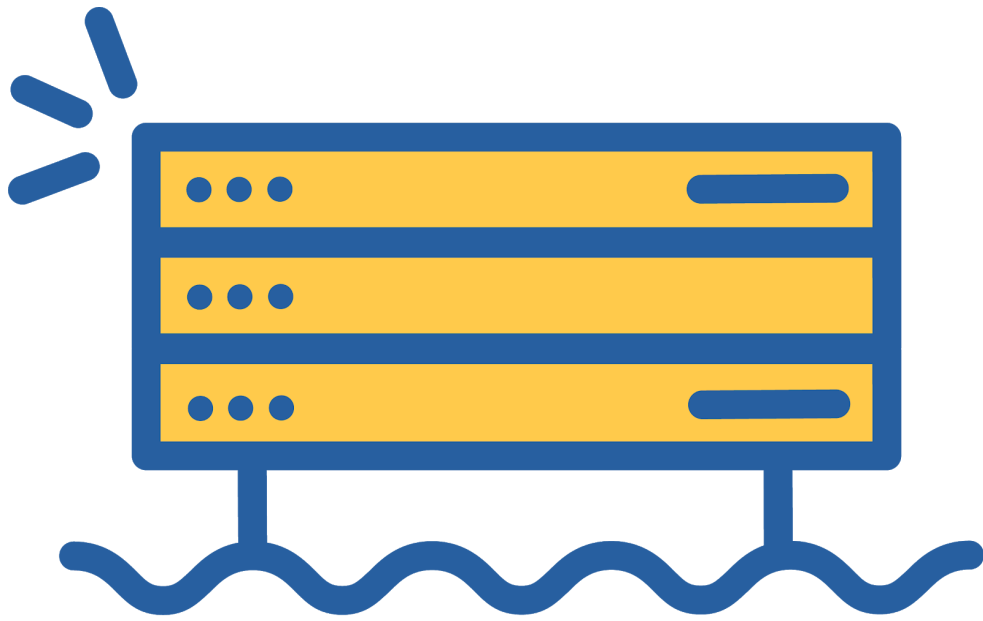

La Plateforme_

Prédisez le prix de votre future
voiture !!



La Plateforme_



Quels sont les objectifs ?

1. Apprendre à écrire un modèle de Machine Learning simple en python simplement avec la bibliothèque d'algèbre linéaire numpy
2. Apprendre à utiliser Jupyter-notebook
3. Charger les bibliothèques et lire les données
 - 3.1. Charger les bibliothèques
 - 3.2 Lire les données
 - 3.3 Visualiser les données
4. Préparer les données en algèbre linéaire (normalisation)
5. Les notions mathématiques qui seront abordées sont les notions permettant l'implémentation d'une régression linéaire, plus spécifiquement :
 - 5.1 Fonction de coût
 - 5.2 Descente de gradient
 - 5.3 Équation normale

Technos utilisées :

1. Jupyter-notebook
2. Bibliothèques Python Pandas, Seaborn, Matplotlib, Scikit-Learn, numpy, scipy, statModel

Vous travaillez depuis un an en tant qu'expert en data et avez économisé assez d'argent pour acheter une voiture. En tant qu'expert en données et étant assez économe, vous voulez en avoir pour votre argent !!!

Imaginez que vous ayez également des données du site Web de voitures CarDekho, qui contient des informations sur une grande variété de voitures, y compris leur prix. Vous réalisez que vous pouvez utiliser ces données pour vous assurer d'obtenir une bonne affaire sur une nouvelle voiture. En particulier, vous pouvez déterminer exactement combien vous devriez payer pour un type de voiture spécifique. Cela peut être particulièrement utile si vous rencontrez un vendeur de voitures difficile!

Mais la question est de savoir comment utiliser les données pour déterminer combien vous devriez payer ?

Vous pouvez utiliser la régression linéaire !

Pour rappel ... La régression linéaire est une méthode permettant de découvrir la relation entre deux variables de l'ensemble de données, telles que le prix de la voiture et l'année de fabrication. Les Data Scientists s'appuient sur cette méthode pour résoudre un large éventail de problèmes, notamment en matière de prédiction.

1. La première étape est la récupération des données. Vous utiliserez la bibliothèque de science des données appelée Pandas pour charger l'ensemble des données. Au travers de Pandas, il est possible de lire le fichier de données (carData.csv). Les données seront ensuite affectées et stockées dans une variable, par exemple : car_data.
2. Chaque ligne du fichier correspond à une voiture spécifique en vente et chaque colonne détaille des informations sur cette voiture.

Vous pouvez considérer ces colonnes comme étant les éléments nécessaires communs de tous les modèles futurs que nous créerons. nous devons connaître chaque variable (colonne) de notre ensemble de données.

- **Car_Name:** Cette colonne doit être remplie avec le nom de la voiture.
- **Year:** Cette colonne doit être remplie avec l'année de fabrication de la voiture.
- **Selling_Price:** cette colonne doit être remplie avec le prix auquel le propriétaire souhaite vendre la voiture.
- **Kms_Driven:** Il s'agit de la distance parcourue par la voiture en km.
- **Fuel_Type:** Type de carburant de la voiture.
- **Seller_Type:** définit si le vendeur est un revendeur ou un particulier.
- **Transmission:** définit si la boîte de vitesse de la voiture est manuelle ou automatique.

-
3. Exploration des données - Maintenant que vous avez chargé en Python les données de CarDekho, vous pouvez commencer à les explorer. Voyez si vous pouvez déjà repérer des informations qui pourraient vous aider à trouver votre voiture idéale. Quelle est la taille du jeu de données. Calculer quelques statistiques de base (moyenne, médiane, quartile, tracer la distribution avec Matplotlib (histogramme)).
 4. Visualisez les données grâce à la librairie Seaborn. Notamment avec le type de tracé catplot, expliquez dans quel cas ce type de tracé est pertinent.
 5. Quantifiez la relation entre l'âge et le prix de vente.
Réaliser une veille sur la régression linéaire. Est-ce que notre jeux de données est adapté à ce type d'algorithme (ou existe-t-il une corrélation linéaire entre les variables ?) Proposer un outil de visualisation Matplotlib permettant d'appuyer votre argumentation.
 - Appliquer l'algorithme de régression linéaire univariée en vous aidant de la librairie Scipy.
 - Appliquer l'algorithme de régression linéaire univariée en vous aidant de la librairie Numpy.
 - Appliquer l'algorithme de régression linéaire univariée en vous aidant de la librairie sklearn
 - Améliorer le modèle en utilisant plusieurs variables d'entrée , telles que Kms_Driven et Transmission (réaliser une régression linéaire multiple en Python avec sklearn).
 - Conclure
 6. **Créer VOTRE Class LinearRegression.** Ne pas utiliser de fonctions de régression linéaire existante (comme précédemment vue). Pouvez-vous évaluer la performance de votre modèle ?
 7. Héberger vos sources sous github.
 8. **Je souhaite acheter un véhicule ayant moins de 7 ans, pas plus de 100 000 km, avec une boîte manuelle, quel devrait-être son prix ?**
- 