



INSA Lyon  
20, avenue Albert Einstein  
69621 Villeurbanne Cedex

---

# Datamining

## « Project Flickr »

du 24 Février 2014 au 10 Mars 2014

---



*Hexanôme H4211 :*  
Rémi DOMINGUES  
Adrien DUFFY-COISSARD

*Enseignants :*  
Mehdi KAYTOUE  
Jean-François BOULICAUT

## Sommaire

1	Introduction	3
2	Analyse et nettoyage des données	4
3	Clustering	5
4	Évaluation des clustering	7
5	Description des clusters	7
6	Visualisation des résultats	8
7	Interprétation des résultats	9
8	Something ?	10

## 1 Introduction

Cette étude doit permettre à terme au Grand Lyon d'améliorer la gestion de ses transports en commun et la vie de ses touristes.

L'objectif de celle-ci réside donc premièrement dans l'identification des points d'intérêts situés dans la ville de Lyon. Pour ce faire, l'utilisation d'algorithmes de clustering dans un contexte de fouille de données doit permettre l'identification des zones visées.

À cet effet, ledit clustering s'effectuera sur la base d'un fichier CSV obtenu grâce à la collecte de photos géolocalisées via l'API Flickr.

## 2 Analyse et nettoyage des données

Dans un premier temps, il est nécessaire d'effectuer un travail préliminaire sur les données, afin de s'assurer de leur validité et cohérence. Nous analysons donc les données à la recherche de problèmes.

### 3 Clustering

Essayer d'obtenir la géolocalisation des photos en fonction de l'heure. Ceci est important, car on peut séparer les photos par horaire : une photo prise de nuit est probablement plus sujette à être une photo de bar ou tout autre lieux d'activité nocturne, contrairement à la journée, ou on retrouve plus les monuments.

FIGURE 1 – Estimation du coût de réalisation de la solution  
TODO DBSCAN

## 4 Évaluation des clustering

## 5 Description des clusters

TODO : Google Places et gestion des tags ?

## 6 Visualisation des résultats



## **7 Interprétation des résultats**

Comment votre analyse peut-elle aider le Grand Lyon ? Quelles connaissances lui apporte-t-elle ?

## 8 Something ?

On a fait des trucs en plus ?  $\Rightarrow$  Passage à l'échelle ? Tâches prédictives ? Analyse dynamique et non statique ? Autres sources ? (Tweets, Instagram, ...)