

PROJET

WEBCRAPPING & DATA PROCESSING



Nicolas CARVAL – Bruno PINCET

https://github.com/NicolasCarval/Scrapy_Meal

+ .

SOMMAIRE

Problématique

Hypothèse

Méthode

Résultats

Bilan



Introduction

De nombreux sites proposant des recettes de cuisine en ligne ont été popularisés ces dernières années.

Cependant il n'est pas forcément facile pour les utilisateurs de s'y retrouver dans la liste des ingrédients à acheter pour réaliser les recettes.

Et après ? Dans quel magasin devrait-il se diriger afin d'être optimisé en termes de temps et de prix ?





PROBLÉMATIQUE

Trouver un moyen de fournir rapidement une liste d'ingrédients à acheter ainsi que le meilleur supermarché où aller faire ses courses (Prix/Temps)

SOLUTION ?



Scrapy Meal

LOOKING FOR A RECIPE ?

type-in a meal...

ex : croissant

FIND



ESILV - Webscraping and Data processing - 2022/2023

Scrapy Meal

- Recherchez une recette
- Sélectionnez les ingrédients dont vous avez besoin
- Partez dans le magasin qui correspond le mieux à vos besoins

HYPOTHÈSES



Hypothèses

Fluidité de la solution

- L'utilisateur est capable de renseigner rapidement une recette et de récupérer des statistiques utiles
Ex : (comparaison prix, trajet)
- Le scrapping sera fait en temps réel

Scrapping des produits sur site marchand

- La recherche des produits sera cohérente même en présence de termes inexacts
ex: (3 cuillères à soupe de sucre -> sucre)
- Les sites marchands seront des grandes surfaces populaires avec de nombreux magasins.



MÉTHODE

Parcours Utilisateur & méthodes de scrapping associées

Méthode

sélection d'une
recette existante



Recherche du
prix des
ingrédients



Calcul du prix du
trajet avec les
magasins les
plus proches

750g



Récupération et sélection d'une recette

deuxième site internet de cuisine en termes de visites, après marmiton.org

750g

1

750g.com/recherche/?q=gateau

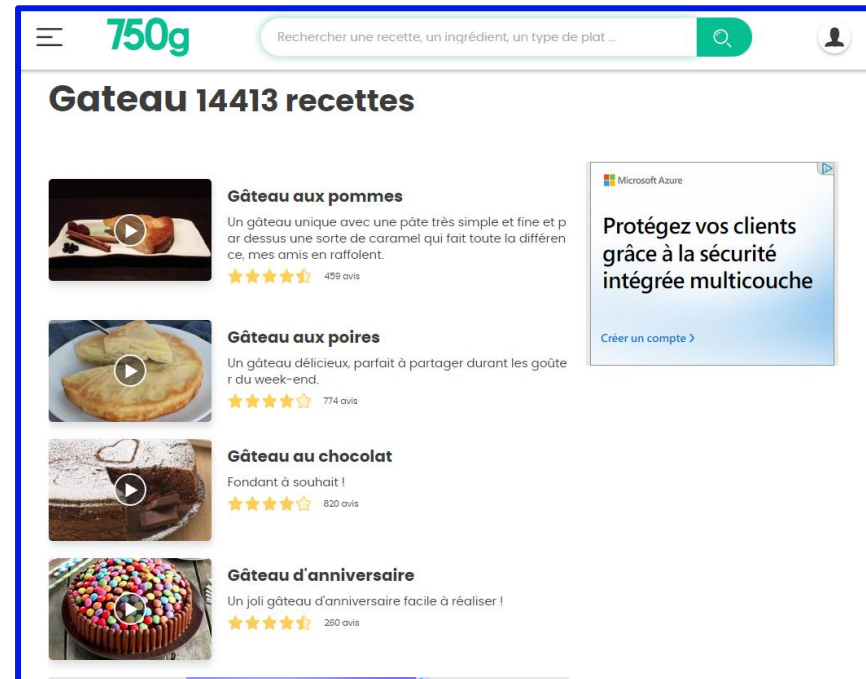
Scrapping du lien où l'on place la requête souhaitée par l'utilisateur en variable

2

Récupération avec les balises et XPATH et noms de classe:

- Nom de la recette
 - Image
- Lien de la recette

L'utilisateur en sélectionne une et on scrappe le lien de la recette sélectionné



Ingrédients

8 personnes

Pour le gâteau

- | | |
|----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
|  1 sachet de levure |  6 pommes à cuire |
|  2 oeufs |  4 c. à s. d'huile |
|  8 c. à s. de sucre |  6 c. à s. de lait |
|  10 c. à s. de farine | |

Pour la sauce

- | | |
|---------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
|  1 sachet de sucre vanillé |  1 oeuf |
|  80 g de sucre |  60 g de beurre |

Récupération de la liste des ingrédients

- 1 Préchauffer de four à 150°C ou Th.5, pas plus.
Mélanger tous les ingrédients "gâteau" dans l'ordre et verser la pâte dans un moule à manqué (24cm) beurré. Disposer les pommes en quartiers dessus.
Enfourner 25 min.
- 2 Pendant ce temps préparer la sauce en faisant fondre doucement le beurre, retirer la casserole du feu et mélanger avec le sucre et l'oeuf faire couler ce mélange sur les pommes et remettre au four 15 minutes.

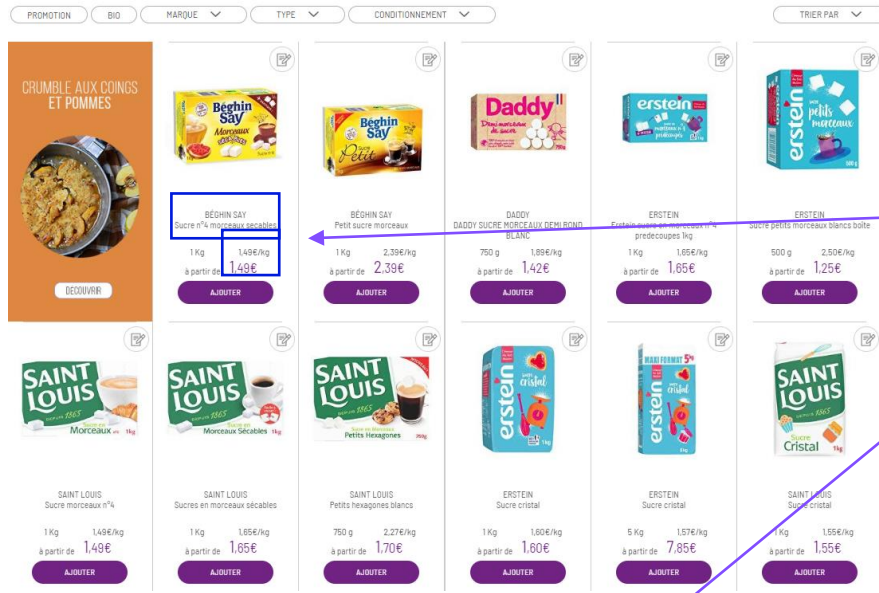
et des étapes de préparation

3

Recherche du prix des ingrédients

L'auteur choisi les ingrédients de la recette qu'il souhaite acheter

Résultats de recherche : sucre (95 produits)



1

```
wd.get(f"https://www.aldi.be/fr/resultats-de-recherche.html?query={produit}")
```

```
wd.get(f"https://www.supermarchesmatch.fr/fr/recherche?recherche={produit}")
```

2

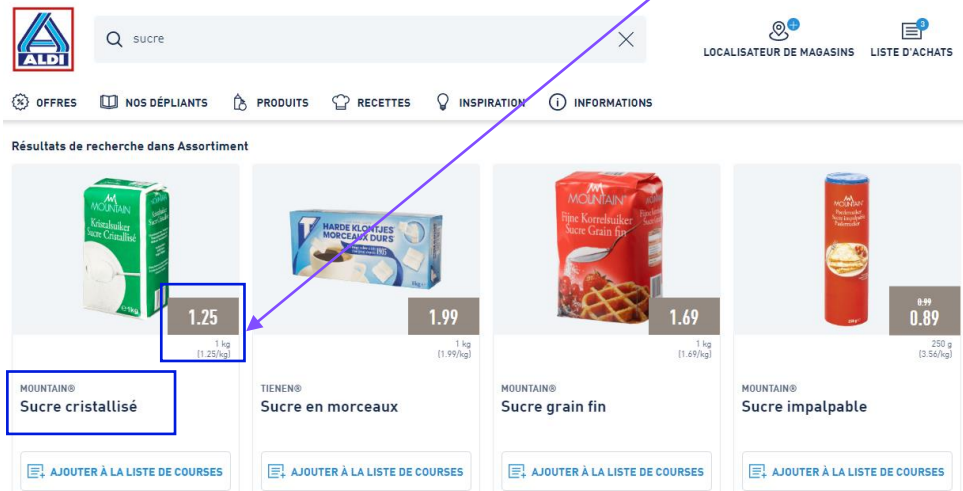
Récupération avec les balises et XPATH et noms de classe des **5 premiers éléments** en filtrant par pertinence et en triant avec la distance de **Damerau-Levenshtein***:

- Les noms des produits
- Le prix à l'unité (/u , /kg, /L)
- Le prix du produit

Présentation à l'utilisateur de plusieurs informations:

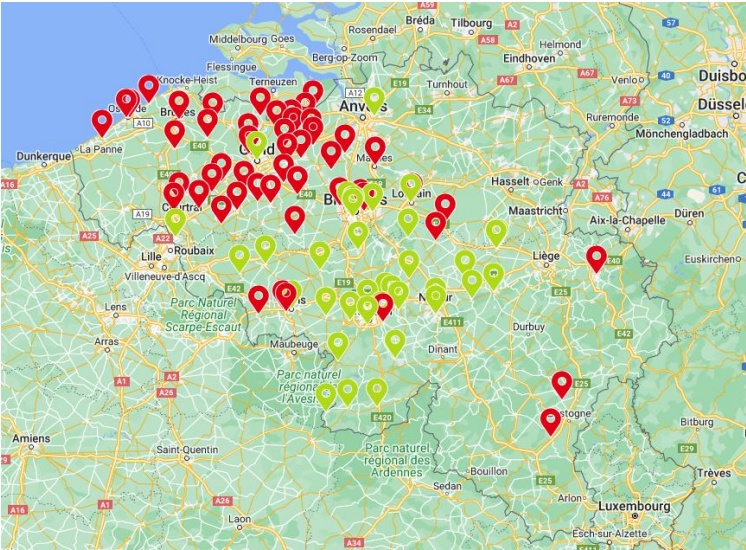
- Calcul du **RPD** (différence relative en pourcentage) sur les ingrédients en communs des 2 sites
 - **Nombre d'ingrédients** présents sur **Aldi**
 - **Nombre d'ingrédients** présents sur **supermarchesmatch**
- **Prix** du panier dans chaque magasin sur les ingrédients en commun

3

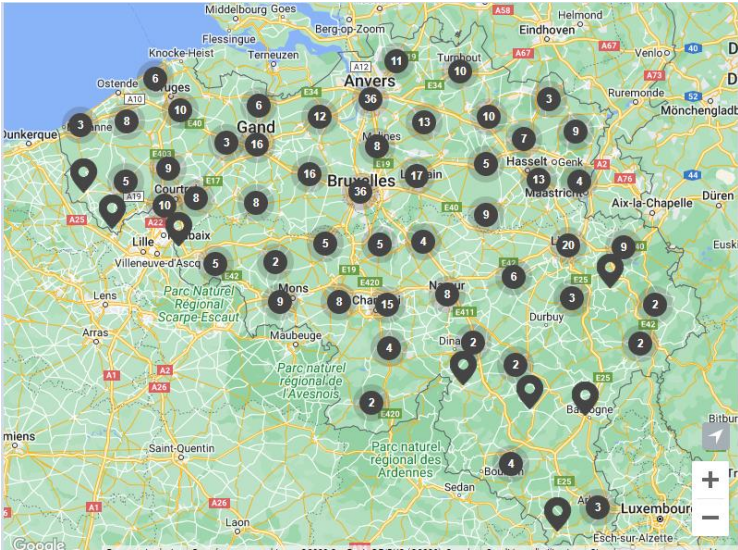


* distance qui consiste à calculer le nombre minimum d'opérations nécessaires pour transformer une chaîne de caractères en une autre ce qui permet de comparer l'ingrédient cherché et les produits proposés par le magasin et maximiser la similarité

Recherche du magasin le plus proche



supermarchesmatch Belgique – 80 adresses



Aldi Belgique – 200 adresses

M Match Andenne

Fermé

Avenue Roi Albert 17
5300 Andenne

les infos complètes >

M Match Ath

Fermé

Chaussée de Bruxelles 62
7800 Ath

les infos complètes >

M Match Auvelais

Fermé

Rue de Falisolle 4
5060 Sambreville

les infos complètes >

443 magasins trouvés

Aldi Supermarkt Kontichsesteenweg 28 2630 Aartselaar Supermarché	Heures d'ouverture: Aujourd'hui ouvert 8:30 - 19:00 Demain ouvert 8:30 - 19:00
Aldi Supermarkt Brugstraat Brugstraat 8-10 2060 Antwerpen Supermarché	Heures d'ouverture: Aujourd'hui ouvert 8:30 - 19:00 Demain ouvert 8:30 - 19:00
Aldi Supermarkt Terninckstraat Terninckstraat 13 2000 Antwerpen Supermarché	Heures d'ouverture: Aujourd'hui ouvert 8:30 - 19:00 Demain ouvert 8:30 - 19:00
Aldi Supermarkt Lange Lozanastraat Lange Lozanastraat 177-179 2018 Antwerpen Supermarché	Heures d'ouverture: Aujourd'hui ouvert 8:30 - 19:00 Demain ouvert 8:30 - 19:00

<https://www.supermarche-match.be/be-fr/storelocator>

https://www.heures-douverture.com/fr/chaine_de_magasins/Aldi?rows=1000&page=1

- 1 Récupération avec scrapping des adresses des 2 chaînes de magasin
- 2 Conversion en longitude et latitude des adresses des magasins et de l'utilisateur avec GeoPy Nominatim qui envoie une requête HTTP
- 3 Calcul du temps de trajet et de la distance kilométrique en voiture avec OSRM API - Route service
- 4 Récupération des prix du carburant sur <https://carbu.com/france/prixmoyens> et calcul du prix avec la consommation de la voiture et distance du trajet

```
GET /route/v1/{profile}/{coordinates}
```

	nom_mag	adresse	lat	long
0	Match Andenne	Avenue Roi Albert 17 5300 Andenne	50.492729	5.103211
1	Match Ath	Chaussée de Bruxelles 62 7800 Ath	50.631359	3.785517
2	Match Auvelais	Rue de Falisolle 4 5060 Sambreville	50.446596	4.632886
3	Match Beaumont	Rue Germain Michiels 46 6500 Beaumont	50.237433	4.240203
4	Match Binche	Rue Zéphirin Fontaine 140 7130 Binche	50.420014	4.170846
...
81	Smatch Woluwe	Priester de l'Epéestraat 11 1200 Sint-Lambrechts Woluwe	50.843147	4.412830
82	Smatch Zele	Gentsesteenweg 228 9240 Zele	51.057032	4.010263
83	Smatch Zeveneken	Drongenastraat 21 9160 Lokeren	51.100933	3.907545
84	Smatch Zottegem	Buke 20 9620 Zottegem	50.876363	3.815831
85	Smatch Zulte	Staatsbaan 161 9870 Zulte	50.920708	3.443827

RÉSULTATS



Résultats

Au final nous avons développé une application web sous Flask qui permet à l'utilisateur de :	Rechercher une recette de cuisine , que nous scrappons en direct (via "750g.fr") et laissons le choix à l'utilisateur parmi plusieurs propositions
	Nous retournons la liste des ingrédients nécessaires en laissant la possibilité à l'utilisateur de supprimer des ingrédients qu'il possède déjà ainsi que les étapes de la recette
	Nous scrappons 2 sites marchands : Aldi et Supermarchés match
	Nous comparons les prix et la distance des magasins, et suggérons le meilleur magasin pour chaque marque à l'utilisateur

Scrapy Meal

Scrapy Meal

Nicolas CARVAL Bruno PINCET

ESILV

LOOKING FOR A RECIPE ?

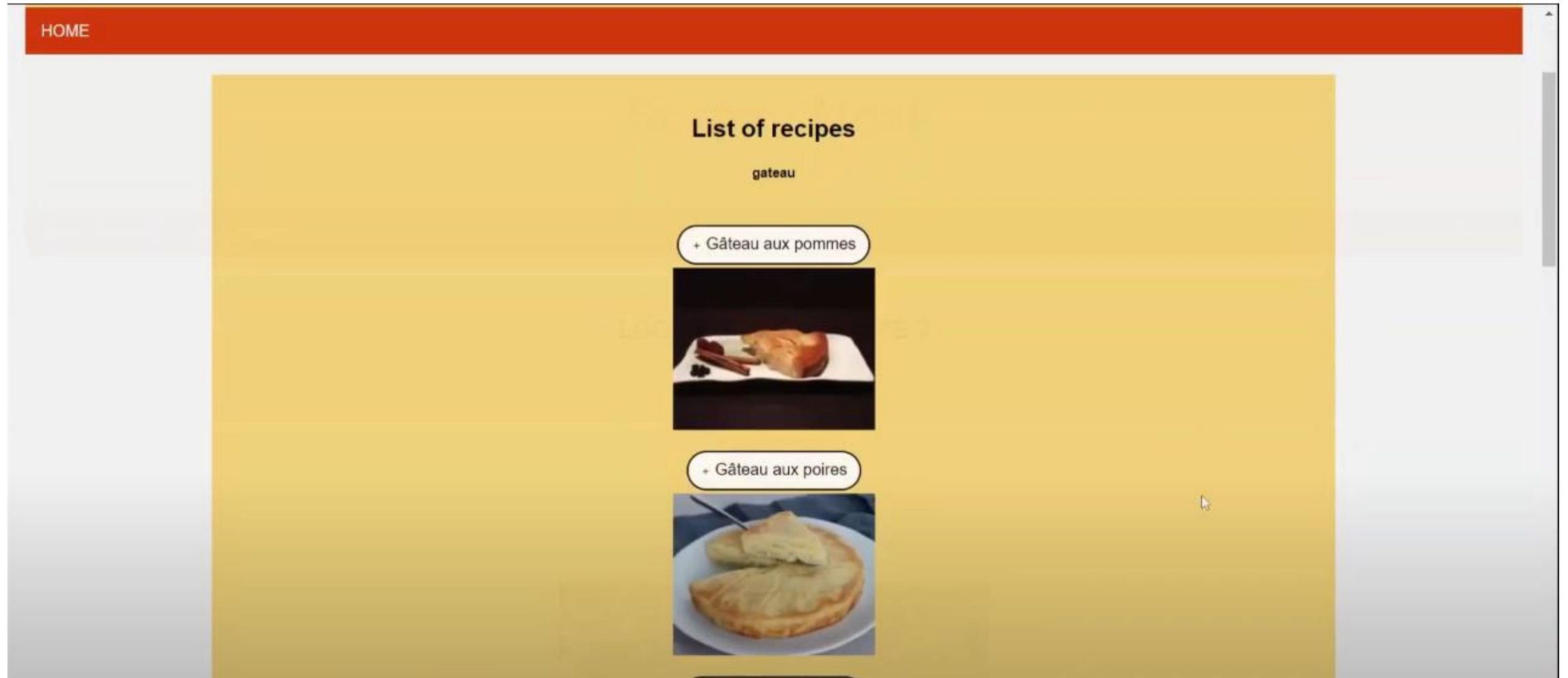
type-in a meal...

ex : croissant

FIND



Scrapy Meal



Scrapy Meal

HOME

List of ingredients

v levure fraîche - 20 - g

v eau froide - 220 - g

+ beurre ferme (pour tourer) - 140 - g

v beurre ramolli - 100 - g

+ c à c de sel - 2 - unités

v farine T45 - 140 - g

v sucre - 60 - g

v poudre de lait - 10 - g

v farine T55 - 240 - g

+ c à s de lait - 1 - unités

+ jaune d'œuf - 1 - unités

GO

ESILV - Webscrapping and Data processing - 2022/2023

SCRAPY MEAL

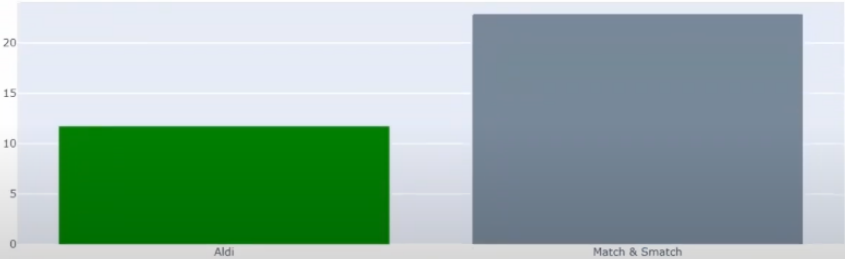
Our suggestions

Total difference : 64.32 %

Available at M&M : 9

Available at Aldi : 10

Prix total le moins cher



	Prix Aldi	Prix Match & Smatch	
beurre	2.95	2.25	
cacao amer	1.89	3.49	
sucre	1.69	2.99	
levure chimique	1.19	0.49	
farine	0.49	7.74	
yaourt brassés	1.59	1.89	
mascarpone	1.39	1.62	
sucre vanillé	0.55	2.40	
	Nom du magasin	Distance km	Prix Trajet
Match	Match Momignies	213.4	16.576912
Aldi	Aldi Supermarché	220.5	17.128440

750g



Scrapy Meal

to PINCET

LOOKING FOR A RECIPE ?

type in a meal...

ex : croissant

FIND



BILAN

- Nous avons pu mettre en pratique nos connaissances de webscrapping et de développement pour arriver à un prototype fonctionnel.
- La recherche et sélection des recettes et des ingrédients a été implémentée
- La comparaison de la distance géographique séparant l'utilisateur des marchands et le coût du trajet a été implémentée
- Le scrapping des grandes surfaces françaises est très compliqué et remplacé par des magasins belges.
- Les recettes avec un nombre d'ingrédients important peuvent être longues à scraper, un scrapping journalier des prix des ingrédients les plus populaires pourrait permettre de limiter cela.

SCRAPY MEAL

+



○



MERCI

Nicolas CARVAL – Bruno PINCET

Annexes

GitHub https://github.com/NicolasCarval/Scrapy_Meal

- Code https://github.com/NicolasCarval/Scrapy_Meal/blob/master/app.py
- Lien de l'application <http://localhost:8989>
- Instructions de déploiement : https://github.com/NicolasCarval/Scrapy_Meal#readme

Sites et sources utilisés :

- <https://www.750g.com/>
- <https://carbu.com/france/prixmoyens>
- <https://www.supermarchesmatch.be>
- <https://www.aldi.be/fr>
- <http://router.project-osrm.org>
- https://www.heures-douverture.com/fr/chaine_de_magasins/Aldi?