IDIADA Automotive Technology SA
L'Albornar – Apartado de correos 20
43710 Santa Oliva (Tarragona) Spain
T +34 977 166 021
F +34 977 166 036
www.idiada.com

SERVICE TO OBTAIN AN ESTIMATE OF THE ANNUALISED NUMBER OF KILOMETRES TRAVELLED BY SPAIN'S VEHICLE FLEET (2024)

REPORT ON THE METHODOLOGY IMPLEMENTED (DELIVERY: GROUP 4)

File no. 3DGT6AP00018

Client reference:
DGT – Directorate-General of Traffic
Josefa Valcárcel, 28
28027 MADRID
SPAIN

Produced by: Applus+ IDIADA

Date delivered: 20/03/2026

This report contains 59 pages including this cover page and the annexes.

CONTENTS

1. INTRODUCTION

In this document we provide a detailed explanation of the methodology used to estimate the annualised number of kilometres travelled by the vehicle fleet in circulation in Spain over the course of 2024. This report incorporates the two deliverables defined in the project specifications for Delivery 4 (namely, the Report on the Model Implemented and the Methodology Document) with the aim of providing a unified, transparent and traceable overview of the complete process by which the results were obtained.

The study was prepared using the odometer readings recorded during vehicle roadworthiness inspections (ITVs) and data from the administrative records for Spain's vehicle fleet. However, these sources of information have inherent limitations, which make it necessary to carry out specific statistical processing in order to ensure the accuracy and reliability of the estimates.

The main limitations identified include new vehicles without historical records, the variable frequencies of ITVs depending on vehicle category and age, the presence of anomalous data in the odometer readings, and irregular coverage with regard to the total number of vehicles in circulation. Moreover, the high degree of heterogeneity among Spain's vehicle fleet requires an analysis differentiated by vehicle category.

To overcome these limitations, a predictive model based on machine-learning techniques (gradient boosting) was developed that brings together technical, geographical and socio-demographic information in the form of a set of 13 explanatory variables. The methodology includes data-scrubbing procedures, cross-validation, sensitivity analysis and specific accuracy metrics.

The aim of this introduction is to provide an overview of the process detailed in the following sections, from the preparation of the data to the final validation of the model and its application to the vehicle fleet in circulation.

2. DATA USED

To develop the predictive model and build the estimates of the annualised number of kilometres travelled by Spain's vehicle fleet, we used three official sources of information, each of which play a specific role in the composition of the final dataset.

ITV odometer readings

The main basis for the study is the history of odometer readings recorded during the ITVs. To this end, we have included only periodic ITVs with a positive outcome, and discarded:

- Records without an odometer reading.
- Records without a valid inspection date.
- Non-periodic ITVs.
- Agricultural vehicles, trailers, semi-trailers, special vehicles, temporary transport, tourist and reserved registration numbers.

To achieve an estimate of the kilometres travelled by the vehicle fleet in 2024, we used only the ITV data corresponding to that year, which comprise the effective basis for the modelling and generation of results.

Register of the active vehicle fleet

We used the database for the active vehicle fleet on 31 December for the years 2023 and 2024, as it incorporates the variable that identifies whether or not a vehicle forms part of the fleet that is in circulation.

This source makes it possible to link the real ITV readings to the technical and administrative characteristics of each vehicle, using anonymised registration numbers as the link point between the two databases.

However, to estimate the number of kilometres travelled during the year only the active fleet corresponding to 2024 was employed. The 2023 fleet was used solely for validation purposes.

Demographic and geographical data from the National Statistics Institute (INE)

To enrich the models with contextual information, we incorporated provincial variables:

- Province size: area covered by the province (km²).
- Province population: number of inhabitants in the province.
- Population density: province's number of inhabitants per km².

Updated to 2024 and obtained from the INE.

3. STRUCTURE OF THE FLEET

The structure of the vehicle fleet used in this study is based on a reorganisation of vehicle type and a detailed characterisation of the active fleet.

3.1. VEHICULAR TAXONOMY

To conduct the analysis, we have adopted the taxonomy established in the previous DGT study (2023), reorganising the "TYPE_VEHI_INT" codes into eight main categories. These categories allow for homogeneous segmentation suitable for training independent models:

- Cars.
- Vans.
- Lorries ≤ 3500 kg.
- Mopeds.
- Lorries > 3500 kg.
- Motorcycles.
- Industrial tractors.
- Buses.

3.2. FLEET DISTRIBUTION

Figure 1 shows the distribution by vehicle category in the dataset used for training. We can see that cars account for nearly 80% of the records available. Along with motorcycles, vans and lorries up to a maximum authorised mass (MAM) of 3500 kg, these four categories include nearly 97% of all of the records available for training the model. In contrast, lorries with a MAM in excess of 3500 kg, industrial tractors, mopeds and buses account for a smaller percentage of the training dataset.

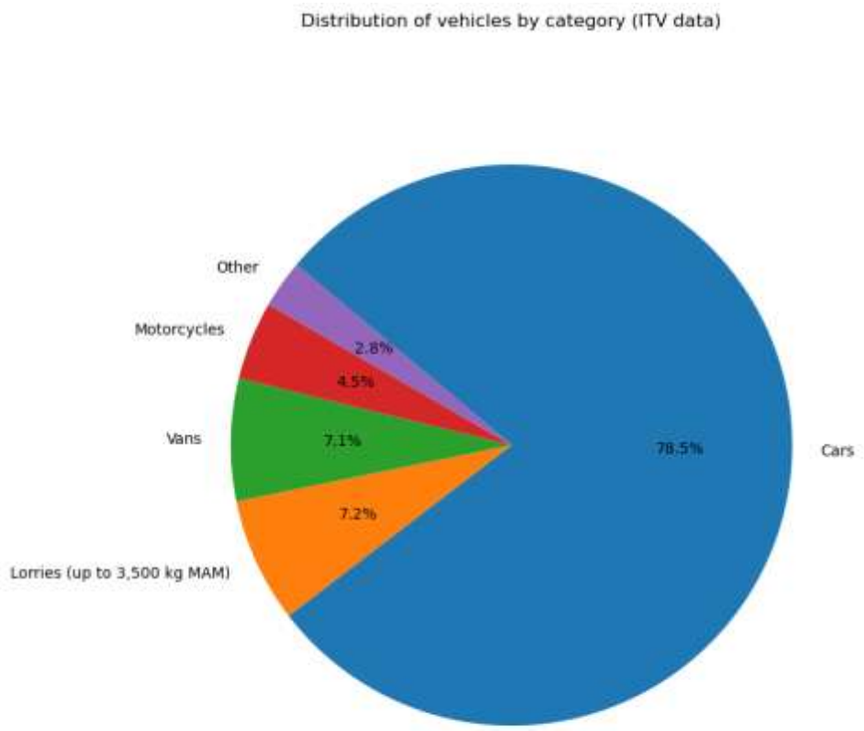

Figure 1. Distribution by vehicle category in the database used for the study.

3.3. LINKING THE FLEET TO THE ITV READINGS

Cross-referencing the active fleet with the ITV readings is vital to building the training database. This integration was achieved using the column containing the anonymised registration number data, which acts as the common identifier across both databases. This process makes it possible to unify both datasets in a single consolidated database combining the register information for the active fleet with the ITV data.

The ITV data include the odometer readings recorded at each inspection, which then enable us to calculate the kilometres travelled over the course of the year by the vehicle in question. To do this, we divide the difference in kilometres compared to the previous inspection (or the total number, if it is the first inspection) by the time elapsed between the two inspections, applying the following definition of variables:

- **DIF_KM**: the difference in kilometres between two consecutive (passed) inspections.
- **DIF_DAYS**: the difference in days between two consecutive (passed) inspections.

Thus, the annualised kilometres (KM_ITV) for each vehicle between two consecutive ITVs are calculated as follows:

$$KM\_ITV = \frac{DIF\_KM}{DIF\_DIAS} \cdot 365{,}25$$

It is possible for the same registration number to appear twice (or more) in the ITV database during the same year, as a vehicle may undergo various annual inspections. In cases where a registration number in the active fleet matches more than one record in the ITV database, we choose one of them at random to establish the link.

As a result of this process, approximately 50% of the registration numbers in the active fleet for 2024 have ITV readings that can be used to train the model. Likewise, it was possible to correctly link 99% of the records in the ITV database for 2024 to the active fleet.

However, as this set of registration numbers with ITV data was subsequently subjected to a process of data cleansing and anomaly elimination, as described in detail in section **4**, **CRITERIA FOR FILTERING AND PROCESSING ANOMALOUS DATA**, the final percentage of active vehicles with usable and valid data will be slightly reduced.

4. CRITERIA FOR FILTERING AND PROCESSING ANOMALOUS DATA

A scrubbing process was used to identify and eliminate anomalous data from the training database, applying a methodology consisting of six consecutive stages.

Stage 1: General filtering by registration status and domestic ownership

We eliminated vehicles that do not form part of the active fleet in circulation according to the "FLG_CIRCULATION" variable. This first stage eliminates vehicles that are de-registered, temporarily inactive or whose administrative status is incompatible with the mobility analysis.

We also excluded records where the owner's province is listed as abroad or has a null value. This enables us to use Spanish population data for the study.

During this stage, 83,997 records were removed from the total of 20,619,359 records in the fleet dataset linked to the ITV.

Stage 2: Filtering by valid vehicle category

We eliminated all records where the vehicle category information for the "TYPE_VEHI_INT" variable is null or incorrect, and selected only the records for the following vehicle categories: cars, vans, lorries ≤ 3500 kg, mopeds, lorries > 3500 kg, motorcycles, industrial tractors, and buses.

During this stage, 135,707 of the records remaining after the previous stage were eliminated, leaving a total of 20,399,655.

Stage 3: Elimination of extreme values

For some of the numerical variables, each vehicle category was cleansed of extreme values. This was done by eliminating records with negative values for these variables, and then eliminating values above the 99.5th percentile and below the 0.5th percentile: in other words, 0.5% of the records with the highest and lowest values.

This filter is applied to the following variables: annual kilometres, vehicle age, power, tare weight, engine capacity and owner age. For the "vehicle age" variable, only the upper percentile filter is applied. Both the upper and lower percentile filters are applied to the rest of the variables.

In addition to this filtering process, the null values for these variables are replaced with the average value for said variable for the same vehicle category.

The extreme values defined for each variable are described in the tables of statistics presented in section **5.1 ANALYSIS OF THE VARIABLES**.

Stage 4: Validation of ITV inspection intervals

We eliminated entries where the number of days between ITVs was outside of the legally permitted range for the vehicle category in question, as shown in **Table 1**.

Table 1. Inclusion intervals (in days) for each vehicle category.

Vehicle category	Minimum interval (days)	Maximum interval (days)
Cars	60	1460 (4 years)
Vans	60	730 (2 years)
Lorries with MAM ≤ 3500 kg	60	730 (2 years)
Mopeds	60	1095 (3 years)
Lorries with MAM > 3500 kg	60	365 (1 year)
Motorcycles	60	1460 (4 years)
Industrial tractors	60	730 (2 years)
Buses	60	365 (1 year)

The cleansing data for this stage were combined with those of stages 3 and 5. The scrubbed records resulting from these joint filtering processes are presented in stage 5.

Stage 5: Validation of intervals for kilometres travelled

We then applied the technically justified limits for annual kilometres travelled (ANNUAL_KM) shown in **Table 2** below.

Table 2. Inclusion intervals (in kilometres) for each vehicle category.

Vehicle category	Lower limit (km)	Upper limit (km)
Cars	1,000	200,000
Vans	1,000	300,000
Lorries with MAM ≤ 3500 kg	1,000	300,000
Mopeds	50	30,000
Lorries with MAM > 3500 kg	1,000	300,000
Motorcycles	50	70,000
Industrial tractors	1,000	300,000
Buses	1,000	300,000

The table below shows the resultant records after applying the various filters for stages 3, 4 and 5.

Table 3. Results of the joint filtering (stages 3-5).

Vehicle category	Records before cleansing	Records after cleansing	Records eliminated
Cars	14,662,359	12,951,461	1,710,898
Vans	1,808,194	1,412,464	395,730
Lorries ≤ 3,500 kg	2,140,509	1,843,755	296,754
Mopeds	153,680	83,477	70,203
Lorries > 3,500 kg	348,151	261,762	86,389
Motorcycles	975,484	744,351	231,133
Industrial tractors	226,450	204,652	21,798
Buses	84,828	71,942	12,886

The total number of records remaining after the cleansing performed during these three stages (3, 4 and 5) is 17,573,864.

Stage 6: Elimination of duplicated vehicles

Lastly, we eliminated records where the registration number was duplicated, in cases where the same vehicle has two or more entries in the ITV database for 2024. A single record, selected at random, was kept for each registration number.

This process resulted in the elimination of 1,165,524 duplicated registration numbers.

Obtaining the final dataset

Following the consecutive application of all of the filters and the scrubbing of anomalous data, we obtained a final dataset consisting of 16,408,340 scrubbed records. This dataset represents approximately 80% of the total number of entries in the dataset for the active fleet that could be successfully cross-referenced with the ITV database; in turn, this implies that 20% of the initial records contained some kind of anomaly or inconsistency that justified their exclusion from the analysis.

5. PREDICTOR VARIABLES

The development of the predictive model for each vehicle category was based on a set of 13 explanatory variables chosen by means of an exploratory data analysis and statistical validation. The final selection includes not only technical vehicle-related variables, but also the socio-demographic and geographical characteristics of the vehicles' owners.

- Technical vehicle-related variables:
  - Vehicle age (numerical): number of full years since first registration.
  - Tare weight (numerical): unladen weight of the vehicle (kg).
  - Power (numerical): maximum power produced by the engine (hp).
  - Capacity (numerical): engine displacement volume (cm^3).
  - Emissions level (ordinal): Euro emissions category (1-6).
  - Propulsion type (categorical): fuel and propulsion system.
- Owner-related variables:
  - Number of owners (numerical): amount of registered owners.
  - Owner age (numerical): age of the main owner in years.
  - Owner gender (categorical): gender of the main owner. Also includes the categorisation of owners by type of legal person.
- Geographical variables:
  - Province size (numerical): area covered by the province (km^2).
  - Province population (numerical): number of inhabitants in the province.
  - Population density (numerical): number of inhabitants per km^2 in the province.
- Use-related variables:
  - Service (categorical): type of use (private, taxi, rental, etc.).

The target variable is the number of kilometres travelled over the year (ANNUAL_KM), calculated based on the differences in odometer readings between consecutive ITVs and annualised based on the amount of time elapsed.

The "vehicle category" variable was used to divide the training process into eight independent models, one for each type of vehicle. We opted to use the same set of explanatory variables for all of the specialised models.

5.1. ANALYSIS OF THE VARIABLES

Below is an analysis of the different variables used in the study. The distributions of annual kilometres shown refer to the values recorded during the ITVs in the dataset linked to the fleet database and subsequently scrubbed.

Annual kilometres

Table 4 contains a summary of the main statistics for annual kilometres.

Table 4. Annualised kilometres by vehicle category for the 2024 ITV (scrubbed data).

	Cars	Vans	Lorries ≤ 3500 kg	Mopeds	Lorries > 3500 kg	Motorcycles	Industrial tractors	Buses
N	12,876,500	1,157,862	1,180,654	82,661	161,136	740,730	166,646	42,151
Average	10,870	13,098	12,190	2,165	25,461	2,655	92,352	52,552
std	7,817	12,344	10,269	2,437	27,597	2,680	53,344	38,948
CV	0.72	0.94	0.84	1.13	1.08	1.01	0.58	0.74
Minimum	1,000	1,000	1,000	50	1,000	50	1,000	1,005
P1	1,238	1,161	1,184	63	1,178	76	2,123	2,495
P5	2,064	1,791	1,908	122	1,938	184	7,952	8,270
P10	2,952	2,561	2,760	211	3,011	324	17,399	13,597
Q1	5,321	4,897	5,163	556	6,989	799	48,621	26,073
Median	9,246	9,505	9,496	1,365	16,875	1,836	95,366	45,076
Q3	14,428	17,063	16,019	2,886	34,402	3,582	131,393	68,071
P90	20,356	27,690	24,551	5,087	57,165	6,005	156,632	97,939
P95	24,825	36,453	31,503	6,814	76,437	7,989	174,529	122,777
P99	37,898	61,063	50,323	12,037	132,828	12,939	223,948	204,600
Maximum	73,315	109,931	100,133	20,173	299,996	20,336	299,974	299,474

The graphs below show the density of the annualised kilometres for each vehicle type.

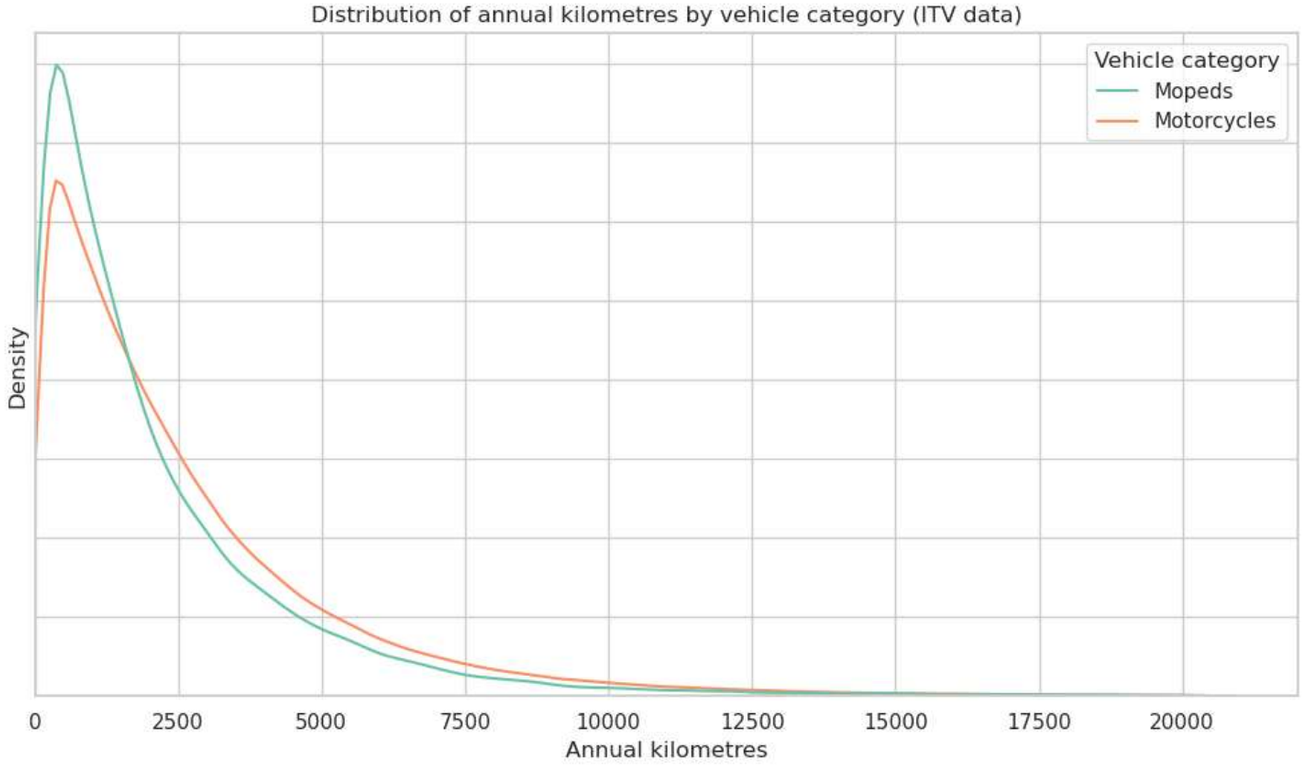

Figure 2. Kernel density estimation of annualised kilometres for mopeds and motorcycles.

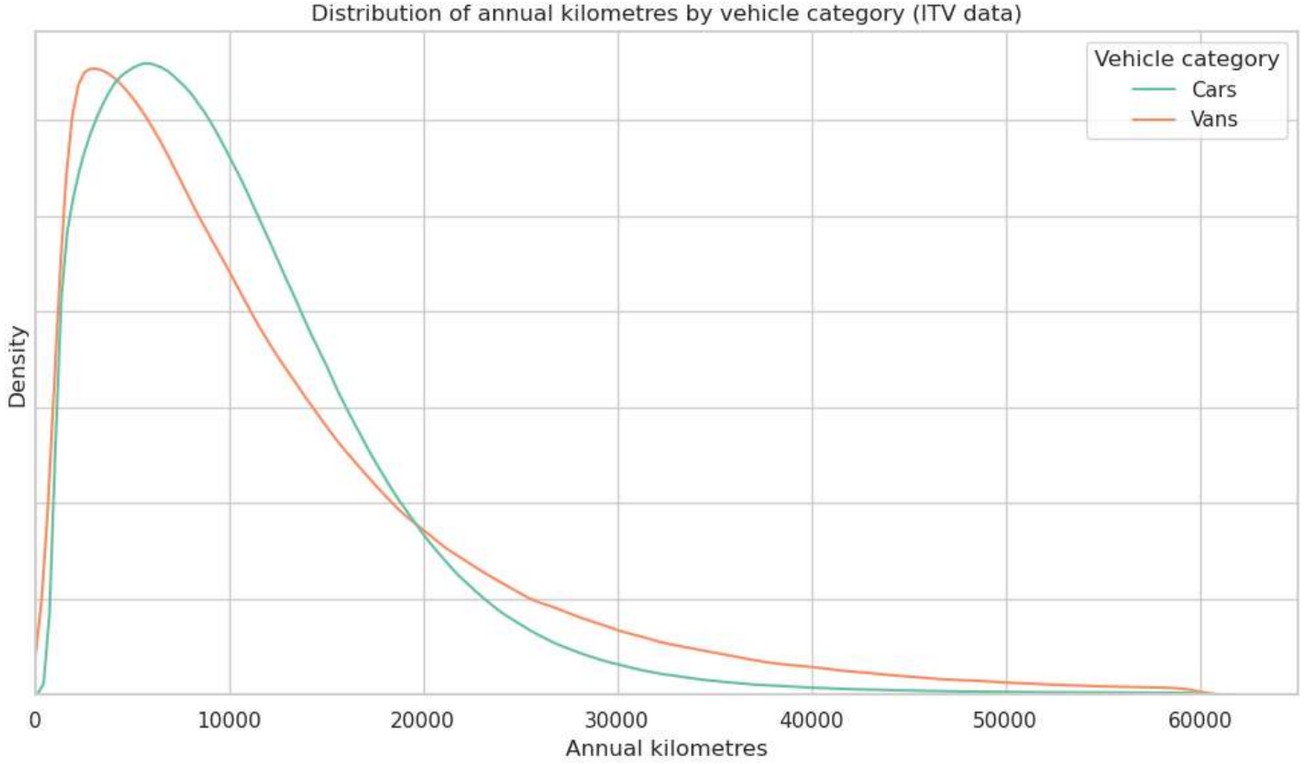

Figure 3. Kernel density estimation of annualised kilometres for cars and vans.

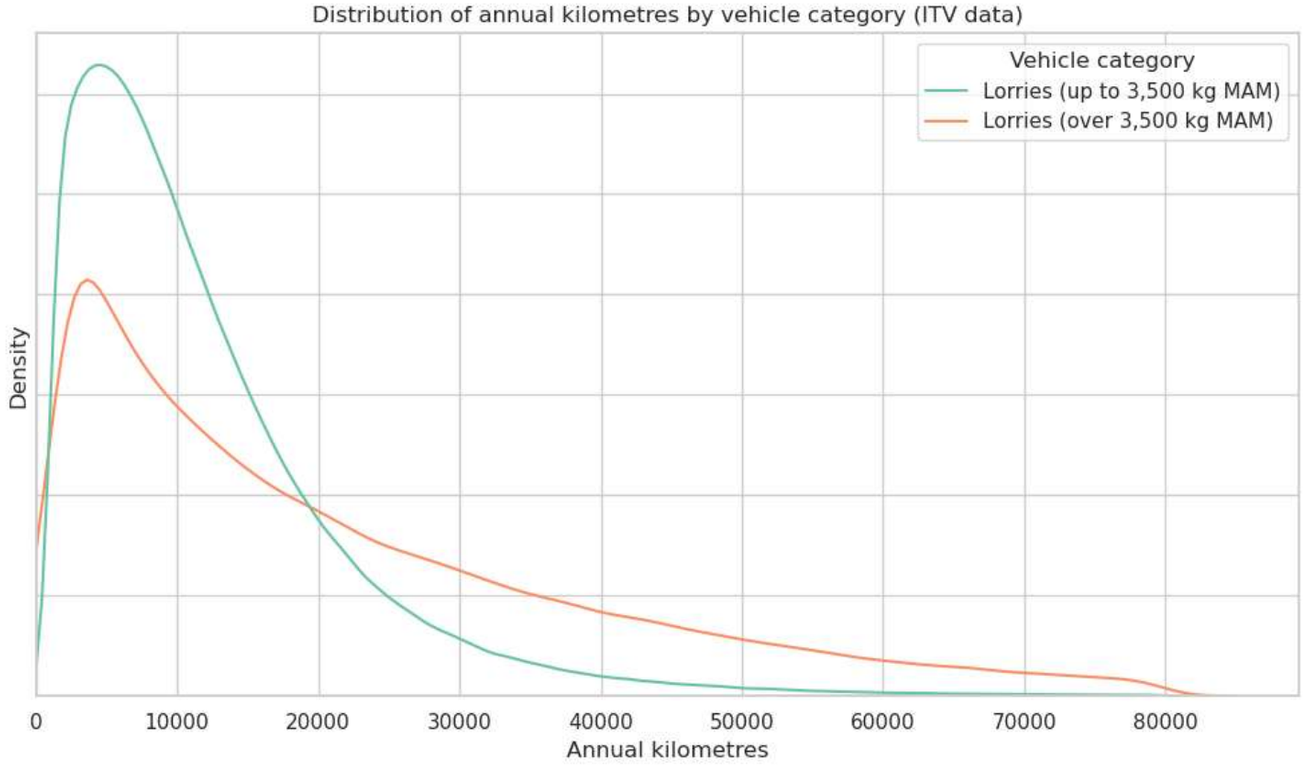

Figure 4. Kernel density estimation of annualised kilometres for lorries.

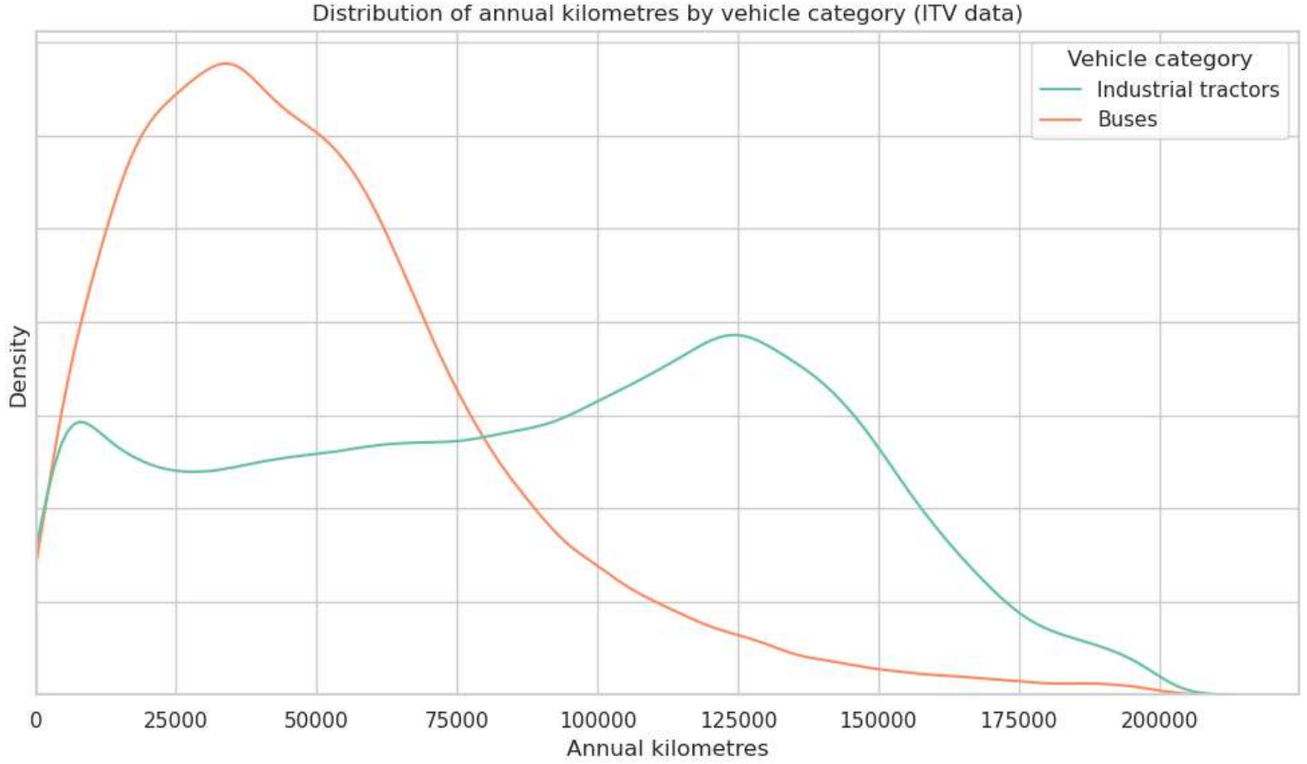

Figure 5. Kernel density estimation of annualised kilometres for buses and industrial tractors.

Vehicle category

Table 5. Number of records and average annual kilometres by vehicle category.

Vehicle category	No. of records	Average annual km
Cars	12,876,500	10,870
Vans	1,157,862	13,098
Lorries ≤ 3500 kg	1,180,654	12,190
Mopeds	82,661	2,165
Lorries > 3500 kg	161,136	25,461
Motorcycles	740,730	2,655
Industrial tractors	166,646	92,352
Buses	42,151	52,552

Vehicle age

Table 6. Statistics for the "vehicle age" variable (expressed in years).

Vehicle category	min	max	average	std
Cars	0.16	36.34	15.36	5.91
Vans	0.16	41.08	17.10	7.59
Lorries ≤ 3500 kg	0.18	36.08	17.56	5.86
Mopeds	0.98	31.73	15.28	6.86
Lorries > 3500 kg	0.19	36.91	17.46	7.28
Motorcycles	0.16	58.40	14.20	7.69
Industrial tractors	0.23	29.17	8.33	6.12
Buses	0.29	28.49	10.17	5.65

Table 7. Number of records and average annual kilometres by age group.

Vehicle category	Age group	No. of records	Average annual km
Cars	0-4	351,703	16,959
Cars	5-9	2,298,082	13,818
Cars	10-14	3,083,368	11,689
Cars	15-19	4,323,902	9,966
Cars	20+	2,819,445	8,198
Vans	0-4	45,521	28,663
Vans	5-9	198,005	21,667
Vans	10-14	207,172	15,619
Vans	15-19	320,123	10,906
Vans	20+	387,041	7,347
Lorries ≤ 3500 kg	0-4	23,771	25,357
Lorries ≤ 3500 kg	5-9	131,111	19,043
Lorries ≤ 3500 kg	10-14	171,753	14,422
Lorries ≤ 3500 kg	15-19	471,924	11,481
Lorries ≤ 3500 kg	20+	382,095	8,892
Mopeds	0-4	5,988	2,684

Mopeds	5-9	18,188	3,064
Mopeds	10-14	11,589	2,308
Mopeds	15-19	21,793	2,062
Mopeds	20+	25,103	1,413
Lorries > 3500 kg	0-4	12,232	44,584
Lorries > 3500 kg	5-9	11,700	44,451
Lorries > 3500 kg	10-14	24,982	32,715
Lorries > 3500 kg	15-19	55,536	23,981
Lorries > 3500 kg	20+	56,686	15,668
Motorcycles	0-4	38,678	4,203
Motorcycles	5-9	214,448	3,355
Motorcycles	10-14	165,304	2,605
Motorcycles	15-19	217,296	2,263
Motorcycles	20+	105,004	1,544
Industrial tractors	0-4	58,912	128,428
Industrial tractors	5-9	55,833	92,542
Industrial tractors	10-14	23,433	68,075
Industrial tractors	15-19	19,449	42,776
Industrial tractors	20+	9,019	25,507
Buses	0-4	6,894	73,590
Buses	5-9	16,465	61,089
Buses	10-14	9,024	42,723
Buses	15-19	7,877	34,421
Buses	20+	1,891	23,951

Tare weight

Table 8. Statistics for tare weight.

Vehicle category	min	max	average	std
Cars	795	2,170	1,300.91	223.02
Vans	820	2,804	1,727.15	379.98
Lorries ≤ 3500 kg	930	3,200	1,676.68	453.09
Mopeds	0	400	116.79	86.29
Lorries > 3500 kg	0	23,220	8,496.83	4,190.60
Motorcycles	0	470	175.48	51.63
Industrial tractors	0	14,480	7,913.44	1,148.57
Buses	0	18,495	10,185.14	5,087.24

Power

Table 9. Statistics for power.

Vehicle category	min	max	average	std
Cars	7.34	21.74	11.98	1.92
Vans	8.72	28.41	14.33	2.76
Lorries ≤ 3500 kg	9.59	19.29	13.45	1.82
Mopeds	0.00	4.42	1.39	1.05
Lorries > 3500 kg	13.88	59.00	32.94	9.44

Motorcycles	1.00	13.10	3.61	2.41
Industrial tractors	33.62	62.01	47.50	3.37
Buses	13.12	53.44	34.72	11.31

Capacity

Table 10. Statistics for capacity.

Vehicle category	min	max	average	std
Cars	0	12,902	8,026.52	3,759.63
Vans	1,248	2,999	2,039.52	438.76
Lorries ≤ 3500 kg	2,148	13,798	7,459.30	2,990.61
Mopeds	0	523	91.88	130.71
Lorries > 3500 kg	1,199	3,996	2,195.09	549.08
Motorcycles	0	1,585	425.23	346.55
Industrial tractors	9,834	16,353	12,602.86	1,057.29
Buses	898	2,996	1,690.97	367.39

Emissions level

Table 11. Average annual kilometres by emissions level (only groups with more than 10 records).

Vehicle category	Emissions level	No. of records	Average annual km
Cars	1	149,984	6,073
Cars	2	733,607	7,629
Cars	3	2,768,059	8,823
Cars	4	4,440,615	10,351
Cars	5	2,257,390	12,167
Cars	6	2,457,075	14,330
Vans	1	63,125	9,221
Vans	2	124,240	7,061
Vans	3	251,931	9,051
Vans	4	284,824	11,993
Vans	5	237,874	18,448
Vans	6	137,310	23,696
Lorries ≤ 3500 kg	1	24,244	9,524
Lorries ≤ 3500 kg	2	147,864	8,668
Lorries ≤ 3500 kg	3	356,889	10,042
Lorries ≤ 3500 kg	4	382,965	12,161
Lorries ≤ 3500 kg	5	156,421	16,451
Lorries ≤ 3500 kg	6	92,616	20,917
Mopeds	1	18,405	1,331
Mopeds	2	49,840	2,253
Mopeds	3	1,185	3,051
Mopeds	4	8,446	3,172
Mopeds	5	1,538	3,322
Lorries > 3500 kg	1	4,854	11,393

Lorries > 3500 kg	2	24,306	15,446
Lorries > 3500 kg	3	45,691	20,715
Lorries > 3500 kg	4	1,552	31,300
Lorries > 3500 kg	5	54,710	28,384
Lorries > 3500 kg	6	23,116	45,171
Motorcycles	1	40,578	1,836
Motorcycles	2	108,766	2,117
Motorcycles	3	382,675	2,663
Motorcycles	4	127,428	3,758
Motorcycles	5	24,783	3,722
Motorcycles	6	118	8,799
Industrial tractors	2	1,006	21,140
Industrial tractors	3	12,708	36,925
Industrial tractors	4	507	52,883
Industrial tractors	5	31,745	59,036
Industrial tractors	6	116,860	110,320
Buses	1	51	13,155
Buses	2	687	20,553
Buses	3	3,736	28,938
Buses	4	194	28,327
Buses	5	14,582	40,342
Buses	6	22,075	66,309

Propulsion type

Table 12. Average annual kilometres by propulsion type (only groups with more than 10 records).

Vehicle category	Propulsion type	No. of records	Average annual km
Cars	Biodiesel	75	11,322
Cars	Butane	81	18,210
Cars	Diesel	8,273,835	11,821
Cars	Electric	49	15,035
Cars	Ethanol	63	9,776
Cars	Liquefied petroleum gas	31,935	21,617
Cars	Compressed natural gas	6,878	26,579
Cars	Liquefied natural gas	58	20,386
Cars	Petrol	4,563,102	9,046
Cars	Hydrogen	18	16,706
Cars	Others	396	15,381
Vans	Diesel	1,076,266	13,537
Vans	Liquefied petroleum gas	2,986	19,207
Vans	Compressed natural gas	731	23,597
Vans	Petrol	77,831	6,687
Vans	Others	34	6,487
Lorries ≤ 3500 kg	Diesel	1,158,743	12,241
Lorries ≤ 3500 kg	Liquefied petroleum gas	951	19,770

Lorries ≤ 3500 kg	Compressed natural gas	451	22,675
Lorries ≤ 3500 kg	Petrol	20,478	8,732
Lorries ≤ 3500 kg	Others	17	10,843
Mopeds	Diesel	8,699	3,631
Mopeds	Electric	4,400	1,977
Mopeds	Petrol	69,561	1,993
Lorries > 3500 kg	Diesel	158,235	25,423
Lorries > 3500 kg	Liquefied petroleum gas	243	32,927
Lorries > 3500 kg	Compressed natural gas	1,912	26,668
Lorries > 3500 kg	Liquefied natural gas	79	110,779
Lorries > 3500 kg	Petrol	642	18,145
Lorries > 3500 kg	Others	18	16,642
Motorcycles	Diesel	50	6,315
Motorcycles	Electric	3,310	2,877
Motorcycles	Liquefied petroleum gas	30	2,843
Motorcycles	Petrol	737,305	2,654
Motorcycles	Others	26	2,356
Industrial tractors	Diesel	166,646	92,352
Buses	Diesel	37,772	52,287
Buses	Electric	906	43,987
Buses	Liquefied petroleum gas	485	58,020
Buses	Compressed natural gas	2,843	58,412
Buses	Petrol	15	21,246
Buses	Hydrogen	13	21,357
Buses	Others	100	43,045

Number of owners

Table 13. Statistics for number of owners.

Vehicle category	min	max	average	std
Cars	1	48	2.61	1.80
Vans	1	59	3.05	2.01
Lorries ≤ 3500 kg	1	101	2.86	1.92
Mopeds	1	15	2.23	1.48
Lorries > 3500 kg	1	86	2.31	1.61
Motorcycles	1	23	2.68	1.90
Industrial tractors	1	17	2.33	1.68
Buses	1	35	1.80	1.42

Owner age

Table 14. Statistics for owner age (years).

Vehicle category	min	max	average	std
Cars	21.40	89.84	54.21	13.99
Vans	23.95	90.08	55.96	11.59
Lorries ≤ 3500 kg	23.95	89.71	55.34	11.36
Mopeds	17.77	89.76	54.53	13.50
Lorries > 3500 kg	30.63	73.33	52.91	4.04
Motorcycles	21.55	80.35	49.92	11.79
Industrial tractors	30.72	71.41	52.28	3.49
Buses	48.48	73.23	61.00	1.07

Table 15. Average annual kilometres by owner's age group.

Vehicle category	Owner's age group	No. of records	Average annual km
Cars	21-24	153,941	12,301
Cars	25-29	402,403	12,519
Cars	30-34	661,621	12,503
Cars	35-39	905,534	12,098
Cars	40-44	1,230,234	11,947
Cars	45-49	1,568,190	11,822
Cars	50-54	1,616,437	11,561
Cars	55-59	1,571,535	11,044
Cars	60-64	1,376,296	10,258
Cars	65-69	1,081,287	9,096
Cars	70-74	791,676	8,015
Cars	75+	1,051,631	6,787
Cars	Vehicle in company's name	465,715	15,228
Vans	21-24	2,170	14,213
Vans	25-29	18,359	13,670
Vans	30-34	35,786	13,290
Vans	35-39	54,720	13,013
Vans	40-44	76,122	13,025
Vans	45-49	97,543	12,516
Vans	50-54	105,433	11,925
Vans	55-59	108,885	11,074
Vans	60-64	104,732	10,179
Vans	65-69	84,143	8,798
Vans	70-74	61,056	7,678
Vans	75+	77,167	6,565
Vans	Vehicle in company's name	331,746	18,806
Lorries ≤ 3500 kg	21-24	2,012	13,782
Lorries ≤ 3500 kg	25-29	17,465	13,375

Lorries ≤ 3500 kg	30-34	35,370	13,122
Lorries ≤ 3500 kg	35-39	58,888	13,037
Lorries ≤ 3500 kg	40-44	87,094	13,000
Lorries ≤ 3500 kg	45-49	112,137	12,729
Lorries ≤ 3500 kg	50-54	115,716	12,218
Lorries ≤ 3500 kg	55-59	114,547	11,437
Lorries ≤ 3500 kg	60-64	104,013	10,705
Lorries ≤ 3500 kg	65-69	78,503	9,560
Lorries ≤ 3500 kg	70-74	53,369	8,485
Lorries ≤ 3500 kg	75+	70,734	7,096
Lorries ≤ 3500 kg	Vehicle in company's name	330,806	14,501
Mopeds	18-20	668	2,772
Mopeds	21-24	1,415	2,377
Mopeds	25-29	2,066	2,149
Mopeds	30-34	2,812	2,107
Mopeds	35-39	3,826	2,115
Mopeds	40-44	5,963	2,098
Mopeds	45-49	9,828	2,245
Mopeds	50-54	11,481	2,256
Mopeds	55-59	10,616	2,172
Mopeds	60-64	8,229	1,987
Mopeds	65-69	6,258	1,954
Mopeds	70-74	4,738	2,018
Mopeds	75+	6,315	1,940
Mopeds	Vehicle in company's name	8,413	2,526
Lorries > 3500 kg	30-34	1,016	20,689
Lorries > 3500 kg	35-39	2,144	19,367
Lorries > 3500 kg	40-44	3,361	22,074
Lorries > 3500 kg	45-49	4,694	23,543
Lorries > 3500 kg	50-54	5,284	24,858
Lorries > 3500 kg	55-59	5,776	26,158
Lorries > 3500 kg	60-64	5,145	26,365
Lorries > 3500 kg	65-69	1,957	21,647
Lorries > 3500 kg	70-74	621	18,035
Lorries > 3500 kg	Vehicle in company's name	131,138	25,803
Motorcycles	21-24	9,367	3,381
Motorcycles	25-29	30,291	3,277
Motorcycles	30-34	54,180	3,056
Motorcycles	35-39	65,213	2,751
Motorcycles	40-44	80,503	2,606
Motorcycles	45-49	106,403	2,616
Motorcycles	50-54	115,556	2,708
Motorcycles	55-59	98,886	2,654
Motorcycles	60-64	75,657	2,476
Motorcycles	65-69	46,232	2,140
Motorcycles	70-74	21,164	1,892

Motorcycles	75+	9,210	1,680
Motorcycles	Vehicle in company's name	28,068	3,039
Industrial tractors	30-34	884	77,748
Industrial tractors	35-39	1,827	77,069
Industrial tractors	40-44	2,987	80,244
Industrial tractors	45-49	4,872	83,653
Industrial tractors	50-54	5,504	82,118
Industrial tractors	55-59	5,473	81,808
Industrial tractors	60-64	4,185	79,689
Industrial tractors	65-69	1,254	74,205
Industrial tractors	70-74	169	69,107
Industrial tractors	Vehicle in company's name	139,491	94,596
Buses	45-49	74	32,756
Buses	50-54	100	34,784
Buses	55-59	157	30,994
Buses	60-64	106	27,125
Buses	65-69	83	28,033
Buses	70-74	189	27,174
Buses	Vehicle in company's name	41,442	52,942

Owner gender

Table 16. Average annual kilometres by gender of the owner or legal person.

Vehicle category	Owner gender	No. of records	Average annual km
Cars	Female	4,648,528	10,284
Cars	Male	7,762,257	10,960
Cars	Legal person	465,715	15,228
Vans	Female	155,107	10,573
Vans	Male	671,009	10,859
Vans	Legal person	331,746	18,806
Lorries ≤ 3500 kg	Female	166,847	11,291
Lorries ≤ 3500 kg	Male	683,001	11,291
Lorries ≤ 3500 kg	Legal person	330,806	14,501
Mopeds	Female	21,770	1,961
Mopeds	Male	52,478	2,191
Mopeds	Legal person	8,413	2,526
Lorries > 3500 kg	Female	2,359	21,754
Lorries > 3500 kg	Male	27,639	24,154
Lorries > 3500 kg	Legal person	131,138	25,803
Motorcycles	Female	103,344	2,416
Motorcycles	Male	609,318	2,678
Motorcycles	Legal person	28,068	3,039
Industrial tractors	Female	2,338	84,681
Industrial tractors	Male	24,817	80,459
Industrial tractors	Legal person	139,491	94,596

Buses	Female	52	15,035
Buses	Male	657	30,935
Buses	Legal person	41,442	52,942

Service

Table 17. Average annual kilometres by service (only groups with more than 10 records).

Vehicle category	Service	No. of records	Average annual km
Cars	PUBL-Not specified	4,348	16,416
Cars	PUBL-Rental without driver	24,518	13,848
Cars	PUBL-Rental with driver	6,475	39,124
Cars	PUBL-Driving lessons	16,952	25,706
Cars	PUBL-Taxi	44,024	43,476
Cars	PUBL-Hazardous goods	24	10,225
Cars	PUBL-School transport	32	27,798
Cars	PUBL-Police	1,126	13,001
Cars	PUBL-Fire service	101	17,131
Cars	PUBL-Civil protection and rescue	94	10,003
Cars	PUBL-Defence	18	8,840
Cars	PUBL-Economic activity	65	15,283
Cars	PUBL-Perishable goods	53	12,048
Cars	PART-Not specified	12,770,890	10,718
Cars	PRIV-Housing	52	10,590
Cars	PART-Economic activity	2,276	14,158
Cars	PRIV-Recreational	93	12,620
Cars	PRIV-Fair vehicle	36	11,766
Cars	PRIV-Historic vehicle	5,315	4,111
Vans	PUBL-Not specified	34,110	25,869
Vans	PUBL-Rental without driver	20,125	22,889
Vans	PUBL-Rental with driver	53	20,938
Vans	PUBL-Driving lessons	63	10,118
Vans	PUBL-Taxi	45	20,393
Vans	PUBL-Roadside assistance	29	23,743
Vans	PUBL-Ambulance	6,583	47,682
Vans	PUBL-Funeral vehicle	251	14,324
Vans	PUBL-Waste collection	19	9,077
Vans	PUBL-Police	62	11,386
Vans	PUBL-Fire service	34	11,089
Vans	PUBL-Civil protection and rescue	85	12,443
Vans	PUBL-Defence	15	8,513
Vans	PUBL-Economic activity	1,093	27,707
Vans	PUBL-Perishable goods	14	19,621
Vans	PART-Not specified	1,085,426	12,240
Vans	PART-Ambulance	205	10,457
Vans	PRIV-Housing	12	8,081

Vans	PART-Economic activity	8,921	19,768
Vans	PRIV-Fair vehicle	24	17,710
Vans	PRIV-Historic vehicle	684	4,760
Lorries ≤ 3500 kg	PUBL-Not specified	26,356	19,955
Lorries ≤ 3500 kg	PUBL-Rental without driver	14,365	19,579
Lorries ≤ 3500 kg	PUBL-Rental with driver	139	30,384
Lorries ≤ 3500 kg	PUBL-Driving lessons	112	11,806
Lorries ≤ 3500 kg	PUBL-Taxi	188	40,949
Lorries ≤ 3500 kg	PUBL-Roadside assistance	1,415	22,408
Lorries ≤ 3500 kg	PUBL-Hazardous goods	27	18,084
Lorries ≤ 3500 kg	PUBL-Waste collection	46	17,861
Lorries ≤ 3500 kg	PUBL-Police	198	10,756
Lorries ≤ 3500 kg	PUBL-Fire service	126	10,061
Lorries ≤ 3500 kg	PUBL-Civil protection and rescue	114	8,700
Lorries ≤ 3500 kg	PUBL-Economic activity	422	24,974
Lorries ≤ 3500 kg	PUBL-Perishable goods	22	30,103
Lorries ≤ 3500 kg	PART-Not specified	1,127,530	11,858
Lorries ≤ 3500 kg	PART-Economic activity	8,615	17,194
Lorries ≤ 3500 kg	PRIV-Fair vehicle	58	11,415
Lorries ≤ 3500 kg	PRIV-Historic vehicle	892	4,705
Mopeds	PUBL-Not specified	27	2,746
Mopeds	PUBL-Rental without driver	3,982	2,021
Mopeds	PUBL-Driving lessons	1,469	650
Mopeds	PART-Not specified	77,017	2,202
Mopeds	PART-Economic activity	27	3,056
Mopeds	PRIV-Historic vehicle	126	1,039
Lorries > 3500 kg	PUBL-Not specified	63,011	33,180
Lorries > 3500 kg	PUBL-Rental without driver	4,249	39,031
Lorries > 3500 kg	PUBL-Driving lessons	621	11,648
Lorries > 3500 kg	PUBL-Roadside assistance	6,100	34,551
Lorries > 3500 kg	PUBL-Hazardous goods	93	38,736
Lorries > 3500 kg	PUBL-Waste collection	1,514	30,260
Lorries > 3500 kg	PUBL-Fire service	779	5,424
Lorries > 3500 kg	PUBL-Civil protection and rescue	11	4,149
Lorries > 3500 kg	PUBL-Economic activity	263	44,989
Lorries > 3500 kg	PUBL-Perishable goods	39	37,519
Lorries > 3500 kg	PART-Not specified	83,386	18,437
Lorries > 3500 kg	PART-Economic activity	750	28,153
Lorries > 3500 kg	PRIV-Fair vehicle	283	8,067
Motorcycles	PUBL-Not specified	385	2,306
Motorcycles	PUBL-Rental without driver	4,026	2,952
Motorcycles	PUBL-Driving lessons	6,592	2,945
Motorcycles	PUBL-Police	858	3,079
Motorcycles	PUBL-Civil protection and rescue	15	381
Motorcycles	PART-Not specified	726,201	2,655
Motorcycles	PART-Economic activity	60	3,951

Motorcycles	PRIV-Recreational	35	2,727
Motorcycles	PRIV-Historic vehicle	2,529	1,150
Industrial tractors	PUBL-Not specified	130,515	94,417
Industrial tractors	PUBL-Rental without driver	16,823	118,208
Industrial tractors	PUBL-Driving lessons	458	13,545
Industrial tractors	PUBL-Roadside assistance	60	44,002
Industrial tractors	PUBL-Hazardous goods	233	104,561
Industrial tractors	PUBL-Economic activity	3,237	139,407
Industrial tractors	PUBL-Perishable goods	95	123,885
Industrial tractors	PART-Not specified	14,623	38,837
Industrial tractors	PRIV-Works	13	32,752
Industrial tractors	PART-Economic activity	115	45,131
Industrial tractors	PRIV-Fair vehicle	454	17,057
Buses	PUBL-Not specified	34,673	52,509
Buses	PUBL-Rental without driver	3,814	69,915
Buses	PUBL-Rental with driver	49	54,042
Buses	PUBL-Driving lessons	384	11,254
Buses	PUBL-School transport	1,821	48,049
Buses	PUBL-Defence	11	38,773
Buses	PUBL-Economic activity	90	57,040
Buses	PART-Not specified	1,288	21,419

6. PREDICTIVE MODELS

A regression model based on gradient boosting was developed and implemented in order to estimate the number of kilometres travelled per year by the vehicle fleet. The model was specifically adapted to each vehicle category, taking as a reference the study validated in 2023 and enriching it by incorporating relevant new variables and optimising the data-scrubbing process.

The gradient-boosting algorithm was chosen based on its capacity to handle heterogeneous variables, capture complex non-linear relationships and provide robust estimates, even with imbalanced distributions of data. Specifically, we used XGBoost Regressor, which sequentially builds multiple decision trees in which each new tree is trained to correct the residual errors from the previous trees.

The analysis was conducted in Python using the Pandas, NumPy and Scikit-learn libraries to manage and transform the data and to evaluate the models. To implement the gradient-boosting algorithm, the specialised XGBoost library was used.

7. ADJUSTMENT AND TRAINING

The scrubbed dataset of 16,408,340 records was divided into two parts, with 80% allocated to training and 20% to validation. This division was applied to each vehicle category.

The training dataset was used to train the model, while the validation dataset was reserved exclusively for validating it.

We trained a separate XGBoost regression model for each vehicle category, using only the records for the category in question as training and validation data. We set the same hyperparameters for all of the models regardless of the vehicle category they specialised in:

- Number of trees (n_estimators): 600. This parameter defines the number of sequential decision trees the model builds.
- Maximum depth (max_depth): 6 levels. This parameter controls the complexity of each individual tree by limiting the number of successive divisions it can make.
- Learning rate (learning_rate): 0.05. This parameter determines the contribution made by each new tree to the model as a whole, acting as a weighting factor that moderates the model's updates.

Additionally, we set a random seed (random_state) to ensure that the training process could be replicated using the same data. We did not use any stopping criteria for the model.

The results of the training process are evaluated in detail in the following section, using aggregated comparative and performance metrics.

8. VALIDATING THE MODELS

8.1. PERFORMANCE METRICS

Table 18. Validation metrics for the predictive models by vehicle category in the 2024 fleet. brings together the main validation metrics obtained in the validation dataset for each specific model developed for the various vehicle categories. The root mean squared error (RMSE) and mean absolute error (MAE) reflect the mean level of error in the estimate of the number of kilometres travelled during the year per vehicle. The MAE is less sensitive to extreme errors than the RMSE.

Table 18. Validation metrics for the predictive models by vehicle category in the 2024 fleet.

Metric	Cars	Vans	Lorries ≤ 3500 kg	Mopeds	Lorries > 3500 kg	Motorcycles	Industrial tractors	Buses
RMSE	6,739.21	9,948.65	9,116.89	2,209.47	22,396.15	2,410.90	36,177.39	28,061.27
MAE	4,902.67	6,553.95	6,339.36	1,426.54	14,376.31	1,660.67	27,325.67	18,788.37
R^2	0.2558	0.3410	0.2083	0.1827	0.3286	0.2038	0.5432	0.4827

The coefficient of determination (R^2) shows the proportion of variance in the data that the model is able to explain in each category.

The following tables contain the performance metrics obtained for the different groupings by vehicle age, electric category, propulsion type and maximum load.

Table 19. Validation metrics for the predictive models by vehicle category and age in the 2024 fleet.

Vehicle category	Vehicle age (years)	RMSE	MAE	R2
Cars	0-4	8,263.45	5,940.09	0.5628
Cars	5-9	7,448.12	5,470.86	0.2707
Cars	10-14	6,929.18	5,141.19	0.1397
Cars	15-19	6,456.83	4,739.05	0.1166
Cars	20+	6,103.69	4,300.97	0.1338
Vans	0-4	14,611.11	11,118.74	0.3006
Vans	5-9	13,632.63	10,039.2	0.1641
Vans	10-14	10,872.88	7,794.75	0.1058
Vans	15-19	8,230.52	5,665.49	0.0932
Vans	20+	7,500.09	4,287.54	0.0717
Lorries ≤ 3500 kg	0-4	13,920.74	10,588.12	0.1837
Lorries ≤ 3500 kg	5-9	11,711.6	8,625.49	0.1319
Lorries ≤ 3500 kg	10-14	9,884.99	7,172.46	0.1070
Lorries ≤ 3500 kg	15-19	8,647.06	6,117.52	0.0740
Lorries ≤ 3500 kg	20+	7,845.22	5,193.93	0.0762
Mopeds	0-4	2,264.93	1,286.79	0.2896
Mopeds	5-9	2,458.04	1,757.1	0.1623

Mopeds	10-14	2,305.1	1,536.81	0.1080
Mopeds	15-19	2,292.56	1,485.64	0.0874
Mopeds	20+	1,858.35	1,118.94	0.0784
Lorries > 3500 kg	0-4	28,370.5	19,981.87	0.4244
Lorries > 3500 kg	5-9	27,083.07	19,523.59	0.4020
Lorries > 3500 kg	10-14	24,322.45	17,239.84	0.2457
Lorries > 3500 kg	15-19	21,275.37	14,152.69	0.1516
Lorries > 3500 kg	20+	19,903.86	11,046.08	0.0863
Motorcycles	0-4	2,856.19	2,057.79	0.2394
Motorcycles	5-9	2,760.26	1,981.88	0.1546
Motorcycles	10-14	2,357.59	1,645.18	0.1412
Motorcycles	15-19	2,217.32	1,520.22	0.1169
Motorcycles	20+	1,877.56	1,175.82	0.1073
Industrial tractors	0-4	35,245.68	26,063.48	0.3378
Industrial tractors	5-9	36,896.16	29,099.98	0.2480
Industrial tractors	10-14	39,585.96	30,889.63	0.2241
Industrial tractors	15-19	34,458.88	25,390.83	0.2308
Industrial tractors	20+	31,815.36	19,534.25	0.1906
Buses	0-4	27,324.13	18,521.35	0.6521
Buses	5-9	30,296.53	20,735.21	0.3694
Buses	10-14	25,598.57	17,855.31	0.2990
Buses	15-19	26,377.07	16,707.8	0.1345
Buses	20+	29,381.02	16,458.75	0.0458

Table 20. Validation metrics for the predictive models by vehicle category and electric category in the 2024 fleet.

Vehicle category	Electric category	RMSE	MAE	R2
Cars	HEV	8,796.55	6,261.8	0.6260
Cars	PHEV	8,122.42	5,864.3	0.5208
Vans	HEV	15,372.75	11,847.94	0.5133
Vans	PHEV	7,936.49	6,127.42	0.0312
Lorries ≤ 3500 kg	HEV	12,623.31	10,188.69	0.2781
Mopeds	BEV	1,642.67	743.52	0.4250
Lorries > 3500 kg	HEV	15,925.93	13,176.52	0.0825
Motorcycles	BEV	2,069.13	1,287.92	0.2762
Buses	BEV	15,095.52	10,133.18	0.3969
Buses	HEV	22,418.55	15,211.81	0.5679
Buses	PHEV	18,790.84	14,490.43	0.6191

Table 21. Validation metrics for the predictive models by vehicle category and propulsion type in the 2024 fleet.

Vehicle category	Propulsion type	RMSE	MAE	R2
Cars	Biodiesel	2,950.07	2,428.83	0.3377
Cars	Butane	6,030.89	4,152.5	0.9205
Cars	Diesel	7,163.4	5,276.92	0.1900
Cars	Liquefied petroleum gas	10,776.03	7,851.94	0.5975

Cars	Compressed natural gas	10,751.99	7,928.67	0.6541
Cars	Petrol	5,840.91	4,198.8	0.2968
Cars	Others	5,560.09	4,187.21	0.8769
Vans	Diesel	10,113.4	6,745	0.3326
Vans	Liquefied petroleum gas	11,951.01	8,702.63	0.2913
Vans	Compressed natural gas	14,329.1	11,172.66	0.1683
Vans	Petrol	7,120.41	3,791.19	0.1280
Lorries ≤ 3500 kg	Diesel	9,148.66	6,365.6	0.2066
Lorries ≤ 3500 kg	Liquefied petroleum gas	11,965.36	8,716.57	0.2854
Lorries ≤ 3500 kg	Compressed natural gas	12,873.1	9,687.21	0.2504
Lorries ≤ 3500 kg	Petrol	6,748.55	4,663.92	0.1277
Mopeds	Diesel	3,239.58	2,320.69	0.0770
Mopeds	Electric	1,661.95	775.8	0.4043
Mopeds	Petrol	2,074.38	1,354.46	0.1535
Lorries > 3500 kg	Diesel	22,471.23	14,408.3	0.3279
Lorries > 3500 kg	Liquefied petroleum gas	26,154.74	18,032.86	-0.0968
Lorries > 3500 kg	Compressed natural gas	16,604.01	12,386.83	0.2826
Lorries > 3500 kg	Petrol	15,468.23	10,471.1	0.3090
Motorcycles	Diesel	3,092.88	2,278.03	-0.1104
Motorcycles	Electric	2,115.43	1,322.73	0.2445
Motorcycles	Petrol	2,412.1	1,662.12	0.2037
Industrial tractors	Diesel	36,177.39	27,325.67	0.5432
Buses	Diesel	28,745.16	19,547.25	0.4853
Buses	Electric	18,825.04	11,858.16	0.3960
Buses	Liquefied petroleum gas	30,779.83	18,795.81	0.3635
Buses	Compressed natural gas	16,786.93	10,666.11	0.5172
Buses	Others	57,590.17	23,832.67	0.0298

Table 22. Validation metrics for the predictive models by maximum load for lorries in the 2024 fleet.

Maximum load	RMSE	MAE	R2
0-10T	10,274.02	6,726.32	0.2133
10-20T	23,263.48	16,353.68	0.3876
20-30T	26,694.44	18,578.57	0.3434
30-40T	24,206.02	15,585.52	0.0450
40+ T	19,932.95	14,317.87	-0.0505

8.2. COMPARISON OF ITV VS ESTIMATE

Table 23 presents the comparison between the annual kilometres recorded by the ITV and the estimate obtained by each individual model in the validation dataset, with an overall error of 0.05%. **Figure 6** presents the results in graph form.

Table 23. Comparison of total annual kilometres according to ITV and estimated data for the 2024 fleet.

Vehicle category	Annual km – ITV	Annual km – estimated	Absolute error	Error %
Cars	27,993,135,446	27,991,543,808	1,591,638	0.01%
Vans	3,027,040,593	3,039,593,984	12,553,391	0.41%
Lorries with MAM ≤ 3500 kg	2,877,337,448	2,877,088,256	249,192	0.01%
Mopeds	35,741,041	35,813,840	72,799	0.20%
Lorries with MAM > 3500 kg	819,405,463	823,060,864	3,655,401	0.45%
Motorcycles	394,733,271	392,887,680	1,845,591	0.47%
Industrial tractors	3,069,666,157	3,078,424,832	577,918	0.10%
Buses	443,067,055	440,560,032	2,507,023	0.57%
Total	**38,660,126,475**	**38,678,970,368**	**18,843,893**	**0.05%**

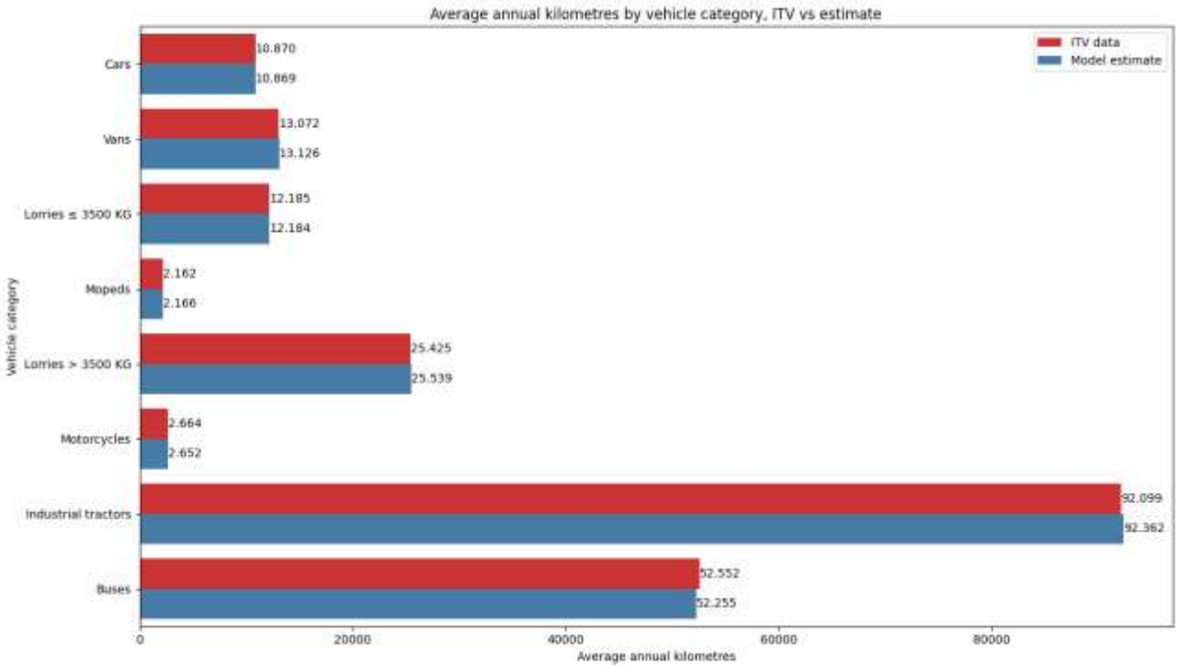

Figure 6. Comparison of annual kilometres according to ITV and estimated data by vehicle category.

The following figures show the average annual kilometres according to ITV data compared to the average annual kilometres estimated by the model, differentiated by vehicle category and age.

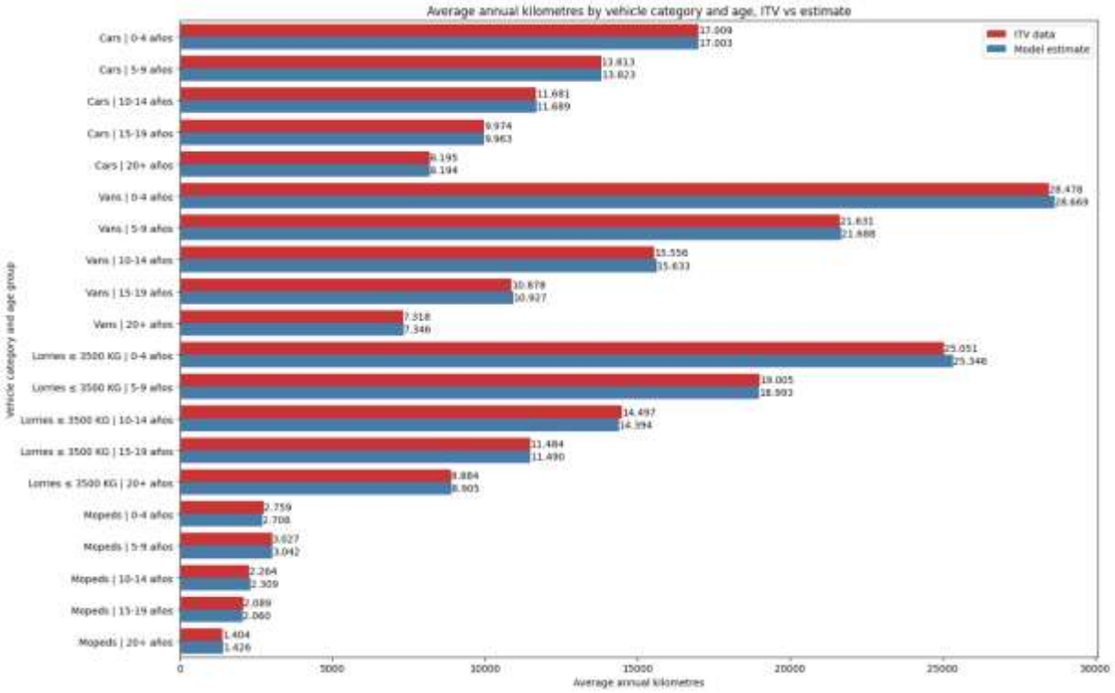

Figure 7. Comparison of annual kilometres according to ITV and estimated data by age for cars, vans, lorries with MAM ≤ 3500 kg and mopeds.

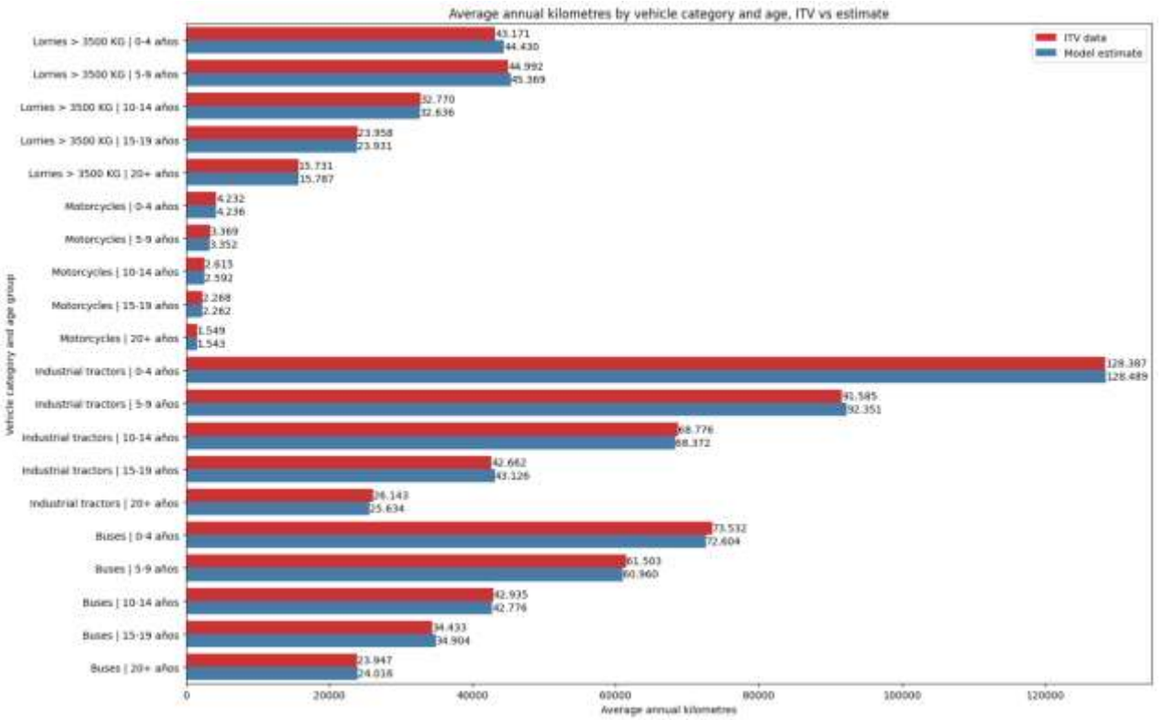

Figure 8. Comparison of annual kilometres according to ITV and estimated data by age for lorries with MAM > 3500 kg, motorcycles, industrial tractors and buses.

The follow figures present a comparison between annual kilometres according to the model and those according to the ITV data, for groupings by electric category, propulsion type and maximum load.

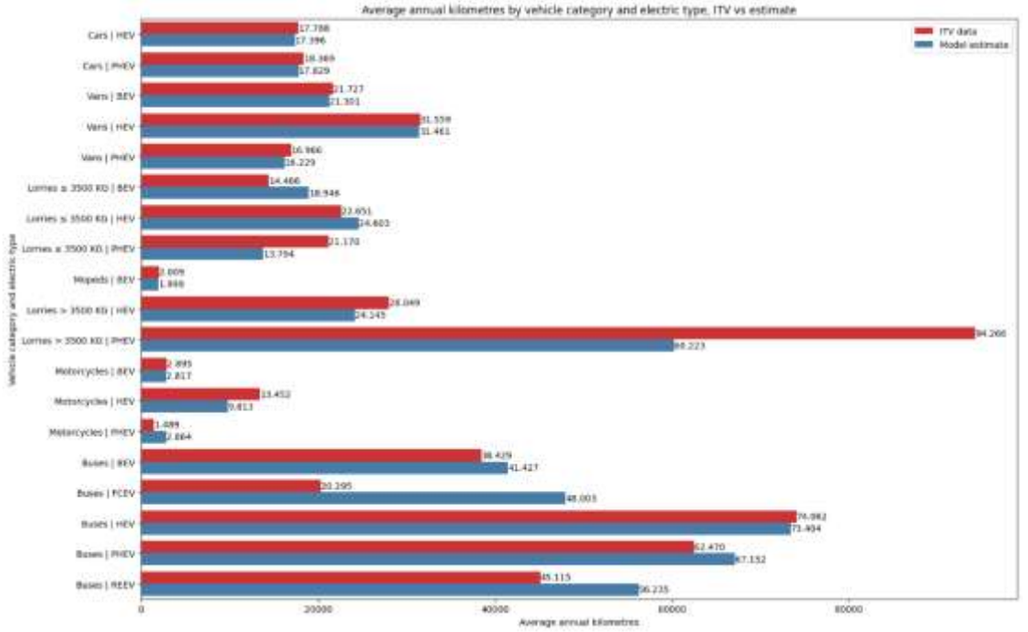

Figure 9. Comparison of annual kilometres according to ITV and estimated data by electric category.

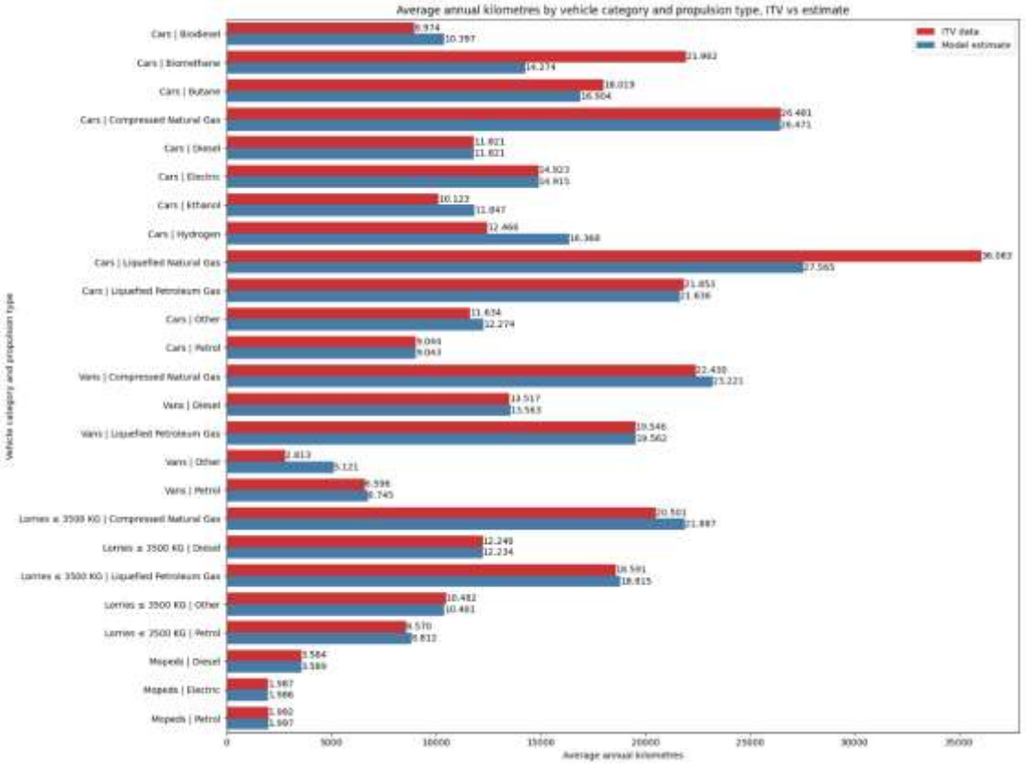

Figure 10. Comparison of annual kilometres according to ITV and estimated data by propulsion type for cars, vans, lorries with MAM ≤ 3500 kg and mopeds.

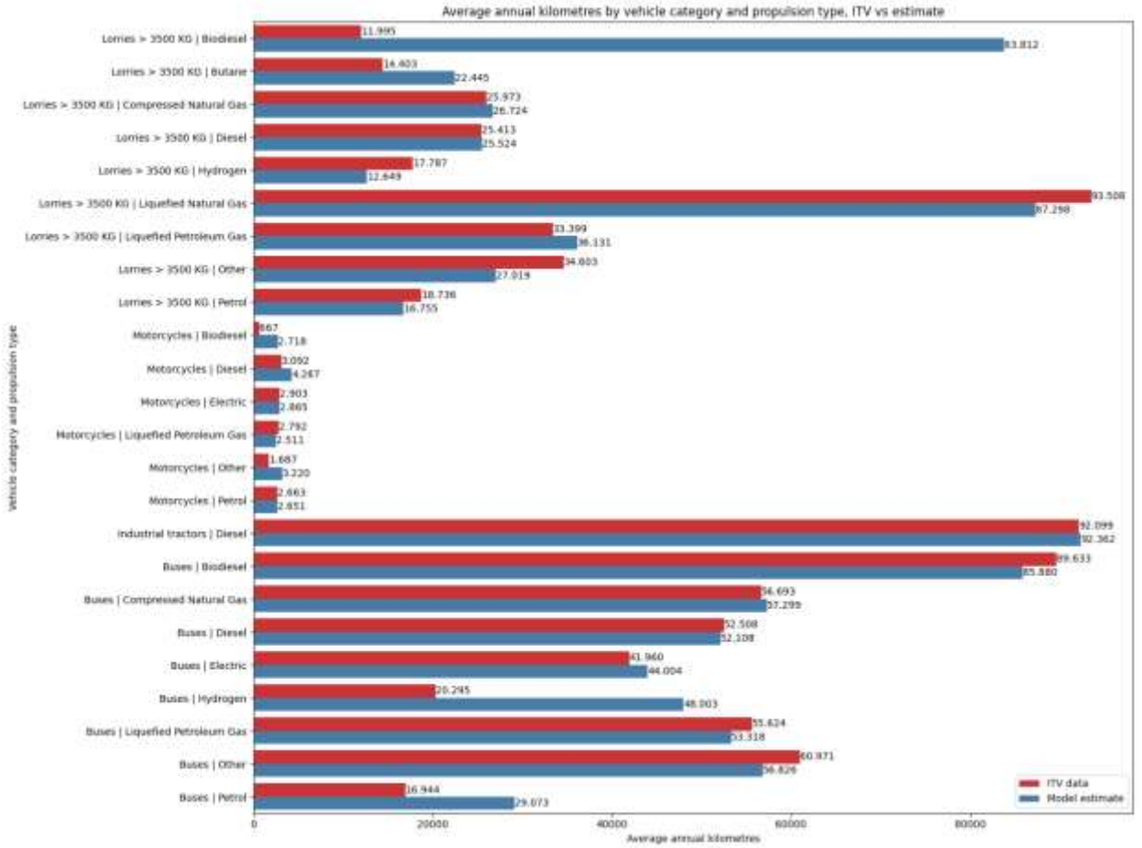

Figure 11. Comparison of annual kilometres according to ITV and estimated data by propulsion type for lorries with MAM > 3500 kg, motorcycles, industrial tractors and buses.

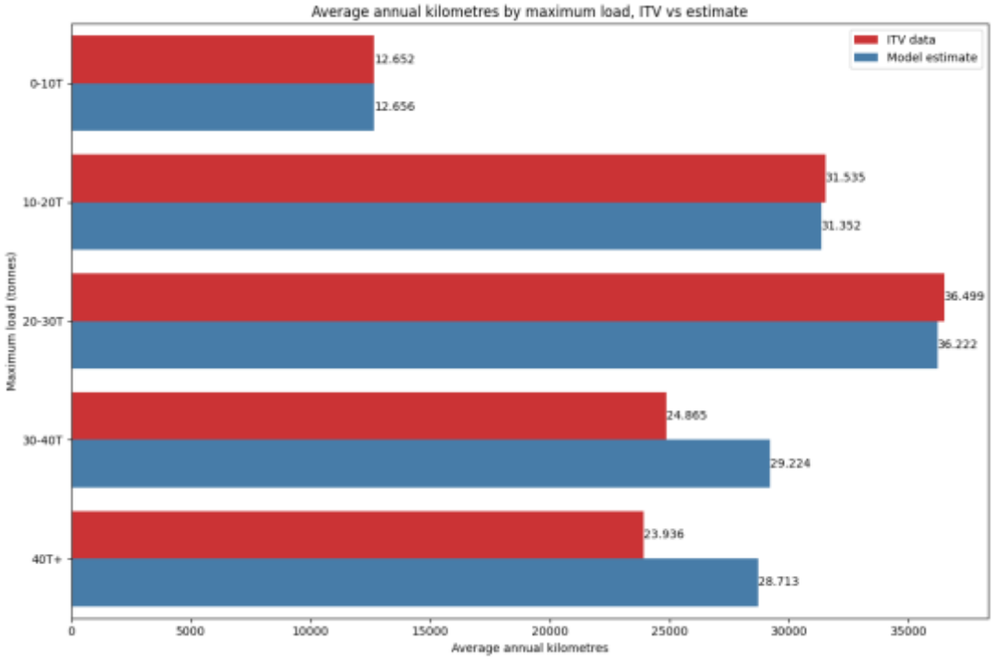

Figure 12. Comparison of annual kilometres according to ITV and estimated data by maximum load (lorries only).

8.3. IMPORTANCE OF THE VARIABLES USED

The levels of importance for the variables were extracted using the "feature_importances_" parameter in the XGBoost Regressor model. This value represents the cumulative gain (total reduction in prediction error) contributed by each variable through all of the divisions in the set of trees. The values are normalised (adding up to 1.0), with a higher importance indicating that the variable plays more of a determining role in predicting the annual kilometres.

The graphs show how different variables become more important depending on vehicle category.

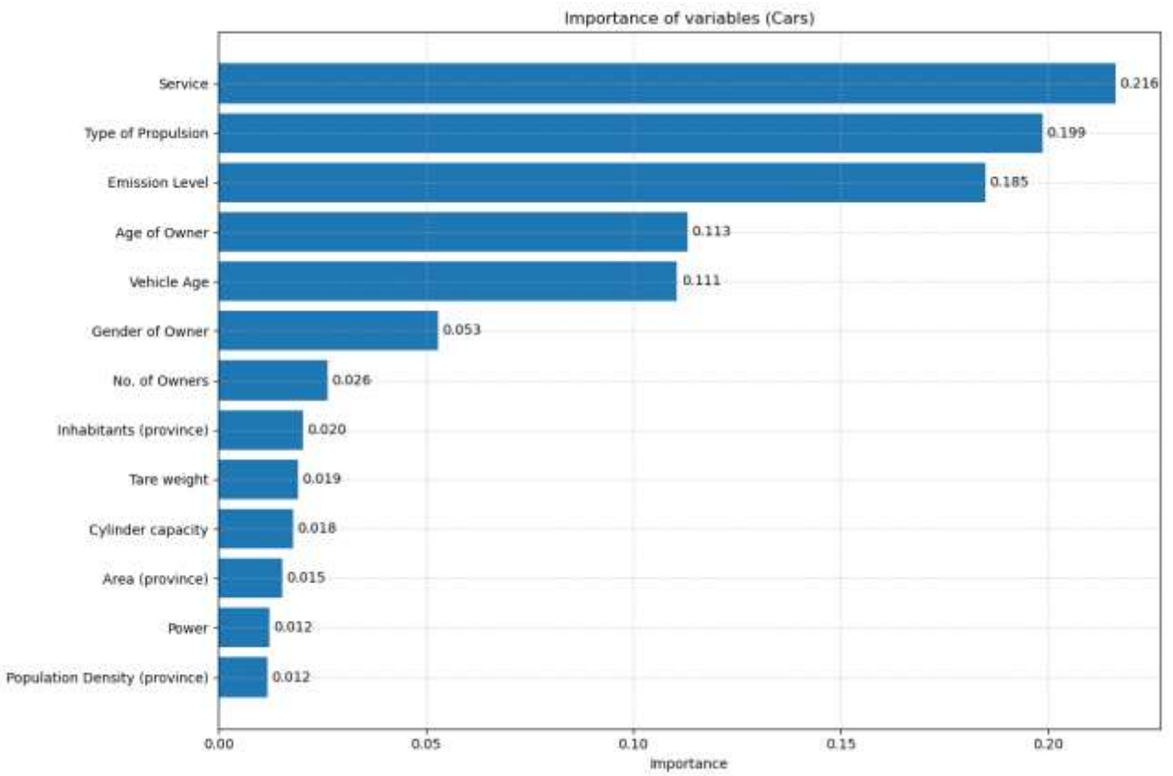

Figure 13. Importance of the variables used in the model for cars.

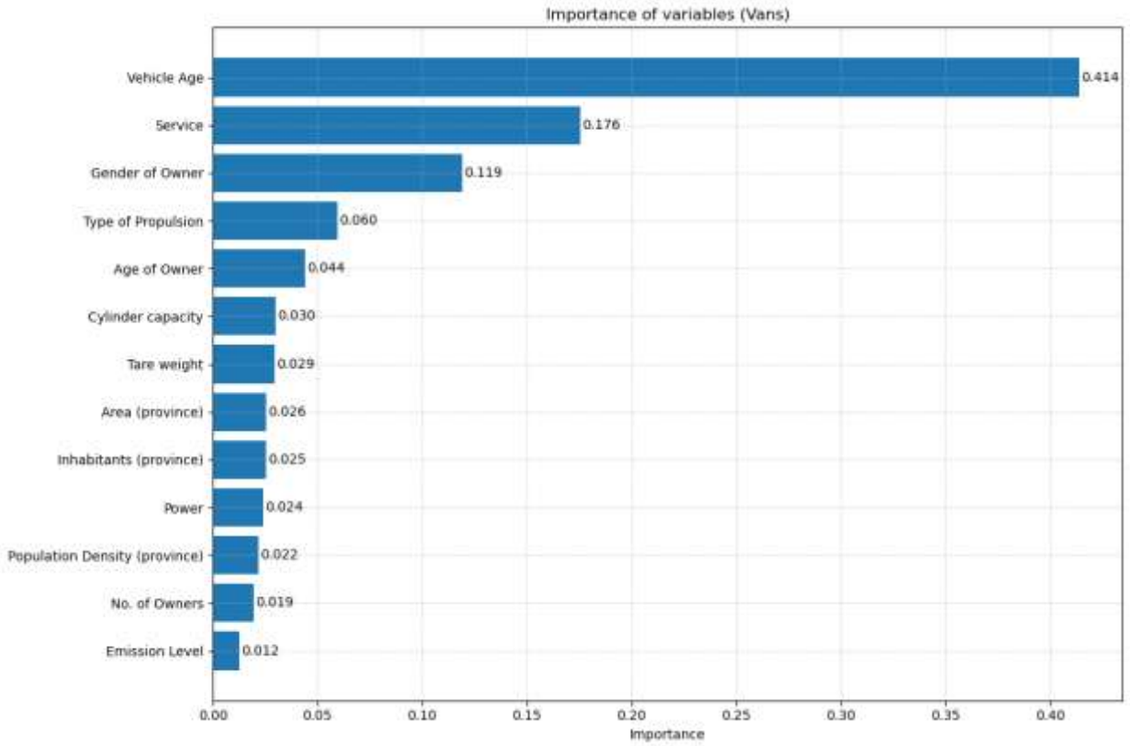

Figure 14. Importance of the variables used in the model for vans.

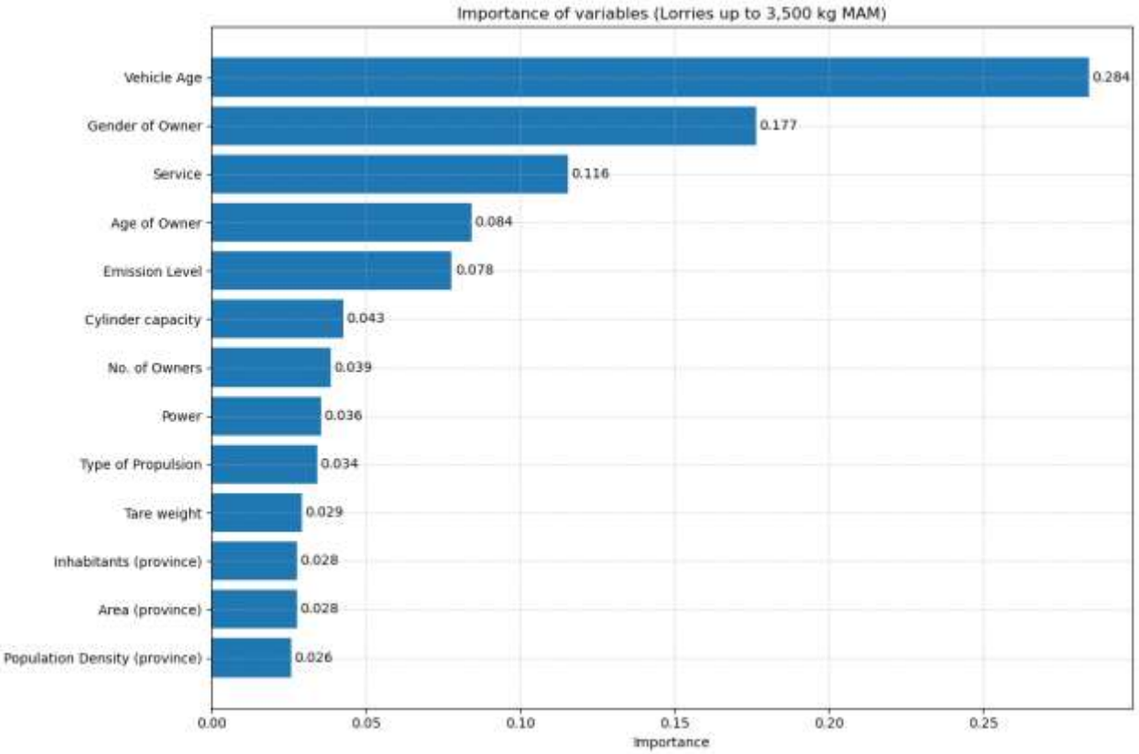

Figure 15. Importance of the variables used in the model for lorries with MAM ≤ 3500 kg.

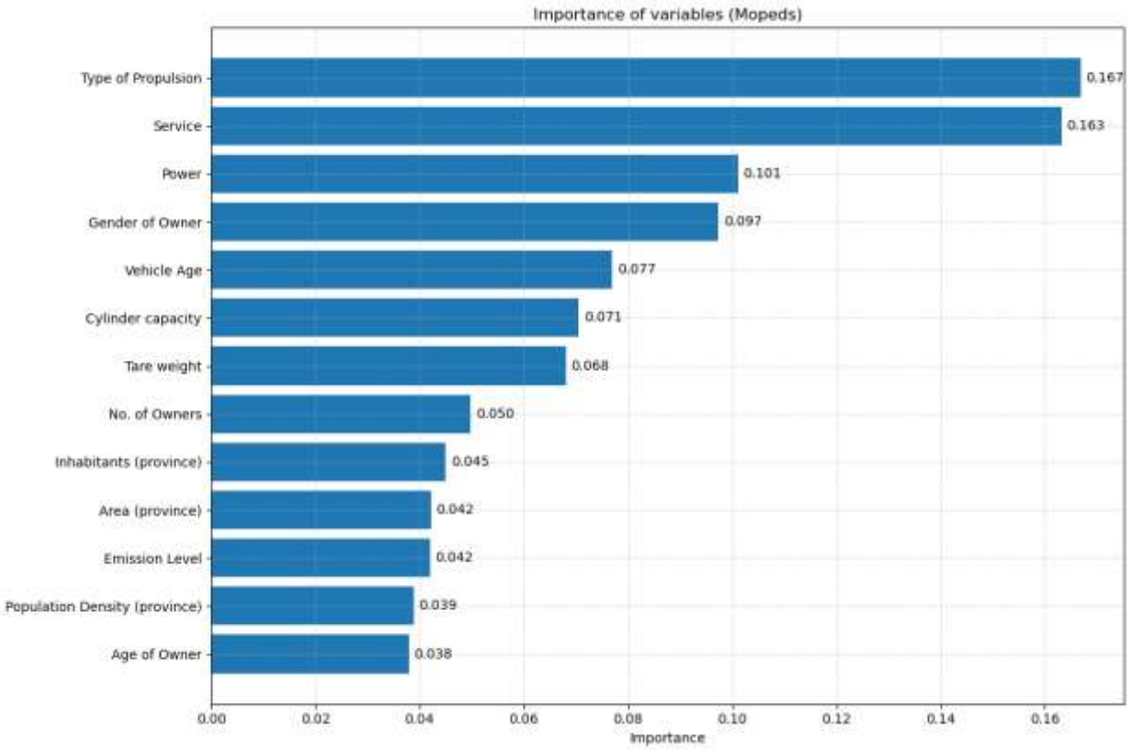

Figure 16. Importance of the variables used in the model for mopeds.

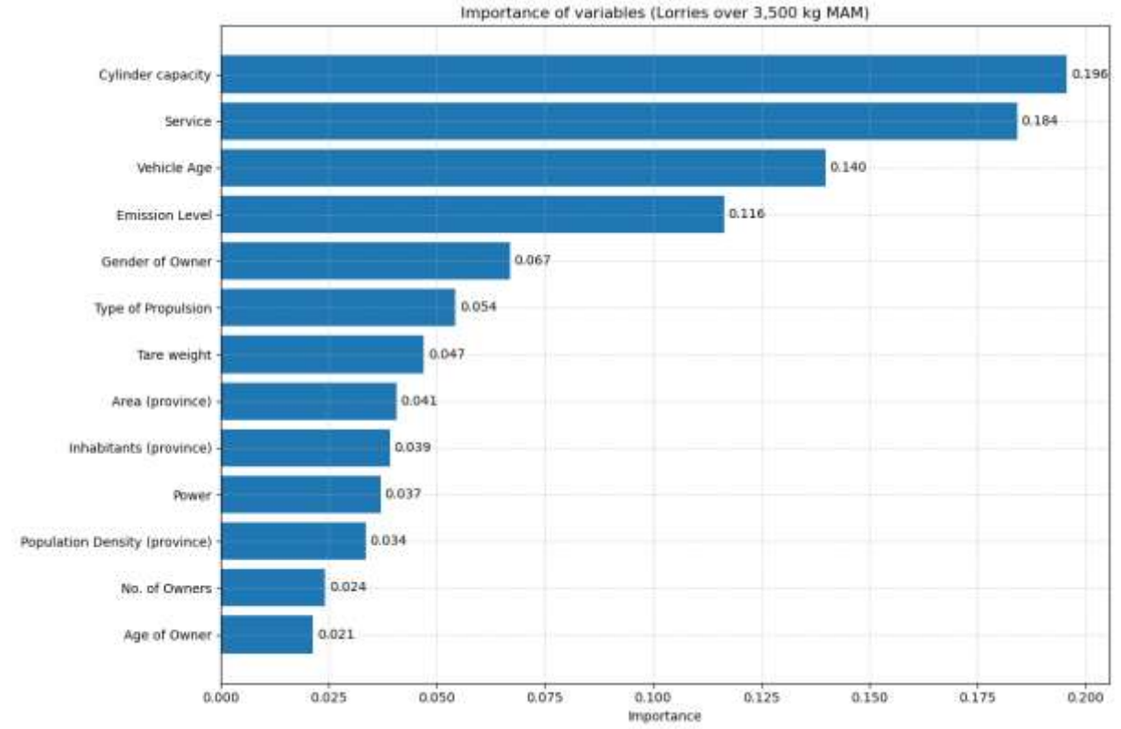

Figure 17. Importance of the variables used in the model for lorries with MAM > 3500 kg.

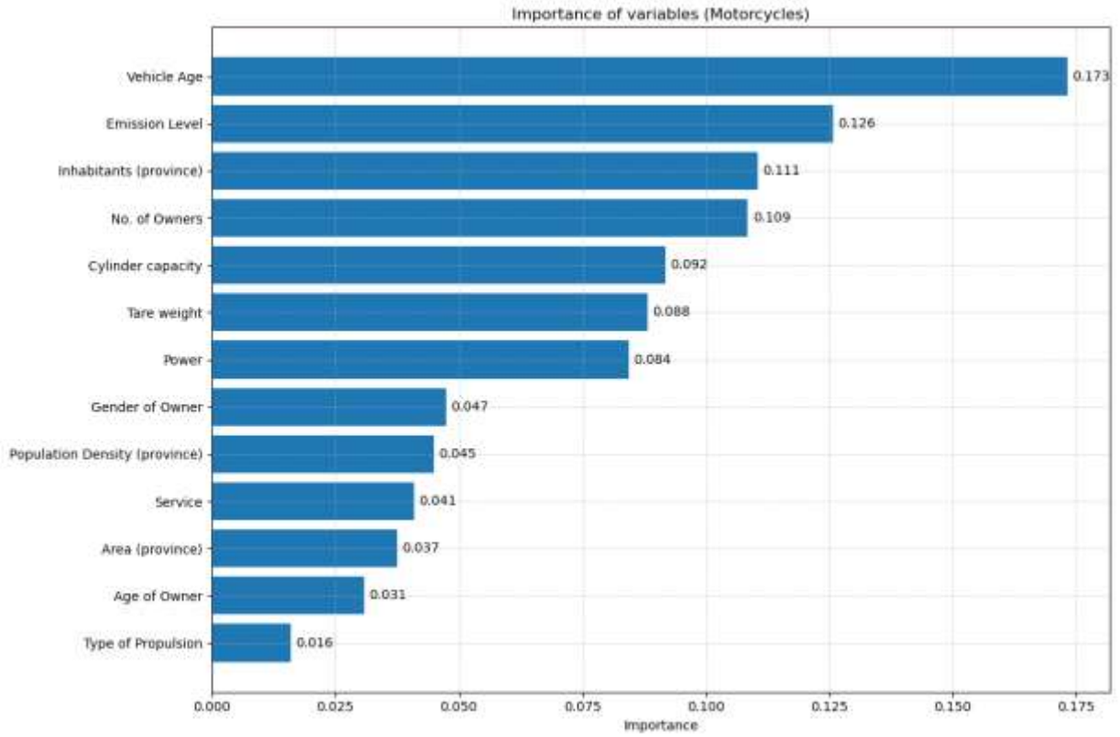

Figure 18. Importance of the variables used in the model for motorcycles.

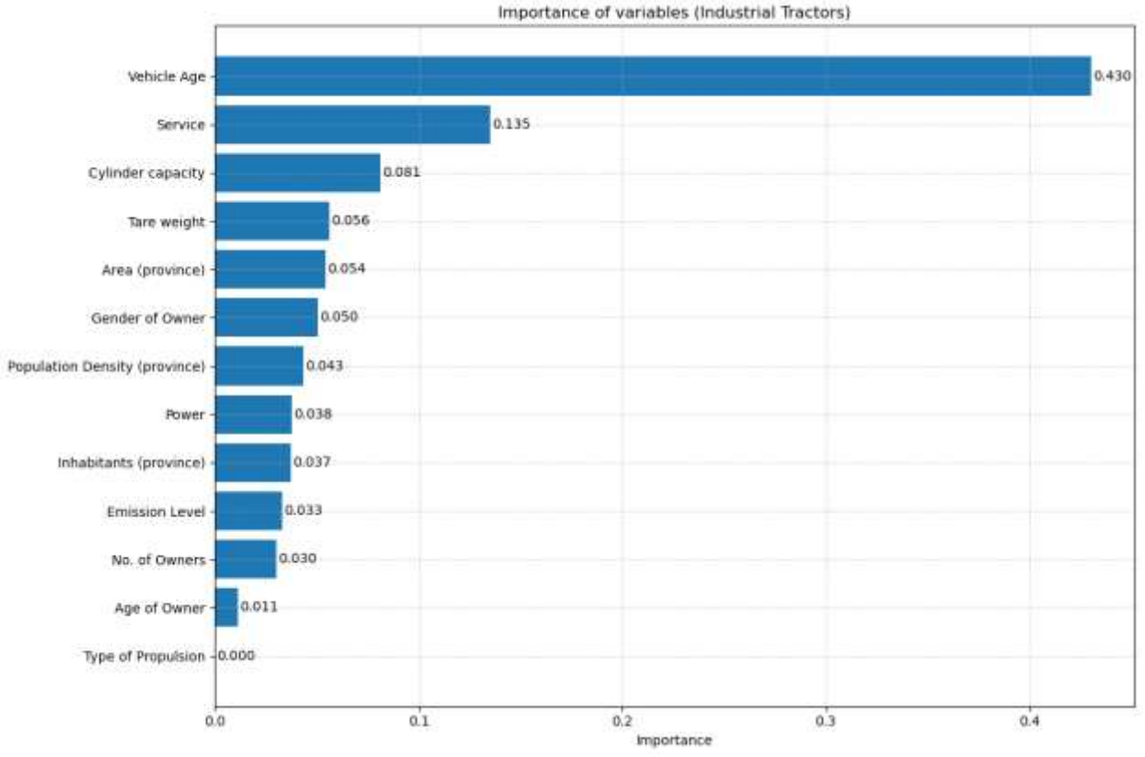

Figure 19. Importance of the variables used in the model for industrial tractors.

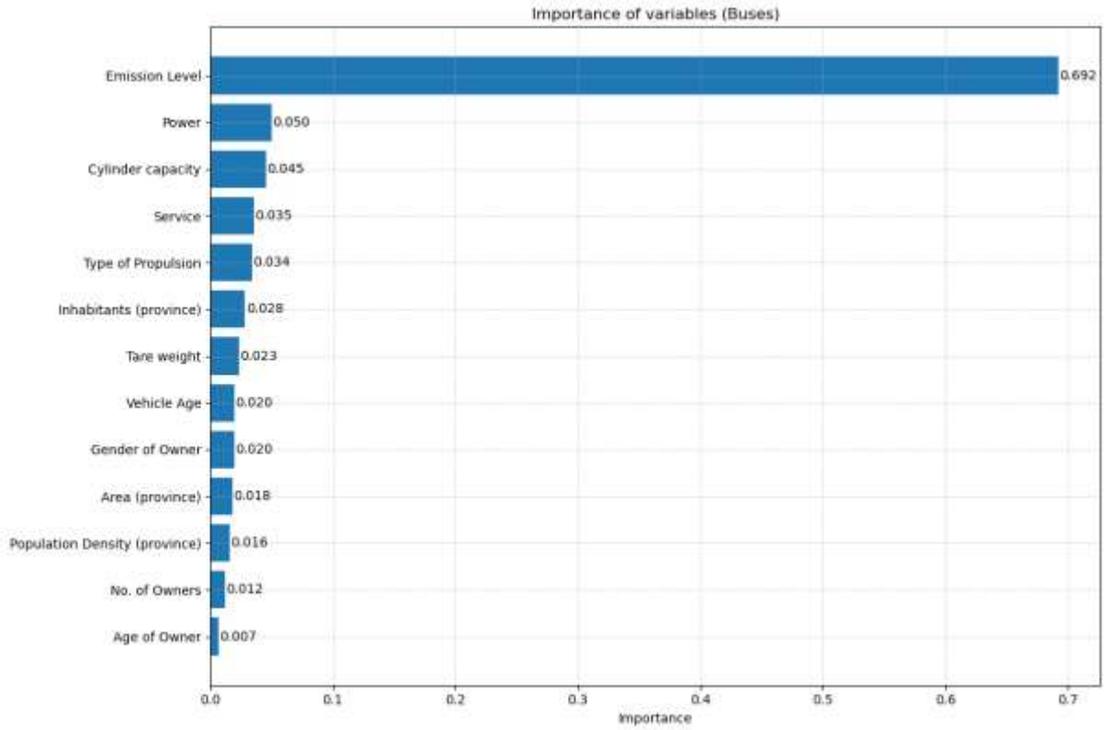

Figure 20. Importance of the variables used in the model for buses.

8.4. EVALUATING THE MODEL USING THE 2023 VEHICLE FLEET

In this section we evaluate the performance of the predictive model when applied to data for Spain's 2023 vehicle fleet. The aim is to test the generalisation capacity of the model trained and validated with 2024 data, using historical records and the same explanatory variables.

To evaluate the model using the 2023 vehicle fleet, we replicated the same process used to generate and prepare the 2024 data. First, we cross-referenced the information in the vehicle fleet database with the ITV records for 2023. We then applied the same criteria for cleansing and scrubbing anomalous data, eliminating inconsistent records, extreme values and entries that did not meet the quality requirements defined in the methodology.

We used the entire scrubbed 2023 dataset, which contained more than 16 million records, as a validation dataset for the model trained on 2024 data.

The following table contains the main validation metrics obtained for each of the vehicle categories analysed, including RMSE, MAE and R^2.

Table 24. Validation and accuracy metrics for the predictive models by vehicle category in the 2023 fleet.

Metric	Cars	Vans	Lorries ≤ 3500 kg	Mopeds	Lorries > 3500 kg	Motorcycles	Industrial tractors	Buses
RMSE	6,763.92	10,497.55	9,608.03	2,304.13	23,981.14	2,417.45	38,694.37	29,122.97
MAE	4,916.88	6,728.38	6,550.12	1,509.19	15,291.46	1,657.74	29,216.69	19,681.93
R^2	0.2408	0.3284	0.1913	0.0941	0.3082	0.1316	0.4783	0.4224

Figure 25 presents the comparison between the total annual kilometres recorded during the ITV and those estimated by the model for each group of vehicles in 2023, along with the absolute error and associated percentage error.

Table 25. Comparison of total annual kilometres according to ITV and estimated data for the 2023 fleet.

Vehicle category	Annual km – ITV	Annual km – estimated	Absolute error	Error %
Cars	139,521,646,962	139,891,097,600	369,450,638	0.26%
Vans	15,105,183,364	15,049,841,664	55,341,700	0.37%
Lorries ≤ 3500 kg	15,198,051,715	15,269,938,176	71,886,461	0.47%
Mopeds	201,991,388	190,437,024	11,554,364	5.72%
Lorries > 3500 kg	4,419,491,239	4,424,227,840	4,736,601	0.11%
Motorcycles	1,920,196,605	1,840,210,432	79,986,173	4.17%
Industrial tractors	15,769,537,158	15,091,160,064	678,377,094	4.30%
Buses	2,151,979,417	2,022,800,128	129,179,289	6.00%
Total	**194,288,077,848**	**193,779,728,384**	**508,349,464**	**0.26%**

The overall error in the estimate of the total number of kilometres travelled by the vehicles in the 2023 dataset is 0.26%.

Figure 21 illustrates the comparison between the average annual kilometres according to the ITV data and those estimated by the model for each vehicle category. In the majority of cases, correlation can be observed between the two sources.

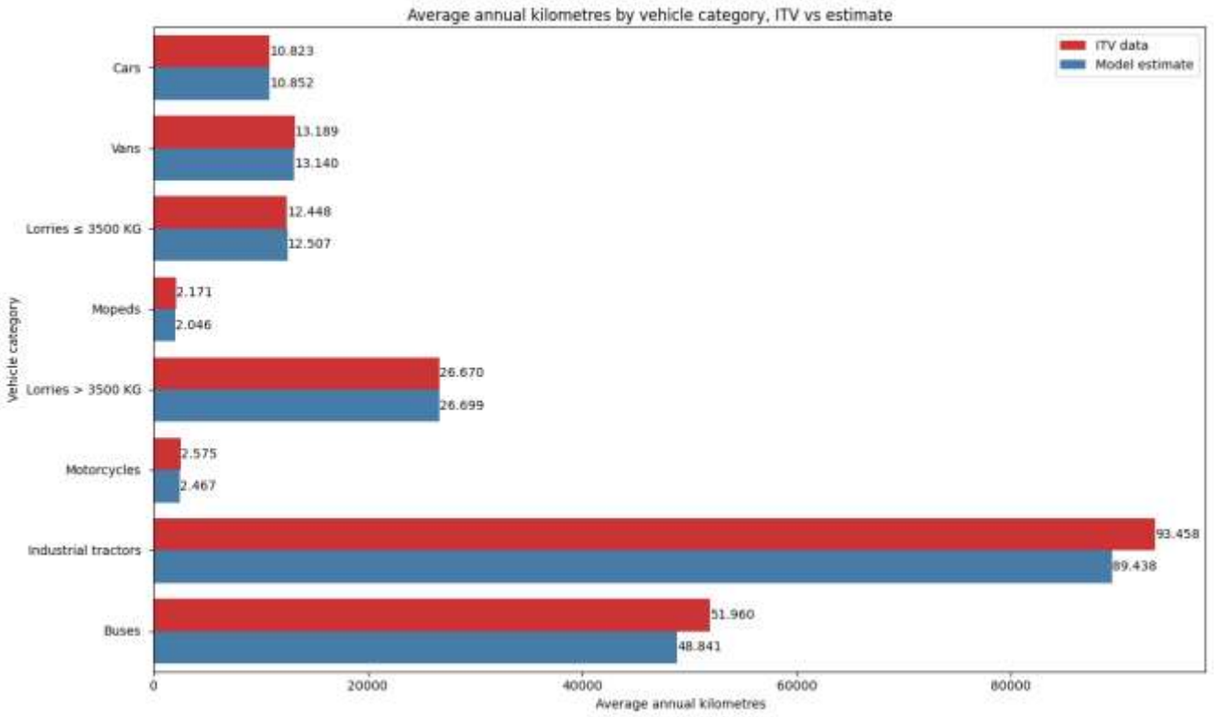

Figure 21. Comparison of annual kilometres according to ITV and estimated data by vehicle category in the 2023 fleet.

9. EXTRAPOLATION TO THE FLEET IN CIRCULATION

After training the model, the final database for the active fleet in circulation was built and a complete estimate of annual kilometres was obtained by combining the recorded data and the estimated values.

9.1. SCRUBBING AND ASSEMBLING THE FLEET

The database for the 2024 fleet is subjected to a scrubbing process in which:

- Records with incorrect vehicle categories are eliminated.
- Vehicles that do not form part of the active fleet in circulation according to the "FLG_CIRCULATION" variable are excluded.

The outcome is 33,504,103 usable records from a total of 37,464,950 for the 2024 fleet.

Subsequently, the model's input variables were adjusted according to the statistics for the numerical and categorical variables, as detailed in section **5.1 ANALYSIS OF THE VARIABLES**:

- The numerical variables are delimited by the minimum and maximum values of the training dataset, with any null values replaced by the mean.
- Categorical variables with null or non-recorded values are replaced by the most frequent entry within their corresponding vehicle category.

9.2. OBTAINING THE ANNUAL KILOMETRES

The annual kilometres are estimated by combining the recorded data and the model's predictions:

- 16,408,340 records (49% of the usable data) contain direct information on annual kilometres from the ITV database for 2024.
- 17,095,763 records (51%) do not contain ITV data and therefore require estimation.

For those records where no information is available, the estimate generated by the predictive model corresponding to the vehicle category is used. As a result, a complete estimate of annual kilometres for the entire fleet in circulation is obtained.

10. MAIN RESULTS

In this section we present the main results obtained from the predictions generated by the models developed.

For 2024, the values for kilometres travelled are obtained from a combination of real and estimated data. Firstly, we use the annual kilometres recorded during the ITV, provided that these records are available and pass the stipulated quality criteria. In cases where there is no valid record, whether due to the absence of data or exclusion during the scrubbing process, the estimate generated by the specific predictive model for the corresponding vehicle category was used.

Table 26 shows the estimated number of kilometres travelled by the vehicle fleet in circulation for 2024.

Table 26. Kilometres travelled by the fleet in circulation (2024). The values are per 10 million.

Cars	Vans	Lorries ≤ 3500 kg	Mopeds	Lorries > 3500 kg	Motorcycles	Industrial tractors	Buses
29,299.1	3,491.1	2,665.6	207.6	779.1	1,128.8	2,123.5	315.1

Figure 22 shows the estimated average kilometres for each vehicle type for 2024.

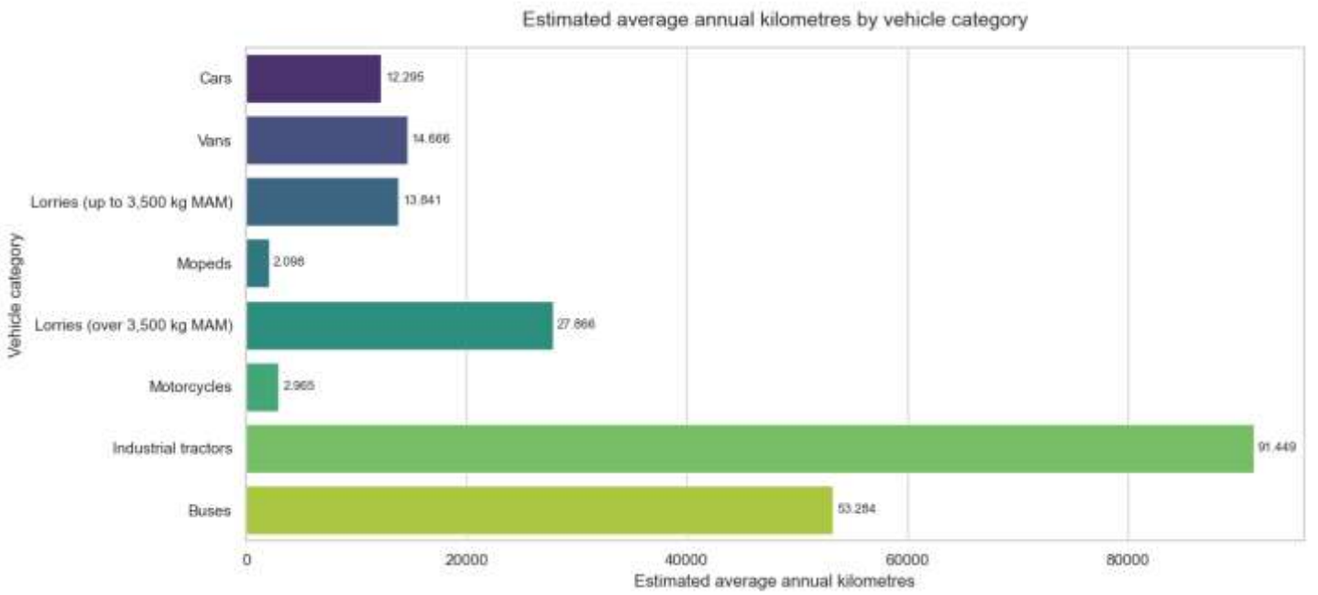

Figure 22. Estimated annualised kilometres by vehicle type (2024 fleet in circulation).

Table 27 presents the estimated average number of kilometres travelled by vehicle type and age, which allows for a comparison of annual use among different segments of the vehicle fleet. Additionally, **Figure 23** provides an illustration in graph form of these estimates for 2024.

Table 27. Estimated average kilometres by vehicle type and age (2024 fleet in circulation).

Vehicle age	Cars	Vans	Lorries ≤ 3500 kg	Mopeds	Lorries > 3500 kg	Motorcycles	Industrial tractors	Buses
0-4 years old	17,522	23,956	25,040	3,750	40,784	4,627	129,791	71,304
5-9 years old	14,368	20,024	19,762	3,056	42,101	3,418	95,448	62,087
10-14 years old	11,956	15,479	14,847	2,377	33,733	2,559	70,786	44,508
15-19 years old	10,070	10,789	11,764	2,113	24,892	2,199	45,982	34,733
20 and older	8,202	6,956	9,199	1,564	18,133	1,689	31,168	29,194

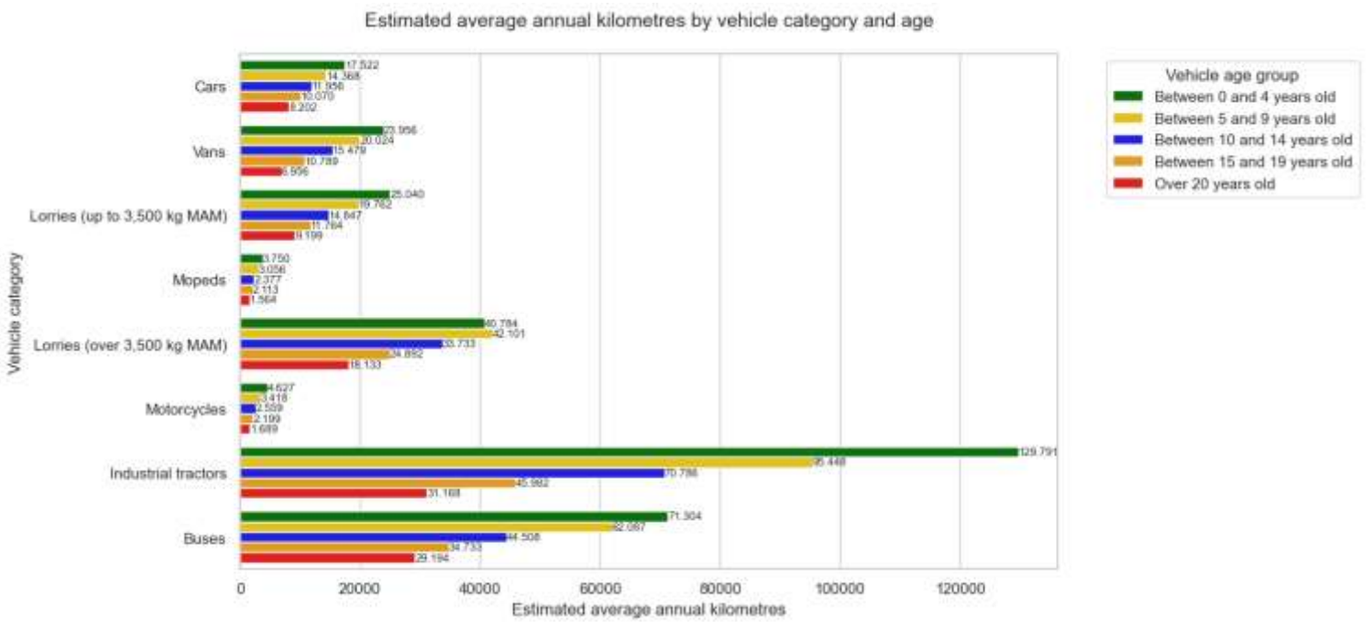

Figure 23. Estimated average kilometres by vehicle type and age (2024 fleet in circulation).

Table 28 below contains the estimated average annual kilometres by propulsion type and vehicle category. Additionally, **Figure 24** and **Figure 25** provide an illustration in graph form of these results for 2024.

Table 28. Estimated average kilometres by vehicle and propulsion type (2024 fleet in circulation).

Propulsion	Cars	Vans	Lorries ≤ 3500 kg	Mopeds	Lorries > 3500 kg	Motorcycles	Industrial tractors	Buses
Petrol	11,242	9,872	10,433	1,980	22,078	2,953		30,204
Diesel	12,925	15,114	13,904	3,932	27,792	6,457	91,449	53,176
Electric	21,926	25,228	22,949	2,304	43,685	3,416		42,266
Other/not specified	14,685	10,144	15,356	1,529	19,266	3,361		42,962
Butane	16,527	9,014	12,724		22,744	2,499		52,546
LPG (liquefied petroleum gas)	19,504	20,220	19,417	1,601	31,980	3,730		63,518
CNG (compressed natural gas)	22,522	21,684	24,200		28,229	7,131		59,340
LNG (liquefied natural gas)	18,676	14,810	28,521	1,409	78,581	3,617		68,709
Hydrogen	18,769	22,175			14,071			24,793
Biomethane	13,664	29,031	5,768		29,181	4,127		55,426
Ethanol	10,078							
Biodiesel	12,110	20,005	32,430		72,674	2,590		86,963
Others	14,685	10,144	15,356	1,529	19,266	3,361		42,962

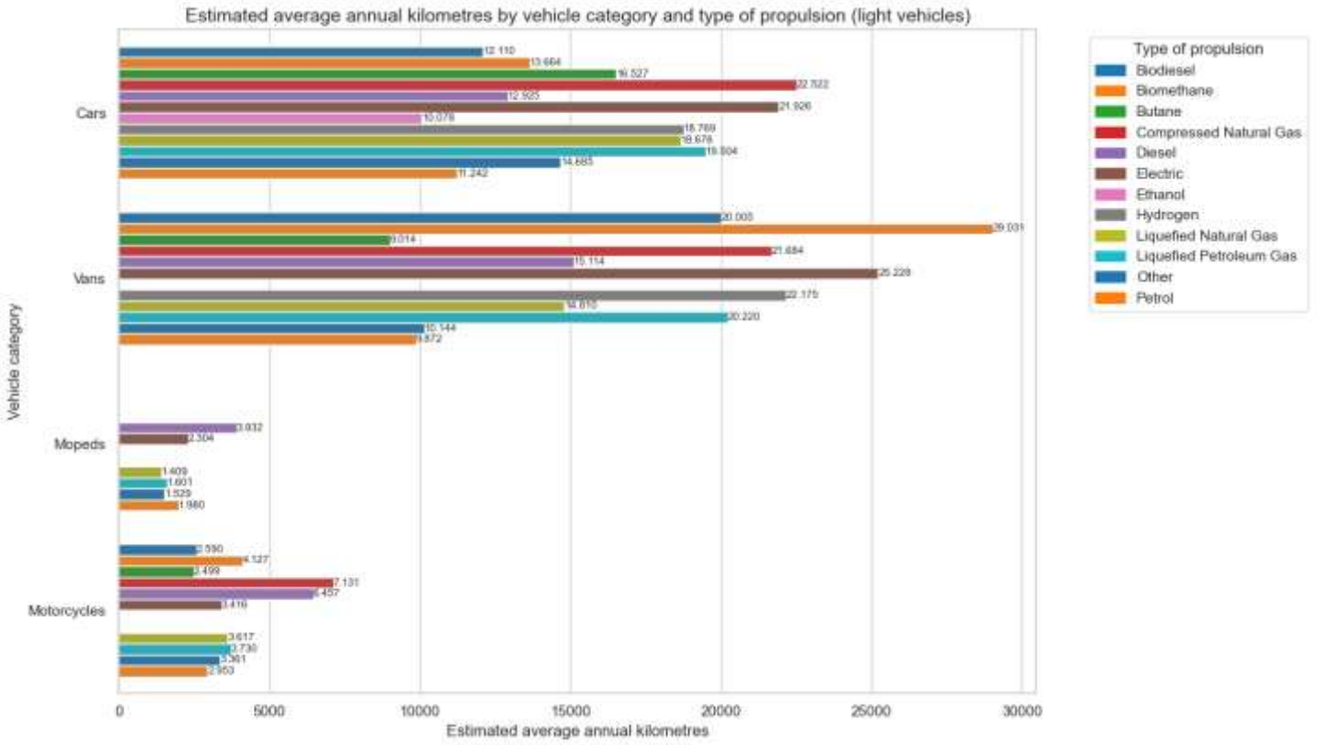

Figure 24. Estimated average kilometres by vehicle and propulsion type for light vehicles (2024 fleet in circulation).

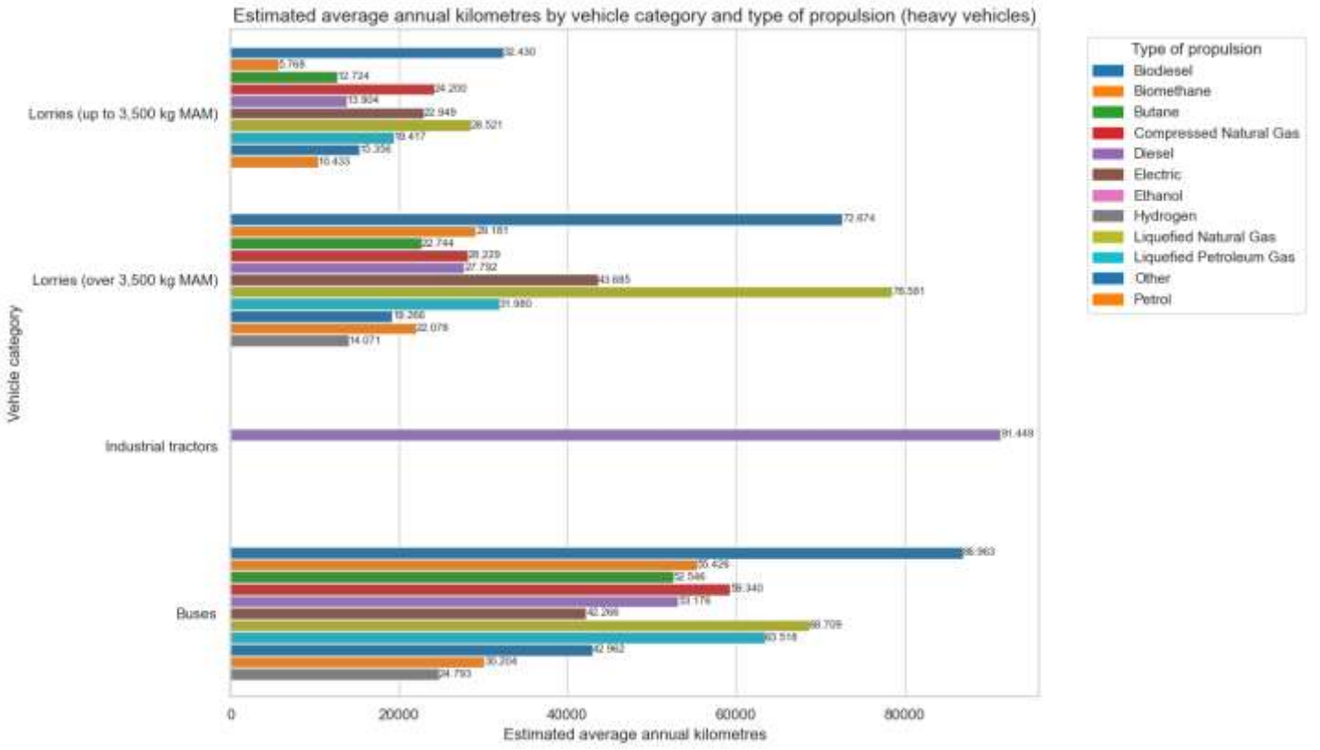

Figure 25. Estimated average kilometres by vehicle and propulsion type for heavy vehicles (2024 fleet in circulation).

Figure 26 and **Figure 27** present the estimated average annual kilometres by owner age for 2024. The table associated with these graphs can be found in section **13. ANNEX**.

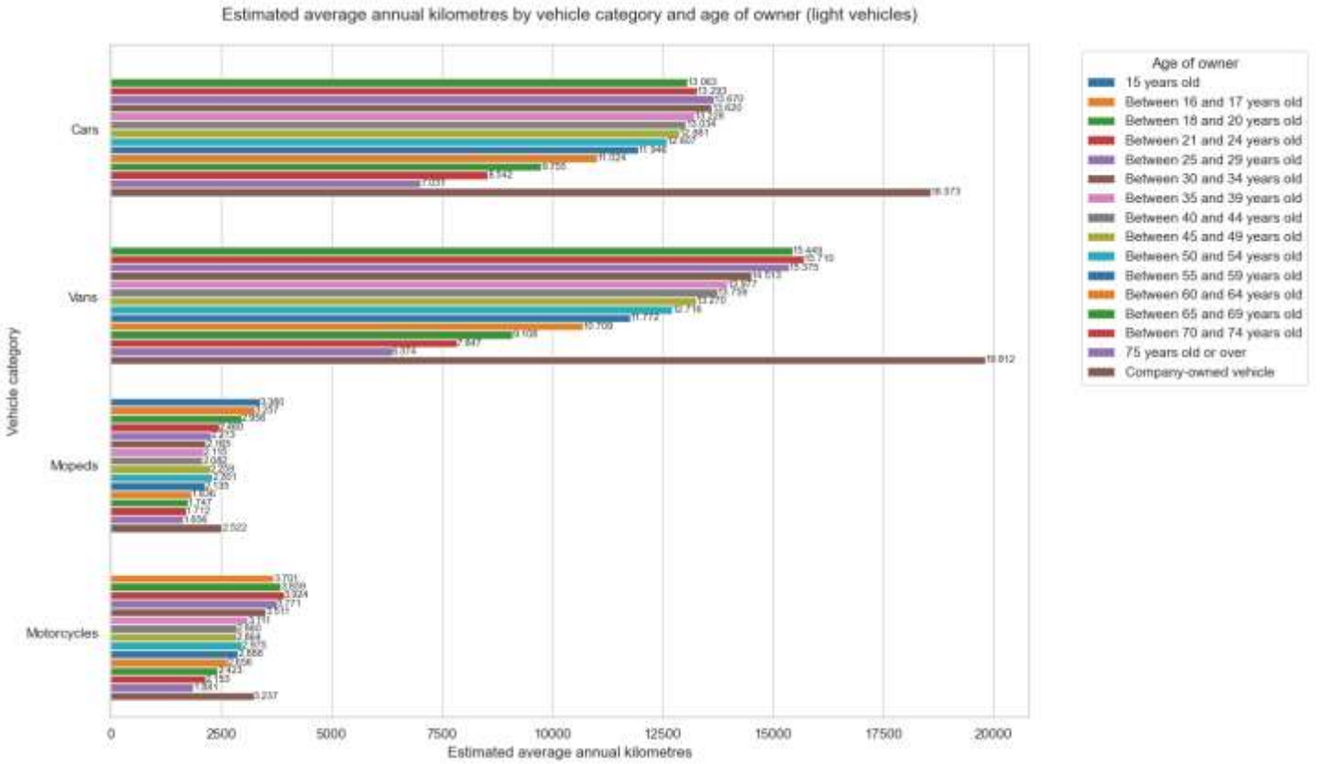

Figure 26. Estimated average kilometres by vehicle type and owner age for light vehicles (2024 fleet in circulation).

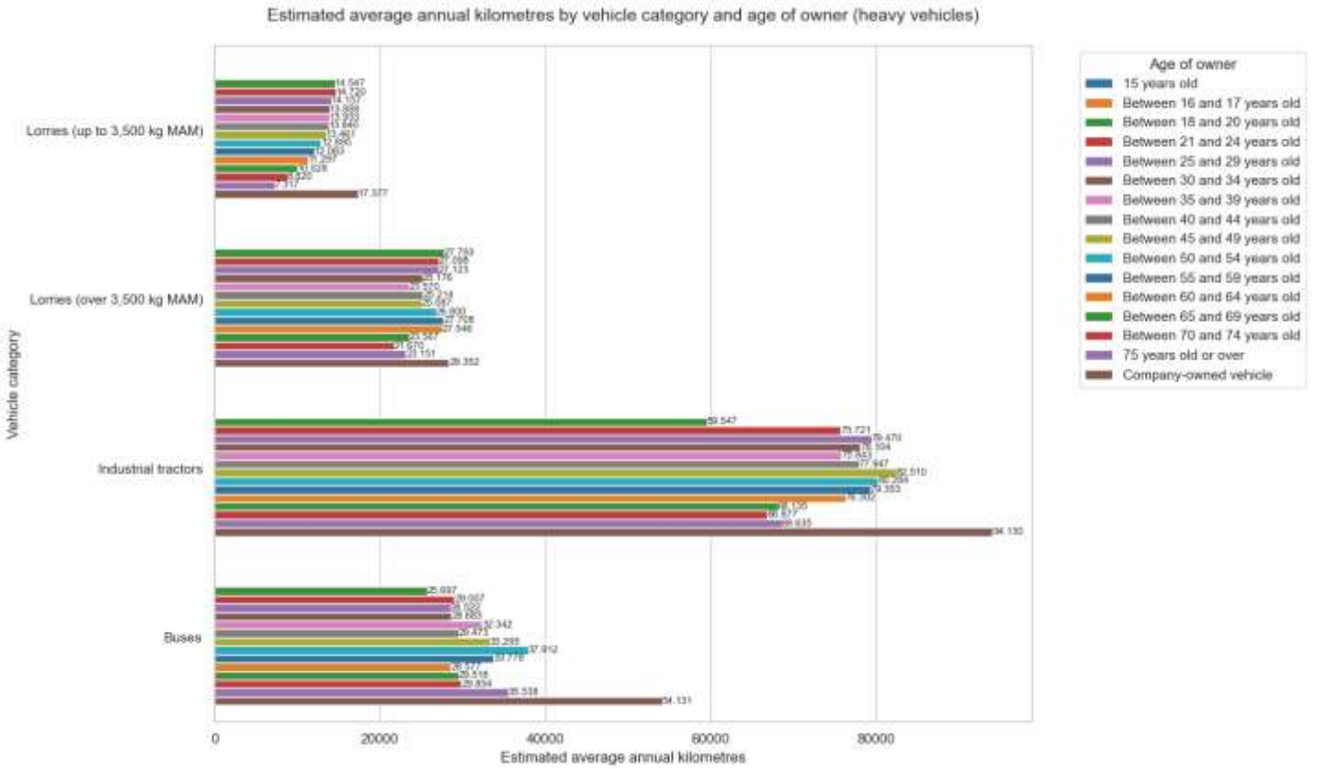

Figure 27. Estimated average kilometres by vehicle type and owner age for heavy vehicles (2024 fleet in circulation).

Table 29 shows the estimated average annual kilometres in 2024 for lorries according to maximum load. **Figure 28** presents these results in graph form.

Table 29. Estimated average kilometres by maximum load (lorries only) (2024 fleet in circulation).

Maximum load	Records	Sum of annual km	Average annual km
0-10T	2,067,433	29,620,544,693	14,327
10-20T	74,816	2,514,694,669	33,612
20-30T	48,032	1,919,081,227	39,954
30-40T	9,214	285,211,988	30,954
40+ T	1,980	58,169,972	29,379

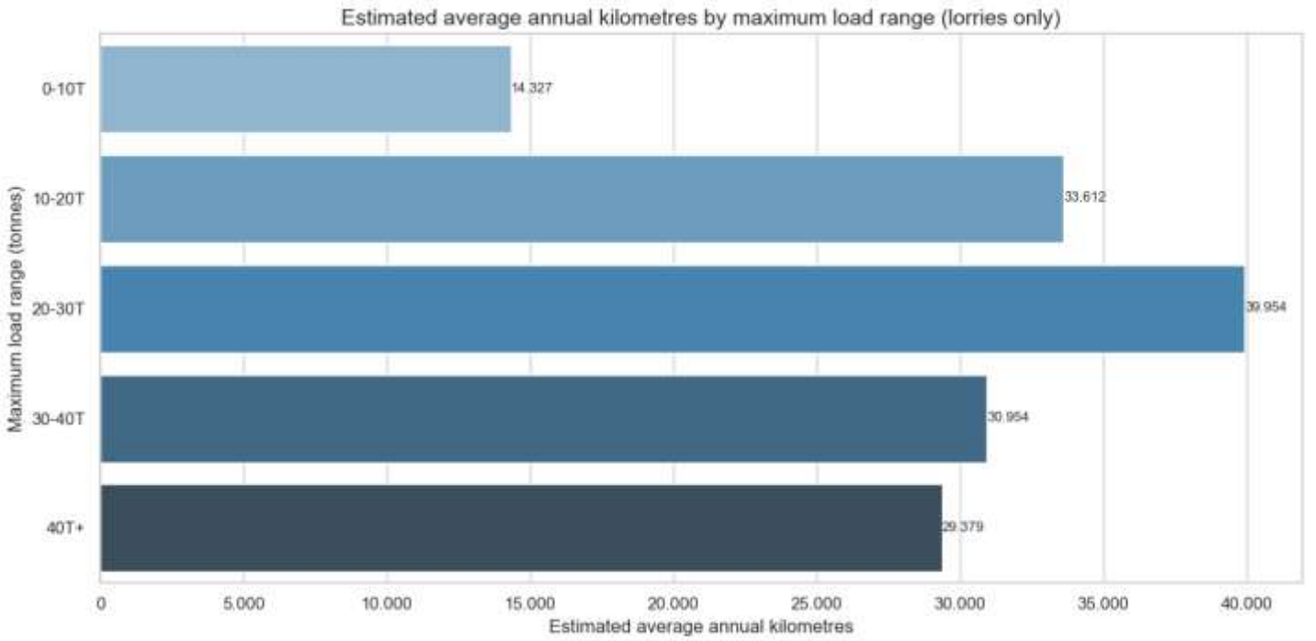

Figure 28. Estimated average kilometres by maximum load (lorries only) (2024 fleet in circulation).

11. SHORTCOMINGS IDENTIFIED

The analysis of the vehicle fleet and databases used reveals a number of limitations affecting both the coverage and consistency of the information available. These shortcomings have an impact on the quality of the estimates and the behaviour of the predictive model, particularly in certain categories of the fleet. The main observations include:

- Incomplete coverage of the active fleet, as only 55% of registration numbers in the 2024 fleet have usable ITV readings, meaning that a significant proportion of the fleet in circulation lacks directly recorded data.
- Highly imbalanced distribution among vehicle categories: cars account for nearly 80% of records, while categories such as mopeds, heavy lorries, buses and industrial tractors have a much smaller presence, making it harder to obtain equally robust models for all of the segments.
- Presence of invalid or incomplete geographical information, such as provinces listed as abroad or with null values. This limits the uniform use of contextual geographical variables such as population density and province size.
- Presence of atypical readings for both kilometres travelled and time elapsed between inspections. In some cases, the inspection intervals do not fall within the legally specified range for each category, thereby hindering reliable estimates of annual kilometres.
- Identification of extreme or physically improbable values for variables such as vehicle age or annual kilometres; these may be due to administrative errors, defective records or possible tampering with the odometer. These cases alter the natural distributions of the fleet.

Although these shortcomings do not prevent the data from being used, they necessitate the application of scrubbing processes, filters and validations in order to ensure the minimum level of consistency required for the analysis and subsequent modelling.

12. RECOMMENDATIONS AND NEXT STEPS

The next steps for the project will focus on optimising and fully automating the workflow, from the acquisition and processing of data onwards. The possibility of incorporating the automatic generation of final reports into that process will also be studied. This automation will provide an effective way to periodically update the model with new data, thereby reducing manual intervention and ensuring the reproducibility of the results.

In parallel, we will continue to work on improving the models' predictive performance by exploring and incorporating new explanatory variables. The candidate variables include vehicle make and model, which would enable us to identify different patterns of use among similar models; and insularity, which would enable us to reflect differences in driving patterns between mainland and insular territories (where the average distances travelled tend to be significantly shorter).

Additionally, we will study the implementation of specialised variables by vehicle category that, although not necessarily applicable to all types, may significantly improve the predictions for specific segments. For example, the maximum number of passengers on a bus may correlate to intensive use of the vehicle (urban services versus inter-urban bus lines).

In future iterations of the model, we will look at progressively incorporating additional ITV data for the years before and after 2024. This temporal expansion of the training dataset will not only increase the amount of data available, but also make it possible to identify time-related trends and variations in patterns of vehicle use over time, thereby improving the robustness and generalisation capacity of the predictive system.

13. ANNEX

This annex contains complementary and detailed information that supports the methodology and the results presented in the main body of the report.

Table 30. Estimated average kilometres by vehicle type and owner age (2024 fleet in circulation).

Vehicle category	Owner's age group	Records	Sum of annual km	Average annual km
Cars	18-20	71,983	940,337,806	13,063
Cars	21-24	319,531	4,247,553,830	13,293
Cars	25-29	828,667	11,328,232,029	13,670
Cars	30-34	1,306,721	17,797,565,804	13,620
Cars	35-39	1,687,108	22,317,596,842	13,228
Cars	40-44	2,178,458	28,394,502,377	13,034
Cars	45-49	2,705,253	34,847,300,282	12,881
Cars	50-54	2,755,883	34,743,593,480	12,607
Cars	55-59	2,633,540	31,459,223,752	11,946
Cars	60-64	2,279,348	25,126,733,640	11,024
Cars	65-69	1,784,623	17,408,472,061	9,755
Cars	70-74	1,286,369	10,987,609,012	8,542
Cars	75+	1,790,244	12,587,692,999	7,031
Cars	Vehicle in company's name	2,176,619	40,426,827,035	18,573
Vans	18-20	2,863	44,231,057	15,449
Vans	21-24	12,672	199,081,274	15,710
Vans	25-29	38,970	599,160,453	15,375
Vans	30-34	74,819	1,085,878,051	14,513
Vans	35-39	112,857	1,577,417,984	13,977
Vans	40-44	150,754	2,074,284,011	13,759
Vans	45-49	184,933	2,454,150,585	13,270
Vans	50-54	191,669	2,437,193,204	12,716
Vans	55-59	186,261	2,192,710,317	11,772
Vans	60-64	170,469	1,825,582,268	10,709
Vans	65-69	132,506	1,206,847,833	9,108
Vans	70-74	94,534	741,853,829	7,847
Vans	75+	139,154	886,965,504	6,374
Vans	Vehicle in company's name	886,259	17,558,416,313	19,812
Lorries ≤ 3500 kg	18-20	2,607	37,924,771	14,547
Lorries ≤ 3500 kg	21-24	9,741	143,388,134	14,720
Lorries ≤ 3500 kg	25-29	30,427	430,745,014	14,157
Lorries ≤ 3500 kg	30-34	60,417	839,055,048	13,888
Lorries ≤ 3500 kg	35-39	97,983	1,365,201,458	13,933
Lorries ≤ 3500 kg	40-44	139,823	1,935,156,423	13,840
Lorries ≤ 3500 kg	45-49	172,147	2,317,266,908	13,461
Lorries ≤ 3500 kg	50-54	170,573	2,199,490,504	12,895
Lorries ≤ 3500 kg	55-59	162,655	1,965,287,805	12,083

Lorries ≤ 3500 kg	60-64	143,600	1,622,191,179	11,297
Lorries ≤ 3500 kg	65-69	105,734	1,060,249,968	10,028
Lorries ≤ 3500 kg	70-74	71,232	628,295,605	8,820
Lorries ≤ 3500 kg	75+	106,638	780,321,578	7,317
Lorries ≤ 3500 kg	Vehicle in company's name	650,361	11,301,591,531	17,377
Mopeds	15	555	1,875,633	3,380
Mopeds	16-17	3,228	10,514,966	3,257
Mopeds	18-20	10,210	30,185,651	2,956
Mopeds	21-24	21,910	53,889,022	2,460
Mopeds	25-29	36,411	82,774,104	2,273
Mopeds	30-34	51,718	111,978,593	2,165
Mopeds	35-39	66,134	139,860,642	2,115
Mopeds	40-44	90,485	188,366,085	2,082
Mopeds	45-49	117,987	266,513,943	2,259
Mopeds	50-54	125,855	289,647,306	2,301
Mopeds	55-59	115,056	245,695,464	2,135
Mopeds	60-64	93,457	171,582,797	1,836
Mopeds	65-69	68,313	119,372,411	1,747
Mopeds	70-74	47,417	81,185,985	1,712
Mopeds	75+	83,876	138,873,203	1,656
Mopeds	Vehicle in company's name	56,457	142,374,586	2,522
Lorries > 3500 kg	18-20	76	2,112,235	27,793
Lorries > 3500 kg	21-24	382	10,351,615	27,098
Lorries > 3500 kg	25-29	1,097	29,754,124	27,123
Lorries > 3500 kg	30-34	2,250	56,646,051	25,176
Lorries > 3500 kg	35-39	4,073	96,000,868	23,570
Lorries > 3500 kg	40-44	6,134	154,664,277	25,214
Lorries > 3500 kg	45-49	8,204	205,814,700	25,087
Lorries > 3500 kg	50-54	8,764	234,871,844	26,800
Lorries > 3500 kg	55-59	9,260	256,572,996	27,708
Lorries > 3500 kg	60-64	7,935	218,577,120	27,546
Lorries > 3500 kg	65-69	3,423	80,670,305	23,567
Lorries > 3500 kg	70-74	1,700	36,838,464	21,670
Lorries > 3500 kg	75+	1,485	34,378,806	23,151
Lorries > 3500 kg	Vehicle in company's name	224,753	6,372,308,271	28,352
Motorcycles	16-17	2,414	8,933,914	3,701
Motorcycles	18-20	15,811	61,007,604	3,859
Motorcycles	21-24	69,043	270,953,637	3,924
Motorcycles	25-29	193,422	729,454,573	3,771
Motorcycles	30-34	303,520	1,065,612,174	3,511
Motorcycles	35-39	346,436	1,077,780,595	3,111
Motorcycles	40-44	412,364	1,179,395,298	2,860
Motorcycles	45-49	524,520	1,502,028,361	2,864
Motorcycles	50-54	548,989	1,632,989,174	2,975
Motorcycles	55-59	457,774	1,321,880,021	2,888
Motorcycles	60-64	339,949	903,047,873	2,656

Motorcycles	65-69	205,450	497,735,926	2,423
Motorcycles	70-74	100,297	215,944,641	2,153
Motorcycles	75+	80,976	152,313,089	1,881
Motorcycles	Vehicle in company's name	205,535	665,417,161	3,237
Industrial tractors	18-20	44	2,620,051	59,547
Industrial tractors	21-24	226	17,113,043	75,721
Industrial tractors	25-29	708	56,264,440	79,470
Industrial tractors	30-34	1,446	112,937,893	78,104
Industrial tractors	35-39	2,643	200,451,837	75,843
Industrial tractors	40-44	4,271	332,910,544	77,947
Industrial tractors	45-49	6,613	545,639,563	82,510
Industrial tractors	50-54	7,382	592,657,154	80,284
Industrial tractors	55-59	7,055	559,834,166	79,353
Industrial tractors	60-64	5,300	404,399,364	76,302
Industrial tractors	65-69	1,809	123,255,816	68,135
Industrial tractors	70-74	820	54,839,187	66,877
Industrial tractors	75+	722	49,554,721	68,635
Industrial tractors	Vehicle in company's name	193,161	18,182,186,929	94,130
Buses	18-20	2	51,393	25,697
Buses	21-24	7	203,046	29,007
Buses	25-29	23	656,010	28,522
Buses	30-34	85	2,438,047	28,683
Buses	35-39	131	4,236,740	32,342
Buses	40-44	242	7,132,526	29,473
Buses	45-49	304	10,121,533	33,295
Buses	50-54	260	9,856,997	37,912
Buses	55-59	289	9,761,712	33,778
Buses	60-64	188	5,372,525	28,577
Buses	65-69	160	4,722,896	29,518
Buses	70-74	303	9,045,641	29,854
Buses	75+	298	10,590,445	35,538
Buses	Vehicle in company's name	56,841	3,076,835,686	54,131

LIST OF TABLES

LIST OF FIGURES

BIBLIOGRAPHY

- Directorate-General of Traffic (2022). *Estudio de estimación de kilómetros anuales recorridos por el parque de vehículos español 2022* [*Study estimating the average number of kilometres travelled by Spain's vehicle fleet in 2022*]. Ministry of the Interior, Spain.
- Directorate-General of Traffic (2025). Dictionary of variables for the databases provided.
- National Statistics Institute (INE) (2024). *Padrón Municipal de Habitantes* [*Municipal Register of Inhabitants*]. Provincial data on population and density.
- XGBoost Python Library. XGBRegressor documentation. Available at: https://xgboost.readthedocs.io/en/stable/python/python_api.html#xgboost.XGBRegressor