

Master 1 – Data Analyst (2025-2026)
KADOUR Djibril et SEU Théotime

Quelles sont les variables qui influencent le plus le prix des véhicules électriques ?

Sous la direction de Mr. SAWADOGO



Introduction

À Paris, la mobilité urbaine évolue rapidement et les habitudes de déplacement sont en pleine transformation. Entre la montée des préoccupations environnementales et la mise en place progressive de restrictions sur les véhicules diesel et essence, la voiture électrique s'impose peu à peu comme une alternative incontournable. C'est dans ce contexte que nous avons choisi de nous pencher sur le prix des véhicules électriques, un sujet qui nous touche directement en tant que conducteurs parisiens, mais aussi en tant que citoyens soucieux des enjeux écologiques.

Le marché des voitures électriques connaît aujourd'hui une croissance fulgurante, portée par les politiques publiques et les avancées technologiques. Cependant, malgré cet essor, les prix de ces véhicules restent encore élevés et très variables selon les modèles. Comprendre les facteurs qui influencent ces prix représente donc un enjeu à la fois économique, technologique et sociétal.

Pour mener à bien cette étude, nous avons choisi d'utiliser une base de données issue du site Kaggle, plateforme reconnue pour la richesse et la fiabilité de ses jeux de données publics. Cette base rassemble un ensemble d'informations détaillées sur différents modèles de voitures électriques, combinant à la fois des caractéristiques techniques (autonomie, capacité de la batterie, temps de charge, type de batterie, etc.) et des caractéristiques commerciales (marque, pays de fabrication, nombre de ventes, garantie, etc.).

Notre démarche s'est déroulée en plusieurs étapes. Nous avons d'abord procédé à la préparation et la structuration des données, incluant leur nettoyage, le renommage et la création de nouvelles variables afin d'obtenir une base exploitable. Nous avons ensuite réalisé une analyse descriptive permettant d'observer la répartition du prix et les principales tendances parmi les caractéristiques techniques et commerciales. Enfin, nous avons mené une étude des relations entre le prix et les autres variables afin d'identifier les facteurs ayant le plus d'impact sur la valeur des véhicules électriques.

L'objectif final de ce projet est donc de répondre à la problématique : **Quelles sont les variables qui influencent le plus le prix des véhicules électriques ?**

Partie 1 — Préparation et structuration de la base de données

1.1 Nettoyage de la base de données

Avant de commencer l'analyse, il a fallu **nettoyer la base de données** pour obtenir des résultats fiables. Cette étape a consisté à **supprimer les variables inutiles** ou sans lien direct avec notre objectif, à savoir l'étude de la relation entre les caractéristiques techniques et le prix des véhicules électriques.

La capture ci-dessous illustre une partie du code utilisé pour cette phase de nettoyage.

```
# suppression des variables inutiles
elec_vehicule$CO2_Emissions_g_per_km <- NULL
elec_vehicule$Vehicle_ID <- NULL
elec_vehicule$Autonomous_Level <- NULL
elec_vehicule$Safety_Rating <- NULL
```

Ce code supprime les variables inutiles à notre analyse.

Par exemple, la colonne `CO2_Emissions_g_per_km` contient uniquement des valeurs nulles puisque les véhicules sont électriques. Les colonnes `Vehicle_ID`, `Autonomous_Level` et `Safety_Rating` n'apportent pas non plus d'informations utiles pour l'étude du prix. Elles ont donc été retirées afin d'alléger la base et de ne garder que les données pertinentes.

Après cette étape de nettoyage, la base de données est devenue plus claire et centrée sur les informations réellement utiles à notre analyse. Nous avons alors pu passer à la phase suivante : le renommage des variables, afin de rendre le jeu de données plus lisible et homogène pour la suite du travail.

1.2 Renommage des variables

Pour faciliter la lecture et l'analyse, certaines colonnes ont été renommées avec des termes plus simples et cohérents. L'objectif était d'obtenir un tableau plus clair, avec des noms uniformes et compréhensibles, notamment en français.

Le code suivant montre la commande utilisée pour effectuer ce renommage dans le dataframe :

```
# Renommer le nom des variables dans le dataframe
names(elec_vehicule) <- c("marque", "modele", "annee", "type_batterie", "capacite_batterie", "autonomie_km", "type_charge",
  "temps_charge_H", "prix_usd", "couleur", "pays_fabrication", "nbr_ventes", "garantie_ans")
```

Ce renommage rend la base de données plus intuitive à manipuler. Par exemple, des variables comme **battery_capacity** ou **range_km** deviennent **capacité_batterie** et **autonomie_km**, ce qui améliore la lisibilité des analyses à venir.

Cette étape conclut la phase de préparation des données, en assurant une base de travail claire, cohérente et adaptée à la problématique.

1.3 Création de variable

Afin d’harmoniser les données et de faciliter leur interprétation, nous avons choisi de convertir le prix des véhicules de dollars américains (USD) en euros (€). Cette transformation permet de travailler avec une unité monétaire plus cohérente avec le contexte européen de l’étude.

Le code ci-dessous illustre la création d’une nouvelle variable appelée `prix_eur`, obtenue en appliquant un taux de change de 0,93 (cours du dollar observé le 8 octobre 2025).

```
# convertir le prix USD en € (taux de change = 0.93, trouvé sur internet le 08/10/2025)
elec_vehicule$prix_eur <- elec_vehicule$prix_usd * 0.93
prix <- elec_vehicule$prix_eur
```

Cette nouvelle variable, exprimée en euros, offre une base plus claire pour comparer les modèles et réaliser les analyses statistiques à venir.

Nous avons également créé une nouvelle variable appelée **“catégorie de prix”**, permettant de classer les véhicules en quatre groupes distincts : ***prix bas, prix moyen-bas, prix moyen-haut et prix haut.*** Cette manipulation a pour but de transformer la variable **“prix”** en une variable qualitative, facilitant ainsi les comparaisons avec d’autres variables.

```
#ranger les prix des véhicules dans des classes (prix bas, moyen bas, moyen haut et haut)
elec_vehicule$catégorie_prix <- cut(elec_vehicule$prix_eur,
                                   breaks = quantile(elec_vehicule$prix_eur, probs = seq(0, 1, 0.25)),
                                   include.lowest = TRUE, labels = c("Bas", "Moyen-Bas", "Moyen-Haut", "Haut"))
catégorie_prix <-elec_vehicule$catégorie_prix
```

1.4 Description des variables

Voici un tableau récapitulant les différentes variables de la base et leur signification, afin de mieux comprendre les données utilisées et de suivre plus facilement la suite du rapport.

Description des variables		
Nom de la variable	Description	Type de variable
marque	Indique le constructeur ou la marque du véhicule	Qualitative
modele	Spécifie le modèle du véhicule au sein d’une marque	Qualitative
annee	Année de fabrication ou de commercialisation du véhicule.	Quantitative
type_batterie	Type de batterie utilisée	Qualitative
capacite	Capacité de la batterie en kWh	Quantitative
autonomie	Distance maximale que le véhicule peut parcourir avec une charge complète (en km)	Quantitative

type_charge	Type de charge supportée	Qualitative
temps_charge	Temps nécessaire pour recharger complètement la batterie	Quantitative
prix_usd	Prix du véhicule en dollars américains	Quantitative
couleur	Couleur principale du véhicule	Qualitative
pays	Pays de fabrication du véhicule	Qualitative
nbr_ventes	Nombre d'unités vendues en 2024	Quantitative
garantie	Durée de la garantie constructeur	Quantitative
prix	Prix du véhicule en euros	Quantitative
categorie_prix	Prix des véhicules rangé par classe (bas,moyen-bas,moyen-haut, haut)	Qualitative

Cette première étape nous a permis de mettre en place une base de données claire, cohérente et prête à être exploitée. Après le nettoyage, le renommage et la création des variables nécessaires, nous disposons désormais d'un jeu de données structuré, plus lisible et adapté à notre objectif d'analyse.

Nous pouvons à présent passer à l'analyse descriptive, afin d'observer les premières tendances et de mieux comprendre la répartition des prix et des caractéristiques des véhicules électriques.

Partie 2 — Analyse descriptive

L'analyse descriptive permet d'examiner les données afin d'en dégager les principales tendances et caractéristiques. Elle offre une vision d'ensemble du jeu de données, met en évidence les valeurs extrêmes ainsi que la répartition des différentes variables. Cette étape constitue une base essentielle pour orienter les analyses statistiques plus approfondies.

Dans cette partie, nous nous concentrons d'abord sur l'étude du prix des véhicules électriques, puis sur la description des autres variables explicatives, afin de mieux comprendre le profil général du jeu de données.

2.1 Analyse descriptive du prix

Le prix constitue la variable principale de notre étude, car il s'agit de la variable expliquée que nous cherchons à interpréter à partir des autres caractéristiques techniques et commerciales du véhicule. L'objectif de cette partie est d'analyser sa répartition, ses tendances générales et sa normalité.

Tout d'abord, nous avons utilisé la commande `summary(prix)` afin d'obtenir les principaux indicateurs statistiques descriptifs.

```
summary(prix)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
27913  56969   84565   84269 111786 139480
```

Le prix minimum observé est de 27 913 €, tandis que le maximum atteint 139 480 €.

La médiane, située à 84 565 €, indique que la moitié des véhicules coûte moins que cette valeur.

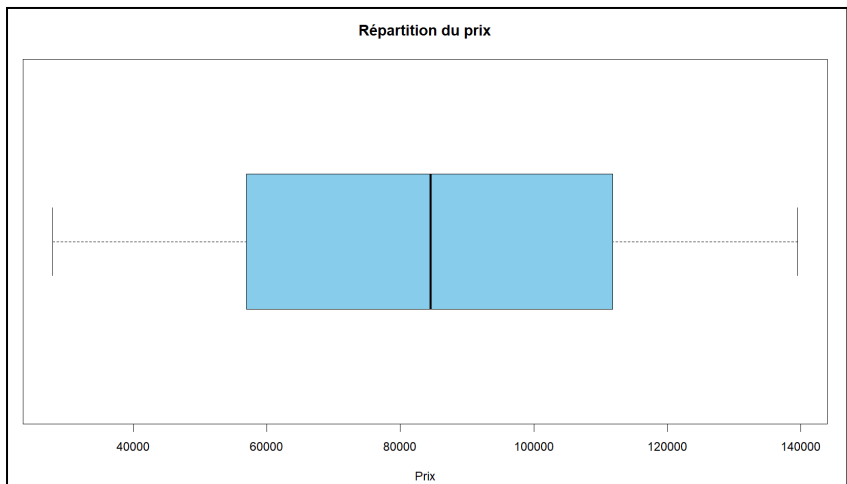
La moyenne est très proche (84 269 €), ce qui montre une répartition relativement équilibrée, sans forte asymétrie.

```
> sd(prix)
[1] 32228.72
```

Cependant, l'écart-type de 32 228 € révèle une grande dispersion des prix, signe que les modèles étudiés couvrent un large éventail — allant des voitures compactes plus abordables aux modèles haut de gamme.

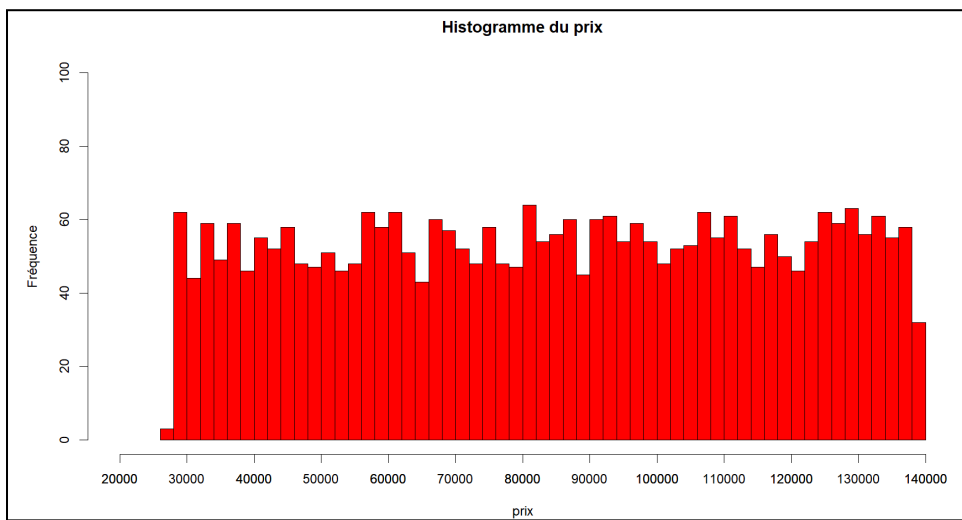
Visualisation de la répartition du prix :

Le boxplot ci-dessous permet de visualiser la répartition générale du prix :



On observe que la majorité des véhicules se situent entre **56 000 € et 111 000 €**, avec peu de valeurs extrêmes. La boîte est relativement large, confirmant la diversité des modèles présents dans le jeu de données. Cette dispersion reflète bien les différences de gamme et de technologie proposées sur le marché.

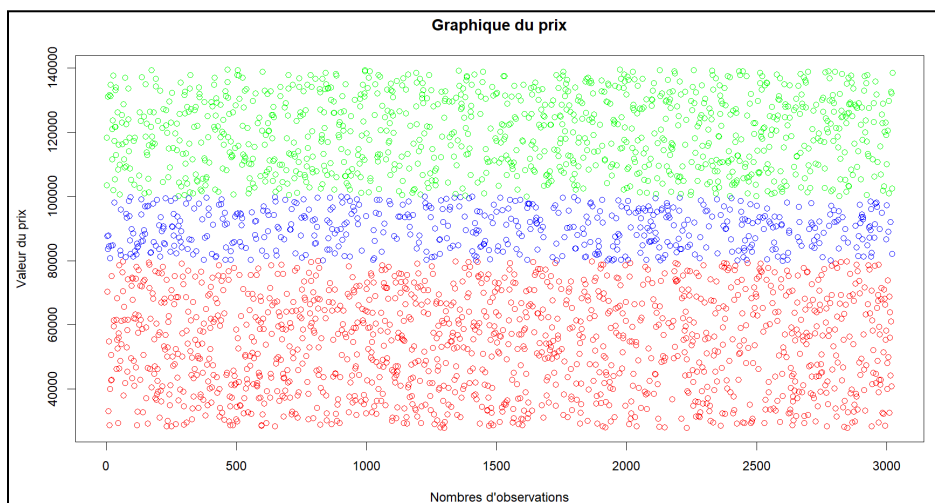
L'histogramme suivant illustre la distribution plus en détail :



La répartition est relativement **étalée et légèrement irrégulière**, ce qui traduit la diversité de l'offre sur le marché — entre modèles d'entrée de gamme et haut de gamme.

Représentation des observations

Le graphique ci-dessous montre la **dispersion des prix selon le nombre d'observations** :



Deux tests statistiques ont été réalisés pour vérifier la normalité de la distribution du prix.

Le test de Kolmogorov-Smirnov affiche une *p-value* de $2,76 \times 10^{-9}$, tandis que le test de Shapiro-Wilk donne une *p-value* inférieure à $2,2 \times 10^{-16}$.

Ces valeurs, nettement inférieures au seuil de 0,05, indiquent que la distribution du prix s'écarte significativement d'une loi normale.

```
> ks.test(prix,pnorm, mean(prix),sd(prix))  
  
Asymptotic one-sample kolmogorov-smirnov test  
data: prix  
D = 0.058095, p-value = 2.767e-09  
alternative hypothesis: two-sided
```

```
> shapiro.test(prix)  
  
shapiro-wilk normality test  
data: prix  
W = 0.95501, p-value < 2.2e-16
```

Dans les deux cas, les *p-values* sont très inférieures à 0,05, ce qui conduit à rejeter l'hypothèse de normalité. Autrement dit, la distribution des prix n'est pas normale. Cette absence de normalité peut s'expliquer par la forte dispersion observée, ainsi que par la coexistence de plusieurs catégories de véhicules au sein de la base. Ce résultat sera pris en compte pour le choix des méthodes statistiques dans les prochaines analyses.

2.2 Analyse descriptive des autres variables

Après l'étude du prix, nous nous intéressons maintenant aux autres variables explicatives de notre base de données.

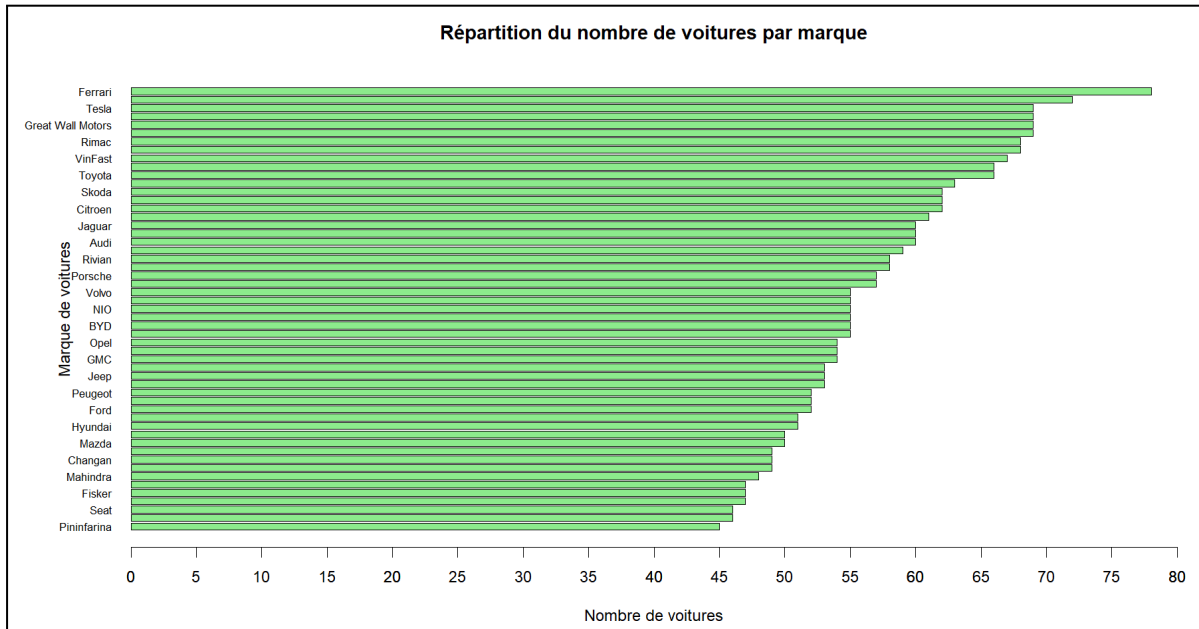
Cette analyse vise à mieux comprendre la répartition des principales caractéristiques techniques et commerciales des véhicules électriques, afin d'obtenir une vision plus complète du jeu de données.

Pour faciliter la lecture, ces variables ont été séparées en deux catégories :

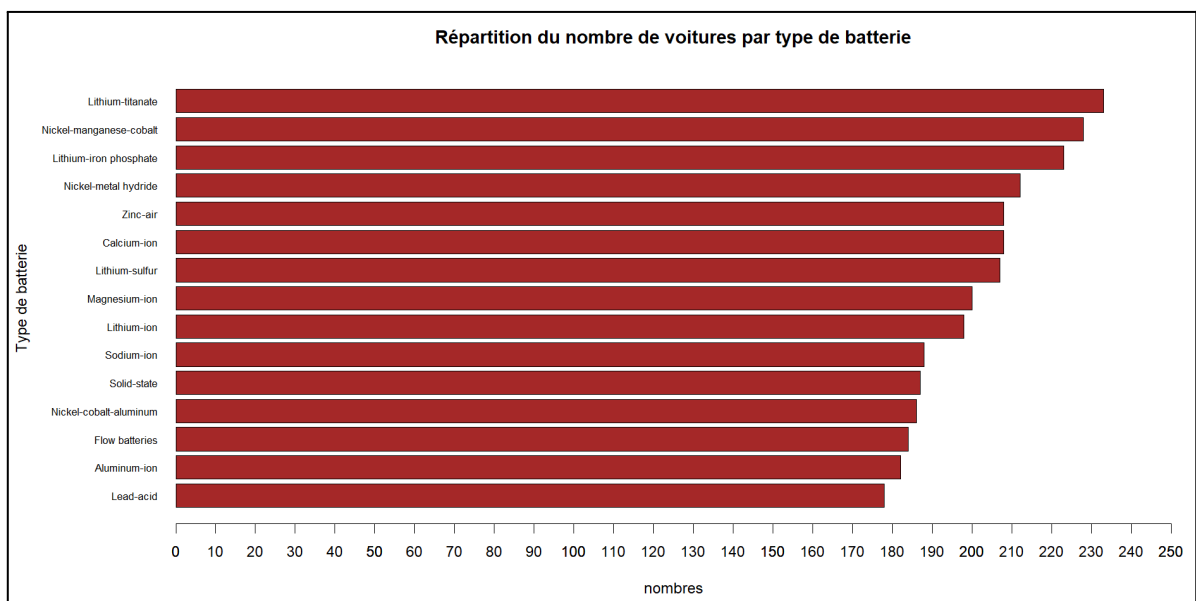
- les **variables qualitatives**, représentées à l'aide de diagrammes en barres ou en points,
- et les **variables quantitatives**, illustrées par des histogrammes permettant d'observer la distribution des valeurs.

2.2.1 Variables qualitatives

Pour mieux comprendre la composition de notre jeu de données, nous avons analysé plusieurs variables qualitatives telles que la **marque**, le **type de batterie**, le **pays de fabrication**, la **couleur** et le **type de charge**. Ces indicateurs permettent d'observer la diversité des véhicules présents dans l'échantillon et d'identifier les grandes tendances du marché.

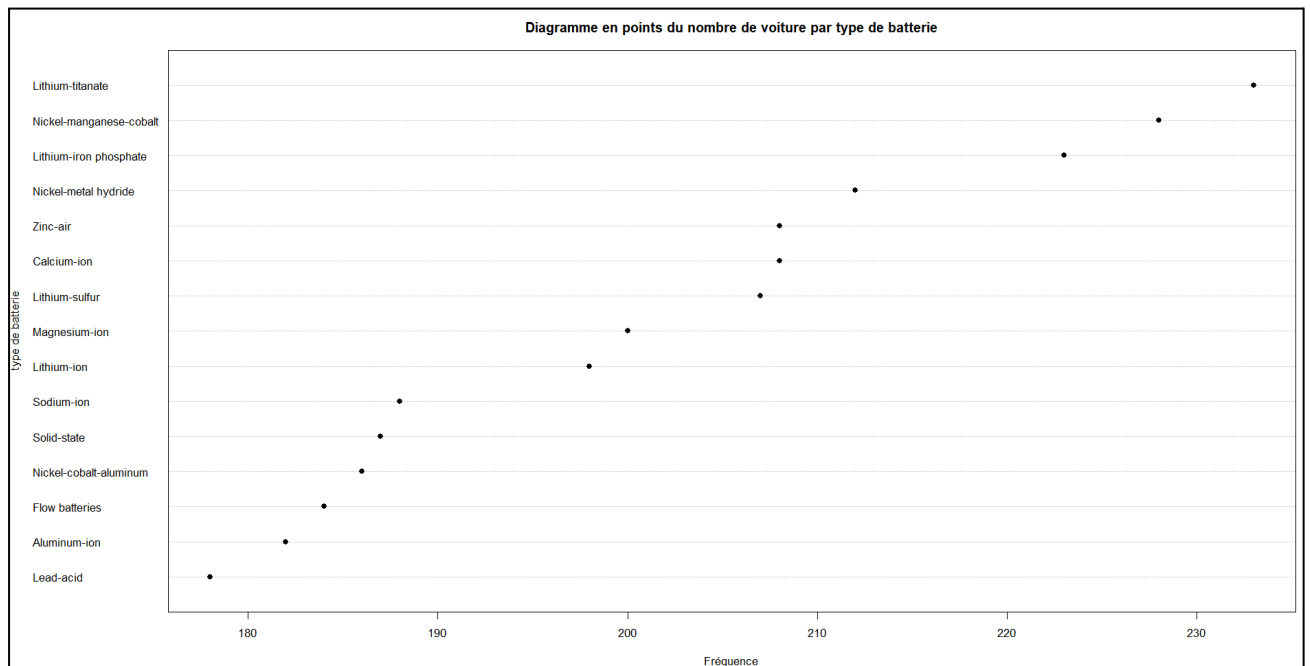


Ce graphique met en évidence une répartition assez homogène entre les différentes marques. Aucune ne domine véritablement le marché dans notre jeu de données : les constructeurs comme BMW, Tesla, Renault, Mercedes-Benz ou Nissan affichent des volumes proches. Cela traduit une forte concurrence entre les principaux acteurs, signe d'un marché dynamique et en pleine diversification.



Les types de batteries les plus représentés sont les batteries lithium-ion et lithium-fer-phosphate, technologies actuellement les plus courantes sur le marché grâce à leur équilibre entre autonomie, coût et durée de vie.

Les technologies plus récentes, comme les batteries à semi-conducteurs (solid-state) ou zinc-air, restent minoritaires, mais leur présence dans la base traduit l'émergence d'innovations dans le secteur.



Le diagramme en points ci-dessus offre une autre manière de visualiser la même information. Il met en évidence les **écarts de fréquence entre technologies** et permet de repérer plus facilement les types de batteries les plus ou les moins utilisés.

Autres variables qualitatives :

D'autres graphiques ont été réalisés pour les variables pays de fabrication, type de charge, modèle et couleur.

Ils montrent une grande diversité géographique et technologique :

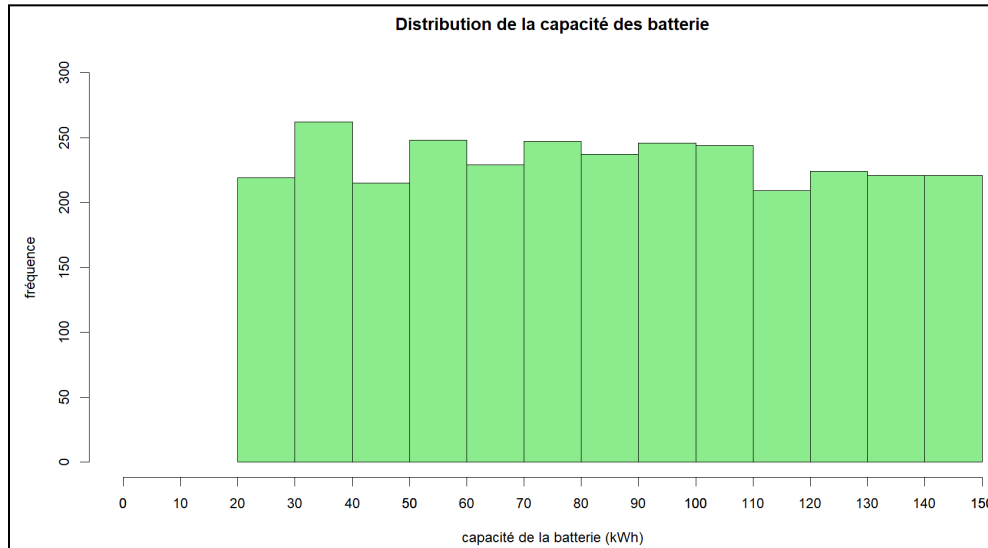
- la production est répartie entre l'Europe, l'Asie et les États-Unis, confirmant la dimension mondiale du marché ;
- les types de charge varient selon les standards (CCS, CHAdeMO, Tesla, etc.), traduisant des différences d'infrastructure selon les régions ;
- les modèles et couleurs présentent une distribution équilibrée, illustrant la variété des gammes proposées.

Remarque : Des graphiques supplémentaires sont disponibles, dans le code, pour consultation sans alourdir le rapport.

2.2.2 Variables quantitatives

Les variables quantitatives permettent d'analyser plus en détail les caractéristiques techniques des véhicules électriques.

Elles offrent des indications précieuses sur la performance, la modernité et la technologie embarquée des modèles étudiés.

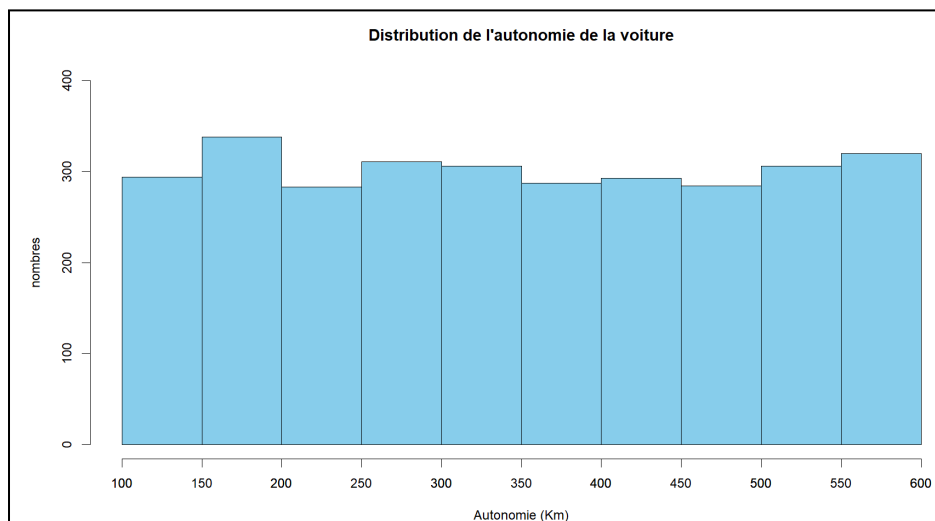


La capacité des batteries est globalement bien répartie entre 20 kWh et 150 kWh, avec une concentration plus marquée entre 40 et 100 kWh.

Cette plage correspond aux véhicules les plus courants du marché, combinant une autonomie suffisante et un coût de production raisonnable.

Les valeurs plus élevées concernent principalement

les modèles premium ou sportifs, tandis que les plus faibles se retrouvent sur les citadines compactes.



L'autonomie s'étend d'environ 100 km à 600 km, avec une majorité de modèles situés entre 200 et 400 km.

Cette dispersion montre la diversité des gammes présentes dans la base, allant des petits modèles urbains aux berlines haut de gamme.

On observe également une légère tendance à l'augmentation de l'autonomie, en lien avec les progrès technologiques et l'amélioration de l'efficacité énergétique des batteries.

Autres variables quantitatives :

Deux autres variables ont également été étudiées :

- **L'année de production**, dont la distribution montre une forte présence de modèles récents, en particulier à partir de 2020, ce qui reflète la montée en puissance du marché électrique.
- **Le temps de charge**, dont la répartition est assez homogène entre 1 et 12 heures, traduisant la coexistence de différentes technologies de recharge (classique, rapide, ou ultra-rapide).

Ces deux histogrammes, disponibles dans le code, complètent l'analyse en soulignant la modernité des véhicules étudiés et la diversité des vitesses de charge proposées par les constructeurs.

Partie 3 – Étude des variables expliquant le prix

Dans cette partie, nous cherchons à identifier les variables qui pourraient expliquer le prix des véhicules électriques.

Pour cela, l'analyse a été divisée en deux volets : d'une part, l'étude des corrélations entre le prix et les variables quantitatives, et d'autre part, l'analyse des relations entre le prix et les variables qualitatives, afin d'obtenir une vision complète des facteurs susceptibles d'influencer le coût des véhicules.

3.1 Corrélation entre le prix et les variables quantitatives

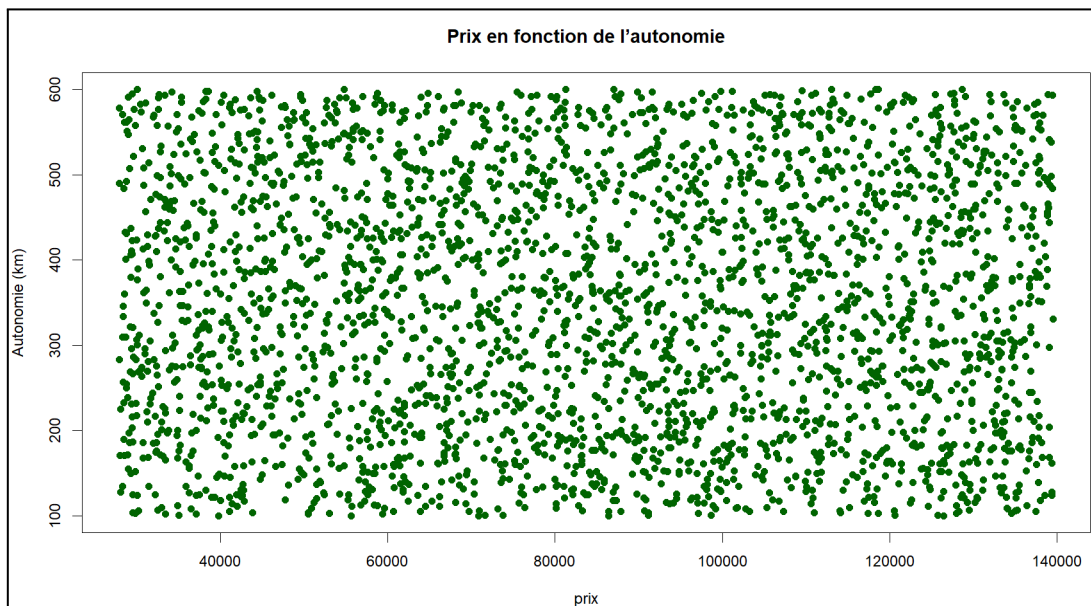
Afin d'étudier les relations entre le prix des véhicules électriques et les autres variables quantitatives, un test de corrélation de Pearson a été réalisé.

Les résultats sont présentés dans le tableau ci-dessous :

Corrélation du prix avec les variables qualitatives		
Variable corrélée avec le prix	Coefficient de corrélation	Interprétation du résultat
annee	-0.031	Corrélation très faible et négative : les véhicules plus récents ne sont pas nécessairement plus chers. L'année n'a donc pas d'influence significative sur le prix.
capacite	-0.017	Corrélation presque nulle : la capacité de la batterie n'a pas d'effet direct sur le prix. Cela peut s'expliquer par le fait que d'autres facteurs comme la marque ou le modèle compensent cet effet.
autonomie	-0.007	Corrélation presque nulle : l'autonomie n'est pas un facteur déterminant du prix dans cet échantillon. Les différences de prix peuvent donc venir d'autres caractéristiques.
temps_charge	-0.013	Corrélation très faible et négative : un temps de charge plus long n'est pas associé à un prix plus bas ou plus élevé. Cette variable semble peu pertinente pour expliquer le prix.
nbr_ventes	-0.013	Corrélation très faible et négative : les véhicules les plus chers ne se vendent pas nécessairement moins, et inversement. Le prix ne semble pas influencer directement le volume des ventes.
garantie	0.039	Corrélation très faible mais positive : les véhicules disposant d'une garantie plus longue sont légèrement plus chers, ce qui reste cohérent mais non significatif statistiquement.

En conclusion, les corrélations observées montrent qu'aucune variable quantitative n'exerce d'influence linéaire marquée sur le prix. On peut donc en conclure que le prix des véhicules électriques dépend plutôt de facteurs qualitatifs tels que la marque, le type de batterie ou encore le pays de fabrication.

Il existe également une autre manière plus visuelle de représenter les corrélations : les plots de corrélation, qui permettent d'observer rapidement les relations entre les variables grâce à une représentation graphique.



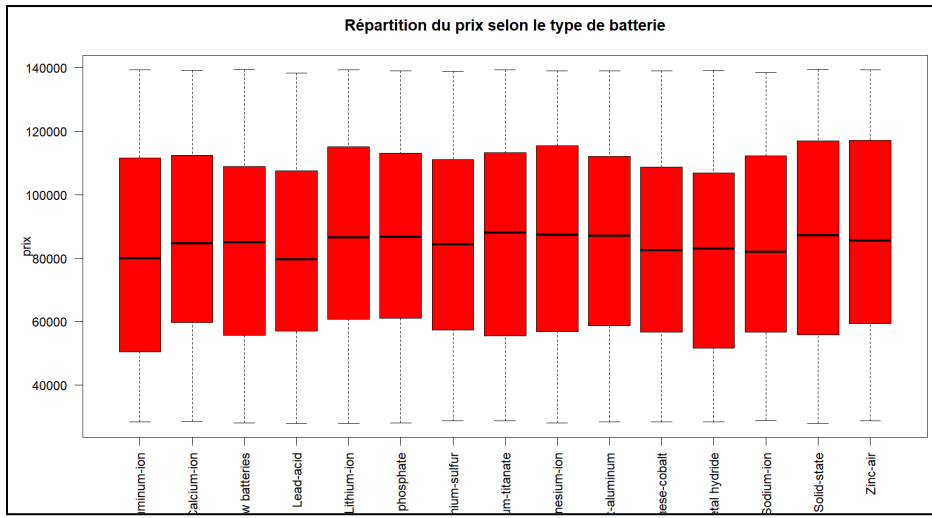
Le nuage de points montre une répartition homogène des véhicules, sans tendance apparente entre le prix et l'autonomie. Les modèles les plus chers ne présentent pas forcément une autonomie plus élevée, ce qui confirme la faible corrélation observée (-0.007).

3.2 Corrélation entre le prix et les variables qualitatives

Afin d'analyser l'influence des variables qualitatives sur le prix, des boxplots ont été réalisées. Ces graphiques permettent de visualiser la distribution du prix selon différentes catégories, telles que la marque, le type de batterie ou le pays de fabrication.

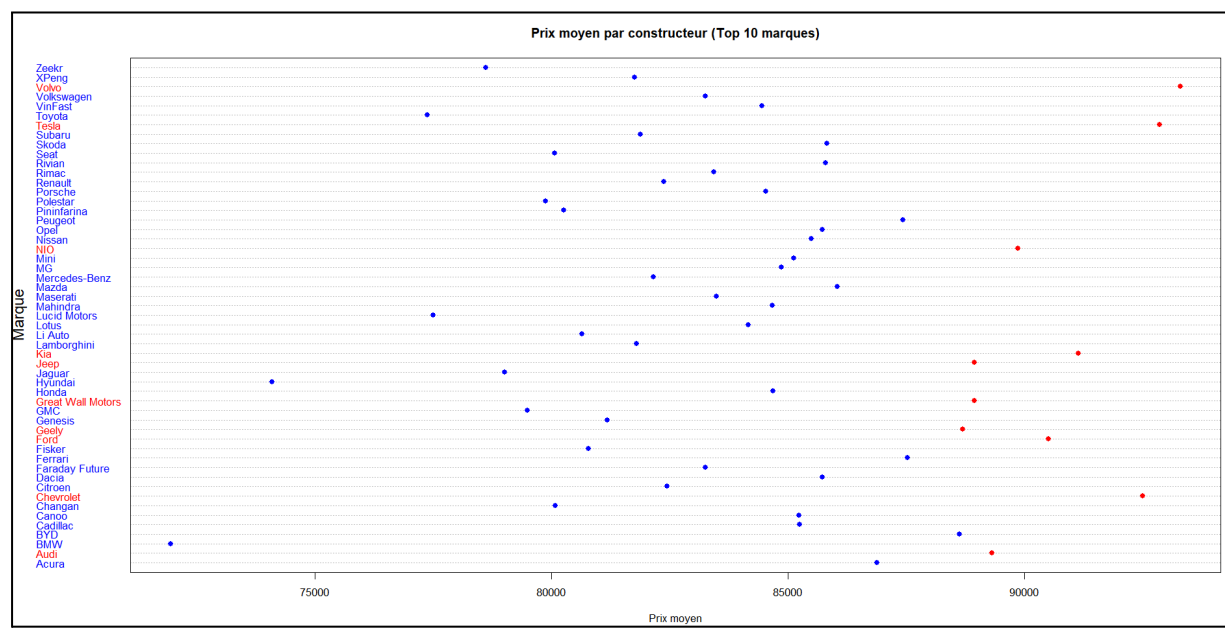
On observe une forte variabilité des prix selon la marque. Certaines marques comme Tesla, Porsche ou Lucid Motors présentent des prix médians nettement plus élevés que d'autres comme Renault, Dacia ou Kia, confirmant que la marque constitue un facteur déterminant du prix.

La répartition du prix selon le type de batterie met en évidence certaines disparités entre les technologies. On observe notamment que les véhicules équipés de batteries à semi-conducteurs (solid-state) et de batteries zinc-air présentent des prix légèrement plus élevés en moyenne, traduisant sans doute leur caractère plus innovant ou leur coût de production plus élevé. À l'inverse, les batteries plus répandues comme les lithium-ion ou lead-acid se situent dans une gamme de prix plus standardisée, reflétant une production plus maîtrisée.



Les boxplots relatifs à la marque et au pays de fabrication, disponibles dans le code, confirment également que certains constructeurs premium (comme Tesla, BMW ou Mercedes-Benz) ou pays à forte capacité technologique (États-Unis, Allemagne, Japon) se distinguent par des prix médians supérieurs. Ces observations soulignent l'importance de ces variables qualitatives dans l'explication du prix des véhicules électriques.

Une question intéressante à examiner est celle de savoir quelles marques dominent le marché des voitures électriques en termes de prix. Pour y répondre, nous avons réalisé un nuage de points représentant le prix moyen de chaque marque.



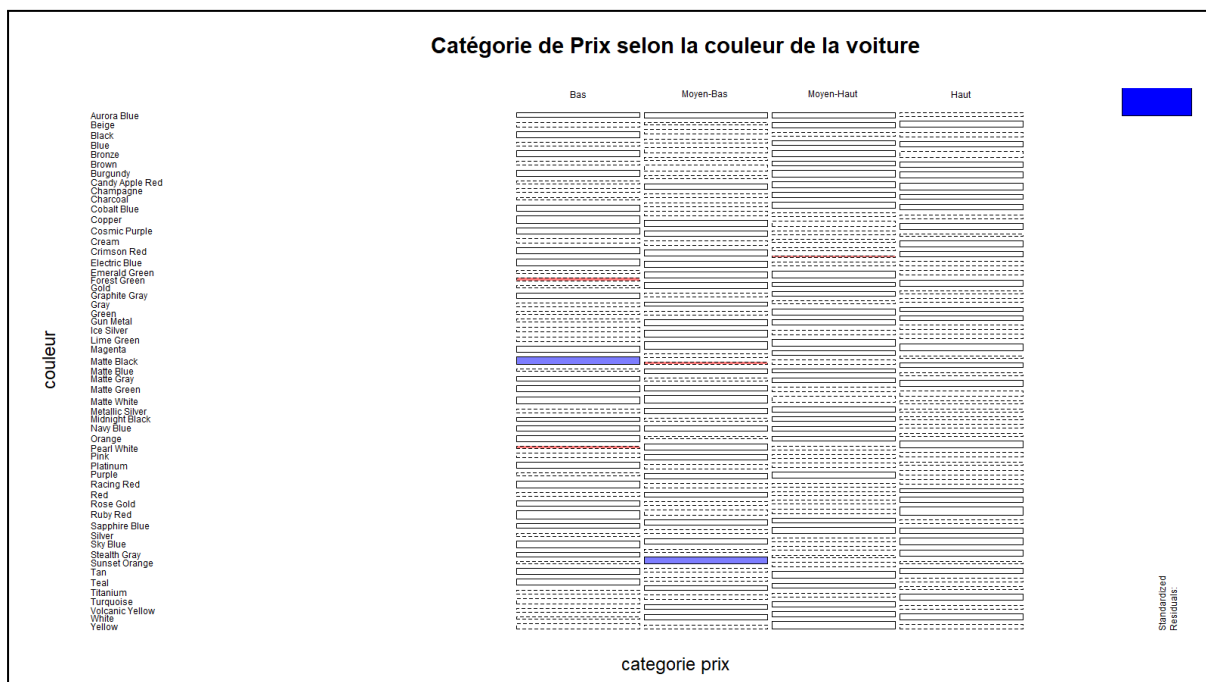
Dans cette visualisation, les 10 marques affichant les prix moyens les plus élevés sont mises en évidence en rouge, ce qui permet d'avoir une vue d'ensemble claire des marques les plus chères du marché.

3.3 Test du khi-deux (χ^2)

Résultat test du khi-deux (χ^2) de Pearson		
variable comparées	p-value	Interprétation
catégorie_prix × marque	0.317	p-value>0.05: Pas de lien significatif entre la catégorie de prix et la marque.
catégorie_prix × couleur	0.156	p-value>0.05: Pas de lien significatif entre la catégorie de prix et la couleur.
catégorie_prix × type_batterie	0.765	p-value>0.05: Pas de lien significatif entre la catégorie de prix et le type de batterie.
catégorie_prix × type_charge	0.685	p-value>0.05: Pas de lien significatif entre la catégorie de prix et le type de charge.
catégorie_prix × pays	0.546	p-value>0.05: Pas de lien significatif entre la catégorie de prix et le pays.

Les résultats du test du khi-deux montrent qu'aucune des variables qualitatives étudiées (marque, couleur, type de batterie, type de charge ou pays) n'entretiennent de lien significatif avec la catégorie de prix. Autrement dit, la répartition des prix semble indépendante de ces variables.

Afin de visualiser plus clairement cette absence de lien, le mosaïcplot ci-dessous illustre la relation entre la catégorie de prix et les principales variables qualitatives.



Ce mosaicplot représente la répartition des couleurs de voitures selon la catégorie de prix (Bas, Moyen-Bas, Moyen-Haut et Haut). On remarque que certaines cases sont colorées en rouge ou en bleu, indiquant respectivement une sous-représentation ou une surreprésentation par rapport à ce qui serait attendu.

Les zones rouges, correspondant à une sous-représentation (moins de cas que prévu), concernent les couleurs Forest Green et Pearl White pour les véhicules à bas prix, Matte Black pour les prix Moyen-Bas, et Crimson Red pour la catégorie Moyen-Haut.

Les zones bleues, indiquent une surreprésentation (plus de cas que prévu), concernent notamment la couleur Matte Black parmi les voitures à bas prix et Sunset Orange dans la catégorie Moyen-Haut. Enfin, les cases blanches traduisent l'absence de différence notable entre les catégories ou un échantillon trop faible pour tirer une conclusion fiable.

L'analyse conjointe des corrélations et des représentations graphiques montre que les variables quantitatives exercent une influence limitée sur le prix des véhicules électriques. En revanche, les variables qualitatives telles que la marque, le type de batterie et le pays de fabrication apparaissent comme des facteurs déterminants dans la formation du prix.

Ces résultats soulignent que le coût d'un véhicule électrique dépend avant tout de facteurs technologiques, de stratégies industrielles et du positionnement des constructeurs sur le marché, plutôt que de simples caractéristiques techniques mesurables.

Partie 4 – Étude statistique des sous-populations

Dans cette partie, nous avons créé deux sous-populations afin de comparer les marques haut de gamme et les marques grand public.

L'objectif est de vérifier si une différence significative de prix existe entre ces deux groupes.

Pour cela, nous avons utilisé plusieurs tests statistiques (Shapiro-Wilk, Student et Wilcoxon) afin d'évaluer la normalité des données et de comparer les moyennes entre les deux sous-populations.

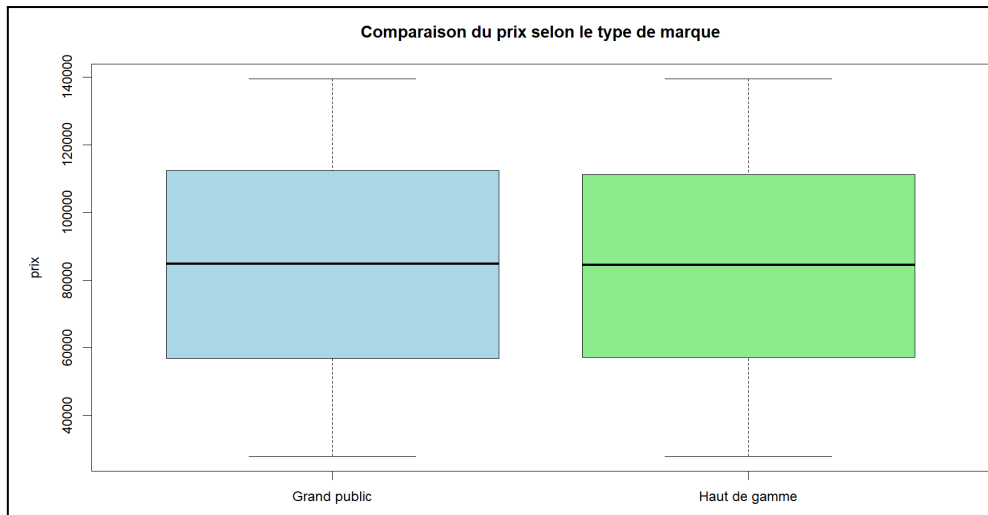
Afin d'évaluer l'influence du prix au sein de ces deux groupes, un test de Student a été appliqué pour comparer les moyennes et déterminer si les différences observées sont statistiquement significatives.

```
> t.test(prix~elec_vehicule$groupe_marque)

Welch Two Sample t-test

data: prix by elec_vehicule$groupe_marque
t = 0.49772, df = 2260.1, p-value = 0.6187
alternative hypothesis: true difference in means between group Grand public and group Haut de gamme
is not equal to 0
95 percent confidence interval:
 -1785.992  3000.966
sample estimates:
mean in group Grand public mean in group Haut de gamme
      84487.99              83880.50
```

La p-value obtenue est de 0.6187, une valeur bien supérieure au seuil de 0.05. Ainsi, on ne rejette pas l'hypothèse nulle d'égalité des moyennes. Cela signifie qu'aucune différence significative n'est observée entre le prix moyen des véhicules électriques des marques haut de gamme (84 487,99 €) et celui des marques grand public (83 880,50 €).



Nous observons sur ce boxplot que les résultats entre les marques haut de gamme et grand public sont assez similaires. Cela confirme les conclusions du t-test, qui indiquent qu'il n'existe aucune différence significative entre les deux groupes en termes de prix.

```
> shapiro.test(grand_public$prix_eur)
      Shapiro-Wilk normality test
data:  grand_public$prix_eur
W = 0.95467, p-value < 2.2e-16

> shapiro.test(haut_de_gamme$prix_eur)
      Shapiro-Wilk normality test
data:  haut_de_gamme$prix_eur
W = 0.95524, p-value < 2.2e-16
```

Comme pour l'étude de la variable prix sur la base de données initiale, nous avons souhaité examiner la normalité de la distribution des prix spécifiquement pour les voitures haut de gamme et grand public. Les résultats obtenus pour les deux groupes sont similaires, avec une p-value de 2.2×10^{-16} , bien inférieure au seuil de signification de 0.05. Nous rejetons donc l'hypothèse de normalité pour les deux distributions, indiquant que les prix ne suivent pas une loi normale dans aucun des deux cas.

```
> wilcox.test(prix ~ elec_vehicule$groupe_marque)
      Wilcoxon rank sum test with continuity correction
data:  prix by elec_vehicule$groupe_marque
W = 1063631, p-value = 0.6294
alternative hypothesis: true location shift is not equal to 0
```

Comme la normalité n'a pas été vérifiée précédemment, nous avons réalisé un test de Wilcoxon, un test non paramétrique adapté lorsque les données ne suivent pas une distribution normale. Les résultats obtenus indiquent une p-value de **0.3882**, valeur largement supérieure au seuil de signification de 0.05. Cela veut donc dire qu'il n'existe pas de différence significative entre les distributions de prix des véhicules des marques haut de gamme et grand public.

Les résultats obtenus à partir de la comparaison entre les marques haut de gamme et grand public doivent être interprétés avec prudence.

En effet, ce regroupement peut être biaisé par la présence de valeurs extrêmes ou de marques grand public proposant des modèles très chers, dont les prix rivalisent parfois avec ceux de constructeurs haut de gamme.

Pour obtenir une analyse plus précise et représentative, il peut être plus pertinent de comparer directement deux marques spécifiques, comme Nissan et Tesla, considérées comme des références majeures sur le marché des véhicules électriques.

Cette comparaison ciblée a d'ailleurs été réalisée dans le code R, afin d'affiner les conclusions de cette étude.

Conclusion

L'objectif de ce projet était de comprendre quelles variables expliquent le prix des véhicules électriques. Pour cela, nous avons croisé le prix avec différentes caractéristiques issues de la base de données, qu'elles soient techniques ou commerciales. À l'issue de cette étude, nous avons constaté que peu de variables exercent une réelle influence sur le prix, ce qui peut sembler surprenant. En effet, on pourrait s'attendre à ce que des éléments comme l'autonomie, la capacité de la batterie ou le temps de charge aient un impact clair, mais les résultats montrent le contraire. Cette absence de lien fort peut s'expliquer par le fait que le prix d'un véhicule électrique dépend souvent de facteurs non mesurés dans la base, comme la stratégie de marque, le positionnement commercial ou encore la perception de qualité et d'innovation. De plus, les caractéristiques techniques ont parfois des valeurs proches, ce qui rend difficile la mise en évidence de différences significatives. Enfin, le marché des véhicules électriques, encore jeune et en constante évolution, reste marqué par des politiques de prix très variables selon les marques et les régions. Pour aller plus loin, il serait intéressant d'intégrer des données plus détaillées sur les modèles, les options technologiques ou les coûts de production, afin d'obtenir une vision plus complète des mécanismes qui façonnent réellement le prix de ces véhicules.