
Etude déplacements vélo ville Bruxelles

lundi 10 décembre 2018

NOM : TAHI
PRENOM : MOHAND TAHAR

NOM : BOUADDA
PRENOM : BAHIYA

Master 1 BigData
Université PARIS 8

Introduction :

Le projet consiste en :

- Constituer une base de données pour étudier les déplacements à vélos dans Bruxelles
- La collecte des données s'est faite à partir du service web de Vélib: <https://developer.jcdecaux.com/#/opendata/vls>
- Réaliser un script en python qui va récupérer quotidiennement les trajets effectués en Vélib, sous la forme de document JSON
- Produire des graphiques montrant les vélos et les places disponibles pour l'heure du jour
- Mettre en œuvre un mini algorithme pour indiquer aux utilisateurs quelles sont les stations qui ont le plus de chances de proposer une disponibilité emprunt/retour, à un horaire donné.

Constitution de la base de données :

Pour la constitution de la base de données, on récupère nos données grâce au site données dans l'énoncé du projet, c'est à dire : <https://developer.jcdecaux.com/#/opendata/vls> , les données sont stockées dans document JSON et ensuite chargées dans la base de données, Cela se fait grâce, à un script sh, sous le nom de « donneesVelo.sh »

Qui comporte :

la commande curl pour récupérer les données du site

la commande mongoimport pour charger dans la base de données

L'automatisation de la récupération et du chargement, se fait grace au programme Crontab, qui permet d'éditer les tables de configuration du programme Cron. Ces tables spécifient les *tâches* à exécuter et leur horaire d'exécution avec éventuellement une périodicité.

On paramètre notre cron, pour qu'il exécute le script toutes les heures, Ce qui fait nous avons de nouvelles données qui s'ajoutent a nos anciennes données stockées dans la base, chaque une heure.

Analyses de données :

Nous analysons nos données dans le fichier python suivant :
« traitement.py », nous avons utilisé les bibliothèques suivantes :

- ❖ math
- ❖ pymongo
- ❖ pprint
- ❖ datetime
- ❖ matplotlib.pyplot
- ❖ numpy

On y trouve nos différente fonction de traitement des données récupéré :

liste_nomStations() : Nous renvoie la liste des stations chargé dans la base de données

statistique() : Une fonction qui nous affiche quelques statistique sur les vélibs à Bruxelles

position_gps() : Fonction qui nous affiche un graphe décrivons la positions GPS de toutes les stations

`dispo_velo(heure)` : Fonction qui nous affiche un graphe décrivant la disponibilité des stations et leurs positions GPS pour une heure donnée

`place_place_dispo(heure)` : Fonction qui nous affiche un graphe décrivant les places disponibles et leurs positions gps

`histo_velo_dispo(heure,taille)` : Fonction qui nous affiche un histogramme décrivant le nombre de vélos disponibles pour chaque station à une heure donnée, `taille` est le nombre de station qu'on veut étudier

`emprunt_retour(station)` : Fonction qui nous affiche un graphe décrivant la disponibilité des vélos pour une station donnée .

`distance(a,b)` : Fonction qui calcule la distance entre un point `a` et un point `b`, qui sont décrits par leurs positions gps (latitude , longitude)

`prediction(action,heure,position)` : Fonction qui nous propose une station adéquate, suivant l'action qu'on veut faire, l'heure et notre position