

Projet BI

Thématique : Ville de Paris

Introduction

Le projet BI vise à mettre bout-à-bout les différentes notions vues dans la première moitié de l'enseignement : ETL, Datawarehouse, OLAP, Reporting. Le projet BI servira aussi de support de TP pour la partie DataMining.

Organisation et contraintes

Le projet est organisé en binômes. Sa thématique est centrée sur la Ville de Paris et doit donc s'y rapporter d'une façon ou d'une autre.

Les contraintes du projet sont les suivantes :

- Le projet doit être centré sur une problématique claire d'étude
- Le projet doit être construit à partir de plusieurs tables de données extraites depuis le site ParisData
- Une partie des données doit être issue d'une capture Web
 - A travers une API
 - A travers d'autres sources de données
 - A travers de l'extraction de pages HTML
 - <http://www.garethjames.net/a-guide-to-web-scraping-tools/>
- Une attention particulière doit être portée à la dimension Localisation qui doit être la plus fine possible (rue, arrondissement, coordonnées GPS, éventuellement quartier, ...)
- Le projet doit viser à la construction de trois schémas (data-mart) + 3 cubes OLAP associés
- Une analyse des données sous forme de graphiques simples (générés à partir de la sortie des cubes OLAP + outil de visualisation) doit être proposée
- Une visualisation cartographique doit être proposée

Thématiques

- Les transports à Paris (Métro, Autolib, Velib, Parking, Fréquentation du métro/RER, Utilisation Velib, Autolib, ...) + les accidents éventuellement
- La santé à Paris : Centres de santé, données de consultation (voir données de la sécurité sociale), extraction des médecins à Paris (site Ameli), etc....

Référence : <http://jplusplus.github.io/guide-du-datajournalisme/>

Date de rendu = 14 mars