

I - Groupe

Participants: OUENADIO Alexandre, GOSSON Pierre, GOURIACH Souad, GORVIEN Mathis

Nom groupe:

II - Organisation de travail

Nettoyage des données airbnb :

- RAW -> CLEAN -> GOLD
- Faire attention au transtypage (ne pas hésiter à faire du print_schema)

Web scraping données événements sur Lyon + géolocalisation (Beautiful Soup):

Alexandre

Mise en relation des données airbnb - géolocalisation:

- Distance événements-bâtiments (studio ou maison)
- Plus gros foyers de location (à voir si c'est simple ou pas)

Visualisation avec Power BI: Pierre

Schéma architecture du projet:

Github: <https://github.com/AlexandreOuenadio/datathon-2024>

Ne pas oublier de faire des sauvegardes intermédiaires !

III - Définition du sujet

Problématique:

Questions auxquelles il faut répondre:

- Est-ce que c'est innovant ?
- Comment seront présentés les résultats ?
- Quelles sont les données à rassembler ?
- Quelles sont les phases de réalisation du projet ?
- Est-ce qu'il y a assez de variations dans les données pour qu'elles soient représentatives

Dataset: [1GB MAX pour être performant]

Datasets de croisement:

-

Pour la restitution ils aiment bien : savoir le volume de données, cmb de fichier au départ, travail sur combien de lignes.

Problématique : Je suis un propriétaire qui souhaite louer mon appartement, mais je ne sais pas quelle période me rapportera le plus d'argent (événements ou autre bonifie le prix de location).

<https://insideairbnb.com/get-the-data/>

On prend les données de Lyon : **23 March, 2024 . 15 December, 2023**

Analyse en deux temps :

- Caractérisation de % profils de propriétaires : Location de Studio ou plutôt maison pour famille, si dans le centre ville ou à l'extérieur. (potentiellement on peut tricher et se permettre de prendre des profils bien renseignés sur la carte, si l'étude de profil n'est pas concluante)
- Mise en place d'une analyse de prix de location en fonction du temps et des événements. Par exemple je veux louer 2 semaines, on peut mettre un slider de deux semaines pour savoir environ où maximiser l'argent
 - retirer événements des sites et transformer la data (beautiful soup) bien le montrer à la restitution.
<https://www.visiterlyon.com/sortir/l-agenda/tous-les-evenements>
 -
 - caractéristique distance de l'événements puis savoir si événement global (pas important la dist) ou la distance impact directement le prix (événements moins importants pas significatifs pour des locations loins)

Marche à suivre (chatGPT):

Fichier **listings.csv**

- **Propriétés** : Les colonnes **property_type**, **room_type**, et **accommodates** permettent de distinguer les studios, appartements ou maisons, ainsi que les propriétés adaptées aux familles.
- **Emplacement** : Les colonnes **neighbourhood** et **latitude/longitude** permettent d'identifier si une propriété est dans le centre-ville ou à l'extérieur.
- **Prix** : La colonne **price** donne le coût de location.

Fichier **calendar.csv**

- **Disponibilité et prix ajustés** : Les colonnes **date**, **available**, et **adjusted_price** permettent d'analyser les variations des prix au cours du temps, notamment pendant des périodes spécifiques (comme des évènements).
- **Nuitées minimales et maximales** : Les colonnes **minimum_nights** et **maximum_nights** permettent d'affiner les recherches.

Fichier **neighbourhoods.csv** et **neighbourhoods.geojson**

- Ces fichiers apportent des informations sur les quartiers, facilitant la catégorisation géographique des locations.

Fichier **reviews.csv**

- Les colonnes **comments** et **date** peuvent fournir des informations qualitatives pour comprendre l'impact des évènements ou des saisons sur les expériences des locataires.