



UNIVERSITÉ PARIS NANTERRE

Rapport de Projet

AutoPredict

Membres du projet :

X. Frédéric

R. Yann

R. Jérémy

8 avril 2025

Table des matières

1	Présentation de AutoPredict	2
1.1	Problématique	2
1.2	Solution	2
2	Caractéristiques Principales	3
2.1	Interface Conviviale	3
2.2	Sources Fiables	3
3	Architecture métier	4
3.1	Frontend	4
3.2	Backend	4
3.3	Base de données	4
4	Architecture distribuée	5
4.1	Application Hosting	5
4.2	Database Hosting	5
5	Pratiques de Collaboration et de DevOps	6
5.1	Project Management	6
5.2	Versionnement	6
5.3	Intégration Continue et Déploiement Continu	6
5.4	Maintenabilité du code	6
5.5	Qualité du code	6
6	Partie Data Analytique	7
6.1	Analyse de données	7
6.1.1	Préparation et nettoyage des données	7
6.1.2	Présentation des graphiques et interprétation	7
6.1.3	Analyse de la diversification de marque et de modèle	12
6.1.4	Analyse de la dépréciation	14

Chapitre 1

Présentation de AutoPredict

1.1 Problématique

Le choix d'une voiture peut s'avérer complexe pour un acheteur, en particulier lorsque de nombreux critères entrent en jeu : budget, type de motorisation, consommation, puissance, style, marque, etc. Les plateformes existantes n'offrent pas toujours une expérience personnalisée ou intuitive pour explorer l'offre de véhicules selon ses préférences réelles. Par ailleurs, les vendeurs ou les analystes souhaitent mieux comprendre les facteurs qui influencent le prix d'un véhicule, et optimiser leur stratégie de vente.

1.2 Solution

AutoPredict est une plateforme web intelligente permettant d'exploiter une base de données automobile pour proposer deux fonctionnalités principales :

- **Recherche assistée de modèles** : à partir d'un budget donné et de certaines préférences (ex. type de transmission, puissance, taille...), l'utilisateur reçoit une liste de véhicules correspondant à ses besoins.
- **Estimation de prix** : à partir des caractéristiques sélectionnées (ex. année, style, consommation...), l'utilisateur obtient une estimation de la fourchette de prix des véhicules correspondants.

Ces fonctionnalités sont rendues possibles grâce à une architecture combinant analyse de données, machine learning, backend intelligent et une interface frontend intuitive.

Chapitre 2

Caractéristiques Principales

2.1 Interface Conviviale

L'interface utilisateur est conçue avec ReactJS pour offrir une navigation fluide et interactive. L'accent est mis sur l'ergonomie, la clarté des résultats et la facilité d'utilisation même pour des utilisateurs non-experts.

2.2 Sources Fiables

Les données utilisées proviennent de datasets publics sur les voitures, incluant des caractéristiques techniques (puissance, consommation, taille, style), économiques (prix, popularité), et temporelles (année de sortie).

Chapitre 3

Architecture métier

3.1 Frontend

Le frontend est développé en ReactJS. Il intègre des composants interactifs permettant à l'utilisateur de :

- Rechercher des modèles de voitures correspondant à ses critères et à son budget
- Estimer la valeur d'un véhicule en fonction de caractéristiques spécifiques
- Visualiser graphiquement les résultats obtenus (filtres, comparateurs, graphiques de prix, etc.)

L'interface dialogue avec le backend via une API REST.

3.2 Backend

Le backend est conçu en Python à l'aide du framework Flask. Il gère :

- L'accès à la base NoSQL contenant les données automobiles
- L'exécution des modèles de machine learning pour la recommandation de véhicules et l'estimation des prix
- L'interface avec le frontend via une API structurée

Les requêtes utilisateurs sont traitées dynamiquement pour retourner des résultats adaptés et rapides.

3.3 Base de données

La base de données utilisée est de type NoSQL, permettant une flexibilité dans la gestion des formats de données hétérogènes typiques du domaine automobile.

Chapitre 4

Architecture distribuée

4.1 Application Hosting

Le projet est conteneurisé afin de faciliter le déploiement, la scalabilité et la portabilité. Docker est utilisé pour packager les composants.

4.2 Database Hosting

La base NoSQL est hébergée dans un environnement compatible cloud. Elle stocke les jeux de données enrichis et traités, accessibles via API.

Chapitre 5

Pratiques de Collaboration et de DevOps

5.1 Project Management

Le projet est géré en équipe de trois membres : Frédéric, Yann et Jérémy. Le suivi des tâches se fait de manière collaborative autour d'outils de gestion agile.

5.2 Versionnement

L'ensemble du code source est versionné via Git, avec des dépôts organisés pour le frontend, le backend et les notebooks d'analyse/ML.

5.3 Intégration Continue et Déploiement Continu

Des pipelines CI/CD seront mis en place pour automatiser les tests, le linting, et le déploiement sur l'environnement de développement.

5.4 Maintenabilité du code

L'utilisation de conteneurs, de frameworks standards (Flask, React) et de pratiques de développement modulaire assure la maintenabilité du projet.

5.5 Qualité du code

Le code est documenté, typé et validé avec des outils de linting et des tests unitaires, notamment sur les scripts de preprocessing et les modèles ML.

Chapitre 6

Partie Data Analytique

6.1 Analyse de données

6.1.1 Préparation et nettoyage des données

La phase de préparation a consisté à rendre les données cohérentes, complètes et prêtes à être visualisées. Elle s’est déroulée comme suit :

- **Standardisation** : Uniformisation des noms de colonnes en minuscules avec des underscores pour assurer une manipulation fluide.
- **Suppression des doublons** : Élimination des entrées redondantes basées sur les identifiants véhicule/modèle.
- **Traitement des valeurs manquantes** :
 - Remplacement par des valeurs par défaut ou par la moyenne (ex. nombre de portes ou de cylindres).
 - Suppression des lignes avec des données critiques absentes.
- **Filtrage des transmissions inconnues** : Les entrées comportant ‘UNKNOWN’ pour la transmission ont été exclues de l’analyse.
- **Création de nouvelles variables** :
 - Quantiles de prix (MSRP) pour catégoriser les véhicules.
 - Plages de puissance moteur pour les regrouper en catégories (‘faible’, ‘moyenne’, ‘élevée’, etc.).
 - Calcul de la consommation moyenne combinée (ville + autoroute).

6.1.2 Présentation des graphiques et interprétation

Dans cette section, nous explorons différentes visualisations de données afin d’extraire des tendances structurelles sur notre flotte de véhicules. Chaque graphique est accompagné d’une interprétation opérationnelle, utile pour guider les recommandations clients.

1. Répartition des types de transmission selon les quantiles de prix

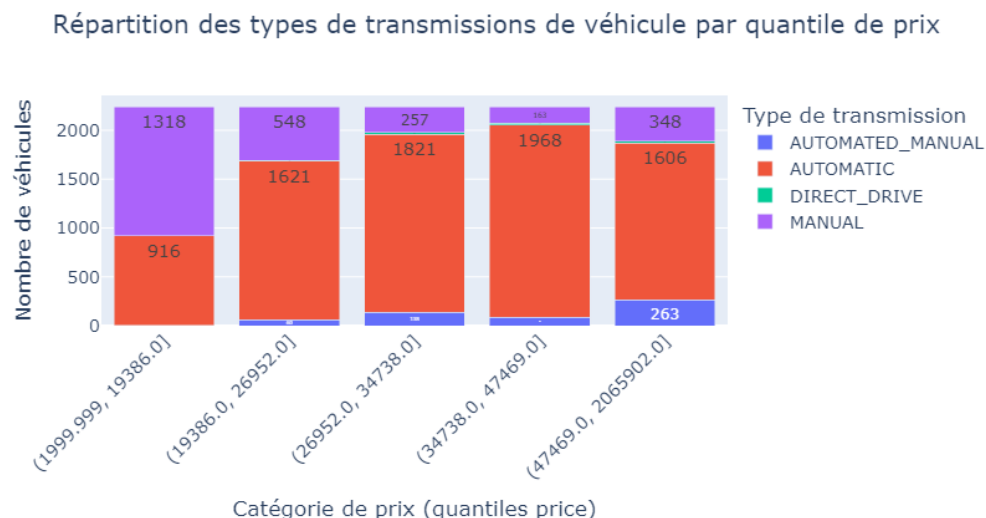


FIGURE 6.1 – Répartition des transmissions par tranche de prix (quantiles MSRP)

Ce graphique à barres empilées montre que :

- Les transmissions **automatiques** dominent très largement dans les gammes de prix *intermédiaires à élevées*, représentant parfois plus de **80 %** des véhicules.
- Les transmissions **manuelles** sont surtout présentes dans les véhicules du *premier quantile de prix*, et disparaissent progressivement à mesure que le prix augmente.
- Les transmissions **automatisées manuelles** sont rares, présentes essentiellement dans des véhicules haut de gamme ou spécifiques.

Cette distribution illustre un lien direct entre la gamme tarifaire et le type de confort/conduite attendu.

2. Répartition des styles de véhicules selon le type de transmission

Répartition des styles de véhicule selon le type de transmission

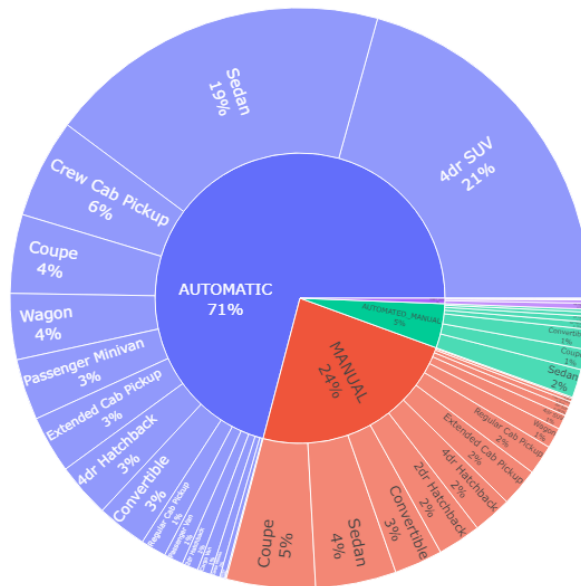


FIGURE 6.2 – Répartition des styles de véhicule selon la transmission

Le diagramme sunburst confirme les tendances précédentes :

- Les **SUV** et **berlines (sedan)** sont principalement associés aux transmissions automatiques, répondant à des besoins de confort et d'usage urbain/familial.
- Les **coupés**, **hatchbacks** ou **pickups** sont souvent en transmission manuelle, adaptés à des usages économiques, sportifs ou professionnels.
- Les transmissions rares comme **DIRECT_DRIVE** restent anecdotiques.

3. Consommation moyenne selon puissance moteur et cylindres

Consommation moyenne par puissance moteur

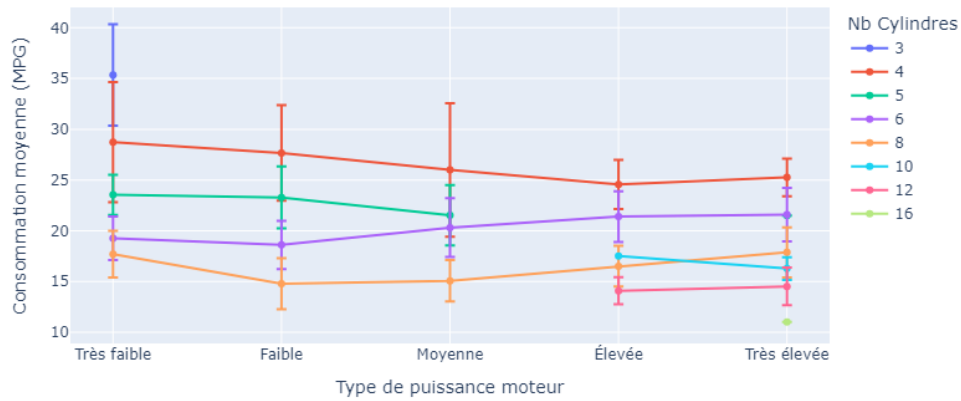


FIGURE 6.3 – Consommation moyenne selon puissance moteur et nombre de cylindres

On observe une relation logique entre la puissance du moteur et la consommation :

- Les véhicules à **puissance très élevée** et **8 cylindres ou plus** affichent une consommation moyenne plus élevée.
- Les modèles à **puissance moyenne à faible** ont des consommations plus stables et optimisées.
- Cette analyse permet de recommander les modèles selon un compromis performance/efficacité.

4. Prix moyen par catégorie de véhicule et puissance moteur

Prix moyen (price) par catégorie de voiture et puissance moteur (avec nombre de véhicul

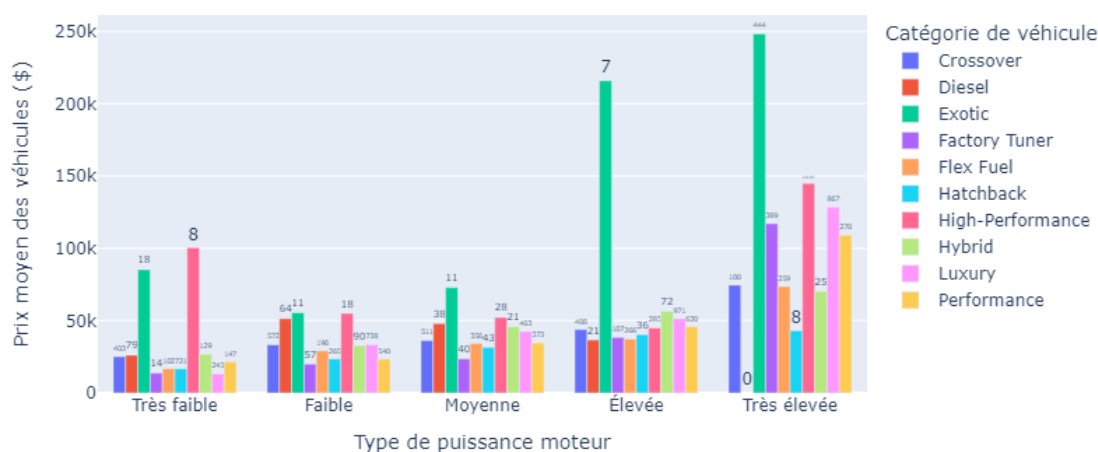


FIGURE 6.4 – Prix moyen par catégorie de voiture et puissance moteur (avec nombre de véhicules)

Cette visualisation en barres groupées révèle :

- Les véhicules **exotiques** dominent dans les plages de puissance élevée avec un prix moyen largement supérieur à 200 000\$.
- La catégorie **haut de gamme (Luxury)** est répartie sur toutes les puissances, mais fortement concentrée sur les plages hautes.
- Les **véhicules hybrides, diesel, flex fuel et compactes (Hatchback)** occupent les plages basses à moyennes, avec des prix accessibles.

Ce graphique complète la compréhension des segments en croisant l’offre produit avec les puissances moteur disponibles.

5. Profil d’achat et recommandations commerciales

À partir de l’ensemble de ces analyses, plusieurs profils-types émergent :

- Un **client urbain familial**, à la recherche de confort et de fiabilité, sera orienté vers un **SUV automatique** de gamme intermédiaire.
- Un **client professionnel ou rural** peut viser un **pickup manuel**, robuste et économique.
- Un **jeune conducteur ou petit budget** sera conseillé vers un **coupé ou hatchback manuel**, situé dans les premiers quantiles de prix.
- Pour les **amateurs de performance ou de véhicules hybrides**, on oriente vers des modèles à forte puissance ou technologies spécifiques, tout en tenant compte du marché (exotic, performance, luxury).

L’approche analytique appliquée à nos données permet donc d’alimenter directement notre moteur de recommandation personnalisé.

6.1.3 Analyse de la diversification de marque et de modèle

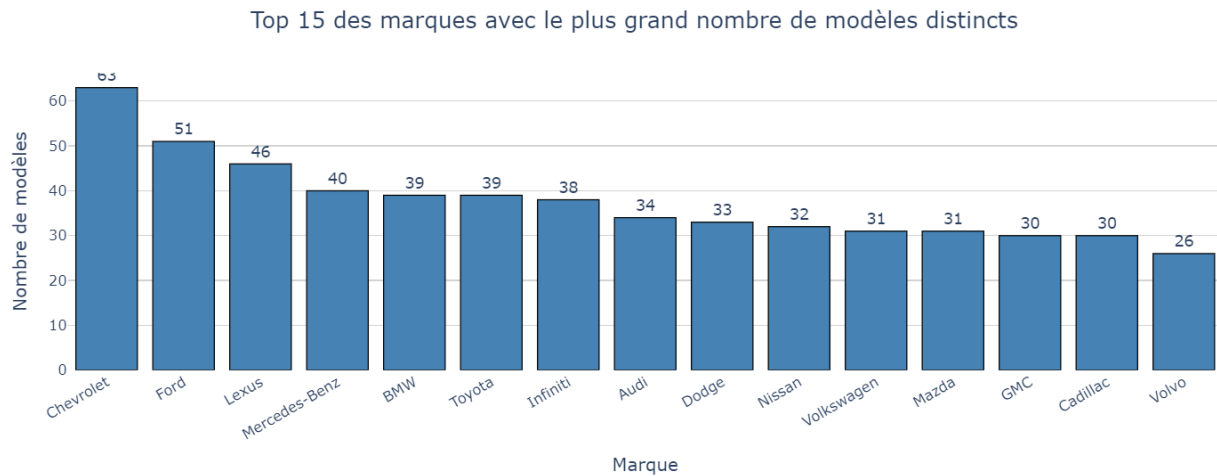


FIGURE 6.5 – Top 15 des marques avec le plus grand nombre de modèles distincts

Ce graphique met en évidence les marques proposant le plus grand nombre de modèles différents. **Chevrolet** et **Ford** occupent les premières places avec respectivement 63 et 51 modèles, témoignant d'une stratégie de gamme très étendue. D'autres marques comme **Lexus**, **Mercedes-Benz** ou **BMW** offrent également un catalogue riche, mais plus orienté haut de gamme. Ce volume important de modèles suggère une forte présence sur le marché et une capacité à répondre à des profils variés de clients.

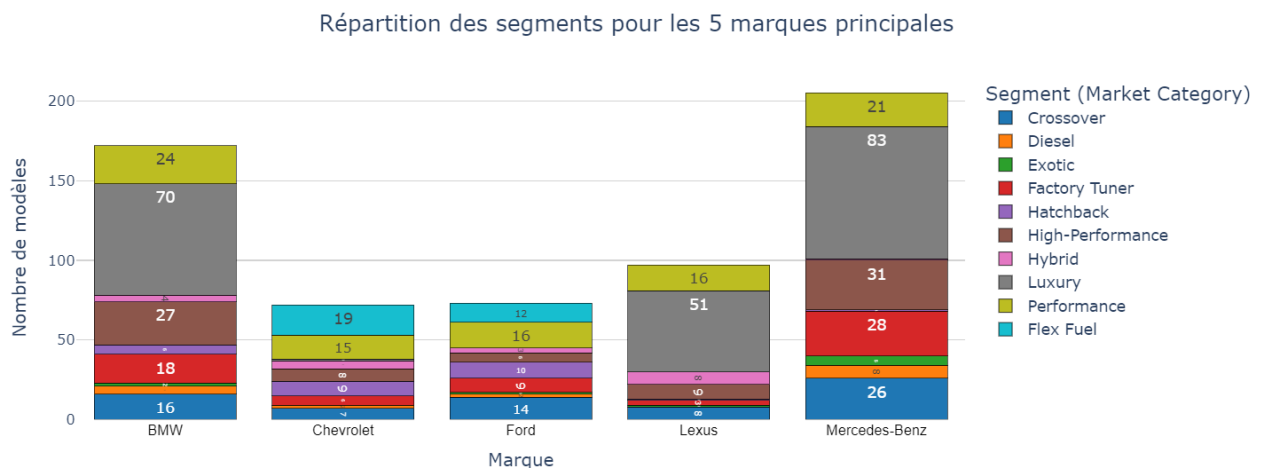


FIGURE 6.6 – Répartition des segments pour les 5 marques principales

Observations principales

- **BMW** et **Mercedes-Benz** se distinguent par une forte concentration de modèles dans le segment *Luxury*, avec respectivement 70 et 83 modèles.
- **Mercedes-Benz** est la marque la plus segmentée, avec des modèles présents dans presque toutes les catégories : Crossover, Performance, Factory Tuner, etc.
- **Chevrolet** et **Ford** affichent une répartition plus équilibrée entre les segments, bien que moins dense en nombre total.
- **Lexus** adopte une stratégie proche de Mercedes-Benz, mais dans une échelle plus réduite : forte présence sur le *Luxury*, accompagnée de quelques modèles Performance, Hybrid et Crossover.

Points clés à retenir

- Les marques comme **BMW** et **Mercedes-Benz** ciblent principalement le haut de gamme avec une concentration marquée sur le segment *Luxury*.
- **Mercedes-Benz** se distingue par une couverture quasi complète des segments, traduisant une stratégie premium mais aussi étendue.
- **Ford** et **Chevrolet** privilégient une stratégie généraliste, en diversifiant leurs modèles sur différents segments.
- **Lexus** suit un modèle proche de Mercedes-Benz, mais avec un catalogue plus restreint.

Implications pour AutoPredict

- Les marques aux offres diversifiées sont pertinentes pour répondre à différents profils clients et alimenter des recommandations plus personnalisées.
- Les marques spécialisées peuvent mieux correspondre à des besoins ciblés ou des utilisateurs ayant des préférences affirmées.
- Ces données enrichissent notre moteur de suggestion, en affinant les correspondances entre segments de véhicules et attentes utilisateurs (prix, performance, consommation, etc.).

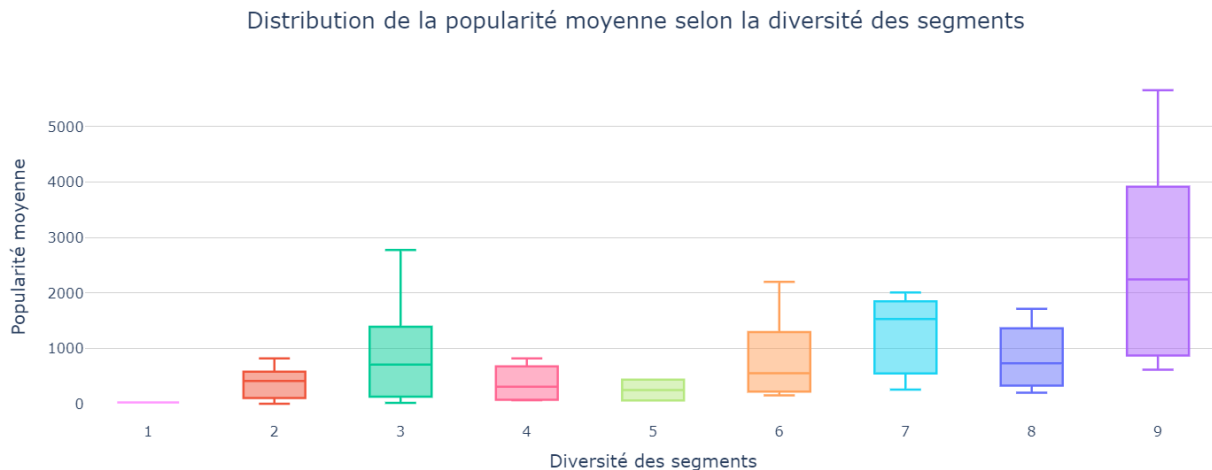


FIGURE 6.7 – Distribution de la popularité moyenne selon la diversité des segments

L'analyse de la relation entre le nombre de segments couverts par une marque et sa

popularité moyenne révèle une tendance globale : les marques les plus populaires sont souvent celles dont l'offre est la plus diversifiée.

Les marques couvrant peu de segments (1 à 3) restent relativement peu visibles, tandis que celles qui atteignent une couverture large (8 à 9 segments) bénéficient en moyenne d'un niveau de popularité plus élevé. Cette observation suggère qu'une stratégie de diversification permet de mieux répondre à la variété des attentes des clients et d'élargir la portée commerciale d'une marque.

Toutefois, cette tendance s'accompagne d'une forte dispersion : certaines marques très diversifiées restent peu populaires, tandis que d'autres, bien que spécialisées, conservent une forte notoriété. Cela montre que la diversité ne suffit pas à elle seule : l'image de marque, le positionnement marketing et la perception client jouent un rôle tout aussi déterminant.

En résumé, la diversité de l'offre constitue un levier important dans la construction d'une marque attractive. Pour AutoPredict, cet indicateur peut être intégré au moteur de recommandation pour ajuster les propositions en fonction de la stratégie produit des constructeurs et du profil de l'utilisateur.

6.1.4 Analyse de la dépréciation

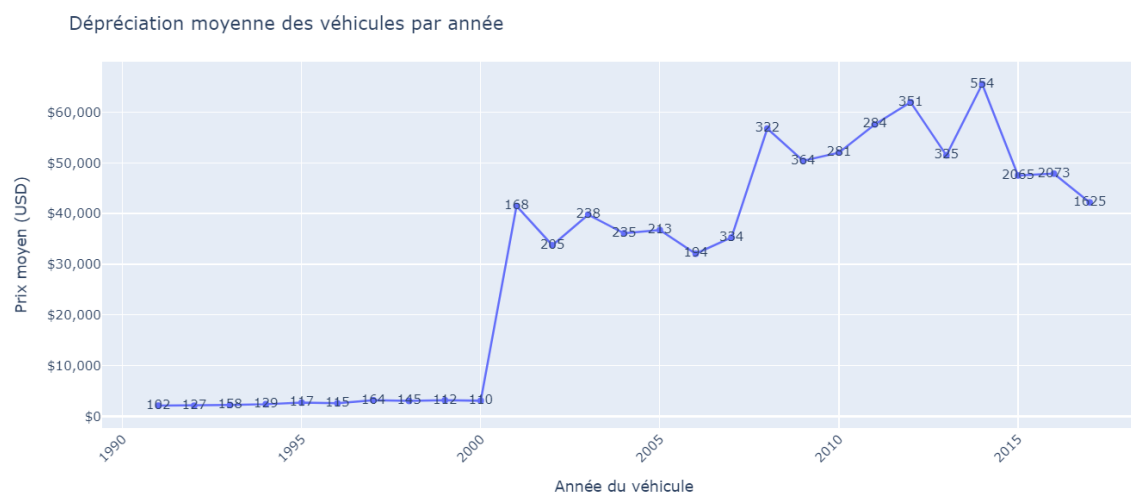


FIGURE 6.8 – Dépréciation moyenne des véhicules pas année

Cette visualisation offre une vue d'ensemble de l'évolution du prix moyen des véhicules selon leur année de fabrication, sur un marché combinant véhicules neufs et d'occasion.

On observe une baisse marquée des prix pour les véhicules produits entre 1990 et 1999, souvent vendus sous la barre des 10 000 dollars. Entre 2000 et 2002, une chute brutale est visible, possiblement liée à l'impact de nouvelles normes environnementales pénalisant les véhicules plus anciens.

À partir de 2005, la courbe devient plus régulière et progresse graduellement jusqu'à atteindre environ 37 000 dollars en 2017. Cette évolution reflète une montée en gamme progres-

sive des véhicules récents, tant en termes de technologie que de confort ou de positionnement commercial.

Dans l'ensemble, les données révèlent une dépréciation rapide des véhicules au cours de leurs 10 à 15 premières années. Un véhicule de plus de 15 ans perd une part importante de sa valeur, ce qui limite sa rentabilité en revente.

Cette première analyse met en lumière la tendance générale du marché. Elle soulève toutefois une interrogation essentielle : tous les styles de véhicules suivent-ils cette même trajectoire de dépréciation ? C'est à cette question que nous répondrons dans la section suivante.

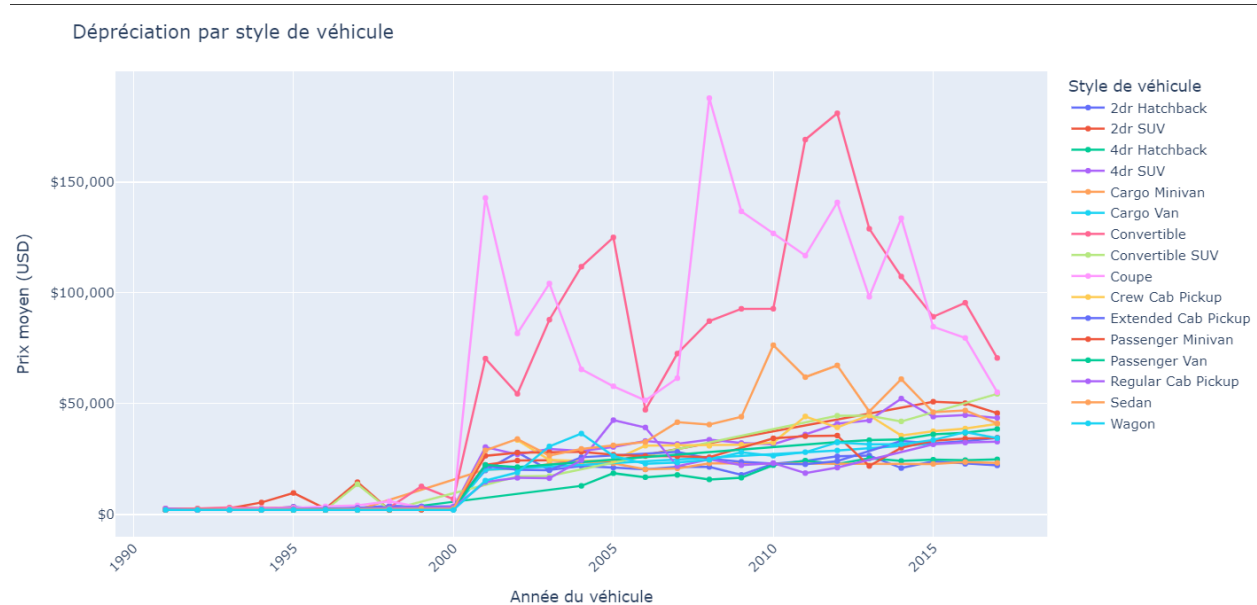


FIGURE 6.9 – Dépréciation par style de véhicule

Cette analyse explore l'impact du style de carrosserie sur la dépréciation des véhicules au fil du temps. Chaque courbe du graphique représente un type de véhicule, avec en ordonnée le prix moyen par année de mise en circulation.

Observations notables :

- Les SUV 4 portes, Convertibles et Crew Cab Pickups conservent des prix élevés, même plusieurs années après leur sortie, ce qui reflète une forte demande et une bonne rétention de valeur.
- À l'inverse, les Cargo Vans, Hatchbacks 2 portes et Sedans affichent des prix plus faibles, même pour des modèles récents, traduisant une faible valorisation sur le marché secondaire.
- Les Convertible SUV, bien que rares, se distinguent par une valeur très élevée, probablement liée à leur rareté et à leur positionnement haut de gamme.

Interprétation générale :

- Les styles utilitaires ou économiques ont tendance à perdre rapidement de la valeur.
- Les styles plaisir ou valorisants (SUV, convertible, pickup) maintiennent mieux leur prix avec le temps.
- Passé 15 ans, la plupart des véhicules, quel que soit leur style, tendent vers une valeur moyenne d'environ 10 000 dollars.

Le style du véhicule, en tant que facteur esthétique et fonctionnel, influence directement sa perception et sa valeur résiduelle. Toutefois, cette dimension visuelle s'inscrit aussi dans une stratégie de positionnement marketing, que nous approfondissons dans la prochaine section à travers l'analyse des catégories de marché.

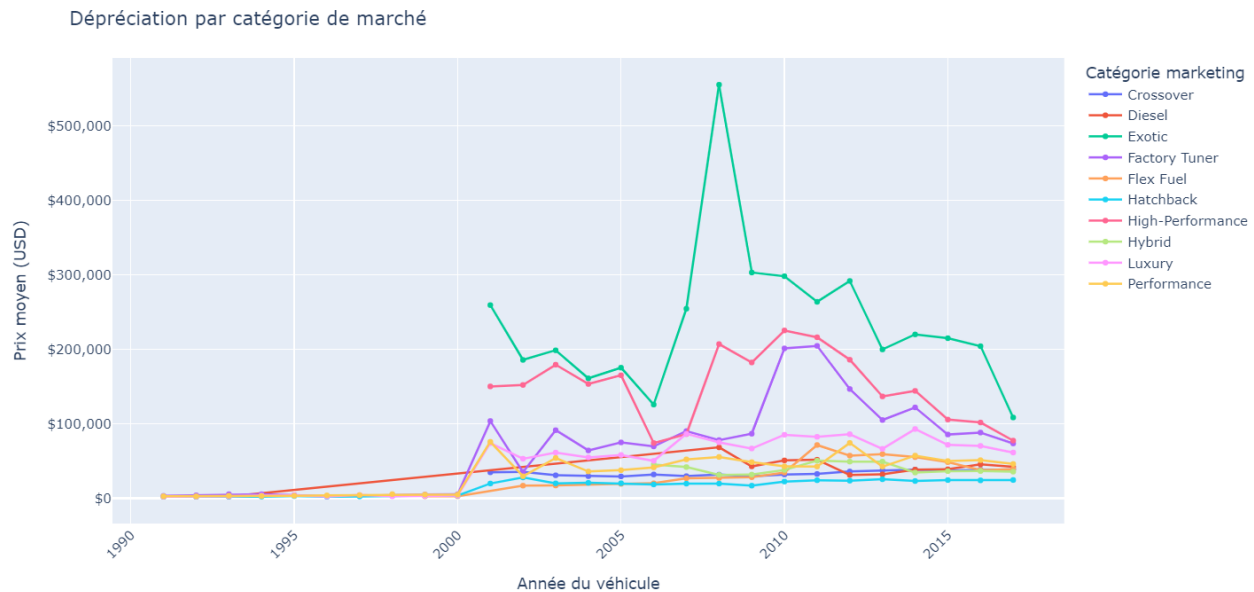


FIGURE 6.10 – Dépréciation par catégorie de marché

Cette analyse s'intéresse à la relation entre le positionnement commercial des véhicules (catégorie marketing) et leur maintien de valeur au fil du temps. Les catégories correspondent à des segments tels que *Luxury*, *Hybrid*, *Diesel*, etc., qui traduisent une stratégie de marque ciblée.

Observations clés :

- Les segments *Exotic*, *High-Performance* et *Luxury* dominent largement, avec des prix moyens dépassant fréquemment les 60 000 dollars, y compris pour des véhicules de 5 à 10 ans.
- Les catégories comme *Diesel*, *Hatchback* ou *Flex Fuel* restent bien en dessous de la barre des 30 000 dollars. Ces modèles économiques ou écologiques subissent une dépréciation plus rapide.
- Le segment *Crossover* montre une très bonne tenue de valeur, en combinant confort, modernité et polyvalence, ce qui en fait un choix populaire et stratégique.

Interprétation globale :

- La catégorie marketing influence fortement la rétention de valeur : les modèles haut de gamme, performants ou hybrides sont mieux valorisés dans le temps.
- À l'inverse, les véhicules à vocation pratique ou économique, bien qu'utiles, perdent plus rapidement de leur valeur.

Ces résultats confirment que la perception du véhicule joue un rôle central dans sa valeur résiduelle. Intégrer cette information dans le moteur de recommandation d'AutoPredict permet d'adapter les suggestions aux attentes des utilisateurs, en tenant compte de leur sensibilité au prestige, à la performance ou à l'économie.