

Rapport Data Science

Émissions de CO2 et de polluants des véhicules commercialisés en France



Efrei, ING3-APP















Réalisé par :

DASSE Aurélien
AZIB Yasmine
DENIS Alan

Encadré par :

Mme HARZI Marwa

Table des matières

1. Introduction 	3
2. Étapes suivies 	3
2.1. Importation des données 	3
2.2. Nettoyage des données 	4
2.3. Chargement des données dans Desktop 	6
2.3.1. Model View 	6
2.3.2. Table View 	7
2.4. Visualisations des données 	7
2.4.1. Émission en fonction des caractéristiques 	7
2.4.2. Caractéristiques des marques 	8
2.4.3. Émission au fil du temps 	8
2.4.4. KPI 	9
2.4.5. Pour aller plus loin 	11
3. Conclusion 	12

1. Introduction

Suite à la sélection d'un jeu de données public sur le site public.opendatasoft.com, nous avons suivi plusieurs étapes afin de représenter aux mieux les informations résident dans ce document. Le nom exact du jeu est “**Emissions de CO2 et de polluants des véhicules commercialisés en France**” et traitera donc des émissions de CO2 des véhicules.

Au cours de ce rapport seront détaillées les étapes de réflexion et procédés utilisés afin d'obtenir un résultat le plus parlant possible. Une phase de réflexion nous a permis d'isoler au préalable les traitements que nous allons faire sur les données. Ensuite, nous utiliserons [PowerQuery](#) pour nettoyer et filtrer les données avant leur exploitation. [PowerBI Desktop](#) sera utilisé pour l'exploitation et la visualisation des données en produisant un rapport graphique, utilisant toutes sortes de graphiques, [BarChart](#), [PieChart](#) ou simplement des [ligneCharts](#).

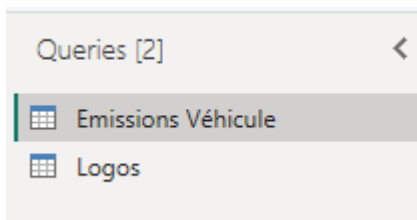
2. Étapes suivies

2.1. Importation des données

Nous avons ouvert l'outil [Power Query](#) et importé deux nouvelles ressources :

- **Excel workbook** contenant le dataset téléchargé depuis l'API
- **Text/CSV** contenant une table que nous avons créé pour renseigner le logo de chaque marque de véhicule

Nous avons ensuite renommé les deux tables importées à : **Émissions Véhicules** et **Logos**



2.2. Nettoyage des données

Dans *Power Query*, nous avons procédé au nettoyage des deux tables pour les rendre plus exploitables et cohérentes. Voici les étapes que nous avons suivies :

table Emissions Véhicule

- Suppression des colonnes non pertinentes
 - **CNIT**
 - **Modèle dossier qui est redondante** avec le **modèle commercial**
 - **Masse vide euro min**
 - **Masse vide euro max**
 - **Désignation commerciale**
 - **Boîte de Vitesse**
 - **CO type I**
 - **HC**
 - **NOx**
 - **HC+NOx**
 - **Particules**
- Fusion de **Modèle commercial** et **TVV** sous le nom **Modèle et Variante**. Les deux valeurs sont séparées avec ' / '.
- Ajout de colonnes calculées : Création de la colonne **Norme EURO** à partir de **Champ V9** en prenant les 5 derniers caractères (Suppression de Champ V9 juste après) :

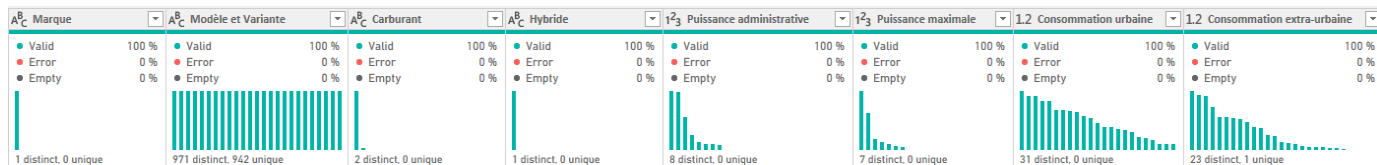
```
= Table.AddColumn(#"Removed Columns4", "Last Characters", each Text.End([Champ v9], 5), type text)
```

```
= Table.RenameColumns(#"Inserted Last Characters",{{"Last Characters", "Norme EURO"}})
```

- Filtrage des données : Filtrage des lignes contenant des valeurs null dans les colonnes **CO2** et **Consommation mixte**
- Normalisation des données pour les uniformiser et éviter les redondances :
 - Remplacer ROLLS-ROYCE par ROLLS ROYCE et ALFA-ROMEO par ALFA ROMEO dans la colonne **Marque**

- Remplacer MOY-INFÉRIEURE et MOY-INFER par MOY-INFER dans la colonne **Gamme**

➤ Validation des données : Affichage de la qualité des colonnes pour s'assurer qu'elles sont correctes et complètes avant de les charger dans Power BI l'analyse :



A gauche le résumé des étapes et à droite l'état final de la table :

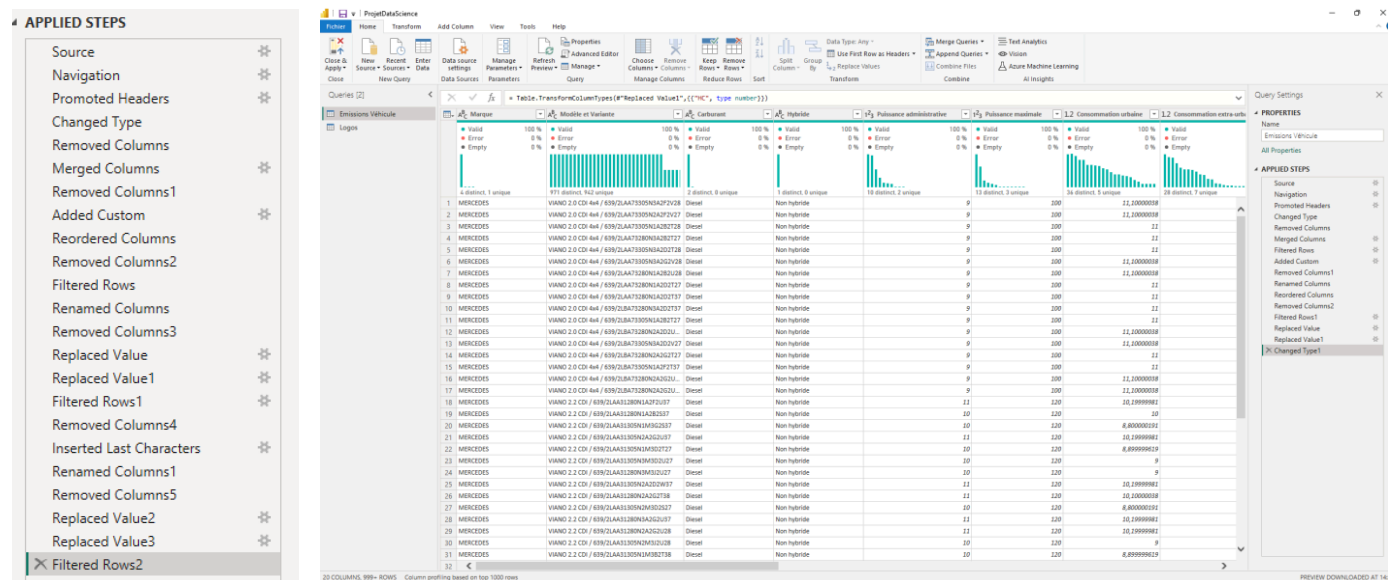


table Logos

Pour la table Logos, nous n'avons appliqué aucun traitement spécifique, car nous l'avons importée à partir d'un fichier CSV que nous avons soigneusement créé avec des données précises. Nous avons simplement utilisé la première ligne comme en-tête, ce qui est une pratique standard pour tous les fichiers CSV que nous chargeons.

Queries [2] | Emissions Véhicules | Logos

Table: RenameColumns(If(Changed Type1, [{"Logo de la Marque", "Logo Marque"}]))

Logo	Marque	Valid	Error	Empty
ALFA ROMEO	https://logodownload.org/wp-content/uploads/2002/03/alfa-romeo-4...	100 %	0 %	0 %
AUDI	https://www.pngall.com/wp-content/uploads/13/Audi-Logo-PNG-File...	100 %	0 %	0 %
BENTLEY	https://logo-walrus.com/wp-content/uploads/2021/06/Bentley-Logo...	100 %	0 %	0 %
BMW	https://walrusdesk.com/wp-content/uploads/2016/05/BMW-Logo-PNG...	100 %	0 %	0 %
BMW I	https://www.hubauer-shop.de/images/logo-header-2020-bmw-i.png	100 %	0 %	0 %
CADILLAC	https://pngimg.com/uploads/cadillac/cadillac_PNG65.png	100 %	0 %	0 %
CHEVROLET	https://listcarbrands.com/wp-content/uploads/2016/02/classic-chev...	100 %	0 %	0 %
CITROEN	https://logos-download.com/wp-content/uploads/2016/03/Citroen_L...	100 %	0 %	0 %
DACIA	https://upload.wikimedia.org/wikipedia/fr/4/4d/Dacia_Logo_Dacia_s...	100 %	0 %	0 %
DANGLER	https://media.cdnvcs.com/_/v46016/4869/2865/72/sickers-dangler-4a...	100 %	0 %	0 %
FERRARI	https://purepng.com/public/uploads/large/purepng.com-ferrari-logof...	100 %	0 %	0 %
FIAT	https://logodownload.org/wp-content/uploads/2014/05/fiat-logo-2.p...	100 %	0 %	0 %
FORD	https://listcarbrands.com/wp-content/uploads/2016/03/ford-new-log...	100 %	0 %	0 %
HONDA	https://logos-world.net/wp-content/uploads/2021/03/Honda-Logo-20...	100 %	0 %	0 %
HYUNDAI	https://purepng.com/public/uploads/large/purepng.com-hyundai-log...	100 %	0 %	0 %
INFINITI	https://gossipevehiculo.com/wp-content/uploads/2023/04/infiniti.png	100 %	0 %	0 %
JAGUAR	https://www.car-brand-names.com/wp-content/uploads/2015/07/jag...	100 %	0 %	0 %
JAGUAR LAND ROVER LIMITED	https://logos-world.net/wp-content/uploads/2023/06/jaguar-Land-Ro...	100 %	0 %	0 %
JEEP	https://logohistory.net/wp-content/uploads/2023/01/jeep-Logo-1970...	100 %	0 %	0 %
KIA	https://wallpaperaccess.com/full/3321956.jpg	100 %	0 %	0 %
LAMBORGHINI	https://pngimg.com/uploads/lamborghini/lamborghini_PNG10709.png	100 %	0 %	0 %
LANCIA	https://logodownload.org/wp-content/uploads/2016/04/Lancia_lo...	100 %	0 %	0 %
LAND ROVER	https://logodownload.org/wp-content/uploads/2015/08/land-rover-l...	100 %	0 %	0 %
LEXUS	https://logos-download.com/wp-content/uploads/2016/02/lexus_log...	100 %	0 %	0 %
MASERATI	https://logos-download.com/wp-content/uploads/2016/03/Maserati_...	100 %	0 %	0 %
MAZDA	https://logodownload.org/wp-content/uploads/2019/11/mazda-logo-...	100 %	0 %	0 %
MERCEDES	https://logodownload.org/wp-content/uploads/2014/04/mercedes-b...	100 %	0 %	0 %
MERCEDES AMG	https://www.carlogos.com/logo/AMG-logo-black-1920x1080.png	100 %	0 %	0 %
MINI	https://www.buscopng.com/wp-content/uploads/2021/03/Mini-Coop...	100 %	0 %	0 %
MITSUBISHI	https://logodownload.org/wp-content/uploads/2014/10/mitsubishi-lo...	100 %	0 %	0 %
NISSAN	https://www.car-brand-names.com/wp-content/uploads/2015/08/Nis...	100 %	0 %	0 %
OPEL	https://www.citypng.com/public/uploads/preview/opel-embliem-logo...	100 %	0 %	0 %

Query Settings

PROPERTIES

Name

Logos

APPLIED STEPS

Source

Changed Type

Promoted Headers

Changed Type1

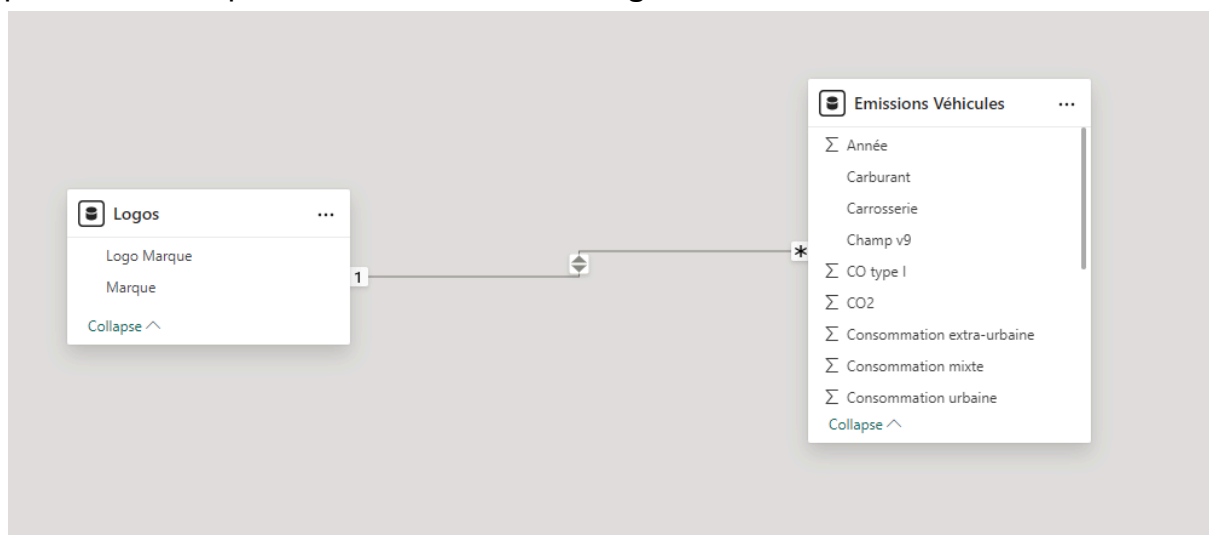
Rename Columns

2.3. Chargement des données dans Desktop

Après la phase nettoyage, nous chargeons nos données sur l'outil [PowerBI Desktop](#)

2.3.1. Model View

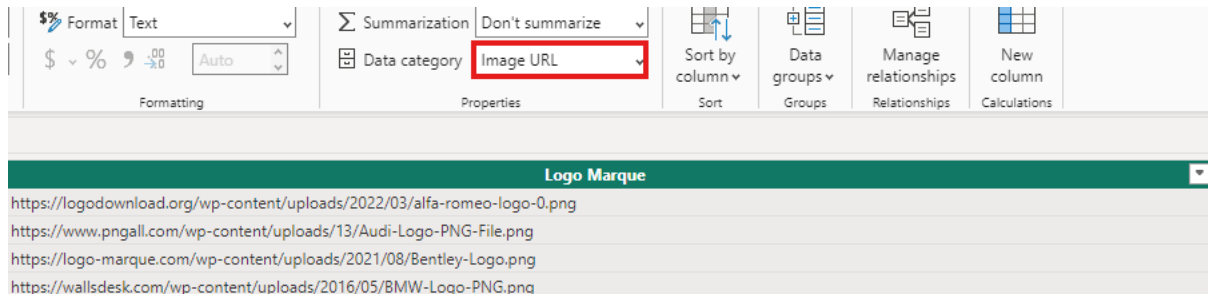
Nous ouvrons [Model View](#) pour voir les tables importées et nettoyées. Nous pouvons remarquer la relation **1..N** entre **Logos** et **Émissions Véhicules** :



⚠ Nous modifions l'association pour permettre le filtrage dans les deux sens

2.3.2. Table View

Nous changeons de vue pour **Table View** afin de spécifier que la colonne **Logo Marque** contient des URL d'images :

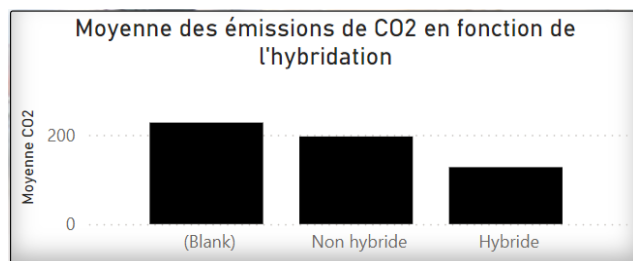
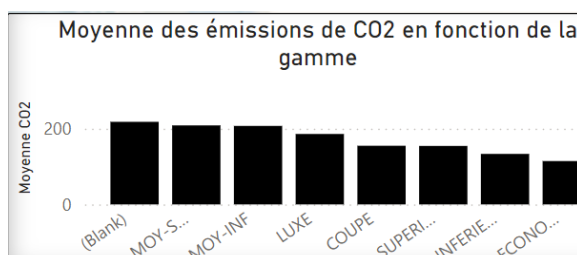
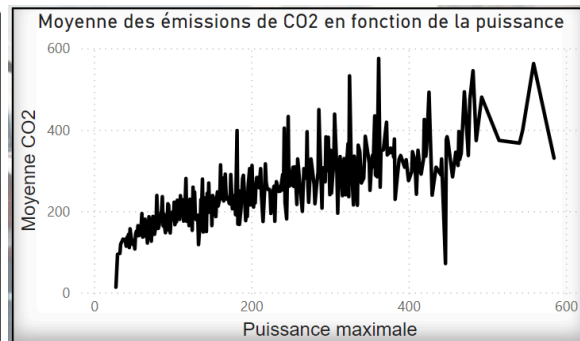
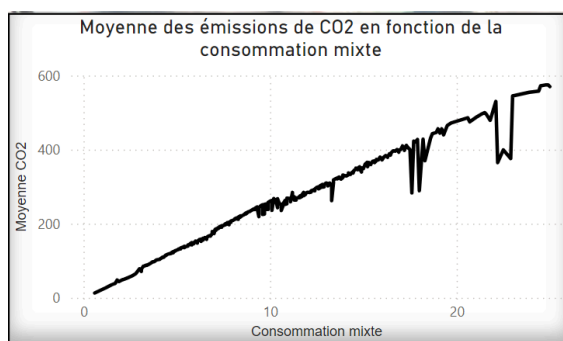


2.4. Visualisations des données

On se place sur la vue **Report** et on ajoute 5 onglets : Émission en fonction des caractéristiques - Caractéristiques des marques - Émission au fil du temps - KPI - Pour aller plus loin. On crée plusieurs mesures (cf powerBI) et on ajoute les visualisations adéquates.

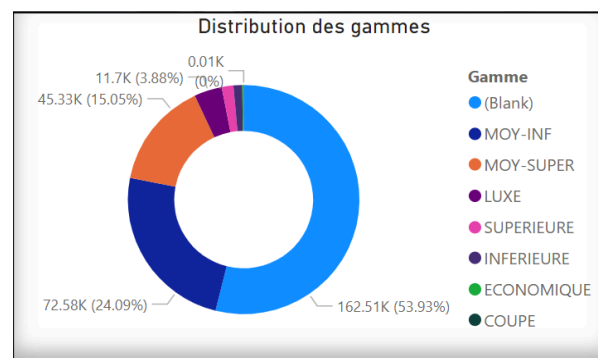
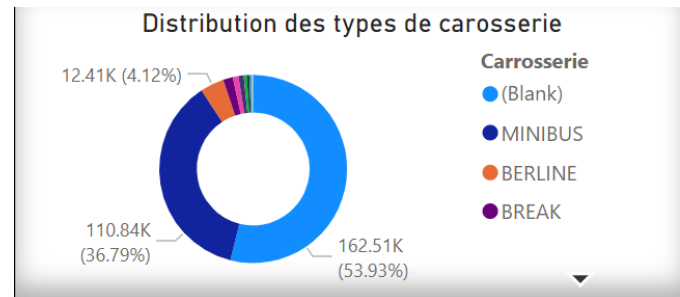
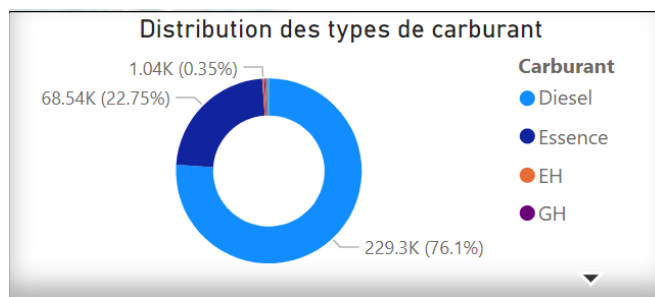
2.4.1. Émission en fonction des caractéristiques

Dans cette section, nous étudions comment les caractéristiques d'un véhicule influencent les émissions de CO2. Les caractéristiques étudiées sont : La puissance du véhicule - La gamme - La consommation - La carrosserie - type de moteur (véhicule hybride ou pas)



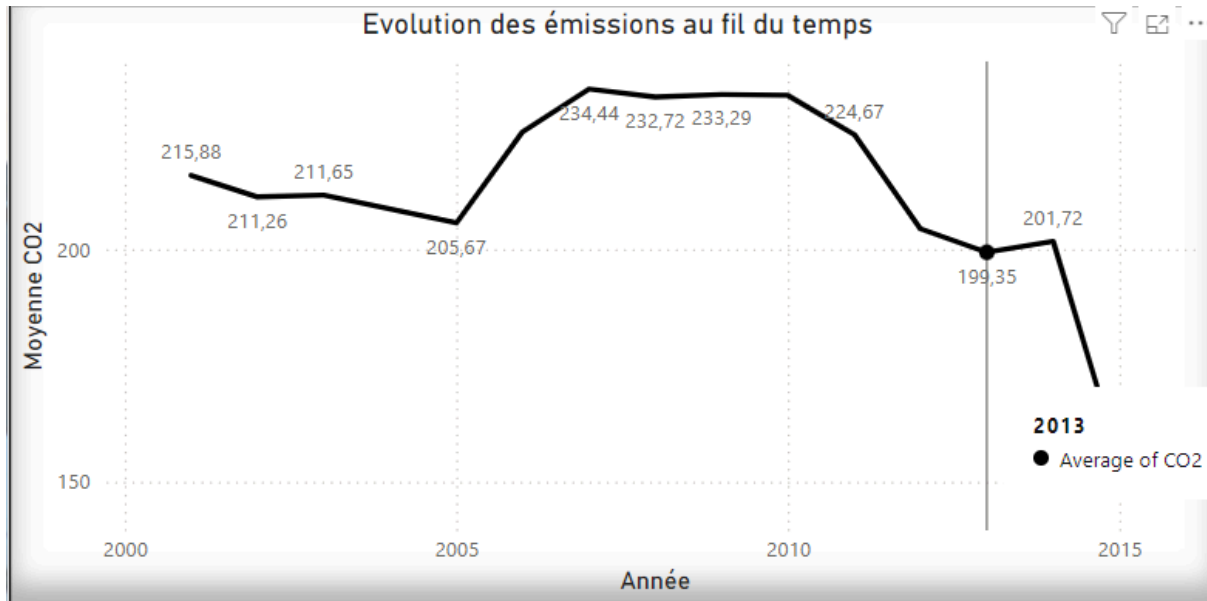
2.4.2. Caractéristiques des marques ⚙️

Nous présentons ensuite une vue d'ensemble des distributions des types de carburant, des types de carrosserie et des gammes de véhicules pour chaque marque.



2.4.3. Émission au fil du temps 🕒

Nous nous sommes également intéressés à l'évolution des émissions de CO2 au fil des années. Comme vous pouvez le constater sur la capture ci-dessous, les émissions ont tendance à décroître entre 2000 et 2005, ensuite elles augmentent légèrement de 2005 à 2010 puis chutent drastiquement jusqu'à atteindre une valeur minimale en 2015. Il est à noter que ces évolutions s'expliquent par les lois et réglementations européennes qui se sont succédées pour réduire les émissions de carbone (cf powerBI) :



2.4.4. KPI 🏆

Enfin, nous avons étudié les indicateurs de performance clés (KPI) afin d'identifier, pour chaque année, le véhicule se distinguant comme le "meilleur" en termes de puissance, de consommation et d'émissions, ainsi que la marque la plus éco-responsable, évaluée sur la base de la moyenne des émissions de CO2 de l'ensemble de ses véhicules pour une année donnée.

KPI 1 : Quel est le "meilleur" véhicule?

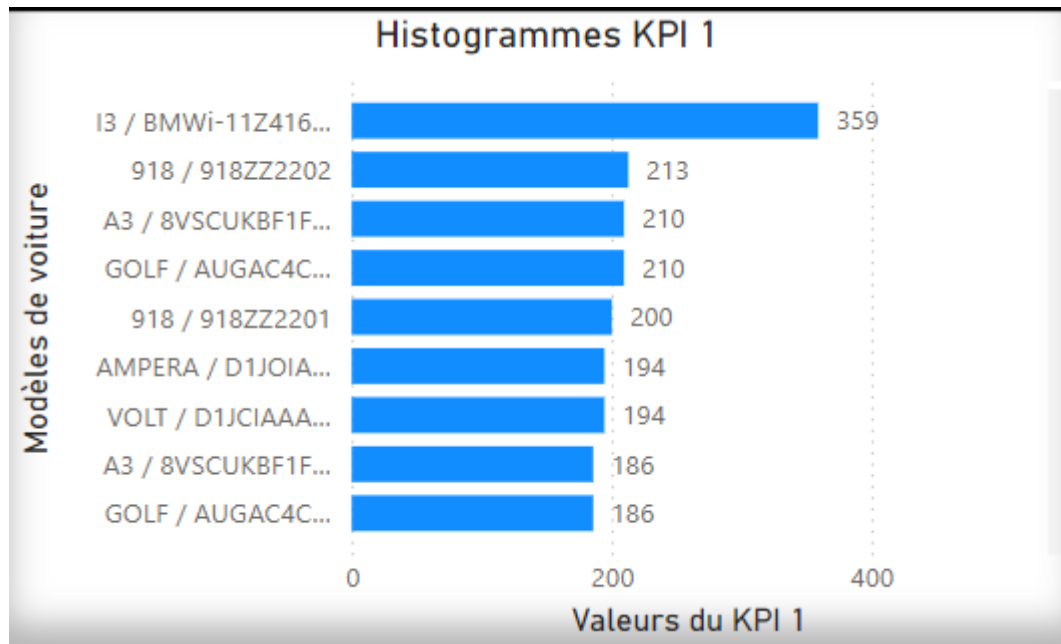
La formule utilisée pour le calcul de cet indicateur clé de performance (KPI) est la suivante :

$$\text{KPI 1} = \text{Puissance} / (\text{Émissions} \times \text{Consommation})$$

Cette formule met en relation trois paramètres essentiels : la **puissance** du véhicule, **ses émissions de CO2** et sa **consommation** énergétique. Plus la puissance est élevée et plus les émissions et la consommation sont faibles, plus la valeur du KPI sera importante, indiquant ainsi une meilleure performance globale.

Ce calcul permet d'évaluer l'efficacité énergétique et environnementale d'un véhicule tout en tenant compte de ses capacités de performance, offrant ainsi une mesure comparative pertinente entre différents modèles.

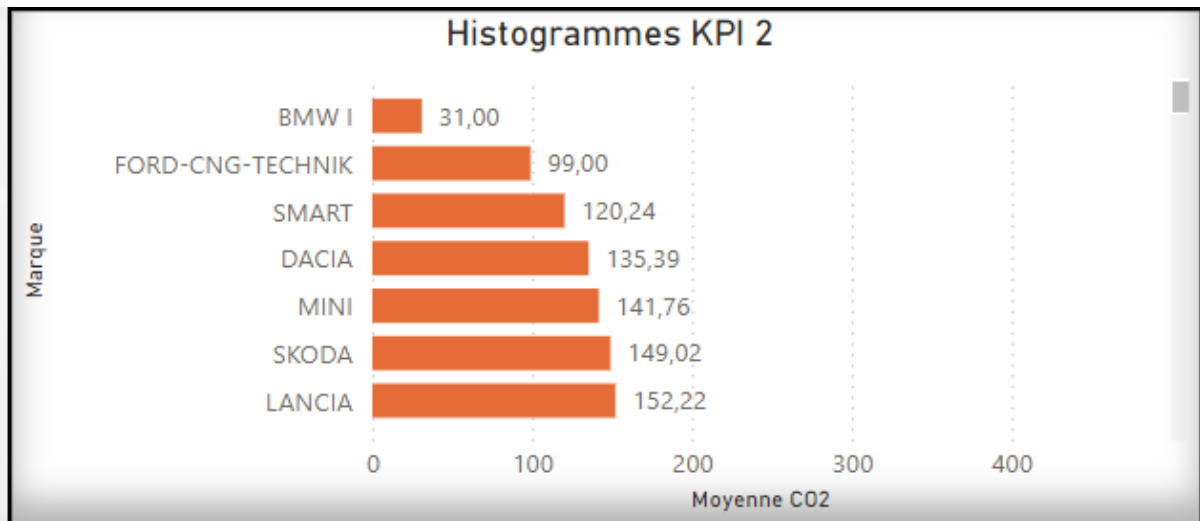
L'exemple ci-dessous donné pour l'année 2015 montre que le **BMW-11Z** est le véhicule offrant le meilleur équilibre entre puissance, consommation et émissions et donc considéré comme le meilleur :



KPI 2 : Quelle est la marque la plus éco-responsable ?

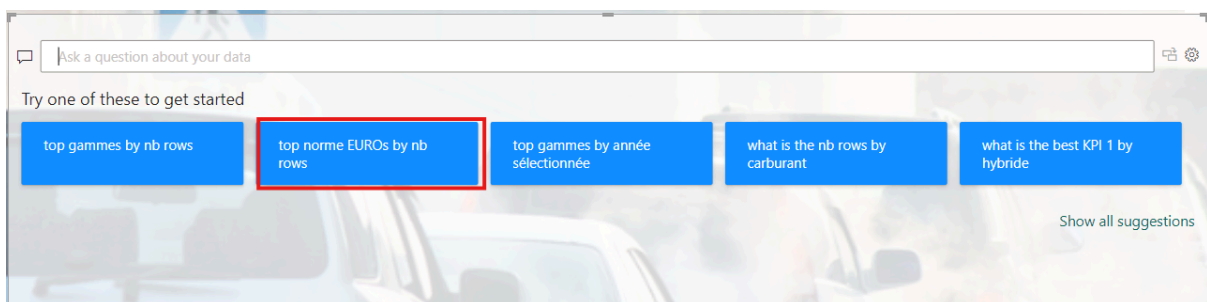
La détermination de la marque la plus éco-responsable repose sur la **moyenne** des émissions de CO2 de l'**ensemble des véhicules** produits par chaque marque au cours d'une année donnée. Cet indicateur permet de mesurer l'engagement global d'une marque en matière de réduction des émissions polluantes et d'optimisation de son empreinte environnementale. Évidemment plus cet indice est faible, plus la marque est éco-responsable.

Sur l'exemple ci-dessous, en 2015, c'est la marque BMW I qui se distingue comme étant la plus éco-responsable, confirmant ainsi le constat fait précédemment selon lequel le modèle **BMW-11Z** de cette même marque est le meilleur véhicule sur l'année 2015 :



2.4.5. Pour aller plus loin 💡

Pour conclure nos analyses sur PowerBI, nous avons tiré parti de sa fonctionnalité d'**intelligence artificielle** dédiée aux questions et réponses, qui génère des insights en se basant sur le jeu de données et le contenu du rapport. Cette fonctionnalité permet d'afficher des réponses sous forme de visuels interactifs en cliquant sur une question spécifique. Par exemple, en sélectionnant la question 2, encadré en rouge, nous avons pu constater que la norme **EURO5** est celle qui prédomine dans le dataset, permettant ainsi une compréhension plus approfondie du jeu de données sans avoir à effectuer les analyses manuellement :





3. Conclusion

Nous sommes convaincus que notre rapport est particulièrement clair et compréhensible, grâce à une palette de couleurs bien choisie, des graphiques simples à interpréter et des métriques pertinentes. L'interface interactive et personnalisable offre une expérience utilisateur stimulante et agréable. De plus, nous avons abordé un sujet d'actualité qui touche tout le monde et est au cœur des débats environnementaux.

Cependant, une des faiblesses de notre projet réside dans la présence de plusieurs valeurs nulles dans le jeu de données. Pour y remédier, nous avons filtré les lignes vides dans la colonne "Émissions CO2", car toute ligne avec une valeur nulle dans cette colonne n'était pas pertinente pour notre étude. En revanche, nous avons choisi de conserver les valeurs nulles dans les autres colonnes, car les filtrer aurait considérablement réduit la taille de notre dataset, ce qui aurait compromis la précision de nos analyses. Nous avons ainsi surmonté cette faiblesse tout en maintenant un jeu de données suffisamment complet.

Pour améliorer notre projet, nous pourrions envisager de chercher un jeu de données plus complet, avec moins de valeurs nulles. De plus, croiser notre table avec d'autres sources, telles que celles sur les prix des véhicules ou les clients, pourrait nous permettre d'approfondir davantage nos analyses. Nous pourrions même intégrer des algorithmes d'apprentissage automatique pour prédire les émissions futures de CO2 en fonction de différents paramètres.

Ce projet nous a appris à manipuler de grands jeux de données et à utiliser les bases de PowerBI. Nous avons également dû réfléchir sur les meilleures métriques à utiliser et sur la manière de les représenter de façon simple et percutante. Apprendre à rendre un grand jeu de données compréhensible a été une expérience intéressante. Cela nous a permis de réaliser que la qualité de la présentation est tout aussi importante que celle des informations elles-mêmes.

En conclusion, nous pensons que ce projet est très réussi, mais qu'il pourrait être encore enrichi en explorant d'autres aspects de la science des données au-delà de PowerBI, afin d'approfondir encore davantage nos analyses.

Merci pour tout 😊