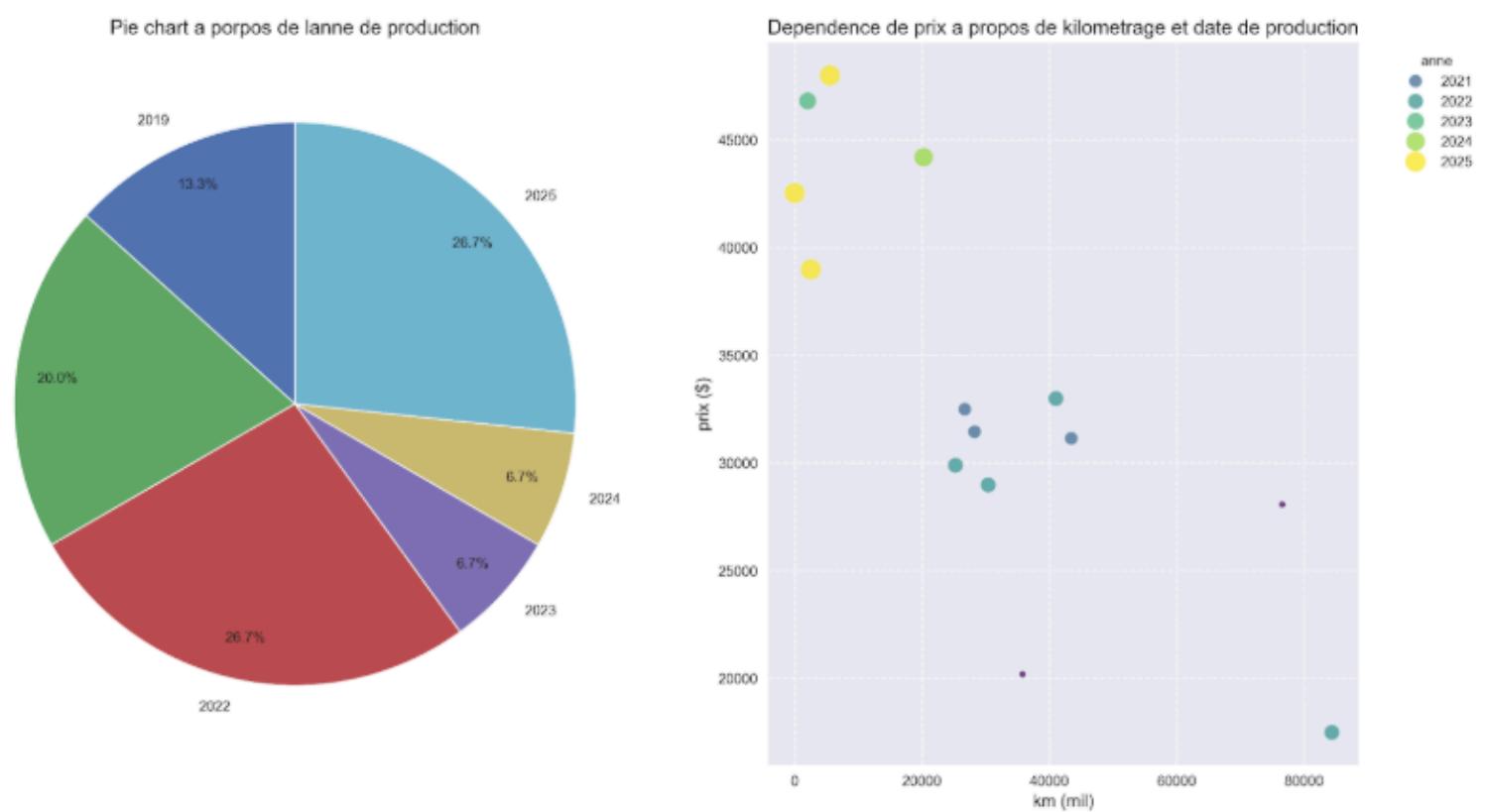
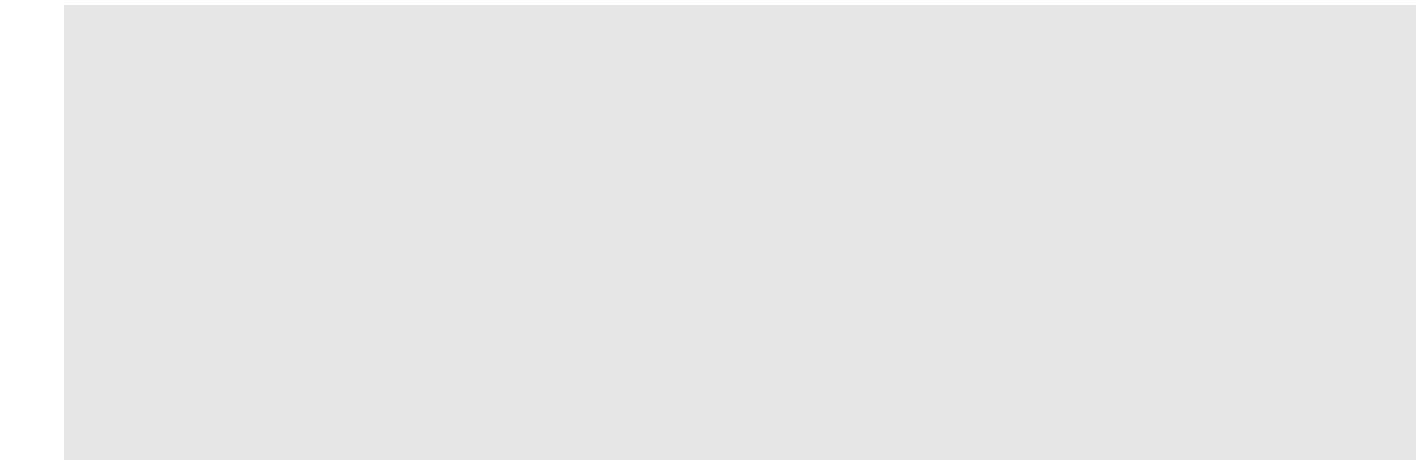


# Projet DATA – B2 Gr2

19/03/2025

ENSEIGNANT : M. TAALBI RABAH

MARTIN TANDERO  
JULIEN YILDIZ  
ARTEM SALNIKOV  
JENNA JARI  
FLORIANE GUILLOU



# Sommaire

---

1.1 Objectif du projet

1.2 Choix du site e-commerce

1.3 Outils & technologie utilisés

1.4 Architecture du projet+schéma de la base de donnée

1.5 Structure de l'équipe

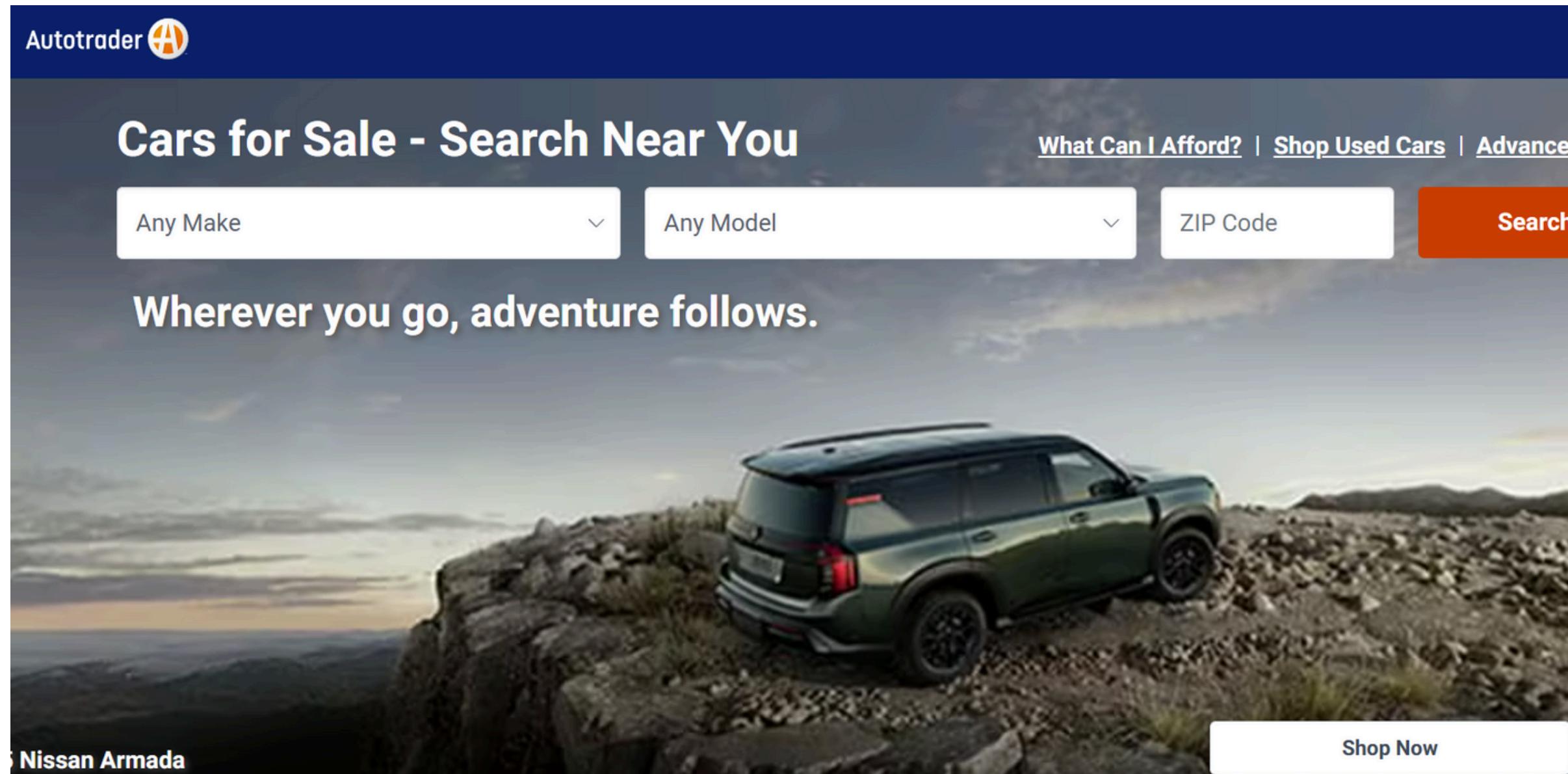
1.6 Conclusion

# 1.1 Objectif du projet

- **Scraping** : Extraction de données à partir d'un site e-commerce réel (Autotrader)
- **Stockage** : Insertion et organisation des données dans une base MySQL structurée
- **Nettoyage** : Conversion des types, traitement des valeurs manquantes, normalisation des catégories.
- **Analyse & Visualisation** : Produire des visualisations pertinentes et dégager des insights.

# 1.2 Choix du site e-commerce

🌐 Site Choisi : Autotrader



Pourquoi ce site ?

- Simplicité d'accès
- Structure HTML claire et cohérente
- Données riches et variées (grand nombre d'attributs)

# 1.3 Outils et technologies utilisés



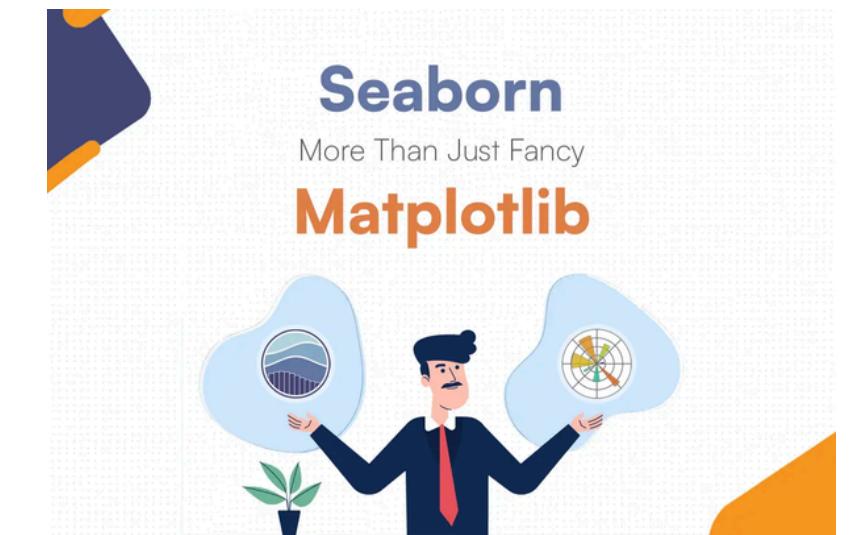
envoyer des requêtes HTTP et analyser le contenu HTML



stocker proprement les données



nettoyage et la préparation



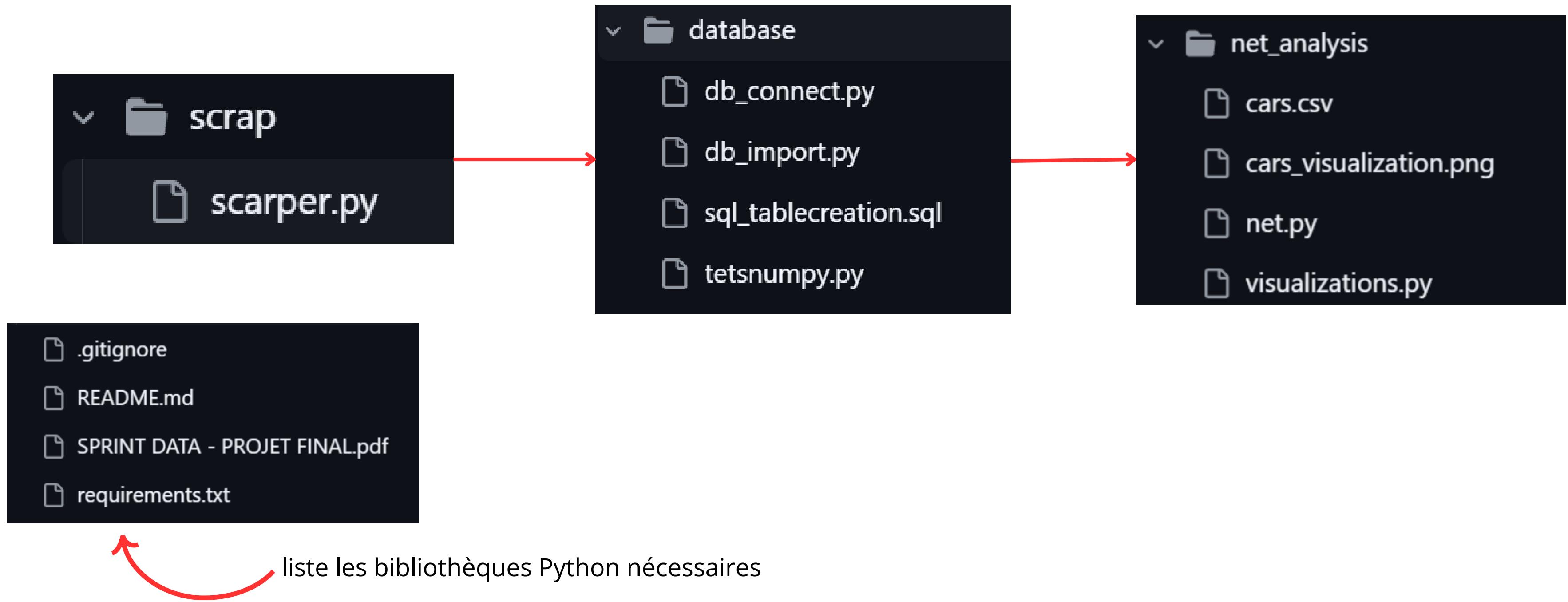
graphiques clairs et exploitables



**GitHub**  
Copilot

# 1.4 Architecture du projet

Scraping → Base MySQL → Pandas → Visualisation



# 1.4 Architecture du projet

Scraping → Base MySQL → Pandas → Visualisation

## Étape 1 : Scraping

```
def scrape_autotrader():
    conn = create_connection() # conn
    if not conn:
        return

    create_table(conn) # table if not exist

    try:
        for page in range(1, 6):
            headers = {'User-Agent': random.choice(USER_AGENTS)}
            params = {
                'searchRadius': 0,
                'sortBy': 'relevance',
                'numRecords': 25,
                'firstRecord': (page - 1) * 25
            }

            try:
                response = requests.get(
                    BASE_URL,
                    params=params,
```

Récupération automatisée des données (prix, modèle, année...) sur Autotrader via des scripts Python.

- requests gère les requêtes HTTP.
- BeautifulSoup parse le HTML
- Une boucle for parcourt les pages et extrait les données dans un dictionnaire.

# 1.4 Architecture du projet

Scraping → Base MySQL → Pandas → Visualisation

## Étape 2 : Connection à MySQL

```
import mysql.connector
from mysql.connector import Error

def create_connection():
    try:
        connection = mysql.connector.connect(
            host="localhost",
            user="root",
            password="",
            database="cars_db"
        )

        if connection.is_connected():
            print("Con success")
            return connection

    except Error as e:
        print(f"fail to con: {e}")
        return None
```

Les données sont enregistrées dans une base MySQL structurée pour garantir leur organisation et leur intégrité.

- Utilisation de mysql-connector-python.
- Création d'une table cars pour stocker les données des voitures.

# 1.4 Architecture du projet

Scraping → Base MySQL → Pandas → Visualisation

## Étape 3 : Nettoyage avec Pandas

```
import pandas as pd
import numpy as np

df = pd.read_csv(
    'cars.csv',
    dtype={
        'id': 'int32',
        'title': 'string',
        'price': 'float32',
        'mileage': 'Int32',
        'year': 'Int64',
        'link': 'string'
    },
    parse_dates=['scraped_at'],
    na_values=['', 'NA', 'N/A', 'NULL'],
    keep_default_na=False
)

# setting de paramétrage de output (terminal)
pd.set_option('display.max_columns', None)
pd.set_option('display.expand_frame_repr', False)
pd.set_option('display.max_colwidth', 30)
```

Utilisation de Pandas pour transformer les données :

- Nettoyage des valeurs nulles
  - Conversion des types
  - Formatage des catégories
- 
- pd.read\_sql charge les données depuis MySQL.
  - fillna et str.title améliorent la cohérence des données.

# 1.4 Architecture du projet

Scraping → Base MySQL → Pandas → Visualisation

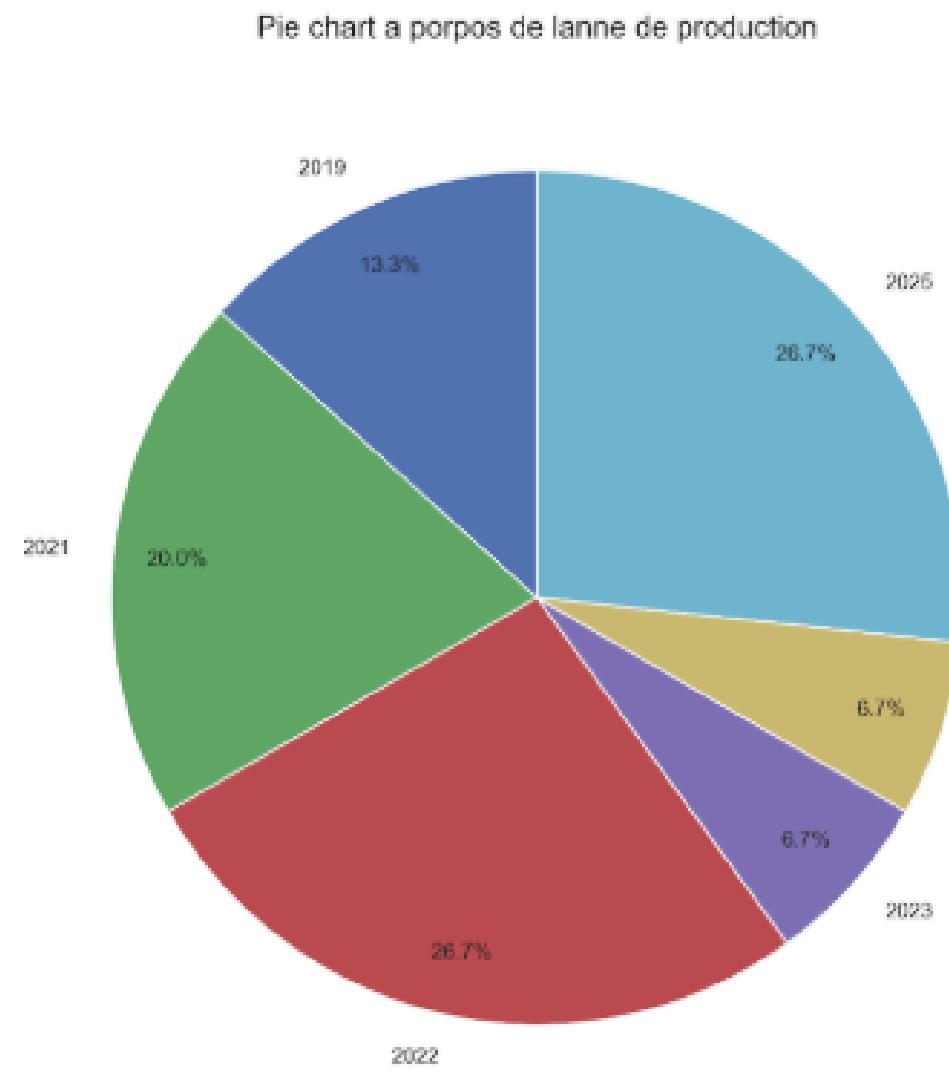
## Étape 4 : Visualisation avec Seaborn & Matplotlib

id	title	price	mileage	year	link
29	Used 2024 Hyundai Palisade Calligraphy	44199.0	20268		<a href="https://www.autotrader.com/cars-for-sale/vehicle/734689554?allListingType=all-cars&amp;firstRecord=100&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/734689554?allListingType=all-cars&amp;firstRecord=100&amp;numRecords=25&amp;searchRadius=0&amp;so</a>
28	Used 2019 BMW X4 M40i	28082.0	76580		<a href="https://www.autotrader.com/cars-for-sale/vehicle/740332079?allListingType=all-cars&amp;firstRecord=100&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/740332079?allListingType=all-cars&amp;firstRecord=100&amp;numRecords=25&amp;searchRadius=0&amp;so</a>
27	Certified 2025 Mercedes-Benz GLA 250	38990.0	2540		<a href="https://www.autotrader.com/cars-for-sale/vehicle/737550692?allListingType=all-cars&amp;firstRecord=75&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/737550692?allListingType=all-cars&amp;firstRecord=75&amp;numRecords=25&amp;searchRadius=0&amp;so</a>
26	Certified 2023 Ford Bronco Outer Banks	46811.0	2053		<a href="https://www.autotrader.com/cars-for-sale/vehicle/739470335?allListingType=all-cars&amp;firstRecord=75&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/739470335?allListingType=all-cars&amp;firstRecord=75&amp;numRecords=25&amp;searchRadius=0&amp;so</a>
25	Used 2021 Lincoln Nautilus Reserve	31150.0	43455		<a href="https://www.autotrader.com/cars-for-sale/vehicle/735451278?allListingType=all-cars&amp;firstRecord=75&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/735451278?allListingType=all-cars&amp;firstRecord=75&amp;numRecords=25&amp;searchRadius=0&amp;so</a>
24	Used 2022 Chrysler Pacifica Touring-L	17500.0	84393		<a href="https://www.autotrader.com/cars-for-sale/vehicle/739062017?allListingType=all-cars&amp;firstRecord=50&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/739062017?allListingType=all-cars&amp;firstRecord=50&amp;numRecords=25&amp;searchRadius=0&amp;so</a>
23	New 2025 Volvo XC60 B5 Core	42535.0	5		<a href="https://www.autotrader.com/cars-for-sale/vehicle/720155760?allListingType=all-cars&amp;firstRecord=50&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/720155760?allListingType=all-cars&amp;firstRecord=50&amp;numRecords=25&amp;searchRadius=0&amp;so</a>
22	Used 2025 Cadillac XT5 Premium Luxury	47997.0	5553		<a href="https://www.autotrader.com/cars-for-sale/vehicle/735786117?allListingType=all-cars&amp;firstRecord=50&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/735786117?allListingType=all-cars&amp;firstRecord=50&amp;numRecords=25&amp;searchRadius=0&amp;so</a>
21	Used 2022 Toyota Tacoma TRD Off-Road	32998.0	41047		<a href="https://www.autotrader.com/cars-for-sale/vehicle/737207463?allListingType=all-cars&amp;firstRecord=25&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/737207463?allListingType=all-cars&amp;firstRecord=25&amp;numRecords=25&amp;searchRadius=0&amp;so</a>
20	New 2025 RAM 1500 Lone Star	51570.0			<a href="https://www.autotrader.com/cars-for-sale/vehicle/729703538?allListingType=all-cars&amp;firstRecord=25&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/729703538?allListingType=all-cars&amp;firstRecord=25&amp;numRecords=25&amp;searchRadius=0&amp;so</a>
19	Used 2022 GMC Acadia SLT	29900.0	25261		<a href="https://www.autotrader.com/cars-for-sale/vehicle/742201503?allListingType=all-cars&amp;firstRecord=25&amp;numRecords=25&amp;searchRadius=0&amp;so">https://www.autotrader.com/cars-for-sale/vehicle/742201503?allListingType=all-cars&amp;firstRecord=25&amp;numRecords=25&amp;searchRadius=0&amp;so</a>

# 1.4 Architecture du projet

Scraping → Base MySQL → Pandas → Visualisation

## Étape 4 : Visualisation avec Seaborn & Matplotlib



Exploitation des données avec Seaborn et Matplotlib pour produire des visualisations statistiques et tirer des insights.

# 1.5 Structure de l'équipe

**Développement et Scraping (Python + MySQL)**

Artem SALNIKOV  
Julien YILDIZ  
Martin TENDERO

**Nettoyage & Analyse des Données (Pandas + Visualisation)**

Artem SALNIKOV  
Floriane GUILLOU  
Martin TENDERO

**Coordination & Rédaction du Rapport**

Martin TENDERO  
Jenna JERI



# 1.6 Conclusion



Ce que nous avons appris

- Automatiser la collecte d'informations web via le scraping
- Structurer et exploiter une base de données
- Nettoyer et normaliser des données avec Pandas
- Produire des visualisations claires pour appuyer une analyse