



Émissions de CO2 par les véhicules en France pour l'année 2022

Objectifs :

- 1) Identifier les véhicules qui émettent le plus de CO2 pour identifier les caractéristiques techniques qui jouent un rôle important dans la pollution.
- 2) Prédire à l'avance cette pollution afin de pouvoir prévenir celle-ci dans le cas de l'apparition de nouveaux types de véhicules (nouvelles séries de voitures par exemple).

SOMMAIRE

Données	2
Contexte et périmètre	2
Jeu de données	3
Analyse	3
Champs inutilisées en 2022	3
Structure	3
Analyse	6
Valeurs nulles	6
Analyse des modalités des variables catégorielles	6
Après avoir supprimé ces colonnes, le jeu de données contient encore des valeurs nulles :	8
Remplacement des valeurs nulles	9
Suppression des lignes dupliquées	9
Le jeu de données comporte 1 333 323 lignes dupliquées. On les supprime pour avoir des véhicules uniques.	9
Jeu de données final	9
Visualisations	10
Variables qualitatives	10
Répartition du type de véhicule	10
Répartition du Fuel mode	10
Variables quantitatives	11
Heatmap	11
Consommation et émissions de CO2	12
Masse et émissions de CO2	12
Puissance et émissions de CO2	13
Preprocessing & Feature Engineering	17
Annexes	19
Fuel mode (Fm) :	19
Définition des champs du jeu de données 2022 :	20

Données

Pour ce projet, deux liens étaient proposés pour choisir le jeu de données.

Depuis le site de l'ADEME :

[Emissions de CO2 et de polluants des véhicules commercialisés en France - data.gouv.fr](https://data.gouv.fr)

Depuis le site de l'[European Environment Agency](https://www.eea.europa.eu/) (EEA)

<https://www.eea.europa.eu/data-and-maps/data/co2-cars-emission-20>

Les données provenant de l'EEA étant plus récentes et complètes, on choisit d'utiliser le jeu de données provenant de l'EEA.

Contexte et périmètre

Le règlement (UE) n° 2019/631 impose aux pays d'enregistrer des informations pour chaque nouvelle voiture particulière immatriculée sur leur territoire. Chaque année, chaque État membre soumet à la Commission toutes les informations relatives à ses nouvelles immatriculations.

Les objectifs d'émissions de CO2 pour l'ensemble du parc automobile de l'UE fixés dans le règlement sont les suivants :

2020 à 2024

- Voitures : 95 g CO2/km
- Camionnettes : 147 g CO2/km

2025 à 2029

- Voitures : 93,6 g CO2/km
- Camionnettes : 153,9 g CO2/km

2030 à 2034

- Voitures : 49,5 g CO2/km
- Camionnettes : 90,6 g CO2/km

À partir de **2035**, l'objectif de réduction des émissions de CO2 pour l'ensemble du parc automobile de l'UE, qu'il s'agisse de voitures ou de camionnettes, est de 100 %, soit **0 g de CO2/km**.

Jeu de données

Analyse

La base de données de l'**EEA** contient toutes les données de tous les pays européens depuis 2010. Pour le projet, seules les données de l'**année 2022** (qui sont les plus récentes) et uniquement pour **la France** seront utilisées afin de limiter la taille du jeu de données (366Mb vs 2Gb).

Les données sont constituées via le protocole WLTP (**W**orldwide harmonized **L**ight vehicles **T**est **P**rocedures) qui est la procédure d'essai mondiale harmonisée pour les véhicules légers.

Il est important de noter que depuis 2010, certaines colonnes ont été ajoutées et d'autres abandonnées en raison de l'apparition de nouvelles spécifications ou de l'obsolescence de certaines normes.

Champs inutilisées en 2022

MMS,
Enedc (g/km),
Ernedc (g/km),
Erwltp (g/km),
De,
Vf

Le jeu de données brut contient **37 colonnes** et **1 638 878 lignes**, mais seuls les champs données en [annexe](#) sont utilisés pour l'année 2022

Structure

#	Column	Non-Null Count	Dtype	#	Column	Non-Null Count	Dtype
0	Country	1638878 non-null	object	19	W (mm)	1638878 non-null	int64
1	VFN	1638878 non-null	object	20	At1 (mm)	1638878 non-null	int64
2	Mp	1570268 non-null	object	21	At2 (mm)	1638878 non-null	int64
3	Mh	1638878 non-null	object	22	Ft	1638878 non-null	object
4	Man	1638878 non-null	object	23	Fm	1638878 non-null	object
5	MMS	0 non-null	float64	24	ec (cm3)	1428317 non-null	float64
6	Tan	1638878 non-null	object	25	ep (KW)	1638878 non-null	int64
7	T	1638878 non-null	object	26	z (Wh/km)	339030 non-null	float64
8	Va	1638878 non-null	object	27	IT	1206951 non-null	object
9	Ve	1638878 non-null	object	28	Ernedc (g/km)	0 non-null	float64
10	Mk	1638878 non-null	object	29	Erwltp (g/km)	1206947 non-null	float64
11	Cn	1638878 non-null	object	30	De	0 non-null	float64
12	Ct	1638878 non-null	object	31	Vf	0 non-null	float64
13	Cr	1638878 non-null	object	32	Status	1638878 non-null	object
14	r	1638878 non-null	int64	33	year	1638878 non-null	int64
15	m (kg)	1638878 non-null	int64	34	Date of registration	1638878 non-null	object
16	Mt	1638878 non-null	int64	35	Fuel consumption	1428317 non-null	float64
17	Enedc (g/km)	297985 non-null	float64	36	Electric range (km)	339030 non-null	float64
18	Ewltp (g/km)	1638878 non-null	int64				

dtypes: float64(10), int64(9), object(18)
memory usage: 475.1+ MB

Émissions de CO2 par les véhicules

Liste explicative des différentes variables ([Details](#)) :

Nom Variable	Définition	Exemple de valeur / Plus utilisé
ID integer	Numéro d'identification unique des données contenues dans le registre national	
Country	Pays	France
VFN varchar(50)	Identifiant de la famille du véhicule (Vehicle family identification number.)	IP-DGY____EAT82552-VR3-0
Mp varchar(50)	Pool du constructeur (Manufacturer pooling)	STELLANTIS
Mh varchar(50)	Nom du constructeur au standard Européen (Manufacturer name EU standard denomination)	PSA
Man varchar(50)	Déclaration OEM du nom du fabricant (Manufacturer name OEM declaration) OEM: Original Equipment Manufacturer	PSA AUTOMOBILES SA
MMS varchar(125)	Nom du fabricant enregistré MS (Manufacturer name MS registry denomination)	Nan (plus utilisé, remplacé par Man)
Tan varchar(50)	Numéro du type d'homologation (Type approval number)	e9*2018/858*11066*03
T varchar(25)	Type	N
Va varchar(25)	Variant	D
Ve varchar(35)	Version	DGYP-A1C000
Mk varchar(25)	Marque (Make)	CITROEN
Cn varchar(50)	Nom commercial (Commercial name)	C5 X
Ct varchar(5)	Catégorie du type de véhicule immatriculé (Category of the vehicle type approved)	M1
Cr varchar(5)	Catégorie du véhicule immatriculé (Category of the vehicle registered)	M1
r integer	Total des nouvelles inscriptions (Total new registrations)	1
m (kg) integer	Masse véhicule chargé (Mass in running order Completed/complete vehicle)	1797
Mt	Masse harmonisée WLTP (WLTP test mass)	1888
Enedc (g/km)	Réduction des émissions grâce à des technologies innovantes (Emissions reduction through innovative technologies)	30.0 (plus utilisé depuis 2019, remplacé par le Ewltip)

Émissions de CO2 par les véhicules

Ewltp (g/km)	Les émissions spécifiques de CO2 (WLTP) (Emissions reduction through innovative technologies (WLTP))	30
W (mm) varchar(35)	Empattement (Wheel Base)	2785
At1 (mm) integer	Largeur de l'essieu directeur (Axle width steering axle)	1600
At2 (mm) integer	Largeur de l'essieu (Axle width other axle)	1605
Ft varchar(25)	Type de carburant (Fuel type)	PETROL/ELECTRIC
Fm varchar(1)	Mode de carburant (Fuel mode)	P
ec (cm3) integer	Cylindrée (Engine capacity)	1598.0
ep (Kw) integer	Puissance du moteur (Engine power.)	132
z (Wh/km) integer	Consommation électrique (Electric energy consumption)	159.0
IT varchar(25)	Technologie innovante ou groupe de technologies innovantes (Innovative technology or group of innovative technologies)	e2 28 29
Ernedc (g/km) float	Emissions spécifiques de CO2 (Specific CO2 Emission. Deprecated value, only relevant for data until 2016)	NaN
Erwltp (g/km) float	Réduction d'émissions spécifiques de CO2 par l'utilisation de technologies spécifiques (Emissions reduction through innovative technologies. Deprecated value, only relevant for data until 2016)	NaN
De	-	NaN
Vf	-	NaN
Status varchar(1)	P: donnée provisoire, F: donnée définitive (P = Provisional data, F = Final data.)	P
Year integer	Année d'enregistrement (Reporting year)	2022
Date of registration	Date d'enregistrement	2022-12-30
Fuel consumption float	Consommation	1.3
Electric range (km)	Autonomie électrique	59

On définit **Ewltp (g/km)** comme étant la **variable cible de notre projet**.

Émissions de CO2 par les véhicules

Analyse

Valeurs nulles

	index	NA	% NA		index	NA	% NA		index	NA	% NA
0	Country	0	0.00	12	Ct	0	0.00	25	ep (KW)	0	0.00
1	VFN	0	0.00	13	Cr	0	0.00	26	z (Wh/km)	1299848	79.31
2	Mp	68610	4.19	14	r	0	0.00	27	IT	431927	26.36
3	Mh	0	0.00	15	m (kg)	0	0.00	28	Ernedc (g/km)	1638878	100.00
4	Man	0	0.00	16	Mt	0	0.00	29	Erwltp (g/km)	431931	26.36
5	MMS	1638878	100.00	17	Enedc (g/km)	1340893	81.82	30	De	1638878	100.00
6	Tan	0	0.00	18	Ewltp (g/km)	0	0.00	31	Vf	1638878	100.00
7	T	0	0.00	19	W (mm)	0	0.00	32	Status	0	0.00
8	Va	0	0.00	20	At1 (mm)	0	0.00	33	year	0	0.00
9	Ve	0	0.00	21	At2 (mm)	0	0.00	34	Date of registration	0	0.00
10	Mk	0	0.00	22	Ft	0	0.00	35	Fuel consumption	210561	12.85
11	Cn	0	0.00	23	Fm	0	0.00	36	Electric range (km)	1299848	79.31
				24	ec (cm3)	210561	12.85				

On remarque que certaines colonnes sont vides à 100%, ce qui correspond aux valeurs qui ne sont plus utilisées pour le fichier 2022 (**MMS**, **Ernedc (g/km)**, **De** et **Vf**)

Quant aux 2 colonnes **Enedc (g/km)** et **Erwltp (g/km)** pour lesquelles il manque 82% et 26% des valeurs, cela correspond à des champs qui ne sont plus complétés depuis 2019.
([MS Guide 22 – Page 19](#))

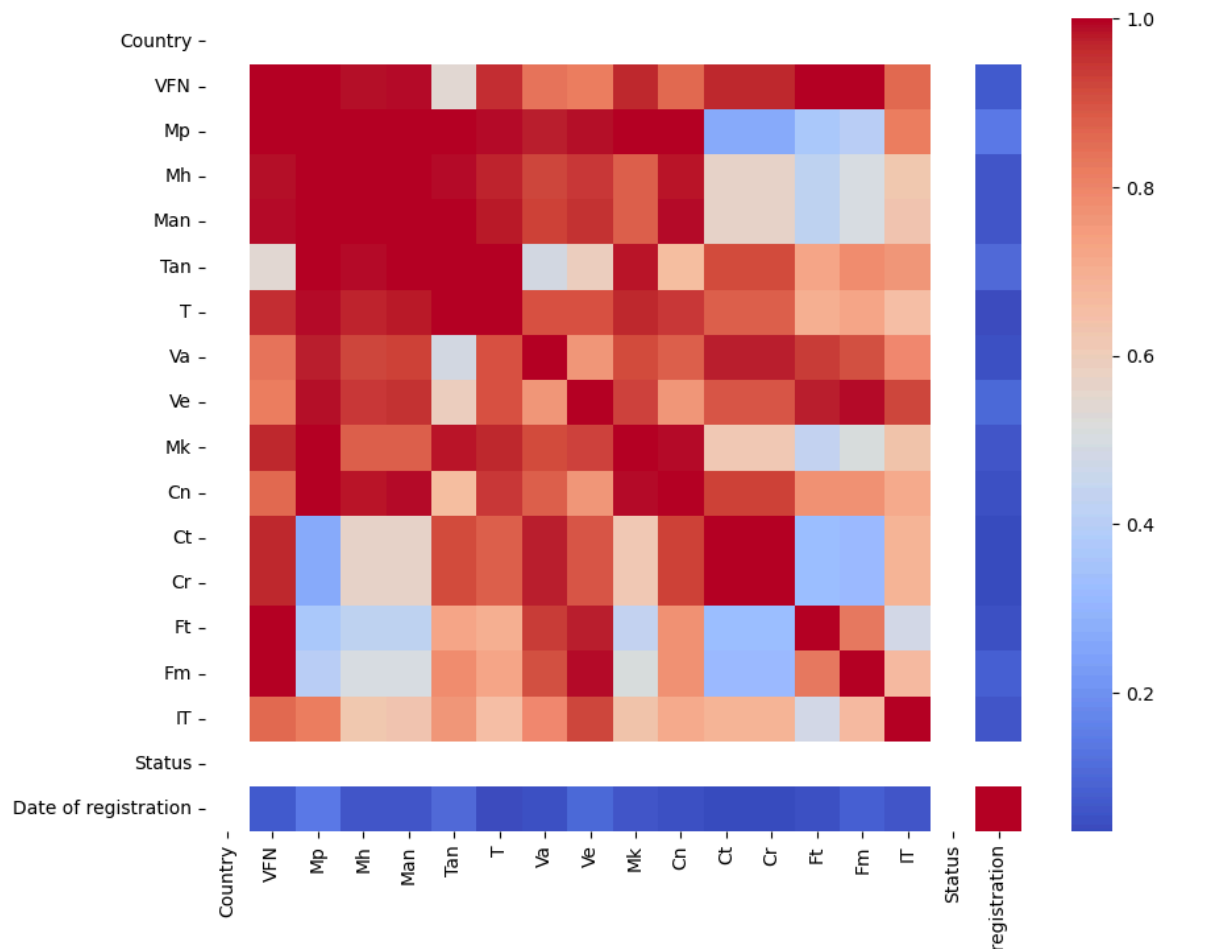
Analyse des modalités des variables catégorielles

Modalités:

0	Country	1	9	Mk	67
1	VFN	2421	10	Cn	932
2	Mp	10	11	Ct	2
3	Mh	69	12	Cr	2
4	Man	68	13	Ft	9
5	Tan	1935	14	Fm	6
6	T	360	15	IT	77
7	Va	1748	16	Status	1
8	Ve	7092	17	Date of registration	356

Émissions de CO2 par les véhicules

Heatmap de la matrice de corrélation des variables catégorielles avec le test du khi-deux :



Les modalités des variables 'Country' et 'Status', étant uniques et corrélées avec aucune variables catégorielles, on supprime ces variables.

D'après la heatmap, on voit qu'il y a une forte corrélation entre les variables VFN, Mp, Mh, Man et T.

En faisant un test ANOVA pour savoir si VFN a une influence sur Ewltip (g/km) on obtient :

	sum_sq	df	F	PR(>F)
VFN	4.007153e+07	1219.0	11625.847402	0.0
Residual	4.288801e+04	15168.0	NaN	NaN

Par ailleurs, avec le test de Fisher, on obtient une p-value = 0.0 (< 5%).

On en conclut que la variable VFN n'a pas d'impact sur les émissions de CO2 et par extension, les variables Mp, Mh et Man non plus ; on peut donc supprimer ces colonnes.

Émissions de CO2 par les véhicules

En faisant d'autres tests ANOVA, on peut aussi supprimer les variables IT.
([MS Guide – page 13](#))

Les variables 'T', 'Va' et 'Ve' étant trois paramètres d'identification d'un même véhicule selon les différents pays européens, elles n'apportent pas d'informations techniques. On les supprime.

Concernant les variables 'Ct' et 'Cr', elles sont identiques. Ainsi, il ne semble pas pertinent de conserver les deux variables. Nous choisissons de garder 'Ct' car elle correspond à la catégorie officielle/approuvée contrairement à 'Cr' qui ne contient que la catégorie enregistrée par le constructeur.

Après avoir supprimé ces colonnes, le jeu de données contient encore des valeurs nulles :

	index	NA	% NA				
0	Cr	0	0.00	7	Ft	0	0.00
1	m (kg)	0	0.00	8	Fm	0	0.00
2	Mt	0	0.00	9	ec (cm3)	210561	12.85
3	Ewltp	0	0.00	10	ep (KW)	0	0.00
4	W (mm)	0	0.00	11	z (Wh/km)	1299848	79.31
5	At1 (mm)	0	0.00	12	Fuel consumption	210561	12.85
6	At2 (mm)	0	0.00	13	Electric range (km)	1299848	79.31

La variable 'Ft' représente le type d'énergie du véhicule avec les valeurs suivantes :

'PETROL', 'DIESEL', '**ELECTRIC**', 'PETROL/ELECTRIC', 'LPG', 'E85', 'DIESEL/ELECTRIC', '**HYDROGEN**' et 'NG'

L'étude concernant les émissions de CO2, les voitures de type **HYDROGEN** ou **ELECTRIC** qui n'en rejettent pas par définition ne sont pas pertinentes.

On décide donc de supprimer les lignes contenant l'une de ces deux valeurs.

Valeurs manquantes après suppressions des colonnes non pertinentes et des véhicules non thermiques :

	index	NA	% NA				
0	Cr	0	0.00	7	Ft	0	0.00
1	m (kg)	0	0.00	8	Fm	0	0.00
2	Mt	0	0.00	9	ec (cm3)	0	0.00
3	Ewltp	0	0.00	10	ep (KW)	0	0.00
4	W (mm)	0	0.00	11	z (Wh/km)	1299655	90.99
5	At1 (mm)	0	0.00	12	Fuel consumption	0	0.00
6	At2 (mm)	0	0.00	13	Electric range (km)	1299655	90.99

Émissions de CO2 par les véhicules

Remplacement des valeurs nulles

Z (Wh/km) : Electric energy consumption

Electric range (km) : Autonomie électrique

On remarque dans notre tableau que les variables 'z' et 'Electric range (km)' ont chacune un taux de valeurs manquantes qui a augmenté puisqu'elles sont dépendantes des valeurs Hydrogen et Electric.

Le moteur des véhicules thermiques ne consommant pas d'énergie électrique, il semble cohérent de remplacer les valeurs manquantes de notre variable 'z' par 0.

Pour la variable 'Electric range' (autonomie électrique), les véhicules thermiques n'ayant pas de réserve en électricité, il semble correct de remplir les valeurs nulles par 0.

Ainsi, à ce stade, il n'y a plus de valeurs nulles dans le jeu de données.

Suppression des lignes dupliquées

Le jeu de données comporte 1 333 323 lignes dupliquées. On les supprime pour avoir des véhicules uniques.

Nombre de lignes dupliquées : 1 333 323

Lignes restantes : **94 994**

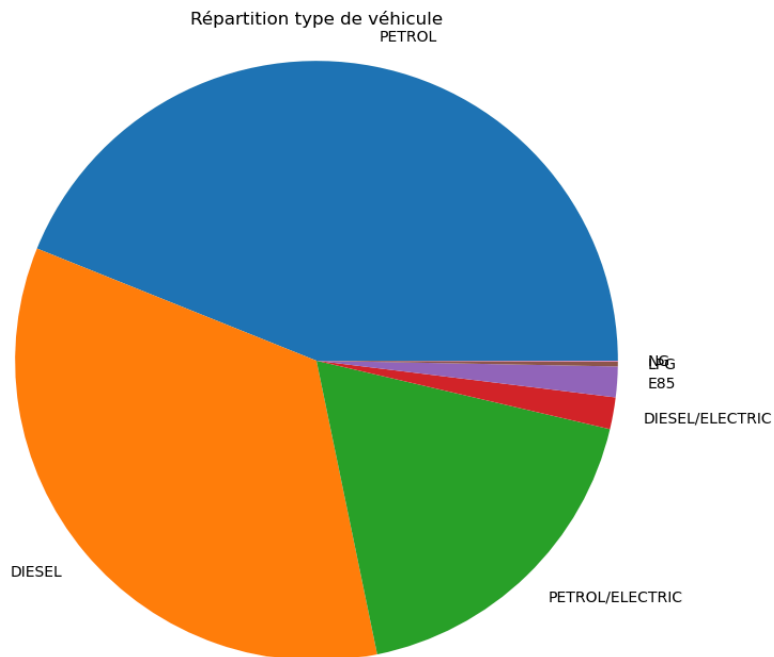
Jeu de données final

```
Index: 94994 entries, 77798939 to 77955102
Data columns (total 14 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Cr                   94994 non-null  object
1   m (kg)               94994 non-null  int64
2   Mt                   94994 non-null  int64
3   Ewlt                94994 non-null  int64
4   W (mm)               94994 non-null  int64
5   At1 (mm)             94994 non-null  int64
6   At2 (mm)             94994 non-null  int64
7   Ft                   94994 non-null  object
8   Fm                   94994 non-null  object
9   ec (cm3)             94994 non-null  float64
10  ep (KW)              94994 non-null  int64
11  z (Wh/km)            94994 non-null  float64
12  Fuel consumption     94994 non-null  float64
13  Electric range (km)  94994 non-null  float64
dtypes: float64(4), int64(7), object(3)
```

Visualisations

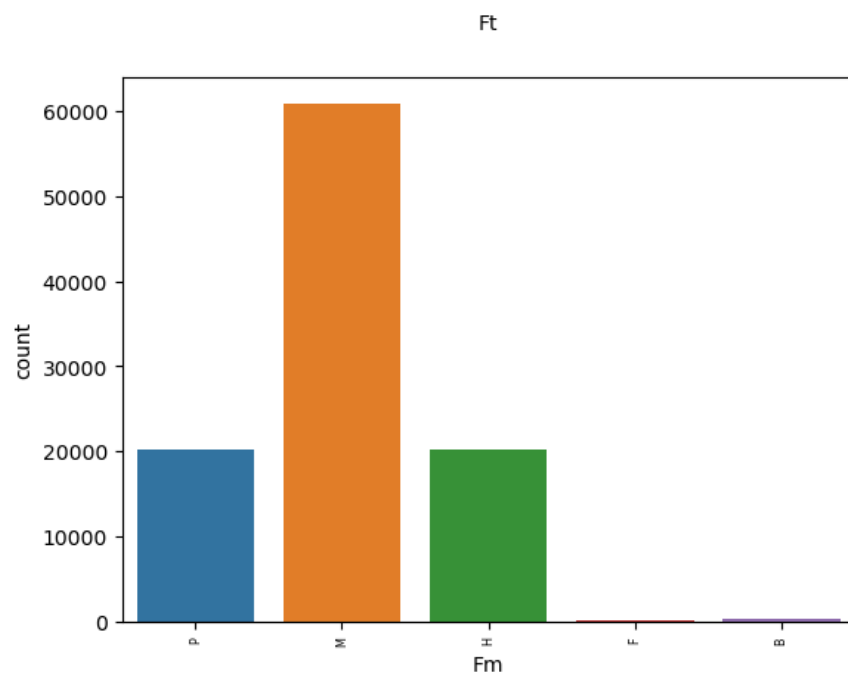
Variables qualitatives

Répartition du type de véhicule



On voit que les 3 catégories de véhicules dominantes dans notre dataset sont les types Essence, Diesel et Hybride.

Répartition du Fuel mode



Émissions de CO2 par les véhicules

Les 3 [modes](#) les plus représentatifs sont :

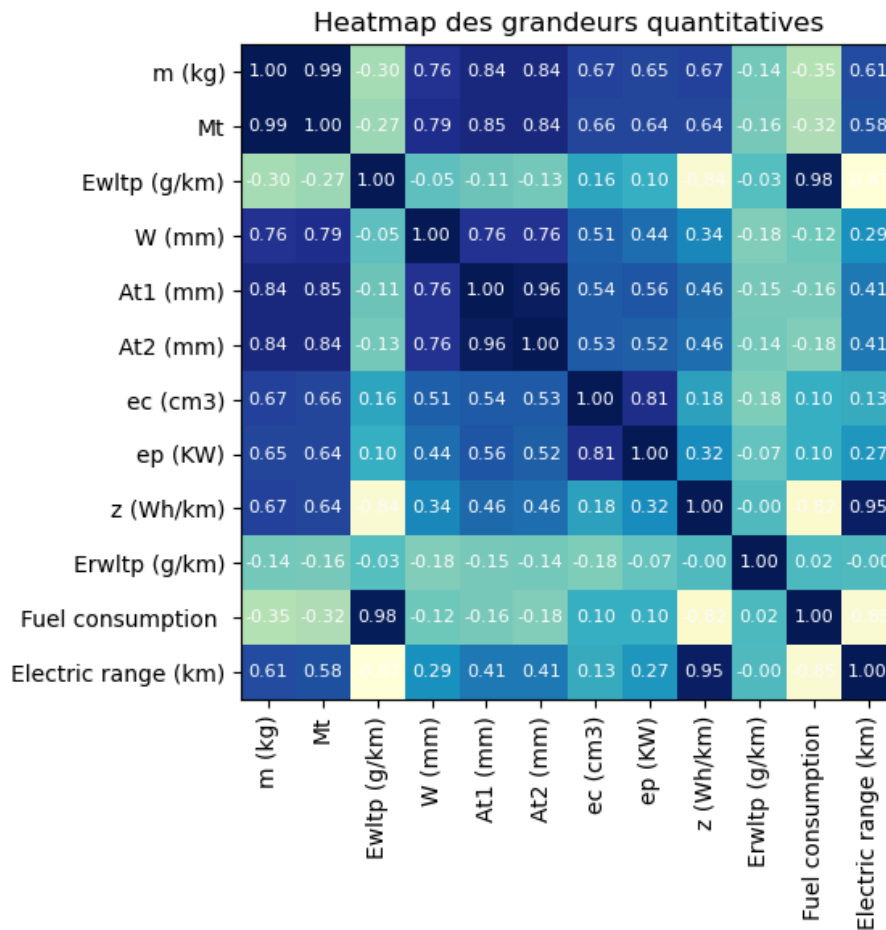
H : Véhicules non électriques

P : Véhicules hybrides acceptant la charge externe

M : Véhicules hybrides autonomes (pas de charge à la borne)

Variables quantitatives

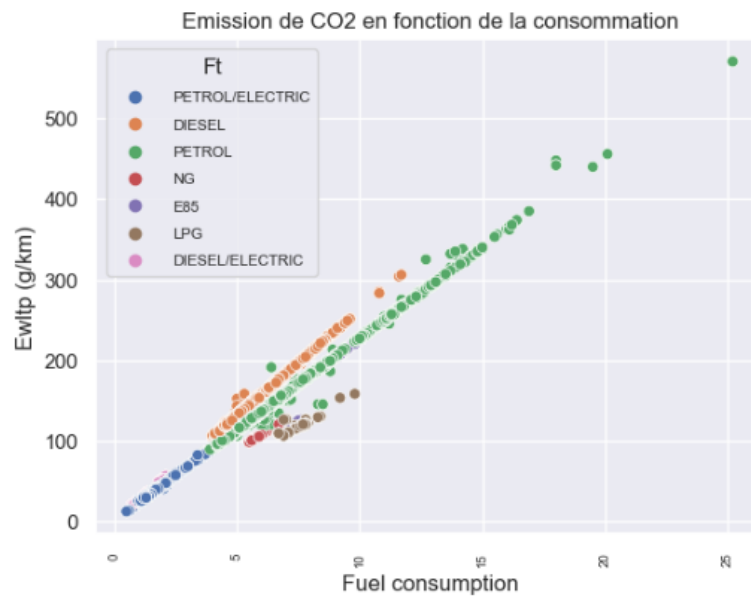
Heatmap



Concernant notre variable cible **Ewltip (g/km)**, on observe immédiatement une très forte corrélation avec la consommation du véhicule, ce qui paraît logique.

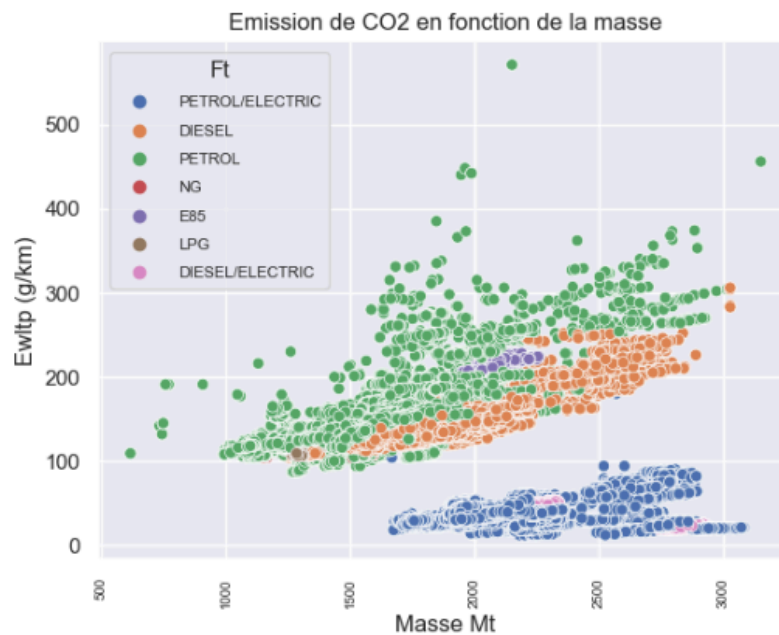
Émissions de CO₂ par les véhicules

Consommation et émissions de CO₂



Le graphique ci-dessus confirme bien qu'il y a une forte corrélation entre la consommation de carburant et les émissions de CO₂

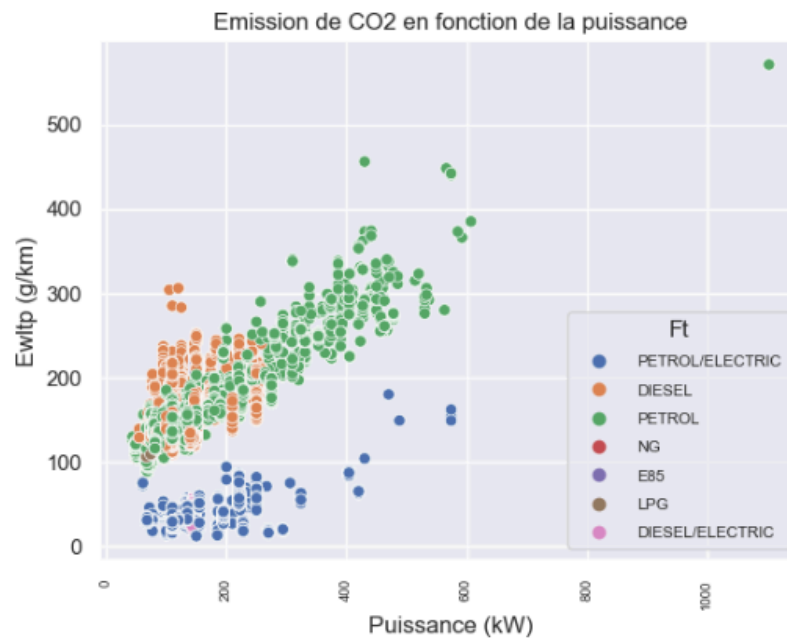
Masse et émissions de CO₂



Il semble également y avoir une corrélation entre la masse et les émissions de CO₂, en particulier pour les moteurs non-hybrides.

Émissions de CO2 par les véhicules

Puissance et émissions de CO2



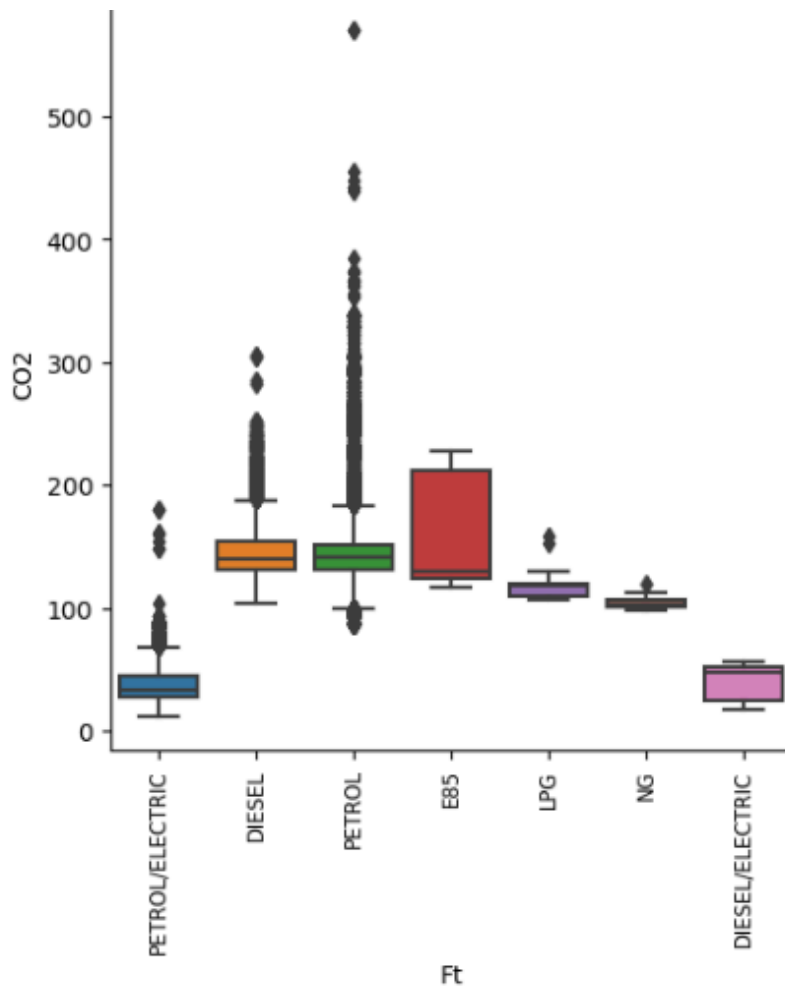
On remarque ici que les véhicules consommant du carburant de type PETROL/ELECTRIC se distinguent des autres concernant la corrélation entre la puissance du moteur et les émissions de CO2.

Lorsque la puissance du moteur augmente, l'émission de CO2 augmente également. La réalisation d'un test de Pearson confirme cela puisque la p-value est inférieure à 5%. Il y a donc une forte corrélation entre la puissance du moteur et l'émission de CO2.

```
PearsonRResult(statistic=0.7853731478970354, pvalue=0.0)
```

Émissions de CO2 par les véhicules

Type de carburant :

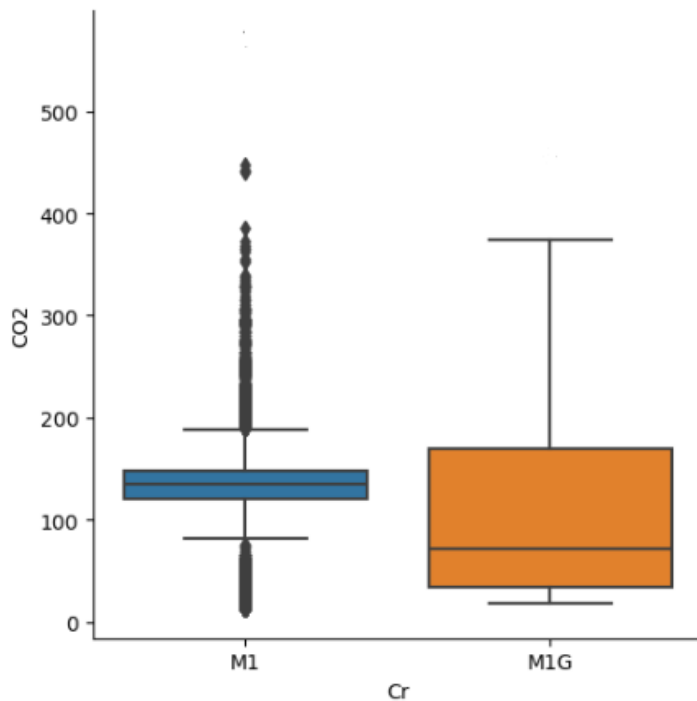


Dans ce graphique nous remarquons que les véhicules qui utilisent du diesel et de l'essence émettent plus de CO2 que les autres types de carburant. En faisant un test ANOVA, on obtient une p-value inférieur à 5%. On rejette donc l'hypothèse H_0 , ce qui signifie que le type de carburant et la quantité d'émission du CO2 ne sont pas indépendants.

	df	sum_sq	mean_sq	F	PR(>F)
Ft	4.0	6.669239e+05	166730.964008	261.417969	1.659581e-223
Residual	74844.0	4.773510e+07	637.794582	NaN	NaN

Émissions de CO2 par les véhicules

Catégorie du véhicule :



On remarque sur le graphique que les véhicules de type M1G, émettent moins de CO2 que les véhicules de type M1. En réalisant un test ANOVA, on obtient une p-value inférieur à 5%.

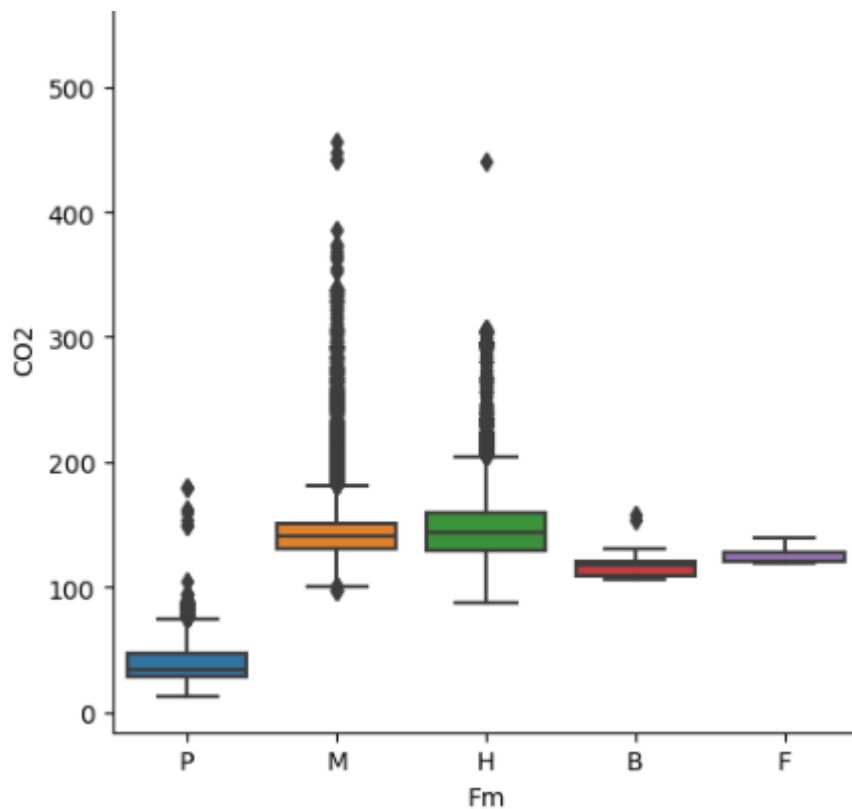
	df	sum_sq	mean_sq	F	PR(>F)
Cr	1.0	6.676778e+06	6.676778e+06	11976.845289	0.0
Residual	74847.0	4.172524e+07	5.574738e+02	NaN	NaN

On en conclut donc que la catégorie du véhicule et la quantité d'émission du CO2 ne sont pas indépendantes.

Mode carburant :

Sur le graphique, on constate que le mode de carburant H (véhicules hybrides) émet plus de CO2 que le mode M (véhicules monocarburant). Mais cet écart n'est pas vraiment significatif. En revanche, ces deux modes émettent plus de CO2 que les autres modes.

Émissions de CO₂ par les véhicules

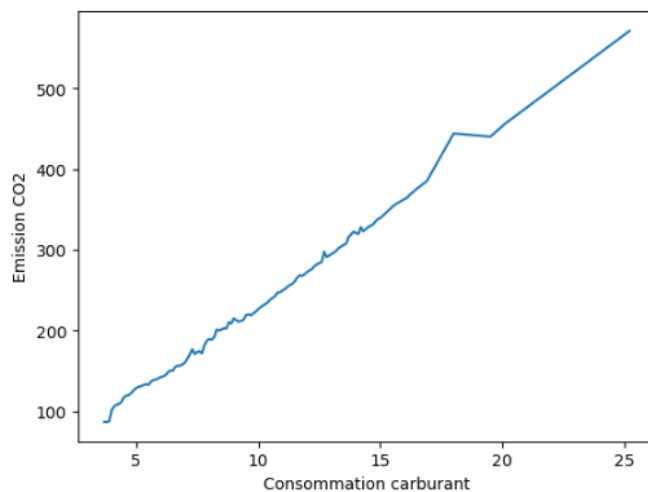


On réalise ici également un test ANOVA et on en conclut que le mode de carburant et la quantité d'émission de CO₂ ne sont pas indépendants puisqu'on obtient une p-value inférieur à 5%.

	df	sum_sq	mean_sq	F	PR(>F)
Fm	3.0	3.635046e+05	121168.211967	188.782573	5.750819e-122
Residual	74845.0	4.803852e+07	641.840028	NaN	NaN

Émissions de CO2 par les véhicules

Emission CO2 en fonction de la consommation du carburant :



Ici, on croise la moyenne d'émissions de CO2 et la quantité de consommation de carburant. On obtient une courbe linéaire croissante. On peut donc interpréter ce graphique en disant que plus une voiture consomme du carburant, plus elle émet de CO2.

On réalise un test de Pearson pour confirmer cette hypothèse :

```
PearsonRResult(statistic=0.9164220148984297, pvalue=0.0)
```

On constate donc qu'il y a bien une corrélation entre la consommation du carburant et l'émission du CO2 puisqu'on obtient une p-value inférieure à 5%.

Preprocessing & Feature Engineering

Le jeu de données est maintenant débarrassé de toutes les valeurs nulles et des variables non pertinentes. On encode dans un premier temps les variables catégorielles pour n'avoir que des données numériques ce qui donne un nouveau jeu de données comportant **94 994 lignes** et **25 colonnes**.

Pour une meilleure lisibilité, on renomme les variables et on sauvegarde le jeu de données pour l'étape de Machine Learning.

```
Index: 94994 entries, 77798939 to 77955102
Data columns (total 25 columns):
#   Column              Non-Null Count  Dtype   12  Cr_M1G              94994 non-null  int32
---  ---                -
0   m                   94994 non-null  int64   13  Ft_DIESEL           94994 non-null  int32
1   Mt                  94994 non-null  int64   14  Ft_DIESEL/ELECTRIC  94994 non-null  int32
2   Ewltip              94994 non-null  int64   15  Ft_E85              94994 non-null  int32
3   W                   94994 non-null  int64   16  Ft_LPG              94994 non-null  int32
4   At1                 94994 non-null  int64   17  Ft_NG               94994 non-null  int32
5   At2                 94994 non-null  int64   18  Ft_PETROL           94994 non-null  int32
6   ec                  94994 non-null  float64  19  Ft_PETROL/ELECTRIC  94994 non-null  int32
7   ep                  94994 non-null  int64   20  Fm_B                94994 non-null  int32
8   z                   94994 non-null  float64  21  Fm_F                94994 non-null  int32
9   Fuel_consumption    94994 non-null  float64  22  Fm_H                94994 non-null  int32
10  Electric_range       94994 non-null  float64  23  Fm_M                94994 non-null  int32
11  Cr_M1               94994 non-null  int32   24  Fm_P                94994 non-null  int32
dtypes: float64(4), int32(14), int64(7)
memory usage: 13.8 MB
```

Après réflexion, on n'a pas besoin de créer de nouvelles variables et on finalise par la sauvegarde du jeu de données nettoyé.

Annexes

Fuel mode (Fm) :

"**M**" for mono-fuel vehicles, i.e. vehicles able to run on only one fuel, either petrol, diesel, LPG, natural gas (NG) or hydrogen. The latter category also covers Fuel Cell electric vehicles, i.e. vehicles equipped with a powertrain containing exclusively fuel cell(s) and electric machine(s) as propulsion energy converter(s).

"**B**" for bi-fuel vehicles, i.e. vehicles with two separate fuel storage systems, which are designed to run primarily on only one fuel at a time. This covers vehicles that can run on petrol and either LPG, NG/biomethane or hydrogen.

"**F**" for flex-fuel vehicles, i.e. vehicles with one fuel storage system that can run on different mixtures of two or more fuels; this concerns more specifically 'flex fuel ethanol vehicles', which can run on petrol or a mixture of petrol and ethanol up to an 85 per cent ethanol blend (E85);

"**E**" for battery electric vehicles (BEV), i.e. "pure" electric vehicles (NOT hybrid vehicles). These vehicles can be identified using section 23 of the certificate of conformity.

"**P**" for Off vehicle charging hybrid electric vehicles (OVC-HEV), i.e. plug-in hybrid vehicles. These vehicles can be identified using section 23.1 of the certificate of conformity. Their weighted average CO₂ values are specified in section 49.1. (NEDC) and section 49.4 (WLTP) of the certificate of conformity.

"**H**" for Not-Off vehicle charging hybrid electric vehicles (NOVC-HEV). These vehicles can be identified using section 23.1 of the certificate of conformity. They cannot take electric energy from external sources and are only fuelled with one of fuel types specified in section 26 of the CoC. The CO₂ values for that fuel shall be reported.

Émissions de CO2 par les véhicules

Définition des champs du jeu de données 2022 :

Name	Définition
ID integer - Cardinality: 1..1	Identification number.
MS varchar(2) - Cardinality: 0..1	Member state.
Mp varchar(50) - Cardinality: 0..1	Manufacturer pooling.
VFN varchar(25) - Cardinality: 0..1	Vehicle family identification number.
Mh varchar(50) - Cardinality: 0..1	Manufacturer name EU standard denomination .
Man varchar(50) - Cardinality: 0..1	Manufacturer name OEM declaration.
MMS varchar(125) - Cardinality: 0..1	Manufacturer name MS registry denomination .
TAN varchar(50) - Cardinality: 0..1	Type approval number.
T varchar(25) - Cardinality: 0..1	Type.
Va varchar(25) - Cardinality: 0..1	Variant.
Ve varchar(35) - Cardinality: 0..1	Version.
Mk varchar(25) - Cardinality: 0..1	Make.
Cn varchar(50) - Cardinality: 0..1	Commercial name.
Ct varchar(5) - Cardinality: 0..1	Category of the vehicle type approved.
Cr varchar(5) - Cardinality: 0..1	Category of the vehicle registered.
M (kg) integer - Cardinality: 0..1	Mass in running order Completed/complete vehicle .
Mt integer - Cardinality: 0..1	WLTP test mass.
Ewltip (g/km) integer - Cardinality: 0..1	Specific CO2 Emissions (WLTP).

Émissions de CO2 par les véhicules

W (mm) integer – Cardinality: 0..1	Wheel Base.
At1 (mm) integer – Cardinality: 0..1	Axle width steering axle.
At2 (mm) integer – Cardinality: 0..1	Axle width other axle.
Ft varchar(25) – Cardinality: 0..1	Fuel type.
Fm varchar(1) – Cardinality: 0..1	Fuel mode.
Ec (cm3) integer – Cardinality: 0..1	Engine capacity.
Ep (KW) integer – Cardinality: 0..1	Engine power.
Z (Wh/km) integer – Cardinality: 0..1	Electric energy consumption.
IT varchar(25) – Cardinality: 0..1	Innovative technology or group of innovative technologies.
Erwltp (g/km) float – Cardinality: 0..1	Emissions reduction through innovative technologies (WLTP).
R integer – Cardinality: 0..1	Total new registrations.
Year integer – Cardinality: 0..1	Reporting year.
Status varchar(1) – Cardinality: 0..1	P = Provisional data, F = Final data.
E (g/km) float – Cardinality: 0..1	Specific CO2 Emission. Deprecated value, only relevant for data until 2016.
Er (g/km) float – Cardinality: 0..1	Emissions reduction through innovative technologies. Deprecated value, only relevant for data until 2016.
Zr integer – Cardinality: 0..1	Electric range.
Dr date – Cardinality: 0..1	Registration date.
Fc float – Cardinality: 0..1	Fuel consumption.