

Projet SAE étude statistique et enquête

Babingui Anaïs
Jacques Virginie
Moulin Sarah

2021-2022

Tables des matières

- I. Résumé
- II. Introduction
- III. Présentation des variables
- IV. Méthode
- V. Analyse et résultats
- VI. Conclusion
- VII. Bibliographie
- VIII. Annexes

I. Résumé

Dans le cadre d'un exercice d'analyse statistique, nous avons exploité une base de données sur différentes catégories de voitures. Nous avons simulé une demande d'un entrepreneur qui souhaite démarrer une concession de voitures en cohésion avec le développement durable. Il veut proposer à ses clients les véhicules les moins polluants. Il possède déjà d'autres entreprises et a donc les fonds nécessaires pour démarrer cette nouvelle activité.

Grâce à la base de données qui nous a été fournie, nous avons détecté quelles variables étaient présentes et avons ainsi pu distinguer les différentes caractéristiques des voitures.

Puis, à l'aide des outils RStudio et Excel, nous avons procédé à la conception de graphiques et à l'analyse des données en utilisant les variables existantes.

Enfin, à partir de nos analyses, nous avons pu tirer des conclusions et proposer des solutions, afin de conseiller au mieux le futur concessionnaire sur les types de voiture à acquérir.

II. Introduction

De nos jours, les voitures représentent un moyen de transport beaucoup utilisé pour leur capacité à aller d'un point A à un point B rapidement et nous permettre d'être autonomes. Mais à cet aspect pratique s'ajoute une nouvelle problématique : le développement durable. Il s'agit alors de déterminer quel type de voiture est le plus adapté à l'écologie.

Nous sommes des statisticiens spécialisés dans le service du transport, un entrepreneur nous a chargés de trouver quel type de voiture serait idéal pour démarrer une concession de voiture en adéquation avec le développement durable.

Des données statistiques nous ont été communiquées. Notre base de données est composée de 3836 individus, ce qui correspond au nombre de voitures. Avec cela, il y a 10 variables, dont 5 quantitatives et 5 qualitatives. Ces 10 variables représentent les différentes caractéristiques des voitures.

Quel type de voiture serait le plus conseillé pour limiter la progression de la pollution, et ainsi avoir un environnement plus écologique ? Faudrait-il retirer un type de voiture du marché pour un avenir plus sûr ?

Dans un premier temps, nous analyserons les données et dans un deuxième temps nous interpréterons les résultats.

III. Présentation des variables

Nous possédons 5 variables quantitatives et 5 variable qualitatives dans notre base de données.

Tout d'abord, voici les 5 variables quantitatives :

La variable "puiss_ch" représente la puissance chevaux d'une voiture. Puis nous avons la variable "conso_urb" qui correspond à la consommation de carburant en zone urbaine, en litre par km. Il y a la variable "conso_exurb" qui à l'inverse correspond à la consommation de carburant en zone ex-urbaine. Ces deux variables permettent d'obtenir la variable "conso_mixte", qui correspond à la moyenne entre les consommations urbaine et ex-urbaine. Enfin, nous avons la variable "col_co" a attrait au CO2 rejeté par la voiture en gramme par km.

Ensuite, voici les 5 variables qualitatives :

La variable "mrq_nom" correspond aux modèles et aux marques des voitures. Ensuite, nous avons la variable "col_carb", qui représente les types de carburant utilisés. La variable "col_hyb" qui permet de savoir si la voiture est hybride ou non. La variable "col_carr" correspond à la carrosserie de la voiture. Enfin, il y a la variable "col_gam" qui représente la gamme de la voiture.

IV. Méthode

Notre base de données nous a été attribuée par notre professeur suite à un tirage au sort.

Tout d'abord, nous avons importé notre base de données sur RStudio et Excel pour connaître le nom de nos variables.

Nous avons par la suite défini leur signification, afin de faciliter la compréhension des données. Nous avons décidé parmi les 10 variables présentes dans la base de données, d'analyser 8 d'entre elles et de laisser de côté la puissance chevaux et la gammes, non considérées comme des facteurs importants de la pollution des voitures.

Nous avons donc réalisé des boxplot, histogrammes, nuages de points et coefficients de corrélation sur Rstudio. Sur Excel, les données ont été filtrées et triées, et des diagrammes en bâton ont été réalisés. Ces éléments nous ont permis de mieux analyser et comparer les différentes catégories de voiture.

Après avoir ces analyses, nous avons pu tirer des conclusions et répondre à notre problématique.

Enfin, nous avons finalisé notre rapport en y rajoutant des documents annexes, et nous avons réalisé un diaporama en retranscrivant l'essentiel de ce rapport pour le partager avec nos camarades.

Nous nous sommes répartis les tâches de façon à progresser sereinement et efficacement, en utilisant la plateforme Nexcloud pour partager nos avancements sur le rapport.

V. Analyse et résultats

Afin répondre à la problématique posée, nous avons analysé notre base de données.

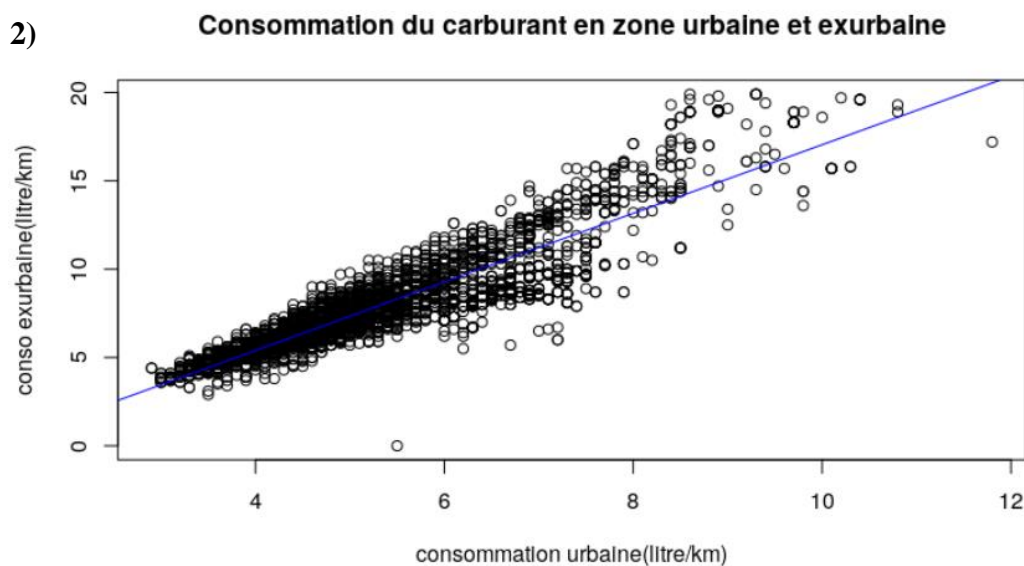
Pour cela, nous avons commencé par comparer la consommation urbaine et ex-urbaine de l'ensemble des voitures. Cela nous permettrait d'évaluer la différence ou la similitude entre ces deux catégories de consommation de carburant.

Nous avons donc effectué le coefficient de corrélation entre les deux variables ainsi qu'un nuage de points.

Voici les commandes Rstudio utilisés:

```
1) plot(conso_urb~conso_exurb,  
      main="Consommation du carburant en zone urbaine et exurbaine",  
      xlab = "consommation urbaine(litre/km)",  
      ylab = "conso exurbaine(litre/km)",  
      pch = 1)  
reg<-lm(conso_urb~conso_exurb)  
abline(reg,col="blue",lwd=1)
```

Cela nous a permis d'obtenir :



On remarque clairement sur ce graphique que les valeurs des deux variables évoluent dans le même sens et à la même allure. Elles sont réparties de façon similaire et se regroupent autour de la droite de régression. Il semblerait donc que la consommation de carburant en zone urbaine et ex-urbaine ne divergeraient que légèrement.

Le résultat du calcul du coefficient de corrélation nous le confirme également. Il a en effet une valeur de 0,92, et est donc proche de 1. Cela montre qu'il y a une forte liaison linéaire entre les deux variables.

Le nuage de points et le coefficient de corrélation nous permettent donc de déduire qu'il n'y aurait pas d'intérêt à utiliser ses deux variables séparément.

Nous en concluons qu'il faudrait alors utiliser la variable consommation mixte (correspondant à la moyenne des deux consommations), plutôt que la consommation urbaine et ex-urbaine.

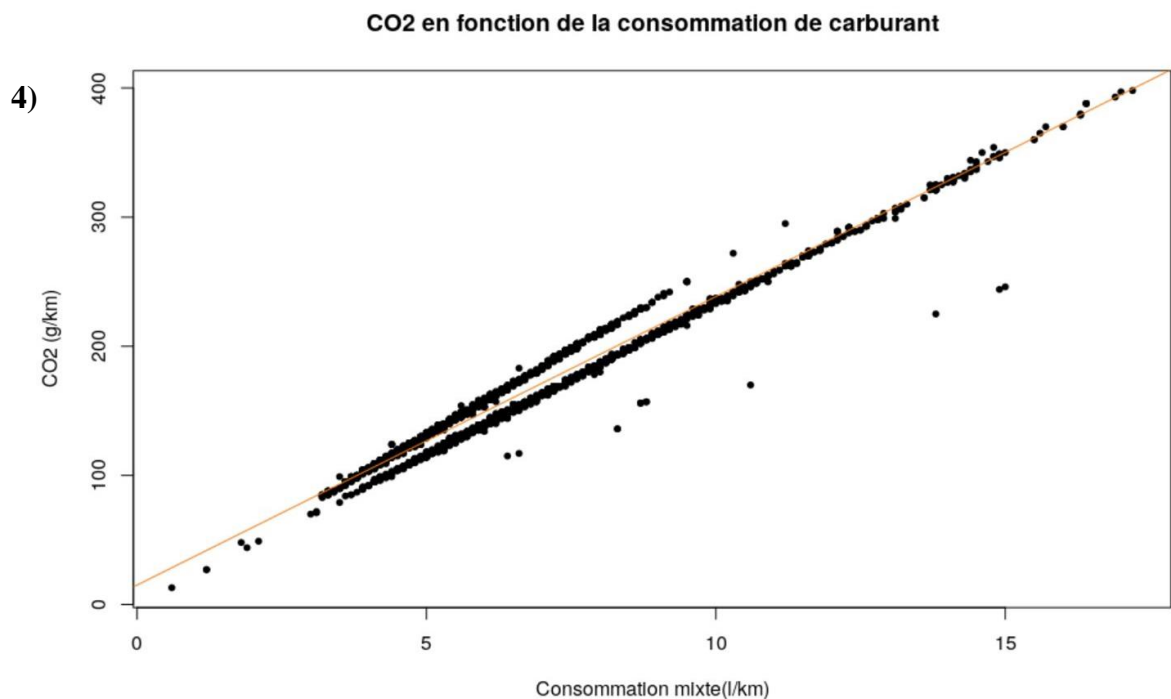
Pour aider le concessionnaire, nous décidons maintenant d'analyser la l'émission de CO2 en fonction de la consommation de carburant, afin de comprendre quelle influence ont les deux variables l'une sur l'autre.

Pour cela nous avons réalisé un nuage de points entre ces 2 nouvelles variables en réalisant ces commandes sur RStudio :

3)

```
plot(co2carb$col_co~co2carb$conso_mixte,  
      xlab="Consommation mixte(l/km)",  
      ylab="CO2 (g/km)",  
      main="CO2 en fonction de la consommation de carburant",  
      pch=20)  
reg1<-lm(co2carb$col_co~co2carb$conso_mixte)  
abline(reg1,lwd=1,col="chocolate1")
```

Nous avons obtenu le graphique suivant :



Le graphique ci-dessus nous montre l'influence de la quantité de la consommation de carburant d'une voiture sur les émissions de CO₂. Les valeurs sont regroupées autour de la droite de régression, ce qui montre une forte liaison linéaire entre les variables. On remarque alors que plus on consomme de carburant, plus on émet de CO₂.

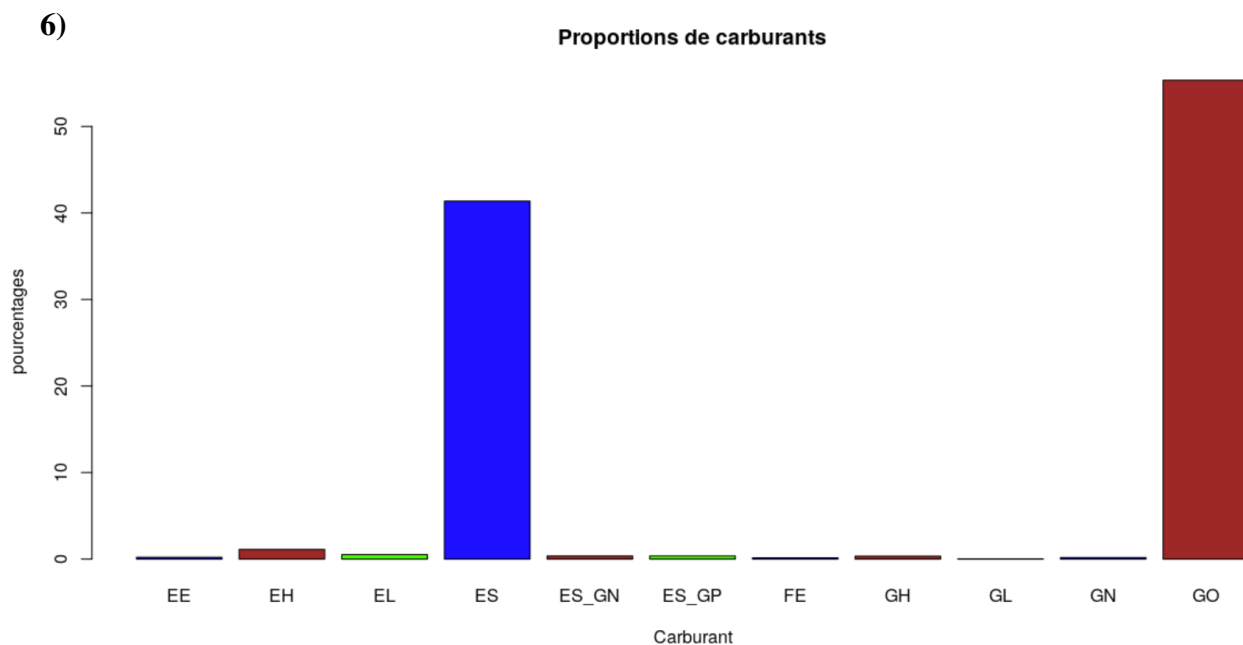
Le concessionnaire devrait donc choisir des voitures qui consomment le moins de carburant possible, afin de réduire les émissions de CO₂.

Pour cela, nous décidons ensuite de déterminer les carburants les plus utilisés. Nous avons réalisé un histogramme sur les proportions de carburants.

Nous avons utilisé les commandes Rstudio suivantes :

```
5) barplot(100*prop.table(table(db3$col_carb)),
           col=c("blue","brown","green"),
           main="Carburants les plus utilisés",
           xlab="Carburante",
           ylab="pourcentages")
n<-100*prop.table(table(db3$col_carb))
round(n, digits=2)
```

Voici l'histogramme obtenu :



Et voici les pourcentages obtenus :

7)

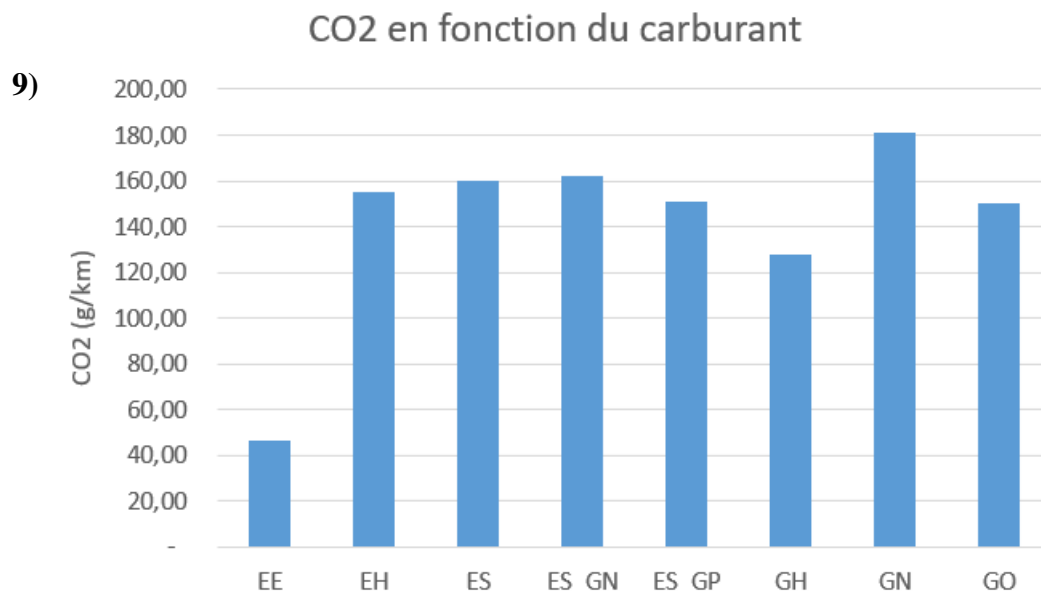
EE	EH	EL	ES	ES_GN	ES_GP	FE	GH	GL	GN	GO
0.2	1.1	0.5	41.4	0.4	0.4	0.2	0.3	0.0	0.2	55.3

- 8)
- EE : essence / électrique (rechargeable)
 - EH : essence / électrique (non rechargeable)
 - EL : électrique
 - ES : essence sans plomb 95
 - ES/GN : bicarburation essence / Gaz Naturel Véhicule
 - ES / GP : bicarburation essence / Gaz de Pétrole Liquéfié
 - FE : véhicule E85 (ou superéthanol-E85)
 - GH : gazole / électrique (non rechargeable)
 - GL : gazole / électrique (rechargeable)
 - GN : monocarburation Gaz Naturel Véhicule
 - GO : gazole

En observant l'histogramme, nous pouvons remarquer que les deux types de carburant principalement utilisés sont l'essence (ES) et le gazole (GO). Le gazole possède la part la plus importante, il représente 55% de la totalité des carburants. L'essence quant à lui représente 41%. Les 9 autres modalités n'ont qu'une part très minime.

Maintenant que nous connaissons les carburant les plus utilisés, nous pouvons déterminer les carburants les plus polluants, c'est-à-dire ceux qui rejettent le plus de CO₂. Nous avons réalisé un histogramme contenant les moyennes d'émission de CO₂ en fonction des types de carburant.

Voici l'histogramme réalisé avec Excel :



Trois types de carburant ne sont pas représentés : le véhicule E85 (FE), le gazole électrique rechargeable (GL) et l'électrique (EL), qui contiennent trop peu de valeurs pour être comptabilisés.

Nous pouvons constater que l'essence électrique non rechargeable (EE) est le carburant émettant le moins de CO₂, avec un taux de 46 g/km. A l'opposé, le monocarburation gaz naturel (GN) possède la valeur la plus haute, qui est de 180 g/km.

Par ailleurs, le Gazole (GO) qui est le carburant le plus utilisé, a une valeur de 152, ce qui correspond à la moyenne de l'ensemble des modalités, qui est de 154. L'essence quant à lui se situe légèrement au-dessus, avec une valeur de 160.

Ainsi, le concessionnaire devrait favoriser l'essence électrique non rechargeable (EE) et afin de tout de même proposer du choix à sa clientèle, proposer des véhicules fonctionnant au gazole.

Nous avons ensuite décidé de comparer les moyennes de consommation de carburant mixte entre les véhicules hybrides et thermiques. Cette comparaison a été effectuée sur la totalité des voitures hybrides, au nombre de 65, ainsi qu'un échantillon de 65 voitures thermiques, prélevées au hasard parmi 3 771 individus à l'aide d'Excel.

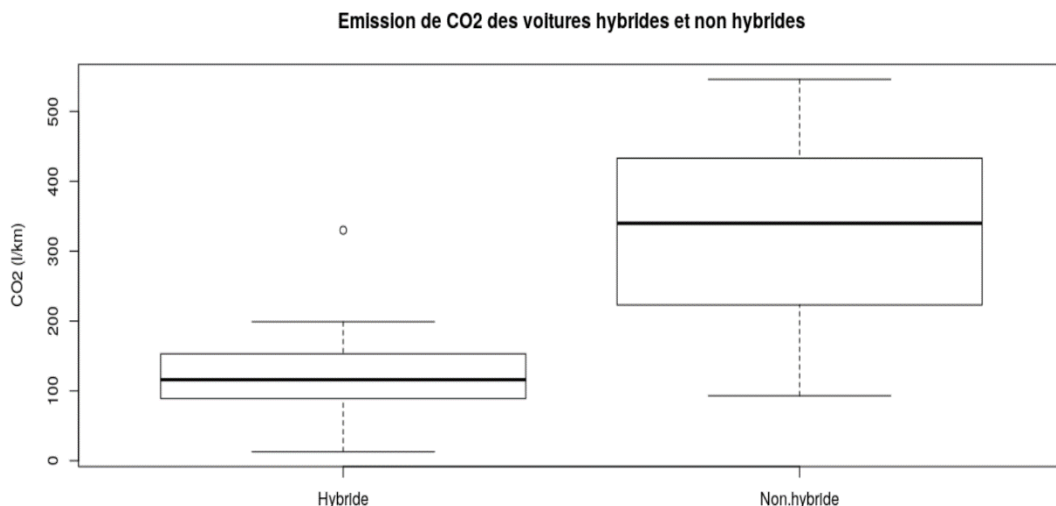
Nous avons utilisé les commandes Rstudio suivantes :

10)

```
boxplot(db, main="Emission de CO2 des voitures hybrides et non hybrides",  
        ylab="CO2 (l/km)")  
mean(db$Hybride)  
mean(db$Non.hybride)
```

Nous avons obtenu le graphique suivant :

11)



Et les données suivantes :

12)

```
> mean(db$Hybride)  
[1] 119.8308  
> mean(db$Non.hybride)  
[1] 318.9385
```

Nous avons obtenu une moyenne de consommation de carburant de 120 litre/km pour les voitures hybrides et 319 litre/km pour les voitures thermiques. Ainsi, les voitures non hybrides émettent 38% de plus de CO2 que les voitures non hybrides. Nous conseillons donc au concessionnaire de favoriser la vente de véhicules hybrides.

Nous allons maintenant déterminer si certaines marques de voitures sont plus polluantes que d'autres. Ainsi, nos résultats nous permettront de proposer au concessionnaire d'éliminer certaines marques de son catalogue, jugées trop polluantes. Pour déterminer quelles marques consomment le plus nous avons utilisé les étiquettes énergie/CO₂, des étiquettes ecolabels permettant d'identifier les émissions de CO₂ mesurées en gramme par kilomètre. Ces étiquettes sont classées de A à G, A étant la catégorie la moins polluante et G la catégorie émettant le plus de CO₂.

Pour cela nous avons filtré sur Excel les voitures émettant le plus de CO₂, c'est-à-dire dépassant le seuil de 250 g/l :

13)

Émissions de CO ₂ faibles	
inférieures ou égales à 100 g/km	A
de 101 à 120 g/km	B
de 121 à 140 g/km	C
de 141 à 160 g/km	D
de 161 à 200 g/km	E
de 201 à 250 g/km	F
supérieures à 250 g/km	G
Émissions de CO ₂ élevées	

Source : www.entretien-auto.com

13)

mrq_nom	col_co	col_gam
ASTON MARTIN ONE_77	555	LUXE
LAMBORGHINI VENENO	398	LUXE
MERCEDES G 65 BREAK LONG BVA7	397	LUXE
BENTLEY MULSANNE	393	LUXE
ASTON MARTIN V12 VANTAGE	388	LUXE
ASTON MARTIN V12 VANTAGE ROADSTER	388	LUXE
ASTON MARTIN V12 ZAGATO	388	LUXE
FERRARI FF	380	LUXE
ALFA_ROMEO 8C SPIDER	379	LUXE
LEXUS LFA	379	LUXE
LAMBORGHINI AVENTADOR	370	LUXE
LAMBORGHINI AVENTADOR ROADSTER	370	LUXE
CADILLAC "CTS_V WAGON 6_2L V8 BVM R"	370	SUPERIEURE
CADILLAC "CTS_V COUPE 6_2L V8 BVM RW	365	SUPERIEURE
MASERATI GRANTURISMO MC STRADALE	360	LUXE
MASERATI GRANTURISMO SPORT	360	LUXE
LAMBORGHINI GALLARDO SPYDER	354	LUXE
FERRARI F12Berlinetta	350	LUXE
LAMBORGHINI GALLARDO SPYDER PERFOR	350	LUXE
AUDI R8 Spyder V10 5_2 FSI (525ch) QUAT	349	LUXE

Nous remarquons que 196 voitures sont dans la catégorie énergétique G, donc la plus polluante. En effet, elles dépassent toutes un taux d'émission de CO₂ de 250 g/km. Parmi elles, les marques les plus redondantes sont Mercedes, qui est présente à 16% ainsi que Cadillac et Chevrolet, présentes toutes deux à 8%.

Dans une optique privilégiant l'écologie, le concessionnaire devrait alors éviter de proposer ces trois marques sur son marché.

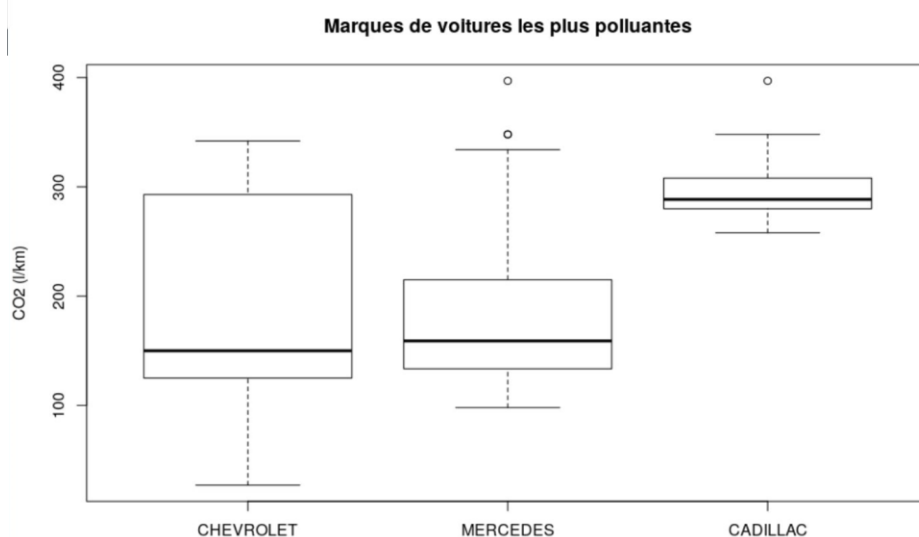
Il s'agit maintenant d'analyser ces 3 marques, et de déterminer si tous leurs modèles de voitures sont très polluantes, ou si c'est seulement le cas de certains modèles. Pour cela, nous avons réalisé 3 boxplots, afin d'analyser la répartition de leurs modèles de voitures.

Voici la commande Rstudio utilisée :

```
14) boxplot(conso1, main="Marques de voitures les plus polluantes")
quantile(conso1$CHEVROLET, na.rm=T)
quantile(conso1$MERCEDES, na.rm=T)
boxplot(conso1$MERCEDES)$out
quantile(conso1$CADILLAC, na.rm=T)
boxplot(conso1$CADILLAC)$out
```

Nos boxplots :

15)



On observe que pour la marque Chevrolet, 25% des voitures émettent moins de 125 g/l de CO2. 50% des valeurs sont entre 125 et 274, et 25% des valeurs sont à plus de 274.

Pour la marque Mercedes, 25% des valeurs sont entre 100 et 135. 50% des valeurs sont comprises entre 135 et 214, donc 25% des valeurs sont à plus de 214. Deux valeurs extrêmes sont notables, représentant respectivement 348 et 397.

Pour la marque Cadillac, 25% des valeurs sont entre 260 et 280. 50% des valeurs émettent entre 280 et 308 g/l, donc 25% des valeurs sont à plus de 308. Une valeur extrême est notable, représentant 397.

Ainsi, la marque Cadillac contient les véhicules les plus polluants, et certains modèles Chevrolet sont peu polluants. Nous conseillons alors au concessionnaire de ne pas proposer les marques Cadillac et Mercedes dans son catalogue, mais il peut proposer certains modèles Chevrolet, qui émettent peu de CO2.

Nous allons maintenant déterminer quelles sont les marques de voitures qui consomment le moins de CO₂. En fonction de nos résultats, le concessionnaire aura une meilleure idée des marques les moins consommatrices et donc à privilégier.

Comme vu précédemment, nous utilisons comme référence les étiquettes écolabels qui classent les voitures en plusieurs catégories selon leur émission de CO₂.

Nous avons filtré sur Excel les voitures émettant le moins de CO₂, c'est-à-dire en dessous du seuil de 100 g/l :

16)

mrq_nom	col_co
ALFA_ROMEO MITO 0_9 Twin Air (105ch) S_S	99
ALFA_ROMEO MITO 0_9 Twin Air (85ch) S_S	98
ALFA_ROMEO MITO 1_3 JTDm (85ch) S_S	90
AUDI A1 1_6 TDI (105ch)	99
AUDI A1 1_6 TDI (90ch)	99
AUDI A3 1_6 TDI (105ch)	99
AUDI A3 1_6 TDI ultra (110ch)	85
AUDI A3 Berline 1_6 TDI (105ch)	99
AUDI A3 SPORTBACK 1_6 TDI (105ch)	99
AUDI A3 SPORTBACK 1_6 TDI ultra (110ch)	88
BMW 116d EfficientDynamics 3P	99
BMW 116d EfficientDynamics 5P	99
BMW i3 avec Prolongateur d'autonomie	13
CHEVROLET Aveo 4P 1_3 VCDI 16v (75ch) BVM5	97
CHEVROLET Aveo 4P 1_3 VCDI 16v (95ch) BVM5	95
CHEVROLET Aveo 5P 1_3 VCDI 16v (75ch) BVM5	99
CHEVROLET Aveo 5P 1_3 VCDI 16v (95ch) BVM6	95

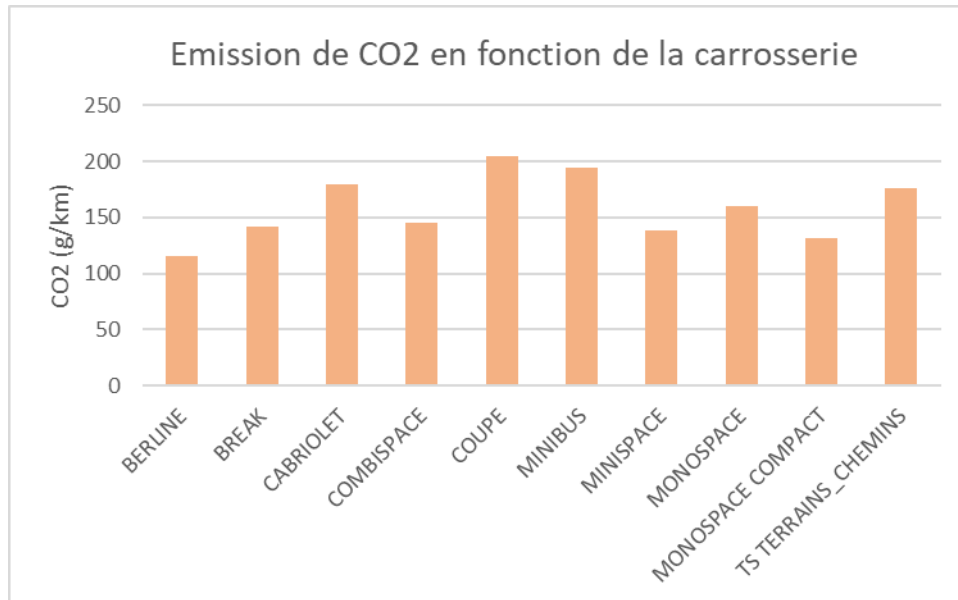
Nous remarquons alors que 212 voitures sont dans la catégorie énergétique A, la moins polluante. En effet, elles ont un taux d'émission de CO₂ inférieur à 100 g/km. Parmi elles, les marques les plus redondantes sont Peugeot, présente à 11%, Citroën à 9% et Toyota à 8%.

Dans une optique d'écologie, le concessionnaire devrait alors privilégier ces 3 marques.

Il s'agit finalement de savoir quel type de carrosserie est le plus polluant, afin de sélectionner, en plus des marques, les carrosseries les moins polluantes. Pour cela, nous avons réalisé un histogramme permettant de comparer les émissions moyennes de CO2 par carrosserie.

Histogramme réalisé à l'aide d'Excel :

17)



On observe que le type Coupé possède la valeur la plus haute, il émet en moyenne 205 g/km de CO2. Le type Minibus est très proche, avec une valeur de 194. A l'opposé, le type Berline a la valeur la plus basse, avec 116. S'ensuit le type Minispace avec une valeur de 138.

Ainsi, nous conseillons au concessionnaire d'éviter les carrosseries Coupé et Minibus, et de favoriser les carrosseries Berline et Monospace Compact, qui sont les moins polluantes.

V. Conclusion

Les analyses réalisées nous ont permis d'établir plusieurs conclusions. Notre principal objectif était de conseiller un entrepreneur qui voulait lancer sa concession de voitures. Il avait besoin qu'on analyse le marché actuel des voitures et qu'on lui propose une bonne stratégie de vente (quelles voitures seraient les plus intéressantes à acquérir, quelles voitures seraient à éviter), dans une optique de développement durable. Pour cela il nous a fallu émettre des hypothèses et produire des analyses, qui répondraient à nos questionnements.

Nous avons pu remarquer que les carburant les plus utilisés sont le gazole et l'essence, et le gazole est le plus polluant. Pour permettre à la concession d'être compétitive sur le marché et de proposer une large gamme de voiture, même si ces deux types de carburant ne sont pas les plus écologiques, nous suggérons à l'entrepreneur de tout de même proposer des véhicules à essence, mais de retirer les véhicules à gazole.

Les voitures hybrides sont beaucoup moins polluantes que les voitures thermiques, c'est pourquoi nous lui suggérons aussi d'en proposer.

Nous avons aussi élaboré un classement, déterminant les voitures les plus polluantes et les moins polluantes. Ce classement nous a permis de mettre en valeur les marques à privilégier, les trois premières étant Peugeot, Citroën et Toyota. En ce qui concerne les marques de voitures les plus émettant le plus de CO₂, trois se démarquent de notre liste initiale. Il s'agit de Mercedes, Cadillac et Chevrolet. Cependant, un quart des voitures Chevrolet sont peu polluantes. Il devrait donc éviter Mercedes et Cadillac, et proposer certains modèles Chevrolet. Enfin, nous avons analysé les types de carrosserie. Il s'avère que le concessionnaire ne devrait pas proposer les Coupé et Minibus, et au contraire favoriser Berline et Monospace Compact, qui sont les carrosseries les moins polluantes.

Pour répondre à la problématique, nous conseillons de privilégier les voitures à essence ainsi que les voitures hybrides. Il serait aussi préférable de favoriser certaines marques à d'autres. Il s'agit de Peugeot, Citroën et Toyota. Enfin, il devrait principalement proposer les carrosseries Berline et Monospace Compact.

VI. Bibliographie

Données :

Base de données voitures

Outils :

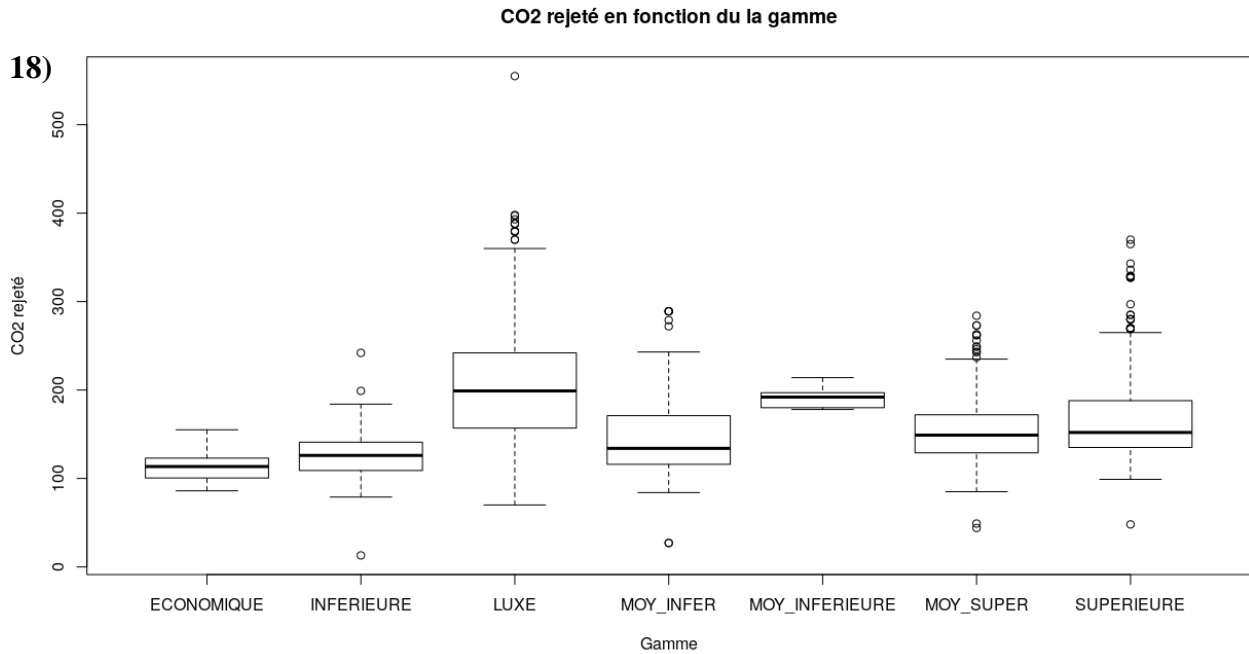
Rstudio

Excel

Sites web :

<https://voitures.com/>
<https://abcdr.thinkr.fr/>
<https://www.larousse.fr/>
<https://www.centralcharts.com/>
<https://www.jmp.com/>
<https://www.ebsi.umontreal.ca/>
<https://www.cnrtl.fr/>
<https://www.ademe.fr/>
<http://forums.cirad.fr/>
<http://www.vedura.fr/>
<https://www.service-public.fr/>
<https://www.europarl.europa.eu/>
<https://www.taxprep.com/>
<https://support.microsoft.com/>
<https://www.excel-downloads.com/>
<https://techniques-pilotage.fr/>
<https://dictionnaire.orthodidacte.com/>
<https://www.auto-moto.com/>
<https://blog.itgs-solutions.ch/>
<https://www.automobile-magazine.fr/>
<https://www.autoplus.fr/>
<https://fr.surveymonkey.com/>
<https://fr.extendoffice.com/>
<https://www.excel-downloads.com/>
<https://formafly.co/>
<http://www.amperiste.fr/>
<https://fr.wikipedia.org/>
<https://support.microsoft.com/>

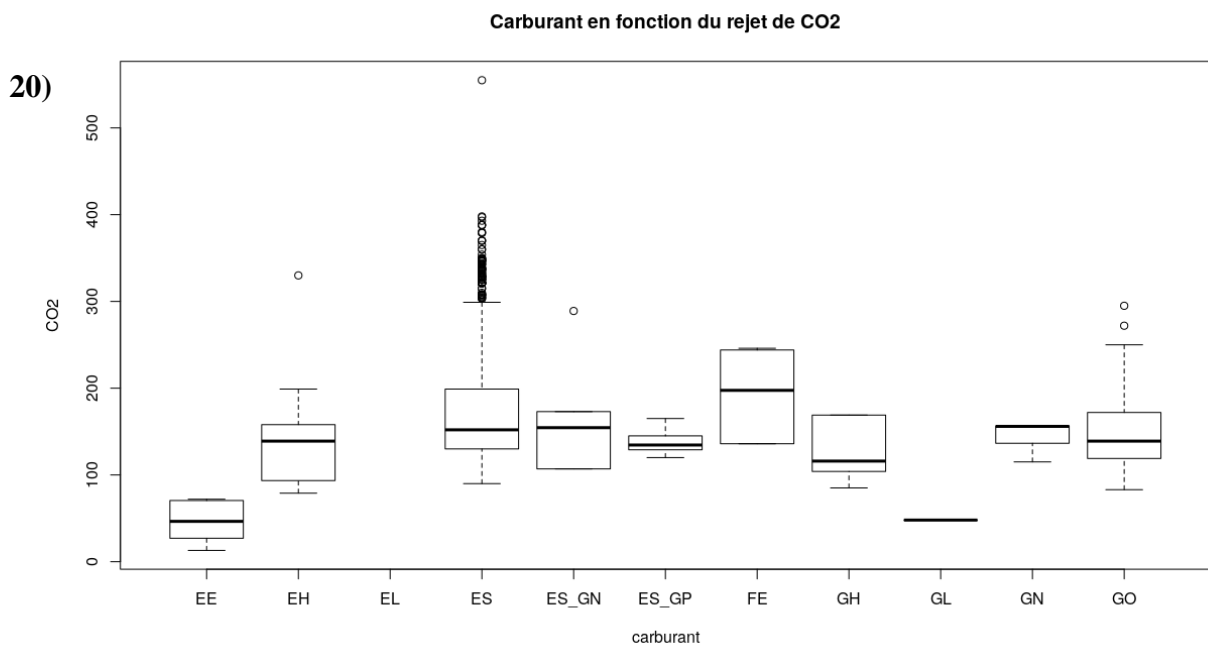
VI. Annexes



Effectifs des carburants

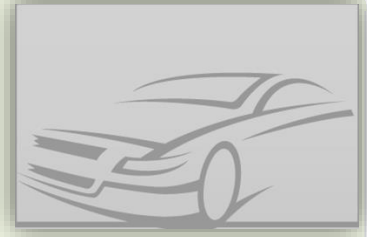
19)

col_carb											
EE	EH	EL	ES	ES_GN	ES_GP	FE	GH	GL	GN	GO	
8	43	20	1587	14	14	6	13	1	7	2123	



QUESTIONNAIRE

Parlez-nous de votre voiture !



Quelle est la marque et le modèle de votre voiture ?

Votre réponse

Quelle est sa gamme ?

- ☐ ECONOMIQUE
- ☐ INFERIEURE
- ☐ LUXE
- ☐ MOY_INFER
- ☐ MOY_INFERIEURE
- ☐ MOY_SUPER
- ☐ SUPERIEURE

Quel est son type de carrosserie ?

Votre réponse

Votre voiture est-elle hybride ?

☐ oui

☐ non

Quel type de carburant utilisez-vous?

Votre réponse

Quelle est sa consommation urbaine (litre/km) ?

Votre réponse

Quelle est sa consommation ex-urbaine (litre/km) ?

Votre réponse

Quelle est sa consommation mixte (litre/km) ?

Votre réponse

Quelle quantité de CO₂ est rejetée par votre voiture (g/km) ?

Votre réponse

Quelle est sa puissance en chevaux (ch) ?

Votre réponse

Merci de votre participation !