



UNIVERSITÉ PARIS NANTERRE

# Rapport de Projet

## AutoPredict

**Membres du projet :**

X. Frédéric

R. Yann

R. Jérémie

7 avril 2025

# Table des matières

<b>1</b>	<b>Présentation de AutoPredict</b>	<b>2</b>
1.1	Problématique . . . . .	2
1.2	Solution . . . . .	2
<b>2</b>	<b>Caractéristiques Principales</b>	<b>3</b>
2.1	Interface Conviviale . . . . .	3
2.2	Sources Fiables . . . . .	3
<b>3</b>	<b>Architecture métier</b>	<b>4</b>
3.1	Frontend . . . . .	4
3.2	Backend . . . . .	4
3.3	Base de données . . . . .	4
<b>4</b>	<b>Architecture distribuée</b>	<b>5</b>
4.1	Application Hosting . . . . .	5
4.2	Database Hosting . . . . .	5
<b>5</b>	<b>Pratiques de Collaboration et de DevOps</b>	<b>6</b>
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
<b>6</b>	<b>Partie Data Analytique</b>	<b>7</b>
6.1	Analyse de données . . . . .	7
6.1.1	Préparation et nettoyage des données . . . . .	7
6.1.2	Présentation des graphiques et interprétation . . . . .	7

# 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



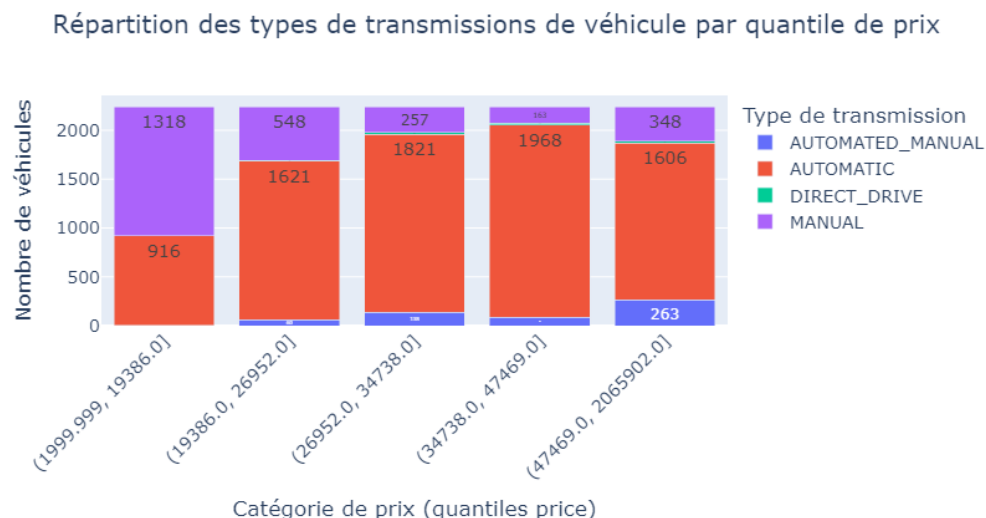


FIGURE 6.1 – Répartition des transmissions par tranche de prix (quantiles MSRP)

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

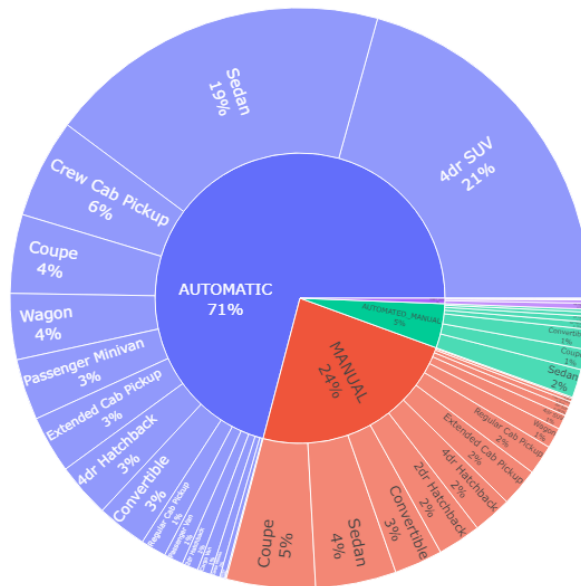


FIGURE 6.2 – Répartition des styles de véhicule selon la transmission

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

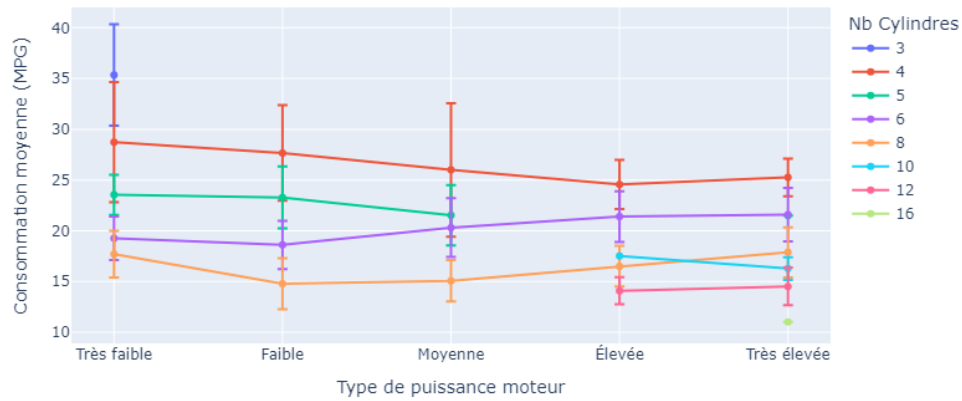


FIGURE 6.3 – Consommation moyenne selon puissance moteur et nombre de cylindres

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

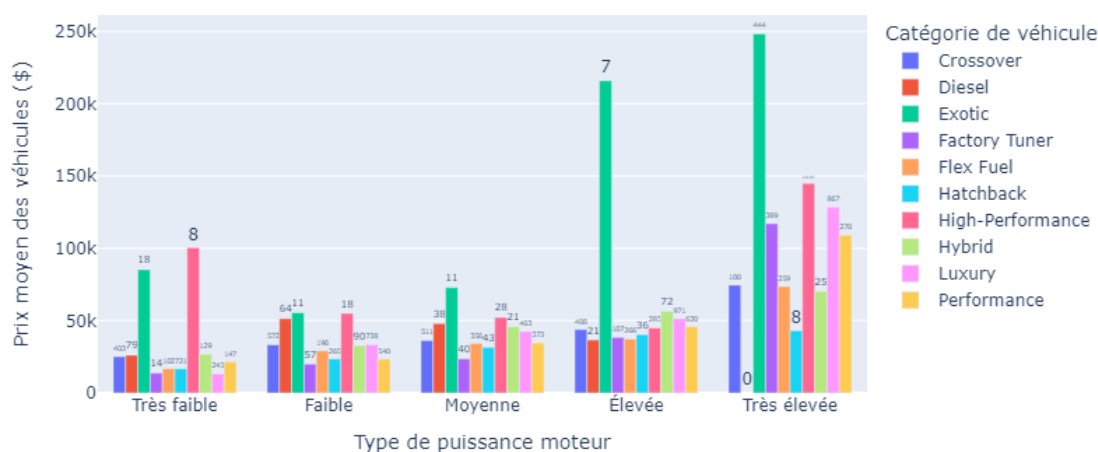


FIGURE 6.4 – Prix moyen par catégorie de voiture et puissance moteur (avec nombre de véhicules)

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.