

Le projet a été réalisé par :

Calligaro Nicolas et Calvet Rodolphe.

### 1. Introduction

L'inspiration est venue du site kaggle où ce cas a été trouvé :

<https://www.kaggle.com/edenau/london-bike-sharing-system-data>

L'auteur a réalisé un site web réalisant une visualisation des données :

<https://towardsdatascience.com/visualizing-bike-mobility-in-london-using-interactive-maps-for-absolute-beginners-3b9f55ccb59>

Les données sources se trouvent ici :

<https://cycling.data.tfl.gov.uk/>

Données qui seront alors nos sources.

Les jupyter ont également été commenté comme le rapport

Le code est disponible et commenté dans le fichier .py

### 2. Description des données.

Les données sont fournies par la ville de Londres. Elles représentent un système de vélo partagé au sein de la ville (style Velib).

On y trouve beaucoup de données diverses sur le système de vélo partagé.

Nous utiliserons les données sur les trajets sur une période d'une semaine (un fichier excel)

Un fichier excel représente une semaine de trajet. Nous choisissons la semaine du 28 Mars au 3 Avril 2018. Car c'est une semaine sans vacances ni jour particulier.

Bien que non réalisé, nous pourrions étendre toutes nos études à l'année entière voir plus. Le seul intérêt aurait été de vérifier si nos conclusions se confirment sur plusieurs période (année, cycle scolaire, vacance).

Nous allons rajouter une donnée : les coordonnées GPS des stations d'arrêt en s'appuyant sur le nom des stations fournis dans le fichier excel.

### 3. Analyse statistique

Nous allons réaliser une première analyse statistique des données et voici les conclusions qui en ont été tirés :

Notre jeu de donnée compte 10.000 Trajets.

La ville compte 785 Stations.

Elles ont en moyenne 100 trajets sur les 10.000 hebdomadaire

Nous avons identifié 42 Stations sur utilisé (+264 fois).

La ville compte 10.000 vélos.

Utilisé en moyenne 10 fois par jour

152 vélos ont été utilisés (+22 fois).

La durée moyenne d'un trajet est de 15 min.

Nous remarquons un étalement bien trop important des trajets : 4301 trajet font plus de 40 min (jusqu'à 4 jours) c'est trajet ne seront écarté de l'étude

Il y a également des trajets qui partent et reviennent sur la même station. Ils seront également écartés.

Suite à une analyse des séries temporelles nous voyons 3 créneaux :

En semaine autour de 8h et 17h sur une période de 2h

Le Week End sur une période plus large 10-17h

#### 4. Choix des sujets à étudier :

Nous nous sommes répartis 1 sujet chacun :

Etude de l'utilisation des stations (les plus utilisés) – Calligaro Nicolas –

Etude de l'utilisation des Vélos (les plus utilisés) – Calvet Rodolphe –

##### 1) Les stations

L'étude s'est concentrée sur l'utilisation de graph et leur connectivité pour mettre en avant l'utilisation de certaine station.

Voir le jupyter pour les graphiques et conclusion à chaque étape

En conclusion :

On voit deux comportement différent entre la semaine et le week end.

Malgré le ratio déséquilibré du nombre de trajet entre Week End et semaine, les stations les plus utilisés apparaissent dans les deux créneaux.

Semaine :

On peut voir que certaines stations desservent d'autre stations.

On ne voit pas de cycle.

On voit que certaines stations au départ le matin coïncident avec des stations du retour le soir.

Les stations sont groupées autour des gares de train et une zone importante au nord de Londres

Week End :

On voit un cycle apparaitre.

On peut voir également certaines stations agir comme des ilots (beaucoup de départ en partance de la station).

Ouverture :

Cette étude peut permettre une étude plus approfondit sur l'utilisation des stations et prévoir leur aménagement.

En rajoutant la circulation on peut prévoir l'aménagement des pistes cyclables en priorisant par rapport aux flux.

En augmentant la période d'étude on pourrait étudier l'évolution de l'utilisation en fonction des différentes période de l'année (vacances, été, hiver)

## 2) Les vélos

L'étude s'oriente autour de 2 questions :

- Question 1 : peut-on trouver une explication ou mieux comprendre le fait que, comme vu en partie 1, une fraction du parc est sur-utilisée au vu de la semaine entière.
- Question 2 : peut-on observer une localité dans cet usage plus intense : ces vélos sont-ils "prisonniers" temporairement d'une zone où la demande est plus/moins forte, y resteraient-ils aléatoirement un certain temps avant d'en sortir?

En conclusion à la question 1, nous dirons :

qu'après variation de tous les paramètres, nous n'avons jamais pu afficher une carte qui présenterait des parcours de vélos (parmi les plus utilisés comme parmi les moins utilisés) vraiment centrés sur une zone : en fait, le bruit est tel que tous les parcours vus sont globalement isotropes ! Les vélos ne demeurent jamais dans la même zone géographique..

que l'affichage ne permet finalement jamais de trouver une explication géographique évidente au sur ou sous-usage affiché : les vélos ne restent pas dans une zone particulière dans les deux cas. Aussi n'avons-nous pas jugé utile d'afficher une carte des 150 parcours d'outliers (utilisés plus de 22 fois), puisque les 10 premiers seuls témoignent d'une très forte non localisation, l'identification des numéros de vélos sur-utilisés le samedi et en semaine montre que ces deux ensembles sont différents, ce qui montre qu'un vélo très utilisé le samedi ne l'est pas en semaine et réciproquement. On peut donc

penser d'après cela que l'aléa prévaut sur l'idée que certains vélos ont des caractéristiques différentes, plus attrayantes.

Conclusion à la question 2 :

Malgré la possibilité de filtrer les données de notre dataframe selon de nombreux paramètres (semaine/weekend, tranche horaire, fréquence d'utilisation des vélos), la représentation sur carte n'a jamais fait apparaître de localité : les vélos ne restent jamais, par hasard, dans une même zone.

## 5. Technique utilisées

Le rajout des données GPS a été réalisé par la bibliothèque Geocoder et l'api arcgis.

L'affichage graphique a été réalisé avec la bibliothèque Ipyleefleat

Les fonctions Ant\_Path et Polyline ont été utilisés. La couleur a été utilisée pour faire apparaître l'élément important (fréquence ou appartenance)

L'étude des graphs a été réalisé avec la bibliothèque NetworkX.

La connectivité des nœuds (nombre de lien sortant) a été utilisé pour différencier les sommets entre eux.

Ainsi on a fait apparaître les notions :

Ilots : Sommet à l'origine de beaucoup d'arrête.

Passerelle : Sommet qui relie deux ilots

Node : Nœuds rattaché aux ilots.