



Présentation

La métropole de Lyon, dans une optique d'amélioration d'une prise en charge des lieux touristiques à Lyon notamment via les transports en commun, a demandé une étude sur les données des photos prises sur Flickr.

L'objectif

Comment rendre plus accessible les lieux touristiques peu visités de Lyon par l'intermédiaire de nos transports en commun ?

La démarche

- Collecter les données
- Préparation des données en respectant le RGPD
- Définition de nos indicateurs clés
- Visualisation et recommandations
- Conclusion

3 questions

- **Quels sont nos lieux touristiques ?**
- **Est-ce que les visites des utilisateurs de Flickr correspondent aux lieux touristiques ?**
- **Est-ce que notre transport en commun permet d'y accéder simplement ?**



Caractérisation des points d'intérêt de Lyon via Flickr

Présentation du dataset

Le dataset

Source : Flickr (source publique)
Nombres de lignes : 83 826
Nombres de colonnes (features) : 16

Après vérification

68 635 doublons
15173 photos uniques
825 utilisateurs uniques

Nom	Type	Description
id	int64	Id de la photo
user	object	Id de l'utilisateur
lat	float64	Latitude
long	float64	Longitude
tags	object	Hashtags
title	object	Description de la photo
date_taken_minute	object	Minute de la prise de photo
date_taken_hour	object	Heure de la prise de photo
date_taken_day	object	Jour de la prise de photo
date_taken_month	int64	Mois de la prise de photo
date_taken_year	int64	Année de la prise de photo
date_upload_minute	int64	Minute de l'upload de la photo
date_upload_hour	int64	Heure de l'upload de la photo
date_upload_day	int64	Jour de l'upload de la photo
date_upload_month	int64	Mois de l'upload de la photo
date_upload_year	int64	Année de l'upload de la photo

Registre RGPD - Catégories de données nécessaires pour nos indicateurs :

lat / long / tags / title / date_taken_hour / date_taken_day / date_taken_month /
date_taken_year / date_upload_hour / date_upload_day / date_upload_month /
date_upload_year

*Les données sont publiques et aucunes informations sensibles seront utilisées donc pas de
précaution à prendre en particulier.*



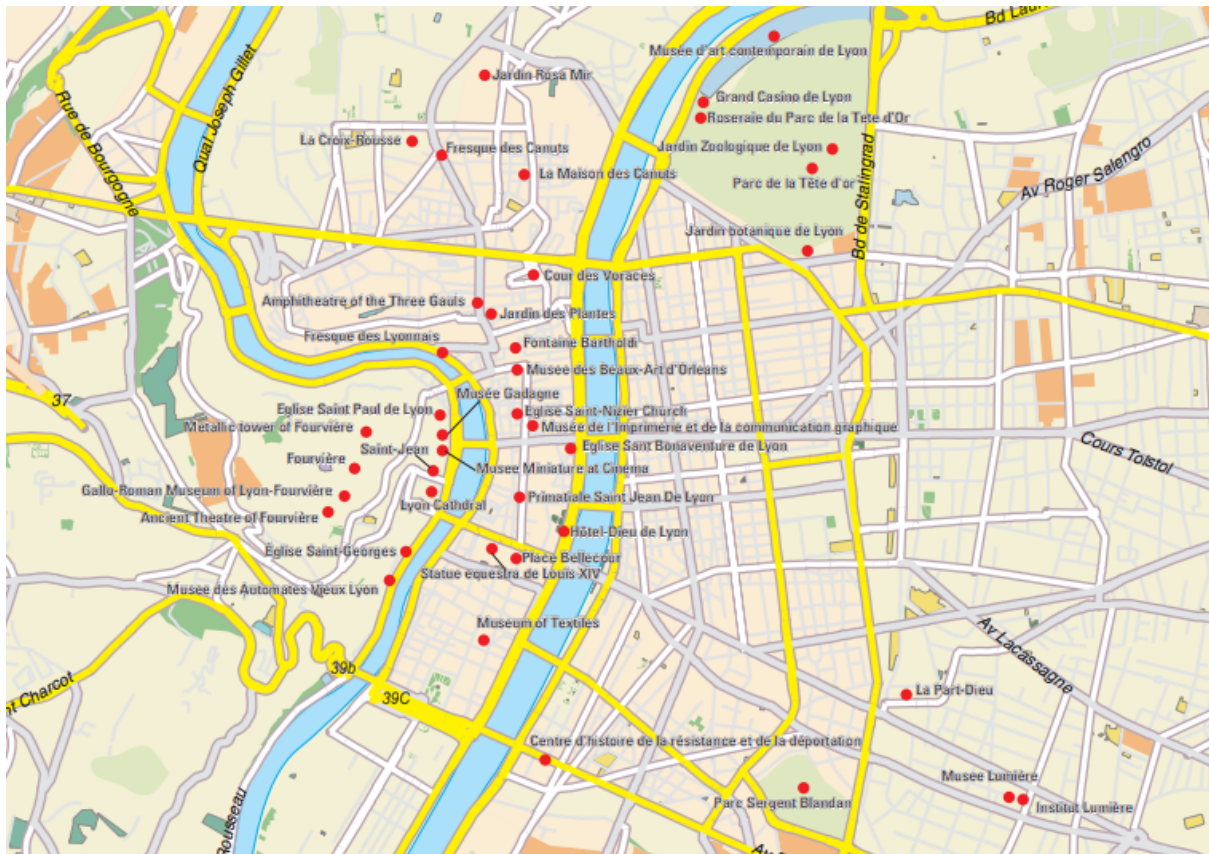
Nos étapes de 'nettoyage' du dataset

- **Suppression de la Ligne 1554 du dataset qui empêche l'import**
- **Enlever les espaces blancs en début des titres des features (colonnes)**
- **Suppression des colonnes 'Unnamed: 16' / 'Unnamed: 17' / 'Unnamed: 18'**
- **Suppression des lignes duplicates (en doublon)**
- **Suppression des 'id' duplicates**
- **Suppression des features 'id' et 'user' pour l'anonymisation**
- **Suppression des features 'date_taken_minute' et 'date_upload_minute'**
- **Suppression des lignes NaN/Null pour 'tags' et 'title'**
- **Suppression des années aberrantes (celles avant la création de Flickr 2004)**
- **Suppression des géolocalisation hors Lyon**
- **Convertir date_taken_upload / Taken en Datetime**



Caractérisation des points d'intérêt de Lyon via Flickr

Quels sont nos lieux touristiques de Lyon ?



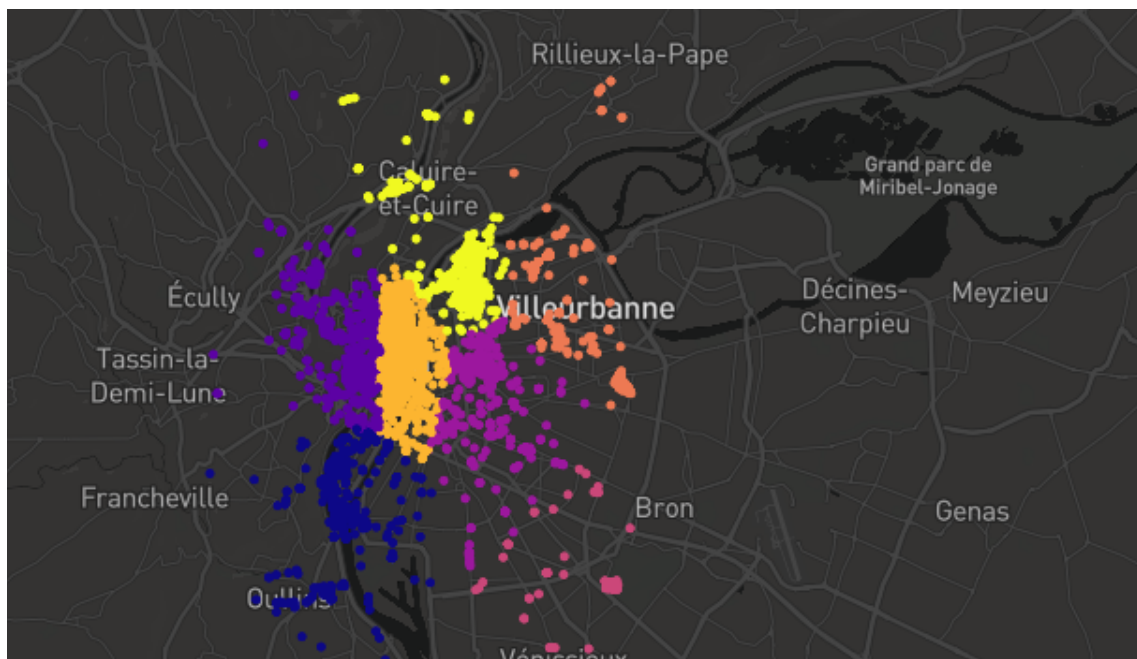
Source : www.cityzeum.com



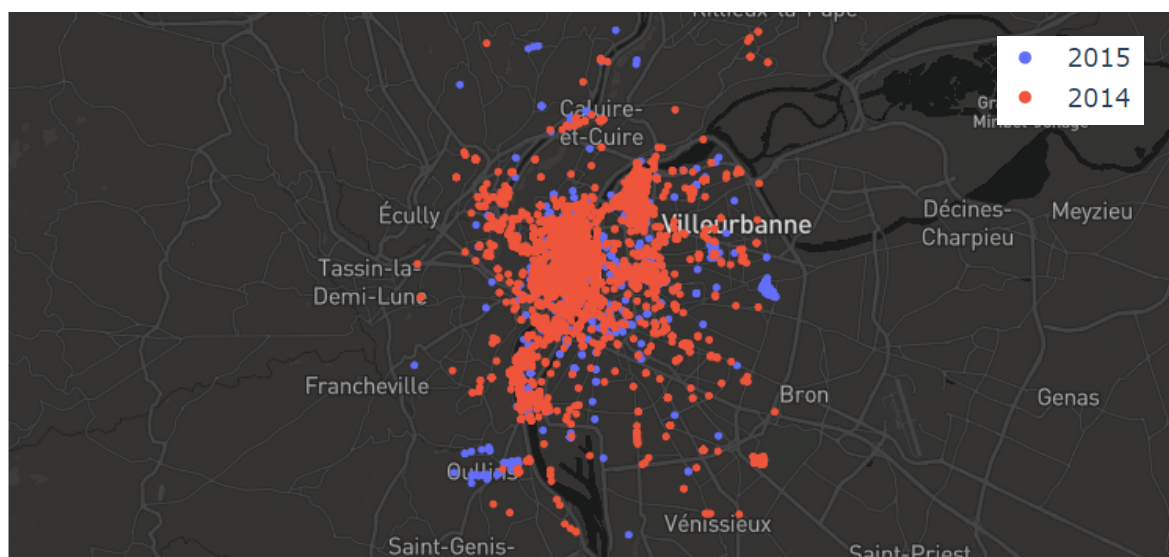
Caractérisation des points d'intérêt de Lyon via Flickr

Est-ce que les visites des utilisateurs de Flickr correspondent aux lieux touristiques ?

Analyse par clustering des photos prises dans Lyon



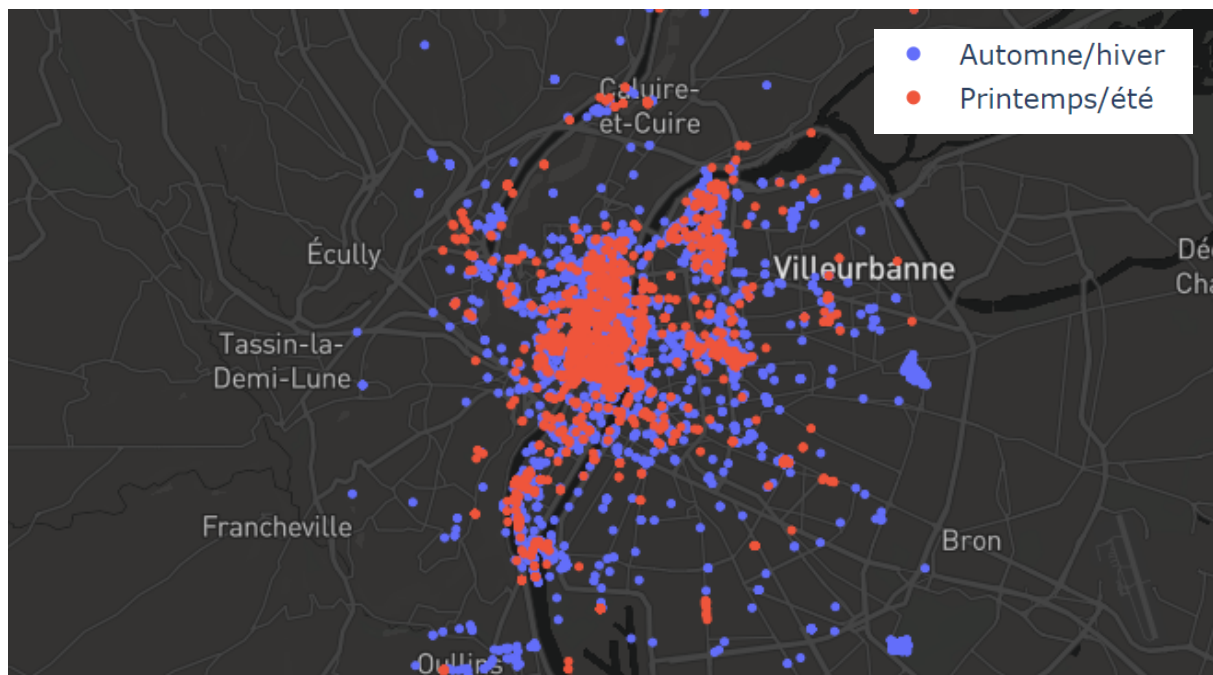
Analyse de l'évolution entre l'année 2014 et 2015 des photos prises à Lyon



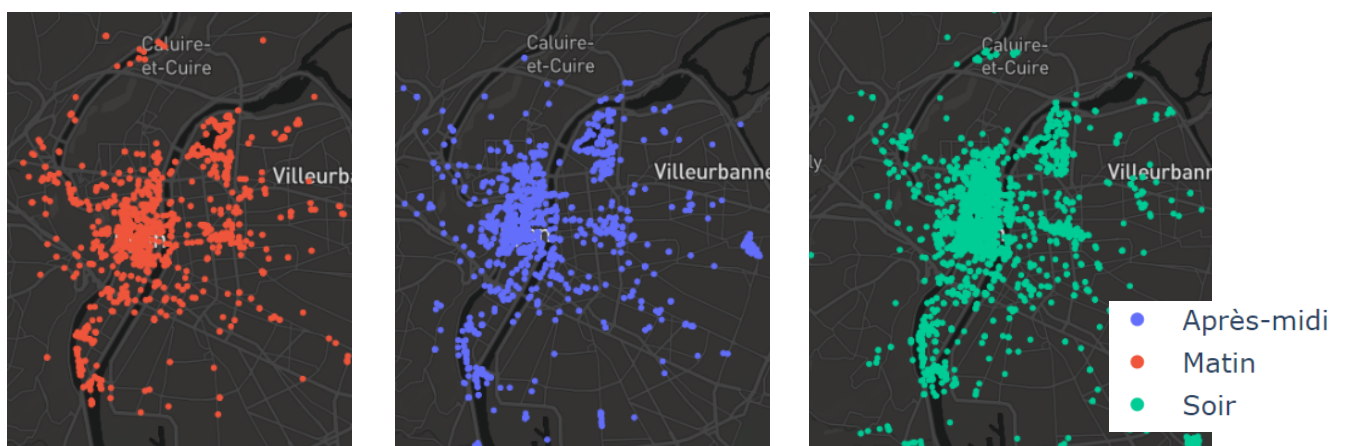


Caractérisation des points d'intérêt de Lyon via Flickr

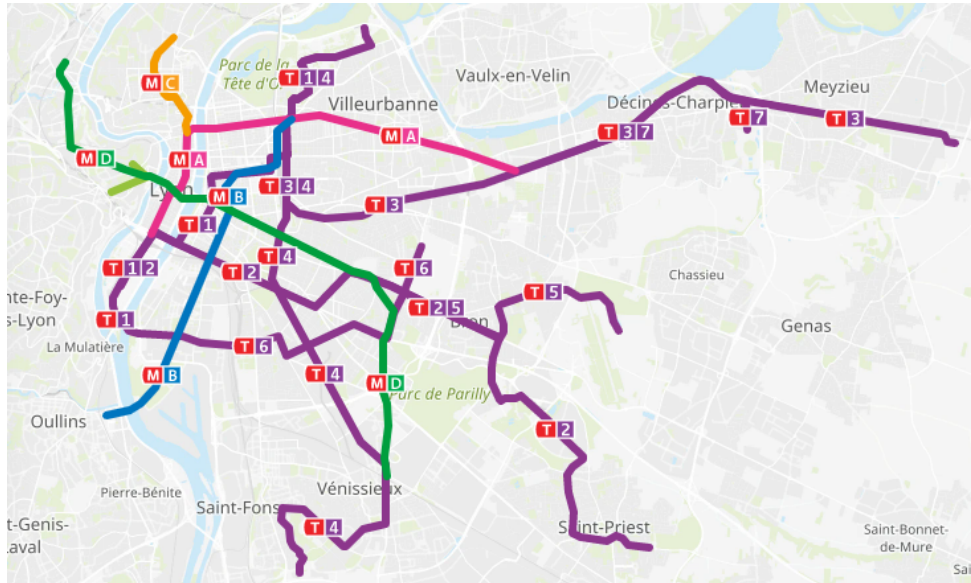
Analyse de l'évolution par saison des photos prises à Lyon
printemps/été (2692 photos) vs automne/hiver (8050 photos)



Analyse de l'évolution par heure des photos prises à Lyon
matin / après-midi / soir



Est-ce que notre transport en commun permet d'y accéder simplement ?



Recommandations

- Le réseau TCL est adapté aux différents lieux touristiques de Lyon.
- Le parcours des utilisateurs de Flickr dans Lyon correspond aux différents lieux touristiques de la ville.
- Globalement, il n'y a pas de problème d'accessibilité aux différents lieux touristiques.



Rétroplanning

Prise en main du sujet / dataset (lundi 9/11 à 17h)

- Premières observations sur la dataset
- Création d'un dossier Github
- Définition des métriques

Datacleaning / export du dataset final (mardi 9/11 à 13h)

- Définition des différentes étapes du datacleaning
- Réalisation du datacleaning
- Réalisation de notre clustering
- Export en .csv

Analyse NLP (mercredi 10/11 à 10h)

Visualisations de nos metrics (mercredi 10/11 à 10h)

Finalisation du dossier + Conclusion (mercredi 10/11 à 12h)