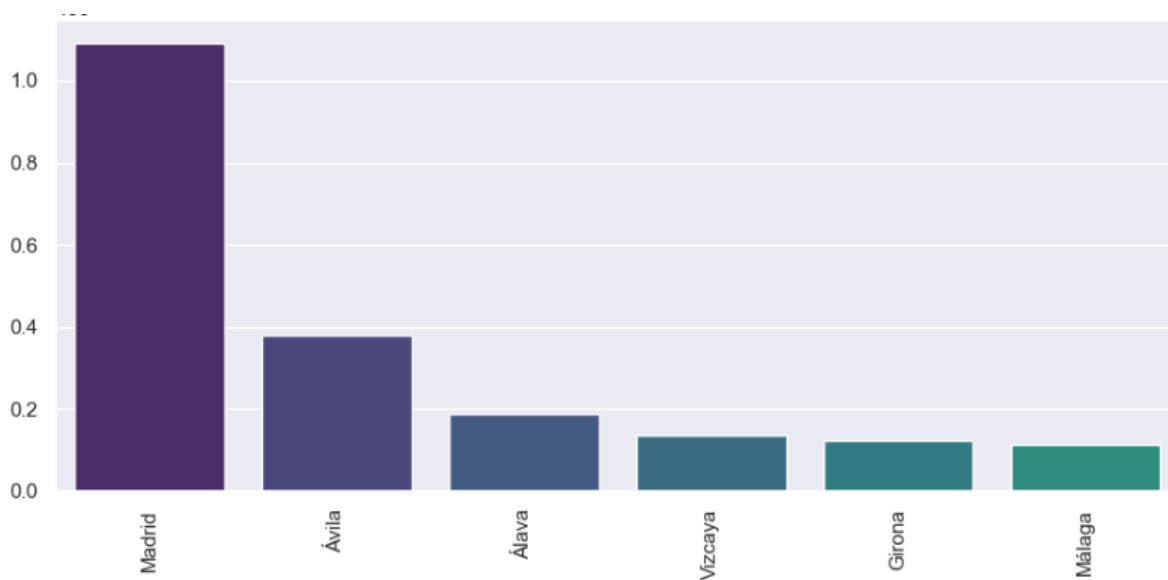
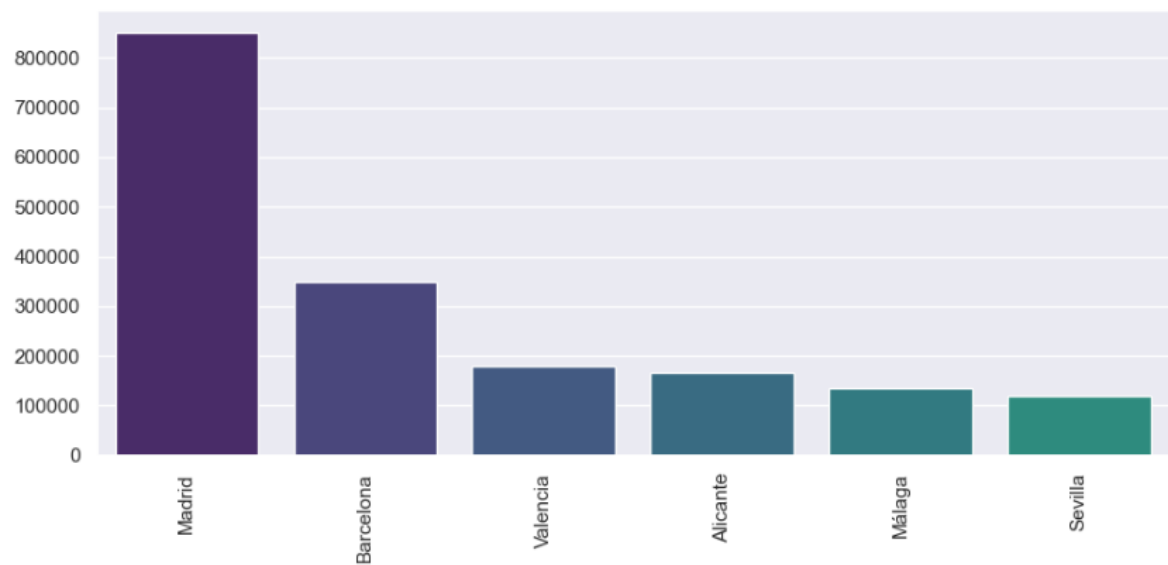
Etiqueta ECO-C

Distribución por provincia de los vehículos

La primera gráfica muestra la distribución de provincias dónde se encuentran domiciliados los vehículos y la segunda la distribución de provincias dónde se matricularon.

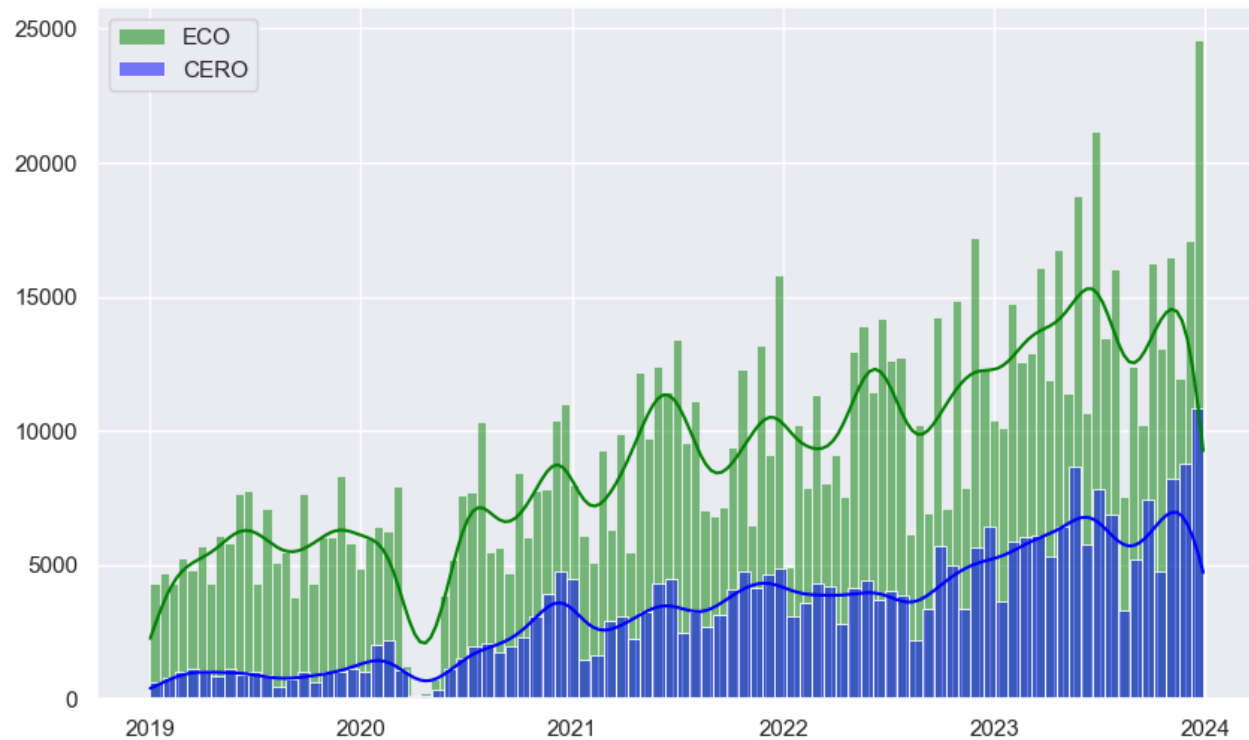
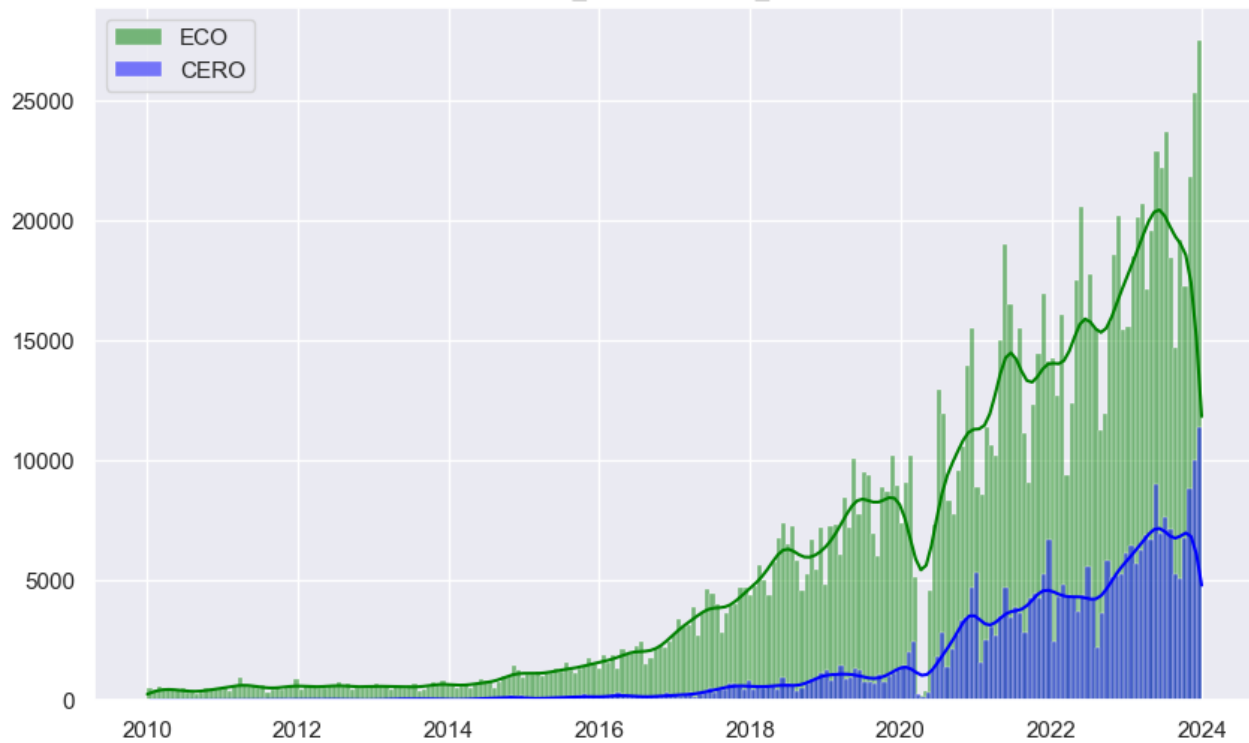
Para la provincia dónde están matriculados los coches, se encuentran en las que están las grandes urbes. Se cuelan Las Palmas e islas Baleares, quizás porque tienen un gran volumen de coches de alquiler.

Parece haber algún tipo de beneficio fiscal en provincias como Ávila, Álava, Vizcaya y Girona, que aparecen en los primeros puestos de provincias de matriculación.

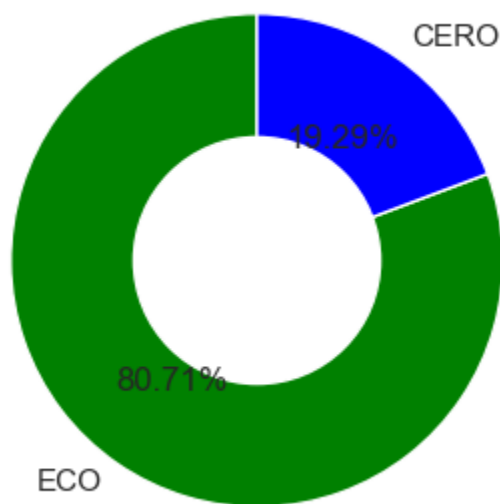


Evolución de matriculaciones ECO-CERO

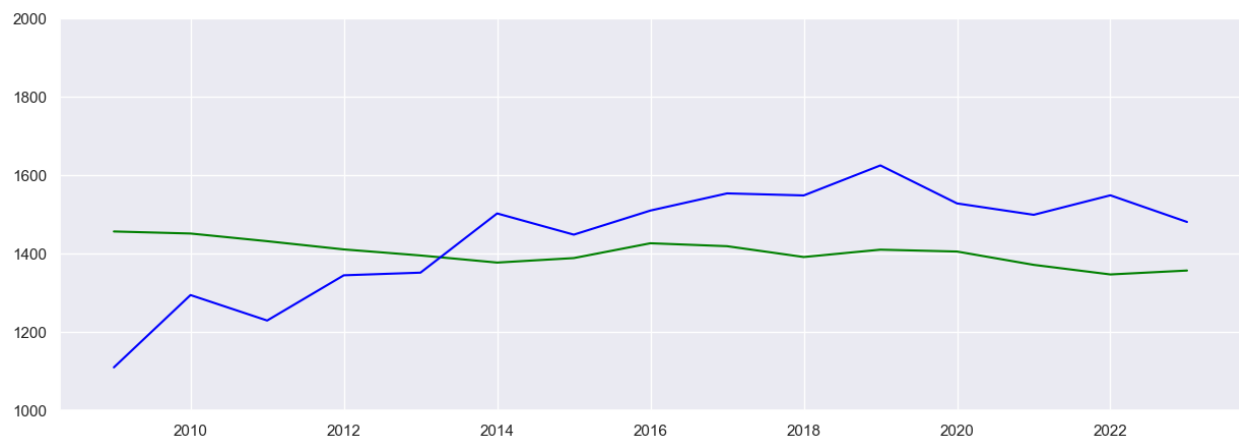
Aumento de matriculaciones para este tipo de vehículos a lo largo de los últimos años. Se puede ver que hay mayor cantidad de coches tipo ECO matriculados que coches CERO.



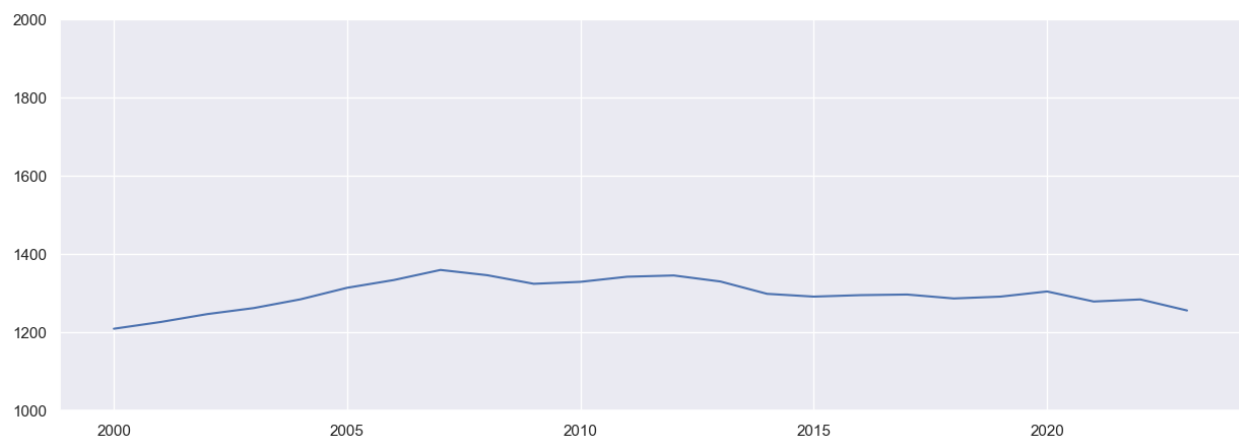
Distribución total de matriculaciones ECO-CERO



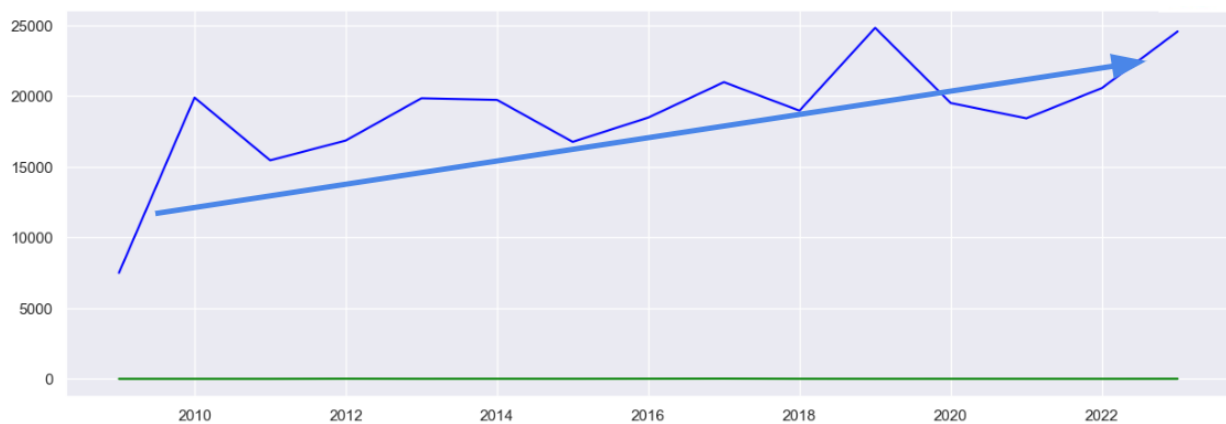
Evolución de la tara para vehículos CERO



Comparado con los vehículos B,C y sin distintivo, no se aprecia ese incremento de peso en este tipo de vehículos.

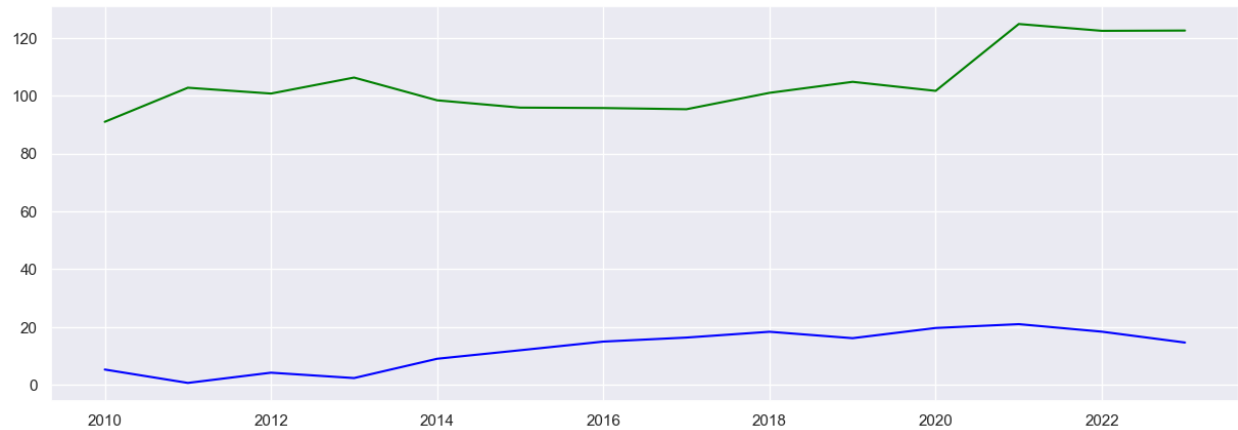


Puede explicarse el incremento en la tara de los vehículos CERO, bien porque se ha identificado un incremento en la autonomía de los vehículos



Evolución de la tara para vehículos CERO

A pesar del aparente incremento en la evolución de los vehículos ecológicos, se ha identificado que desde el 2021 ha habido un incremento para vehículos etiqueta 0. Podría ser provocado por el incremento en la fabricación de vehículos estilo SUV, más pesados, pero no podemos asegurarlo al 100%



Comparando con la evolución de las emisiones de vehículos C,B y sin distintivo, se aprecia un incremento muy similar al de 2021 en vehículos ECO.

