



AIRPLANE PRICE PREDICTION

Objectif :

Prédire le **prix** d'un avion à partir de ses caractéristiques techniques.

Kaggle : [Plane price dataset](#)

Dataset : env. 500 lignes, 15 colonnes

Etapes:

- Traitement des valeurs manquantes, conversion de certaines métriques, filtrage de valeurs aberrantes.
- EDA & Visualisation
- Modèles de prédictions



E 95 Travel Air (Bonanza - turbojet)



100 Darter (S.L. Industries - piston)



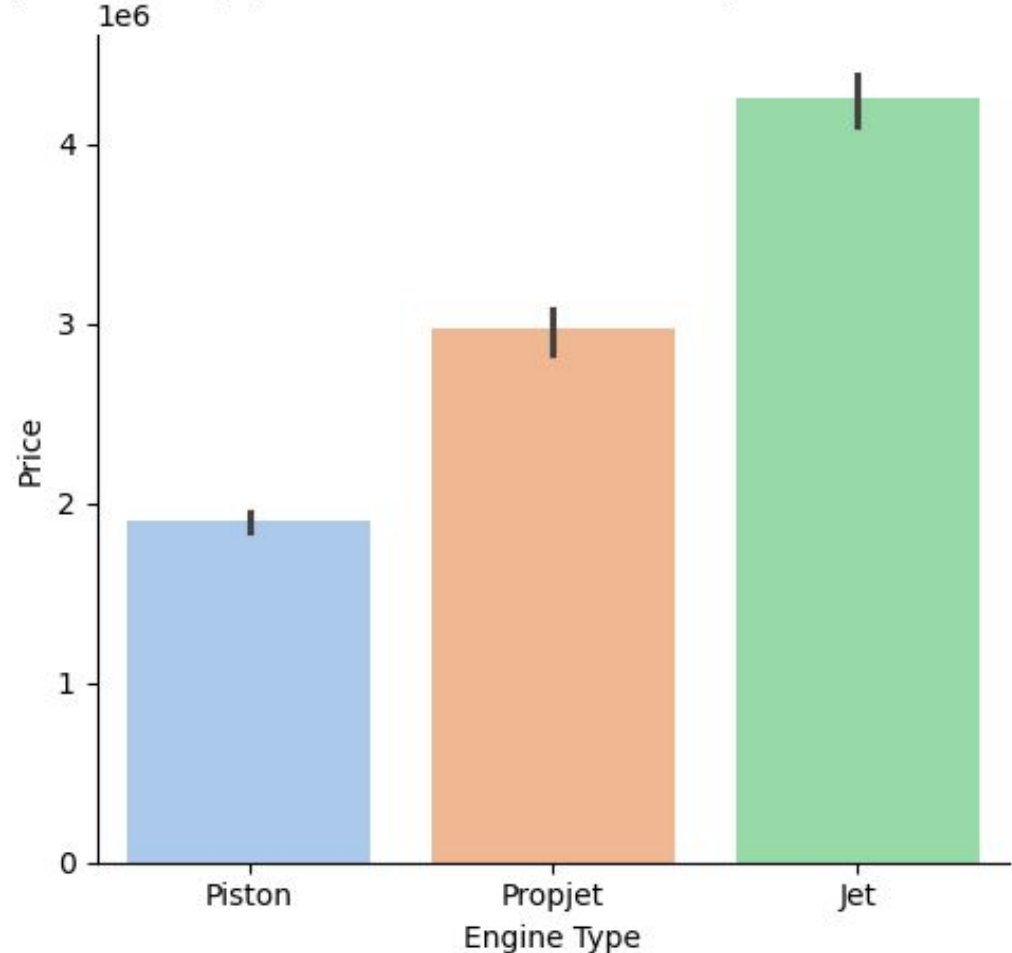
402C Business Liner (Cessna - Jet)

Peut-on visuellement détecter un impact entre des caractéristiques et la détermination du prix ?

**Variable
technologique**
Caractérise la
typologie de l'avion

Le type de moteur influence
le prix des avions et les
moteurs à piston sont
nettement moins chers que
les moteurs à propulsion.

Impact du type de moteur sur le prix d'un avion

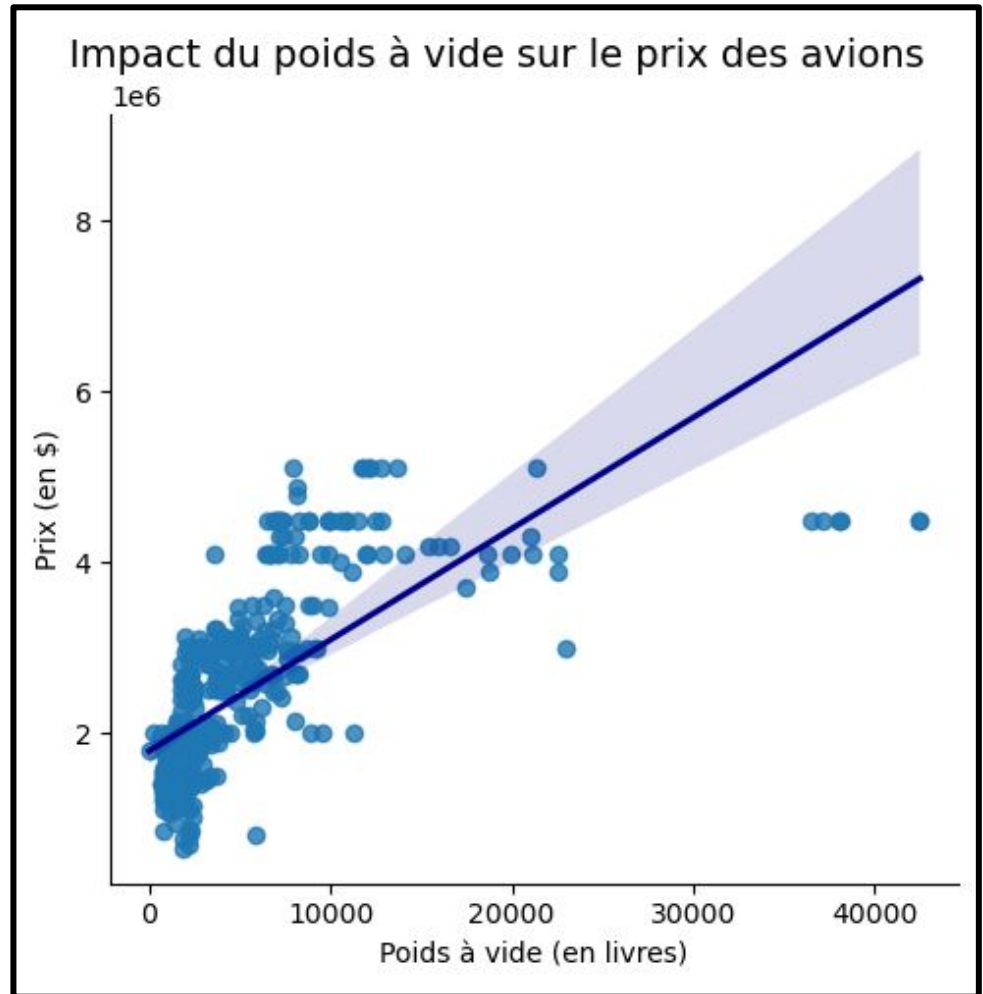


Variable physique / structurelle

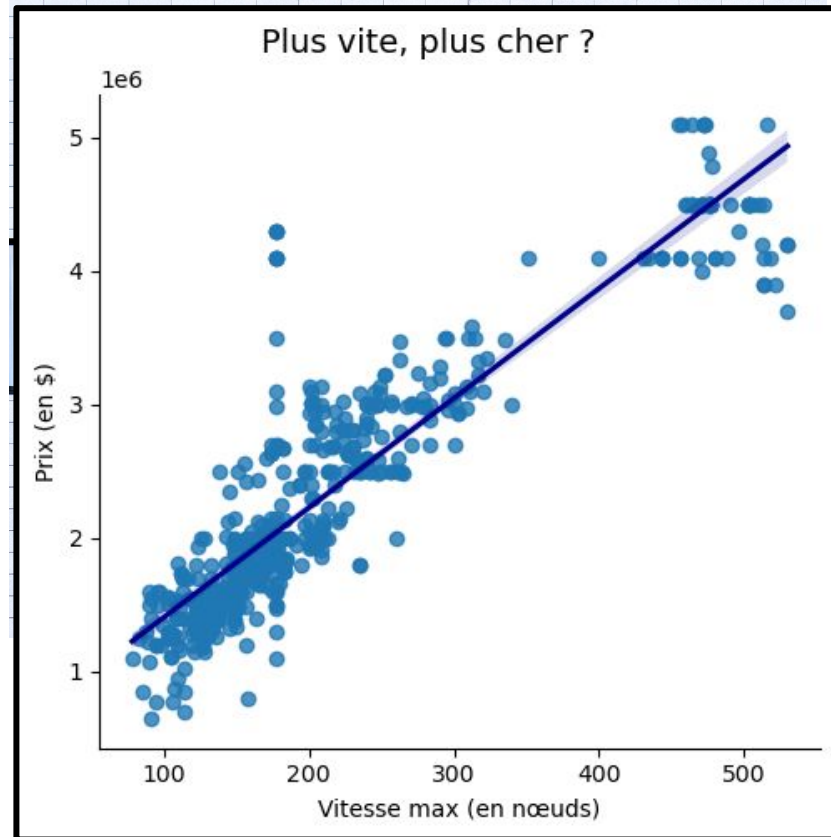
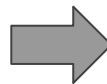
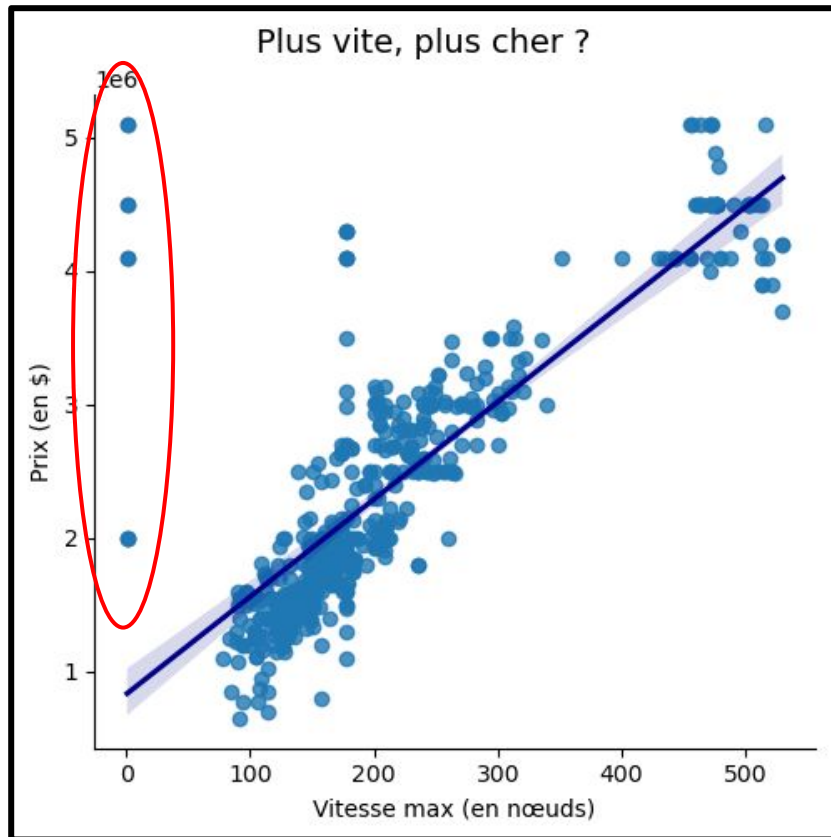
Caractérise le poids
de l'avion

Les avions **plus lourds** ont
tendance à **coûter plus cher**,
bien que la relation présente
une certaine dispersion.

Cela est cohérent : un poids
plus important peut être
associé par exemple à une
capacité plus élevée.



Variables de performances (la vitesse)

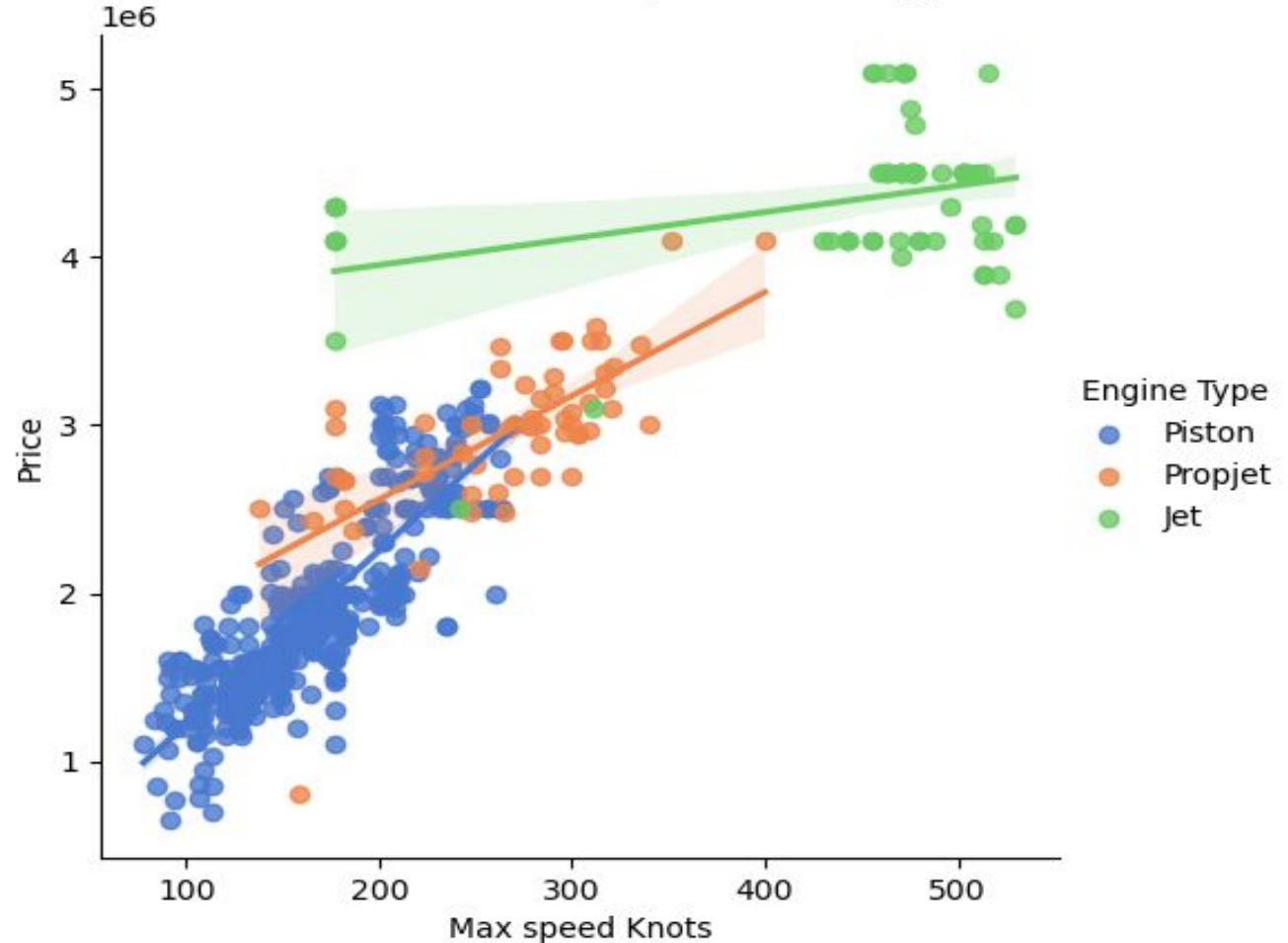


Quelques valeurs aberrantes (4) avec une vitesse maximum proche de 0 apparaissent. On choisit de les ignorer

Performance + Technologie

Cette tendance est renforcée lorsqu'on segmente les données par **type de moteur**, ce qui suggère que la vitesse est un critère **particulièrement déterminant** dans cette catégorie.

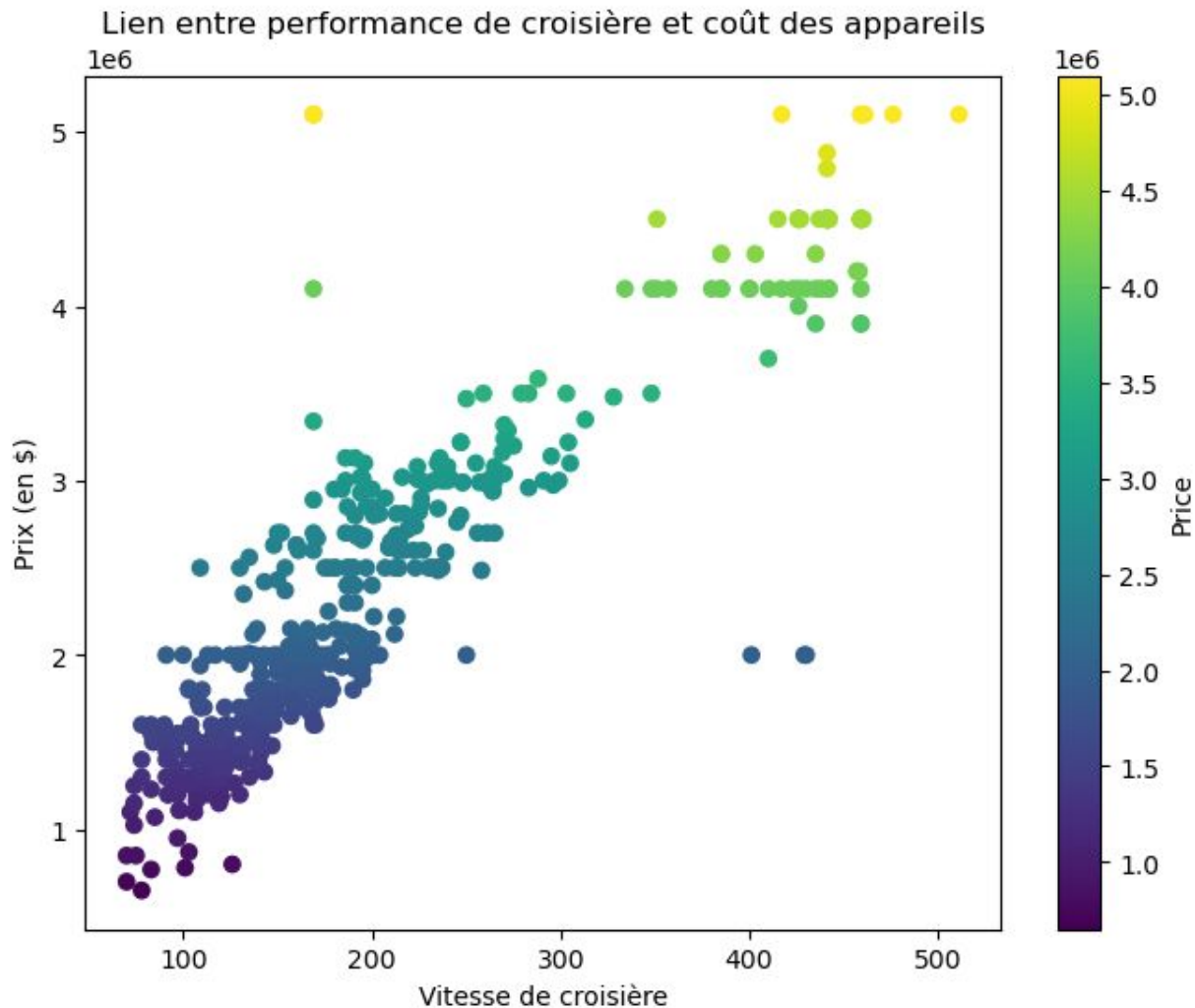
Relation entre vitesse maximale et prix selon le type de moteur



Variables de performances
Caractérisent la
vitesse et
l'efficacité en vol

Une vitesse de croisière élevée est souvent associée à un prix plus important.

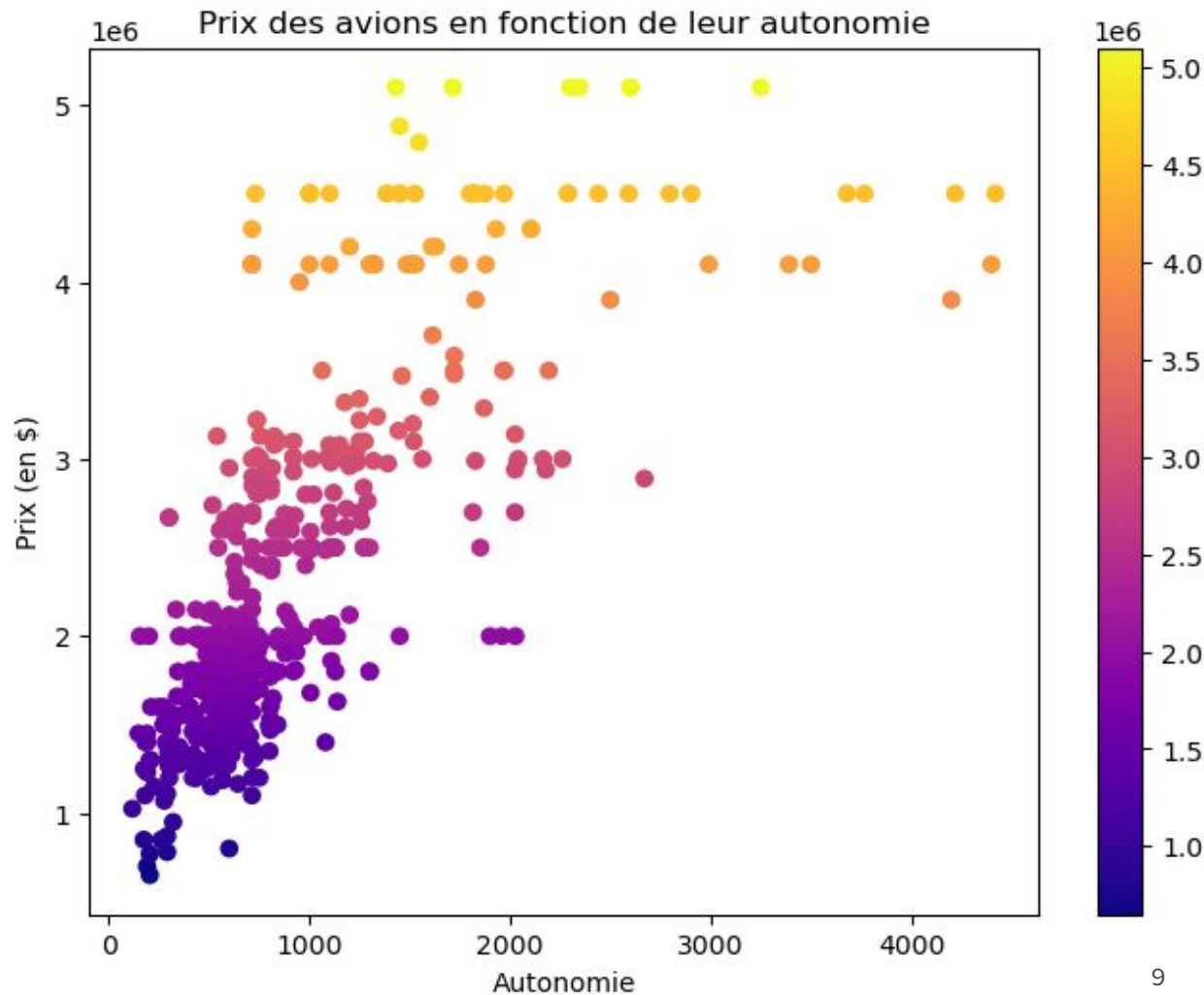
Cependant, la corrélation semble un peu **moins forte** que pour la vitesse maximale.



Variables de performances
Caractérisent la
vitesse et
l'efficacité en
vol

Une dernière
visualisation suggère
une tendance globale :

Les avions ayant une
autonomie plus
importante tendent à
être plus onéreux , bien
que des exceptions
existent.



Preprocessing / Modèles de régression

Preprocessing

- **Nettoyage** des **valeurs numériques stockées en texte** (ex. "18,300" → 18300).
- **Conversion unités** : colonnes ft/in → **metres**.
- **Valeurs manquantes** : **médiane** pour les numériques ; **mode** pour les catégorielles.

Modèles de régression

- **RandomForest Regressor**

Capte des **relations complexes**, **robuste**, peu de réglages.

Permet d'identifier les features qui ont le plus d'importances

- **Lasso**

Coefficients (direction de l'effet) & **sélection de variables**.

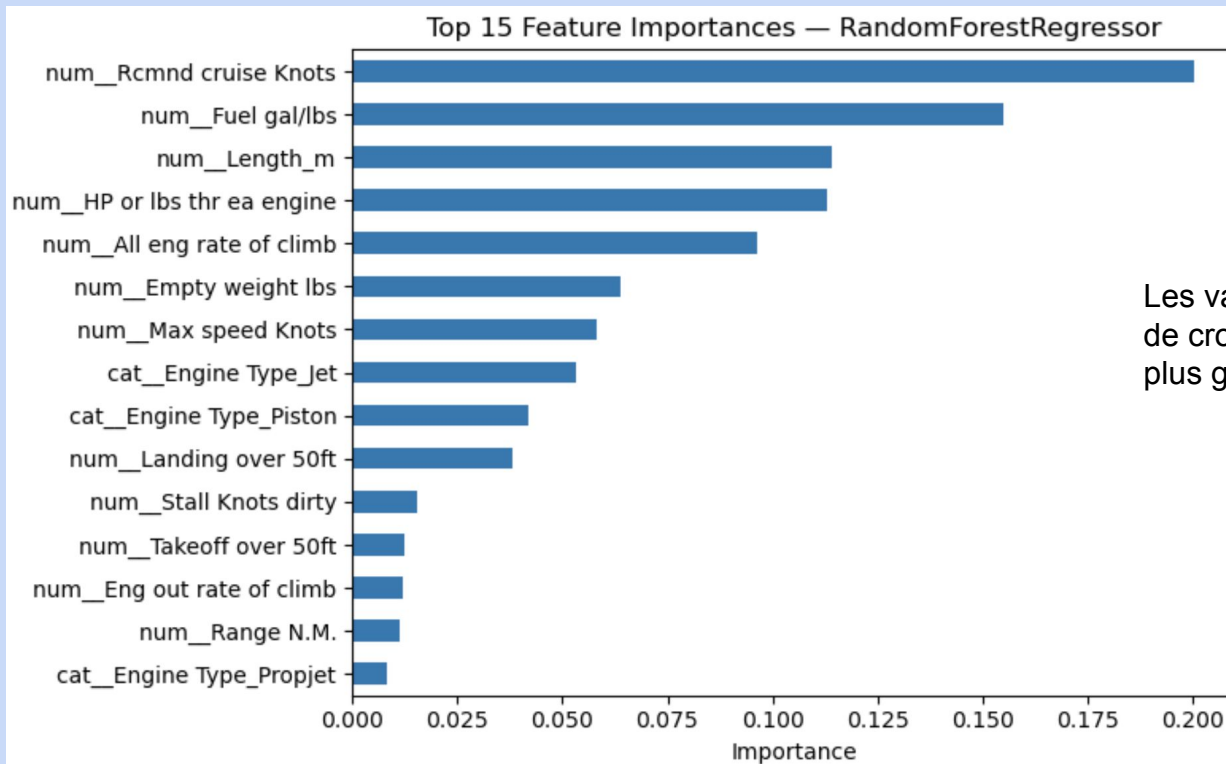
Modèle **linéaire** (approximation) ; sensible aux **features corrélées**.

Résultats RandomForest Regressor

R^2 test: 0.927 → 93%
MAE test: 195, 393
RMSE test: 286, 444

R^2 train: 0.926 → 93%
Train score: 0.981
Test score: 0.926

→ Model performant, équilibré



Les variables de **performance** (vitesse de croisière, puissance) expliquent la plus grande part des variations de prix.

Résultats Lasso

Matrice de corrélation: on a retiré certains doublons pour avoir des coefs plus stables et lisibles

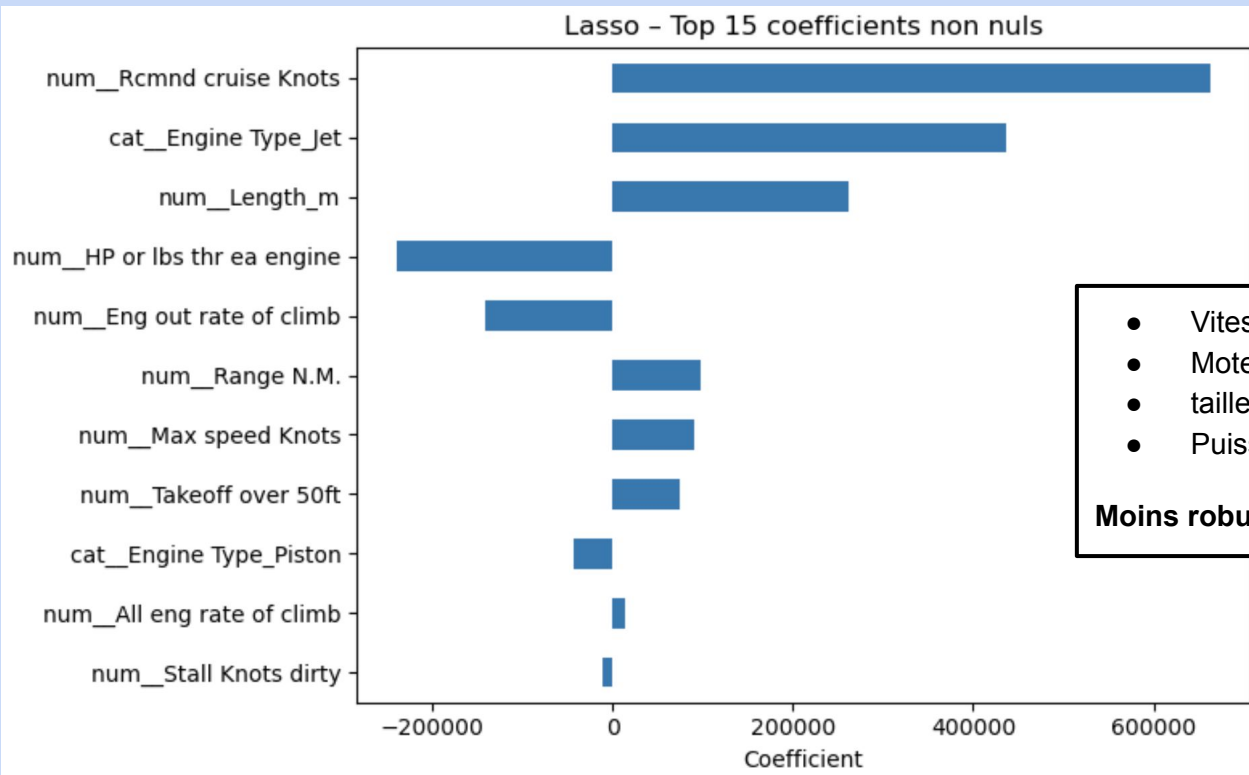
Alpha choisi : 10.0

R^2 test : 0.802 → 80%

R^2 train : 0.888 → 89%

MAE test : 310,485

RMSE test : 470,923



- Vitesse croisière ↑ ⇒ pred. **prix** ↑
- Moteur de type "Jet" ⇒ pred. **prix** ↑
- taille ↑ ⇒ pred. **prix** ↑
- Puissance ↑ ⇒ pred. **prix** ↓ (effet conditionnel)

Moins robuste que RandomForest Regressor

Applications possibles

— — —

Coté **constructeur** : lors de la conception, paramètres techniques et de performance à prioriser pour valoriser mon appareil.

Coté **vendeur** : positionnement prix en fonction des performances / caractéristiques techniques

Coté **acheteur** : Test de la viabilité d'un projet.

- Projet **individuel** : à partir d'un budget, à quoi peut-on s'attendre en termes de spécifications (vitesse, performances) ?
- Projet **professionnel** : cohérence du business plan (ressources financières) avec le business model (Quels types de liaisons ?)

