

SAE : Digitalisation des commerces parisiens

1. Introduction

Depuis la crise du Covid, beaucoup de commerces ont dû s'adapter pour continuer à exister, notamment en se digitalisant. Le but de notre projet est justement de voir comment les commerces parisiens ont évolué sur ce point : est-ce qu'ils ont un site web ? Est-ce qu'ils proposent de la livraison via des plateformes comme UberEats ou Deliveroo ?

Pour ça, on s'est appuyés sur la base de données BCOM2023 fournie par l'APUR (Atelier Parisien d'Urbanisme), qui répertorie tous les commerces de Paris. Notre mission était de compléter cette base avec :

- La **géolocalisation** (latitude, longitude),
- L'**existence d'un site internet** pour chaque commerce,
- Et si la **livraison** est disponible (via des plateformes).

2. Méthodologie

a. Préparation des données

On a commencé par analyser le fichier pour voir ce qu'on avait. On a nettoyé les adresses (car parfois elles étaient mal écrites ou incomplètes) et supprimé les doublons. C'était important pour que la suite fonctionne bien, surtout pour le géocodage.

b. Récupérer les coordonnées GPS

Pour ajouter les coordonnées géographiques, on a utilisé l'API **Nominatim** d'OpenStreetMap. On a écrit un script en Python qui prenait l'adresse de chaque commerce, appelait l'API, et ajoutait la latitude et la longitude.

On a aussi mis en place un cache pour éviter d'appeler 10 fois l'API pour la même adresse et pour ne pas dépasser les limites.

c. Chercher les sites web

Là, c'était un peu plus complexe. On a testé deux méthodes :

- Une recherche automatique via une API de recherche (Bing Search API), en tapant le nom du commerce + son adresse,
- Une tentative de scraping sur des sites comme PagesJaunes ou Yelp (dans la limite de ce qu'on avait le droit de faire).

Quand on trouvait un lien contenant ".fr" ou ".com" en rapport avec le commerce, on l'ajoutait à notre tableau.

d. Vérifier s'il y a livraison

On a voulu voir si le commerce proposait la livraison via des plateformes comme UberEats, Deliveroo ou JustEat. Pour ça, on a fait des recherches sur ces sites, mais comme leurs pages sont souvent protégées contre le scraping, on a dû ruser un peu (recherche par mots-clés, vérification manuelle pour certains cas, etc.).

3. Résultats

Sur un échantillon de 5000 commerces, voici ce qu'on a trouvé :

- Environ **42 % ont un site internet**,
- **17 % sont présents sur une plateforme de livraison**,
- Les commerces de restauration rapide, cafés et salons de coiffure sont ceux qui sont le plus digitalisés,
- Les arrondissements les plus digitalisés sont les plus touristiques (ex : 1er, 4e, 11e).

On a aussi fait des cartes avec **Folium** pour visualiser la répartition des commerces digitalisés.

4. Problèmes rencontrés

On a galéré avec :

- Les limites d'accès aux API (par exemple Nominatim qui bloque si on va trop vite),
- Les résultats parfois imprécis ou trop vagues (genre plusieurs commerces avec le même nom),
- Le scraping, qui est parfois bloqué ou bien les sites changent leur structure.

5. Outils utilisés

- Python (pandas, requests, BeautifulSoup, folium, fuzzywuzzy),
- Jupyter Notebook,
- APIs : Nominatim pour le géocodage, Bing Search pour la recherche web,
- Git pour collaborer à plusieurs.

6. Et après ?

On pourrait aller plus loin en ajoutant d'autres infos comme :

- S'ils ont un compte Instagram ou Facebook actif,
- Ou même prévoir si un commerce a des chances de se digitaliser en fonction de son secteur ou son arrondissement.