Projet Final Webscrapping

Lien vers notre projet sur GitHub

Par Maud Tissot, Arthur Scelles, Aurélien Pouxviel





Objectifs du projet

Au cours de ce projet nous représentons la ville de **Vannes** dont les objectifs pour 2024 sont :

- Promouvoir un tourisme éthique
- Améliorer la vie de ses citoyens
- Redynamiser la ville et attirer les jeunes ménages
- Donner de la visibilité aux établissements qui s'engagent dans une politique durable

Ainsi, nous avons décidé de créer une application de recommandation de restaurants interactive et simple d'utilisation, qui, en plus de s'appuyer sur des critères conventionnels, permet de personnaliser le choix du restaurant selon les envies et l'humeur de l'utilisateur. Nous souhaitons notamment mettre en avant des restaurants qui respectent nos critères RSE (Responsabilité Sociétale des Entreprises) d'inclusion et de développement durables : accès aux fauteuils roulants, menus végétariens, végans, options pour les intolérances, prix abordables etc...



Methodologie



1 – Choix du Sujet

2 – Extraction des données

3 – Visualisation des données

4 – Résumé des commentaires

5 – Recommandations de restaurants

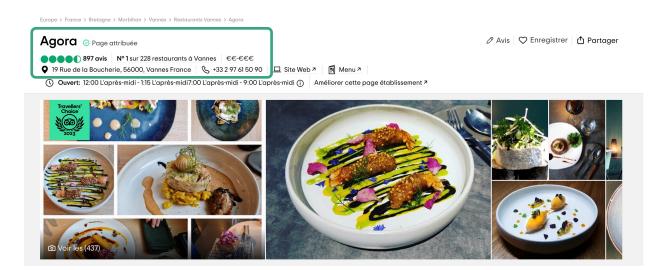
6 – Merge des différentes

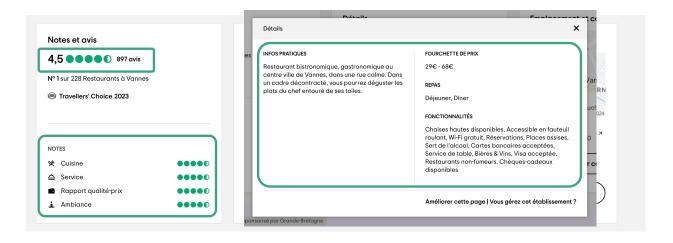
étapes

Étapes simultanées

PARTIE I – Extraction des données

1 – Trip Advisor







Allen, Texas

图 30 凸 11

Avis écrit le 6 août 2023 par mobile Hidden Gem in Brittany!

Google Traduction

Magnificent! I had the most exquisite dining experience at Agora. From the moment I stepped inside, I was transported to a world of elegance and sophistication. The chef, not only creates the food, but is an artist and his paintings were displayed, tastefully throughout. The ambiance was a perfect blend of timeless charm and modern luxury, setting the stage for an unforgettable evening.

The six course lobster tasting menu was our choice and it was extraordinary. Each course was a masterpiece, crafted to tantalize both the palate and the senses. Chef Guillaume prepared the courses with subtle contrasts in textures, flavors, and ingredients.

Delphine chose the perfect wine for us, to compliment the food pairing. Her warm welcome and attention to our table made us feel like A-List celebrities.

Chef Guillaume and Delphine have created an experience that's not just a meal; it's an immersive journey into the heart of French gastronomy, a symphony of flavors and sensations that will linger in my memory for years to come. If you are visiting Brittany, and want an exceptional experience, let Chef Guillaume and Delphine speak to you in the universal language of fine dining.

PARTIE I – Extraction des données

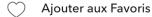
2 – Guide Michelin



La Tête en l'air

43 rue de la Fontaine, Vannes, 56000, France

€€€ · Cuisine créative



L'avis du Guide MICHELIN



Une étoile : une cuisine d'une grande finesse

Un jeune couple dynamique et accueillant, qui a bel et bien la tête... sur les épaules. Dans une ambiance décontractée, Clément Raby le parisien et Estelle Mercier la gardoise pratiquent, comme ils l'indiquent sur leur carte de visite, "une cuisine libre". C'est-à-dire une cuisine créative maîtrisée et originale, avec des associations pertinentes, des recettes qui tombent juste! Dans leurs menus à l'aveugle, les plats ne sont annoncés qu'après leur dégustation - pour mieux mettre en éveil les sens des convives...

1		name	chef	cuisine	distinction
	0	Ryoko - Comptoir à ramen		Spécialité Ramen	
	1	Bvañ		Cuisine moderne	
	2	La Gourmandière - La Table d'Olivier	Olivier Samson	Cuisine moderne	1 star
	3	La Tête en l'air	Clément Raby	Cuisine créative	1 star
	4	Roscanvec	Thierry Seychelles	Cuisine moderne	
	5	La Gourmandière - Le Bistr'Aurélia		Cuisine moderne	
	6	L'Annexe		Cuisine moderne	
	7	Le Sous-sol		Cuisine moderne	bib
	8	lodé		Cuisine créative	
	9	Empreinte	Baptiste Fournier	Cuisine du marché	

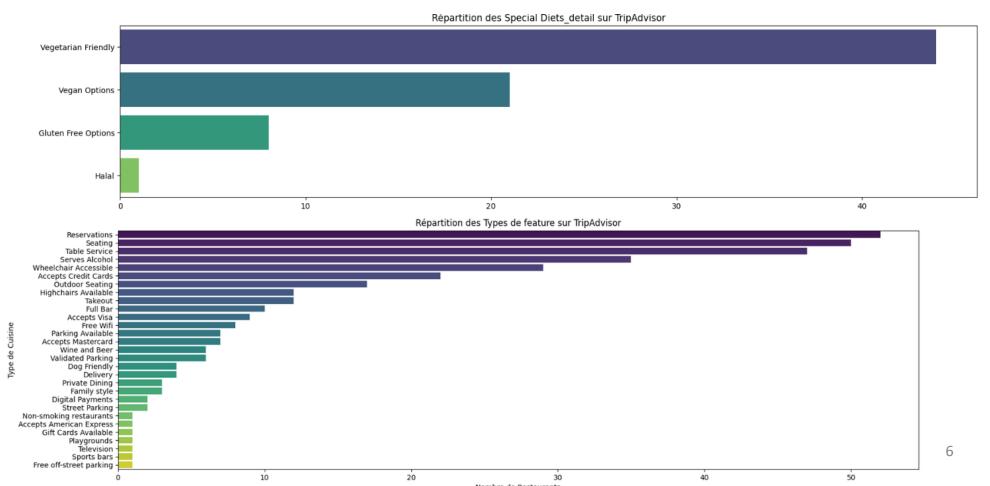
PARTIE II - Visualisations

1 – Exploration des données

Nous avons affiché la répartition des restaurants qui proposent au moins une option de menu, puis les features spéciales qui pourraient nous intéresser pour définir un restaurant comme respectant certains critères RSE.

Inclusion et développement durable

Inclusion comme accès handicapé



PARTIE II - Visualisations

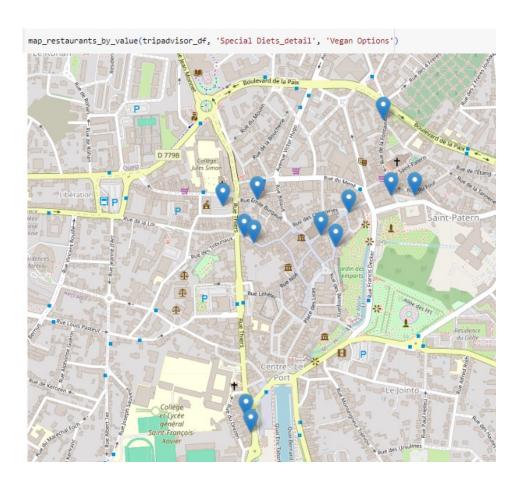
2 – Visualisation par critère

En récupérant la localisation des restaurants à partir de leurs adresses, nous avons obtenu des cartes selon des critères tels que la qualité et le type de restaurant (notes, étoiles, commentaires, cuisine, cadre, etc...)

3 - Permettant au futur utilisateur d'orienter facilement sa recherche vers des restaurants éthique avec une politique sociale et durable.

Nous avons défini les critères qui permettent de définir un restaurant engagé:

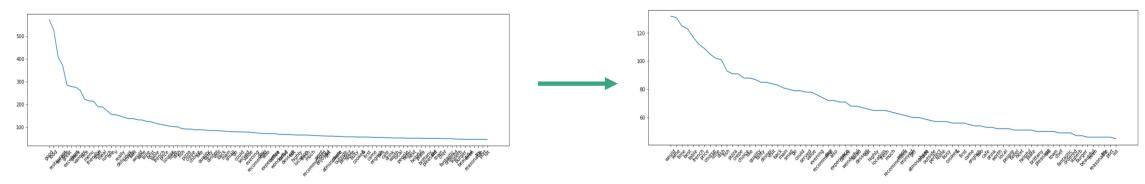
- Alimentation végan ou végétarienne
- Restaurants proposant un accès aux fauteuils roulants
- Etc...



Carte des restaurants proposant une option Vegan

PARTIE III - Resumé des commentaires

1 – Traitement des commentaires



Nombre de mots uniques dans les commentaires :

- Avant traitement: 8500

- Après traitement : 4700

Sélection des 4600 mots les moins utilisés nous avons fait attention à garder les mots valorisants un restaurant engagé.

2 – Création des résumés

En utilisant le LLM pré-entrainé t5-base, pour chaque restaurant nous avons généré un résumé des commentaires par restaurant afin d'avoir une cohérence et d'extraire les informations nécessaires à la recommandation de restaurants.



PARTIE IV – Recommandations de restaurants

1 – Déroulement pour l'application

2– Etat final de notre application python

- Pour chaque résumé de commentaires
- On analyse le sentiment de l'utilisateur
- On garde les restaurants dont le sentiment est positifs

	input	output
0	the chef is an artist and his paintings were	POSITIVE
1	the restaurant is run by a husband and wife t	POSITIVE
2	the food was excellent with a tasty amuse Bou	POSITIVE

 On applique le modèle TF-IDF sur ces commentaires pour faire de la recommendation de restaurants.







Live demo : streamlit application

Problèmes majeurs et solution

PROBLEMES

Manques de données pour certain restaurant

Imbriquer les différentes parties du projet

Risque de manque de temps en clôture du projet



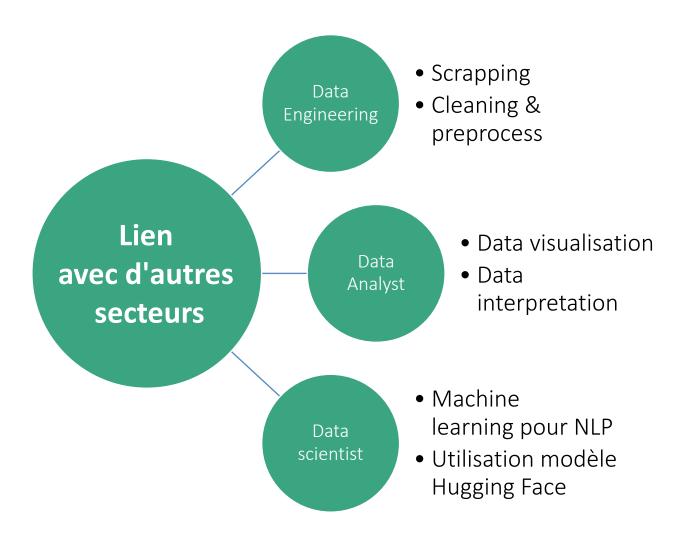
SOLUTION

Scraping complémentaire

Définir clairement les inputs et output de chaque partie

Répartition des travaux et deadline à l'avance

Feedbacks



Appris à poser une problématique claire avec enjeux RSE

Scrapper dans un but final -> scrapper précisément ce qu'on veut

Faire un projet "End-to-End", c'està-dire du scrapping jusqu'à l'utilisation de modèles de ML