

Després, Alexy
Hisier, Yvan
Kouassi, Louis-Gérard
Groupe 4318

Projet 2
Travail présente à
Nicolas Payre
Département Informatique
pour le cours
Exploration de nouvelles technologies

15 octobre 2025

Table des matières

1.	Choix du site et des données.....	1
2.	Dépôt GitHub.....	2
3.	Outil de Webscraping	2
4.	Données extraites.....	3
5.	Importation dans MongoDB Atlas	4
6.	Requêtes MongoDB	5
7.	Tableau de contribution	7
8.	Conclusion	7

1. Choix du site et des données

Nom du site : Kijiji

URL du site : Kijiji - Buy, Sell & Save with Canada's #1 Local Marketplaces | Kijiji

Description des données extraites :

Les données proviennent de la section « Autos et véhicules → Voitures d'occasion » du site Kijiji. Chaque annonce contient plusieurs informations comme :

- Le titre de l'annonce (ex. « 2017 Honda Civic LX »),
- La marque et le modèle du véhicule,
- La ville où le véhicule est vendu,
- Le prix,
- Le kilométrage,
- Le type de carburant,
- La transmission (manuelle ou automatique),
- Une description détaillée du véhicule,
- Et plusieurs images (URL directes vers les photos de l'annonce).

Justification du choix du site et des données :

Le site Kijiji a été choisi, car il contient une grande quantité de données publiques et structurées, faciles à extraire et à analyser.

La catégorie des voitures d'occasion a été sélectionnée parce qu'elle offre :

- Un volume suffisant pour atteindre le seuil des 2 500 items exigés,
- Des champs variés (texte, nombres, liens, images),
- Et un intérêt concret pour des analyses futures (comparaison de prix selon la marque, le kilométrage ou la localisation).

2. Dépôt GitHub

Lien vers le dépôt GitHub : https://github.com/LouisG1er/Projet2_Web scraping_Kijiji

Noms des collaborateurs et description de leurs contributions :

- Louis Gérard : recherche du site, configuration de Web Scraper, création du plan du site et ajout du fichier JSON sur GitHub.
- Yvan : création du sitemap, extraction automatique des données, conversion en JSON et importation dans MongoDB Atlas.
- Alexy : structuration du fichier final, création des requêtes MongoDB et rédaction du rapport final.

3. Outil de Webscraping

Nom de l'outil : Web Scraper (extension Google Chrome)

Description de son fonctionnement :

Web Scraper est une extension gratuite de Google Chrome qui permet d'extraire automatiquement des données à partir d'un site web. L'utilisateur crée un sitemap (plan de navigation) qui indique au robot quelles pages visiter et quelles informations extraire. L'outil permet :

- La pagination automatique pour parcourir toutes les pages de résultats,
- La détection automatique des éléments (grâce à l'assistant *AI Wizard*),
- La navigation dans plusieurs niveaux de catégories,
- Et l'exportation des données en CSV ou JSON.

Méthode d'extraction utilisée :

Nous avons utilisé la méthode automatique (AI Wizard) qui détecte et sélectionne automatiquement les éléments à extraire sur la page.

L'outil a récupéré différents champs comme :

- Le titre de l'annonce
- Le prix
- La ville
- Le kilométrage

- Le type de carburant
- La description (description),
- Ainsi que les images.

Une fois le scraping terminé, les données ont été exportées en CSV, puis converties en JSON en utilisant ce site : <https://csvjson.com/csv2json>.

CSV or TSV > JSON

To get started, upload or paste your data from Excel (saved as CSV or TSV).

Upload a CSV file

Choisir un fichier Aucun fichier n'a été sélectionné

Or paste your CSV here

```
web-scraper-order,web-scraper-start-url,pagination-selector,Image véhicule-src,Prix,Description,Carburant,Ville,Titre,Kilometrage
"1760567122-1","https://www.kijiji.ca/b-autos-camions/quebec/c17419001?view=list","", "https://media.kijiji.ca/api/v1/autos-prod-ads/images/2b/2bc22877-2094-49b4-bc9e-3bd73198663c?rule=kijijica-400-webp", "17 990 $", "AM/FM/HD, écran couleur large de 8,8 po avec MAZDA CONNECT, commutateur HMI Commander, Mazda Harmonic Acoustics avec 8 haut-parleurs, Apple CarPlay et Android Auto, 2 ports USB, commandes ...", "Essence", "Longueuil/Rive Sud", "", "128570 km"
"1760567122-2","https://www.kijiji.ca/b-autos-camions/quebec/c17419001?view=list","", "https://media.kijiji.ca/api/v1/autos-prod-dealer-ads/images/04/04649831-1f53-43dd-8fb8-a4484950691a?rule=kijijica-400-webp",
```

Separator

Auto-detect ▾

☒ Parse numbers

☒ Parse JSON

☐ Transpose

Output: ☒ Array ☐ Hash ☐ Minify

JSON

```
{
  "web-scraper-order": "1760567122-2",
  "web-scraper-start-url": "https://www.kijiji.ca/b-autos-camions/quebec/c17419001?view=list",
  "pagination-selector": "",
  "Image véhicule-src": "https://media.kijiji.ca/api/v1/autos-prod-dealer-ads/images/04/04649831-1f53-43dd-8fb8-a4484950691a?rule=kijijica-400-webp",
  "Prix": "39 222 $",
  "Description": "Moteur 4 cylindres i-VTEC de 2.0L, 16 soupapes, injection dans l'admission DACT, développant 158 chevaux, 138 lb-pi de couple, transmission à variation continue, suspension indépendante aux quatre ...",
  "Carburant": "Essence",
  "Ville": "Longueuil/Rive Sud",
  "Titre": "2026 Honda HR-V SPORT AWD, TOIT OUVRANT, CARPLAY ET ANDROID AUTO",
  "Kilometrage": "10 km"
}
```

Cette étape a permis d'obtenir un fichier structuré et lisible, prêt à être importé dans MongoDB Atlas.

4. Données extraites

Nombre total d'items extraits :

Environ 2 720 annonces ont été extraites à partir du site Kijiji dans la catégorie Voitures d'occasion.

Structure du fichier CSV (nom des colonnes et types de données) :

Les données initiales exportées par Web Scraper étaient au format CSV, avec plusieurs colonnes. Après la conversion en JSON, la structure finale des données est la suivante :

Nom du champ	Type de donnée	Description
Image véhicule	Texte (String / URL)	Lien direct vers l'image principale du véhicule sur Kijiji.
Prix	Texte (String)	Prix affiché dans l'annonce (incluant le symbole \$ et les espaces). Exemple : "17 990 \$".
Description	Texte (String)	Description complète du véhicule (options, caractéristiques, état, etc.).
Carburant	Texte (String)	Type de carburant utilisé par le véhicule (ex. "Essence", "Diesel", "Hybride", "Électrique").
Ville	Texte (String)	Localisation de l'annonce, souvent sous la forme "Longueuil/Rive Sud" ou "Montréal/Nord".
Titre	Texte (String)	Titre de l'annonce sur Kijiji (souvent le modèle du véhicule, ex. "Mazda CX-5 GS 2018").
Kilométrage	Texte (String)	Kilométrage total du véhicule avec unité ("km"). Exemple : "128 570 km".

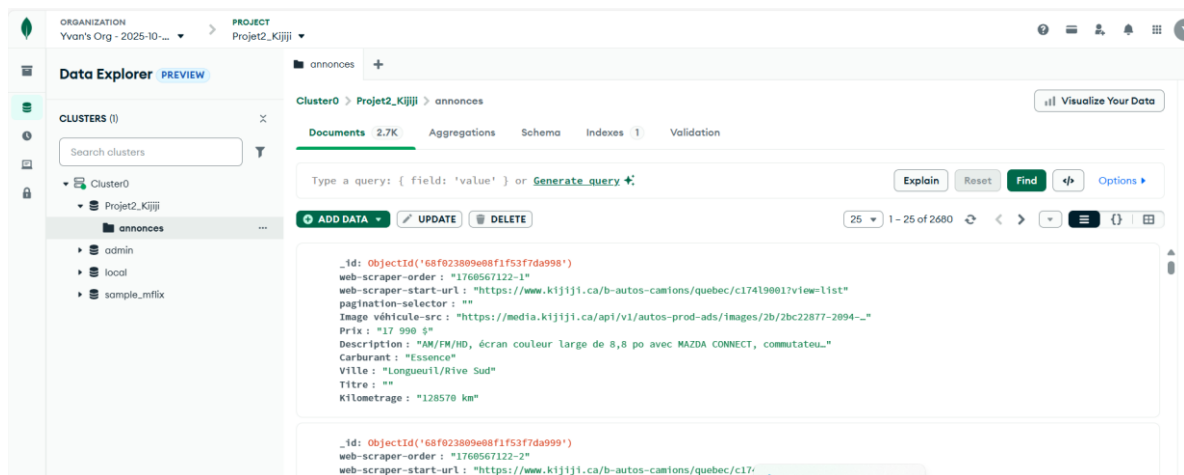
5. Importation dans MongoDB Atlas

Lien vers la base MongoDB Atlas : Cluster0 Data | Cloud: MongoDB Cloud

Nom du projet : Projet2_Kijiji

Nom de la collection : annonces

Aperçu :



Méthode d'importation utilisée :

Les données ont été exportées depuis Web Scraper, converties en JSON, puis importées dans MongoDB Atlas à l'aide de la fonction “Add Data → Insert Document”.

Description du schéma de la collection (champs et types) :

Champ	Type	Description
Image véhicule	Texte	URL de l'image principale du véhicule.
Prix	Texte	Prix du véhicule (ex. “17 990 \$”).
Description	Texte	Description de l'annonce.
Carburant	Texte	Type de carburant (ex. Essence).
Ville	Texte	Ville du vendeur.
Titre	Texte	Titre de l'annonce.
Kilometrage	Texte	Kilométrage du véhicule (ex. “128 570 km”).

6. Requêtes MongoDB

Requête 1 : { "Ville": /Longueuil/i }

Explication : Trouver les annonces d'une ville précise (ex. Longueuil)

Résultat :

```
{ "Ville": /Longueuil/i }
```

Generate query Explain Reset Find Options

ADD DATA UPDATE DELETE

25 1 - 25 of 327

```
_id: ObjectId('68f023809e08f1f53f7da998')
web-scraper-order: "1760567122-1"
web-scraper-start-url: "https://www.kijiji.ca/b-autos-camions/quebec/c174l9001?view=list"
pagination-selector: ""
Image véhicule-src: "https://media.kijiji.ca/api/v1/autos-prod-ads/images/2b/2bc22877-2094-..."
Prix: "17 990 $"
Description: "AM/FM/HD, écran couleur large de 8,8 po avec MAZDA CONNECT, commutateur..."
Carburant: "Essence"
Ville: "Longueuil/Rive Sud"
Titre: ""
Kilometrage: "128570 km"
```

Requête 2 : { "Kilometrage": { \$regex: "^[0-9]{1,2}[0-9]{3,}" } }

Explication : Trouver les voitures dont le kilométrage est inférieur à 100 000 km

Résultat :

Documents 2.7K Aggregations Schema Indexes 1 Validation

```
{ "Kilometrage": { $regex: "^[0-9]{1,2}[0-9]{3,}" } }
```

Generate query Explain Reset Find Options

ADD DATA UPDATE DELETE

25 1 - 25 of 2254

```
_id: ObjectId('68f023809e08f1f53f7da998')
web-scraper-order: "1760567122-1"
web-scraper-start-url: "https://www.kijiji.ca/b-autos-camions/quebec/c174l9001?view=list"
pagination-selector: ""
Image véhicule-src: "https://media.kijiji.ca/api/v1/autos-prod-ads/images/2b/2bc22877-2094-..."
Prix: "17 990 $"
Description: "AM/FM/HD, écran couleur large de 8,8 po avec MAZDA CONNECT, commutateur..."
Carburant: "Essence"
Ville: "Longueuil/Rive Sud"
Titre: ""
Kilometrage: "128570 km"
```

```
_id: ObjectId('68f023809e08f1f53f7da99a')
web-scraper-order: "1760567122-3"
web-scraper-start-url: "https://www.kijiji.ca/b-autos-camions/quebec/c174l9001?view=list"
pagination-selector: ""
Image véhicule-src: "https://media.kijiji.ca/api/v1/autos-prod-ads/images/09/09d2ffdd-41b-..."
```

Requête 3 : { "Carburant": /Essence/i, "Ville": /Sherbrooke/i }

Explication : Trouver les véhicules à essence situés à Sherbrooke

Résultat :

The screenshot shows the MongoDB Atlas web interface. At the top, the breadcrumb navigation is "Cluster0 > Projet2_Kijiji > annonces". There are tabs for "Documents" (2.7K), "Aggregations", "Schema", "Indexes" (1), and "Validation". A "Visualize Your Data" button is on the right. Below the tabs, a query is entered in the search bar: `{ "Carburant": "/Essence/i, "Ville": "/Sherbrooke/i }`. Buttons for "Generate query", "Explain", "Reset", "Find", and "Options" are visible. Below the query bar, there are buttons for "ADD DATA", "UPDATE", "DELETE", and a help icon. The results section shows two documents. The first document is a JSON object with fields: `_id`, `web-scraper-order`, `web-scraper-start-url`, `pagination-selector`, `Image véhicule-src`, `Prix`, `Description`, `Carburant`, `Ville`, `Titre`, and `Kilometrage`. The second document is a similar JSON object with slightly different values.

7. Tableau de contribution

Nom du coéquipier	Tâches réalisées	Pourcentage de participation
Hisier Yvan	Création du sitemap Web Scraper, extraction automatique des données, conversion en JSON, et importation dans MongoDB Atlas.	33 %
Kouassi Louis Gérard	Recherche du site et des données, vérification des champs extraits, tests et validation des requêtes MongoDB.	33 %
Alexy Després	Structuration du fichier final, organisation du dépôt GitHub, rédaction et mise en page du rapport.	34 %

8. Conclusion

Ce projet nous a permis de découvrir plusieurs nouvelles technologies, comme le web scraping, MongoDB Atlas et la manipulation de fichiers JSON.

Nous avons appris à extraire automatiquement des données d'un site web, à les structurer proprement et à les importer dans une base de données en ligne.

Le site Kijiji s'est avéré un bon choix, car il contient beaucoup d'informations variées et exploitables.

Dans l'ensemble, le projet s'est bien déroulé, et nous sommes satisfaits du résultat obtenu.