

Lynn Ghanem
20195110

LES HABITUDES D'UTILISATIONS DU RÉSEAU BIXI MONTRÉAL

Travail individuel

Dans le cadre du cours FAS 1001
« Introduction aux mégadonnées en sciences sociales »

Enseignant : Nadjim Fréchet

Département de science politique
Université de Montréal

19 avril 2024

Question de recherche

Les transports publics sont essentiels à l'économie et à la santé publique d'une métropole. À Montréal, les modes de transport alternatifs gagnent en popularité. Bixi Montréal est un organisme à but non lucratif créé par la ville en 2009 dans le but d'encourager l'utilisation du vélo par ses habitants. Le réseau Bixi comprend plus de 10000 vélos et de 900 stations, faisant de lui un des réseaux de vélopartage les plus larges du monde.

Le travail suivant est une recherche exploratoire visant à estimer les habitudes des utilisateurs de Bixi. Je réponds toutefois à certains questionnements précis tout au long du travail, notamment : les arrondissements dans lesquels les Bixi sont majoritairement utilisés; l'existence potentielle d'une corrélation entre l'âge moyen par arrondissement et l'utilisation des Bixi; et finalement, les heures de la journée durant lesquelles les Bixi sont majoritairement utilisées et pourquoi.

Données utilisées

Les données principales analysées dans le travail suivant ont été tirées du site web www.bixi.com/fr/donnees-ouvertes. La base de données en question contient des données relatives aux vélos Bixi depuis 2014. Pour ma recherche, j'ai choisi de prendre les données de l'année 2023, pour avoir la base complète la plus récente, étant donné que les bases sont régulièrement mises à jour. La base contient les champs suivants :

[1]	"x"	"STARTSTATIONNAME"	"STARTSTATIONARRONDISSEMENT"
[4]	"STARTSTATIONLATITUDE"	"STARTSTATIONLONGITUDE"	"ENDSTATIONNAME"
[7]	"ENDSTATIONARRONDISSEMENT"	"ENDSTATIONLATITUDE"	"ENDSTATIONLONGITUDE"
[10]	"STARTTIMEMS"	"ENDTIMEMS"	

Aux fins de mon travail, les variables que j'ai utilisées sont les suivantes :

- STARTSTATIONNAME (nom de la station de début du trajet)
- STARTSTATIONARRONDISSEMENT (arrondissement dans lequel la station de début se situe)
- ENDSTATIONNAME (nom de la station de la fin du trajet)
- ENDSTATIONARRONDISSEMENT (arrondissement dans lequel la station de fin de trajet se situe)
- STARTIMEMS (heure de début du trajet)
- ENDTIMEMS (heure de fin du trajet)

Les autres données utilisées concernent l'âge moyen par arrondissement de la ville de Montréal. Elles ont été tirées d'un diagramme à barres effectué pour le Service du développement économique de la ville, dont les données sont tirées de Statistiques Canada pour l'année 2016. Il ne s'agit pas d'À partir de ce diagramme, j'ai manuellement créé une base de données pour pouvoir l'intégrer à mes données Bixi.

Méthodologie

Mis à part le nettoyage et la manipulation de données dans R, détaillés dans la section « Analyse des données » de ce travail, j'ai choisi de présenter mon analyse et mes résultats par l'intermédiaire de plusieurs diagrammes à barres. Je travaille principalement avec des fréquences et des catégories (utilisations de Bixi par arrondissement, par heure, âge par arrondissement, etc.) et j'ai donc trouvé que les diagrammes à barres seraient la meilleure façon de visualiser mes informations. Pour citer Midway (2020), « A good use of a bar plot might be to show counts of something. »

Analyse de données

Mon analyse des données commence par le chargement de celles-ci dans un objet R qui s'intitule « databixi. » Ensuite, j'ai créé un objet secondaire, « bixiarr » contenant les données regroupées selon l'arrondissement dans lequel le trajet commence pour les utilisateurs de Bixi. Je fais la somme des déblocages par arrondissement avant de les transformer en pourcentage à l'aide des fonctions « mutate » et « count » de tidyverse et base R, respectivement. Après avoir organisé mes données de cette manière, j'arrive à la visualisation de l'information recherchée, le nombre de Bixi débloquentes par arrondissement, par l'intermédiaire d'un diagramme à barres. La fonction « ggplot » du tidyverse permet de créer le diagramme en question, nommé « plotarr » par souci de description et présenté dans la section « résultats » de ce travail.

```
2
3 library(readr)
4 library(tidyverse)
5
6 databixi <- read.csv("databixi.csv")
7
8
9 bixiarr <- databixi %>%
0   na.omit() %>%
1   group_by(STARTSTATIONARRONDISSEMENT) %>% summarize(count = n()) %>%
2   mutate(pourcentage = (count / sum(count)) * 100)
3
4
5 plotarr <- ggplot(bixiarr, aes(x = reorder(STARTSTATIONARRONDISSEMENT, pourcentage), y = pourcentage)) +
6   geom_col(fill = "pink2") + theme_minimal() +
7   coord_flip() +
8   labs(title = "Déblocage de Bixi par arrondissement, en % du total", y = "Pourcentage",
9        x = "Arrondissement") +
0   scale_y_continuous(breaks = seq(0, 40, by = 5)) +
1   theme(panel.grid.minor.x = element_blank(),
2         panel.grid.major.y = element_blank(),
3         axis.title.y = element_blank(),
4         axis.title.x = element_text(vjust = -1, size = 10),
5         axis.text = element_text(face = "bold", size = 9),
6         plot.title = element_text(hjust = 0, vjust = 3, size = 13, color = "pink4"))
7
8
9 plotarr
0
```

La deuxième étape était de déterminer si l'utilisation des Bixi est liée à l'âge moyen par arrondissement. Pour cela, le code suivant est utilisé, similairement au précédent, pour créer un diagramme. La deuxième ligne du code ci-dessous montre la manière que j'ai utilisé pour créer

la colonne « âge moyen » et l'ajouter à l'objet « datage. » J'ai créé la colonne manuellement en juxtaposant les âges aux arrondissements (voir section « Données utilisées » de ce travail). J'ai ensuite utilisé ggplot pour construire un autre diagramme, cette fois représentant l'âge moyen par arrondissement.

```

1 datage <- bixiarr %>% select(STARTSTATIONARRONDISSEMENT)
2
3 datage$agemoyen <- c(41.6, 43.9, NA, 38.7, 42.2, 41.8, NA, 37.4, 38.9, NA, NA, 41.2, 41, 39.6, 41.2,
4                      38.5, 42.2, 40.4, 39, 41.1, NA, 41.4, 40, 38.3, 45.1 )
5
6 datage <- datage %>%
7   na.omit()
8
9 plotage <- ggplot(datage, aes(x = reorder(STARTSTATIONARRONDISSEMENT, -agemoyen), y = agemoyen)) +
10   geom_col(fill = 'darkgreen') + theme_minimal() + coord_flip() + |
11   labs(title = "Âge moyen par arrondissement", y = "Âge") +
12   scale_y_continuous(breaks = seq(0, 45, by = 5)) +
13   theme(panel.grid.minor.x = element_blank(),
14         panel.grid.major.y = element_blank(),
15         axis.title.y = element_blank(),
16         axis.text.y = element_text(face = "bold", size = 9, hjust = 1),
17         axis.text.x = element_text(face = 'bold', size = 9, vjust = 2),
18         plot.title = element_text(hjust = 0, vjust = 0.5, size = 13, face = "bold"))
19
20 plotage

```

Étant donnés les résultats des analyses ci-dessous, détaillées dans la section suivante de ce travail, j'ai choisi de déterminer le nombre de stations Bixi par arrondissement. L'approche est la même, mis à part le fait d'avoir eu à omettre les données répétées dans la base à l'aide de la fonction « distinct » :

```

3 uniquedat <- distinct(databixi, STARTSTATIONNAME, .keep_all = T)
4
5 stationpararr <- uniquedat %>% select(STARTSTATIONNAME, STARTSTATIONARRONDISSEMENT) %>%
6   group_by(STARTSTATIONARRONDISSEMENT) %>% na.omit() %>% summarize(numstations = n())
7
8 plotstations <- ggplot(stationpararr, aes(x = reorder(STARTSTATIONARRONDISSEMENT, numstations),
9                                                  y = numstations)) +
10   geom_col(fill = 'blue') + theme_minimal() + coord_flip() +
11   scale_y_continuous(breaks = seq(0, 200, by = 10)) +
12   labs(title = "Nombre de stations bixi par arrondissement") +
13   theme(panel.grid.minor.x = element_blank(),
14         panel.grid.major.y = element_blank(),
15         axis.title.y = element_blank(),
16         axis.title.x = element_blank(),
17         axis.text.y = element_text(face = "bold", size = 9, hjust = 1),
18         axis.text.x = element_text(face = 'bold', size = 9, vjust = 2),
19         plot.title = element_text(hjust = 0, vjust = 0.5, size = 13, face = "bold"))
20
21 plotstations
22

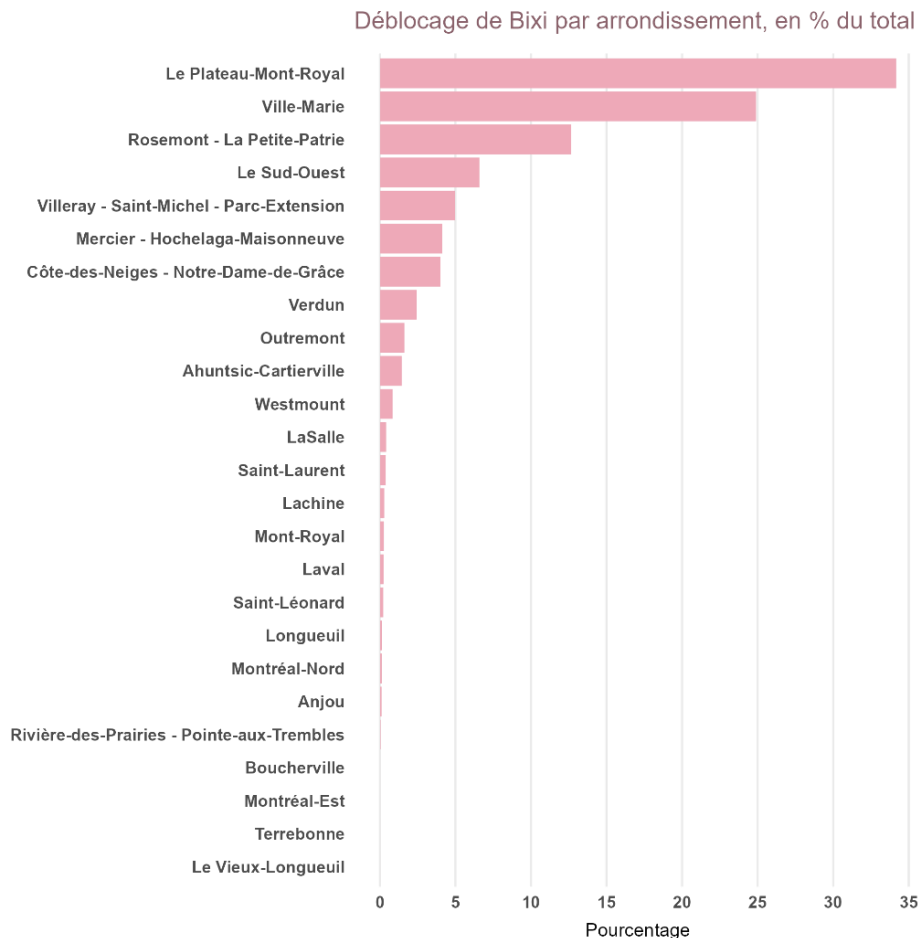
```

L'étape suivante consistait à déterminer l'utilisation de Bixi par heures de la journée. Pour cela, j'ai adopté une approche similaire aux deux autres : un regroupement par heure, suivi d'un sommaire des valeurs et d'une visualisation :

```
bixistarttime <- databixi %>%  
  na.omit() %>%  
  group_by(hour) %>% summarize(count = n())  
  
plottime <- ggplot(bixistarttime, aes(x = hour, y = count / 1000)) + geom_col(fill = "darkgreen") +  
  theme_minimal() +  
  labs(title = "Déblocage de bixi par heure de la journée", y = "Bixis débloquées (en milliers)",  
       x = "Heure") +  
  scale_x_continuous(breaks = seq(0, 23, by = 1)) +  
  theme(panel.grid.minor = element_blank(),  
        axis.title.x = element_text(vjust = -1, size = 10),  
        axis.text = element_text(face = "bold", size = 9),  
        axis.title.y = element_text(vjust = 3.5, size = 10),  
        plot.title = element_text(hjust = 0.5, vjust = 3, size = 13, color = "grey3"))  
  
plottime
```

Résultats, discussion et conclusions

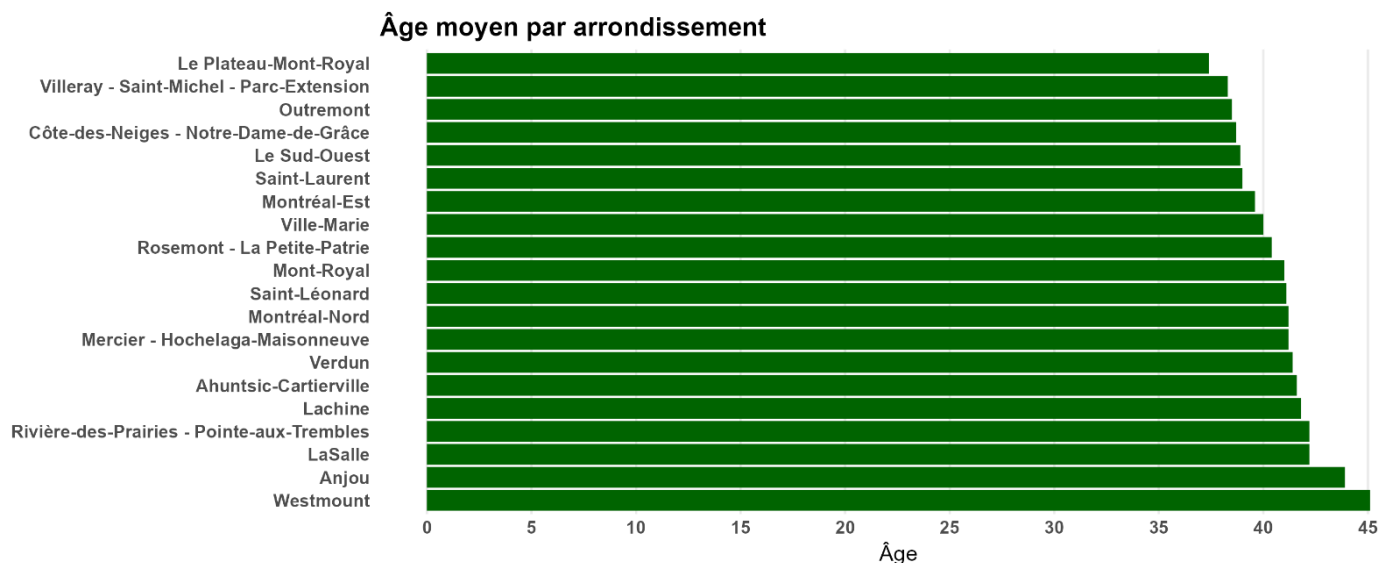
L'utilisation de Bixi par arrondissement



Le graphique ci-dessus montre clairement que le Plateau Mont-Royal compte pour 35% de tous les déblocages de Bixi effectués en 2023. Ce diagramme pose plusieurs questions concernant l'explication de ce chiffre.

L'âge moyen par arrondissement

Le 35% pourrait-il s'expliquer pas l'âge moyen des habitants du Plateau? Le vélo serait-il une activité de jeunes?

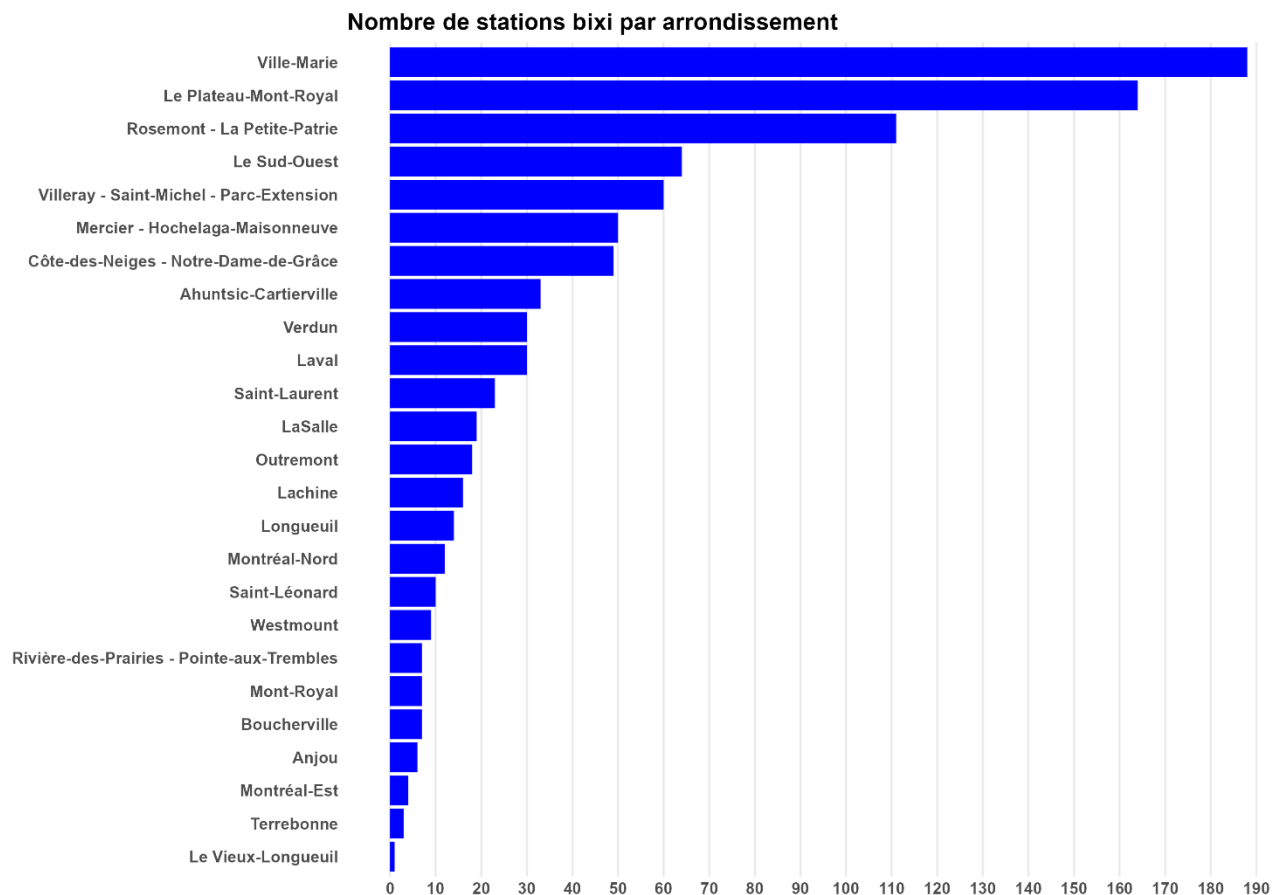


Le Plateau-Mont-Royal est bel et bien l'arrondissement le plus jeune de la ville. La différence entre celui-ci et les autres ne semble toutefois pas être assez drastique pour justifier la majorité d'utilisation de Bixi se faisant sur le plateau.

Le nombre de stations Bixi par arrondissement

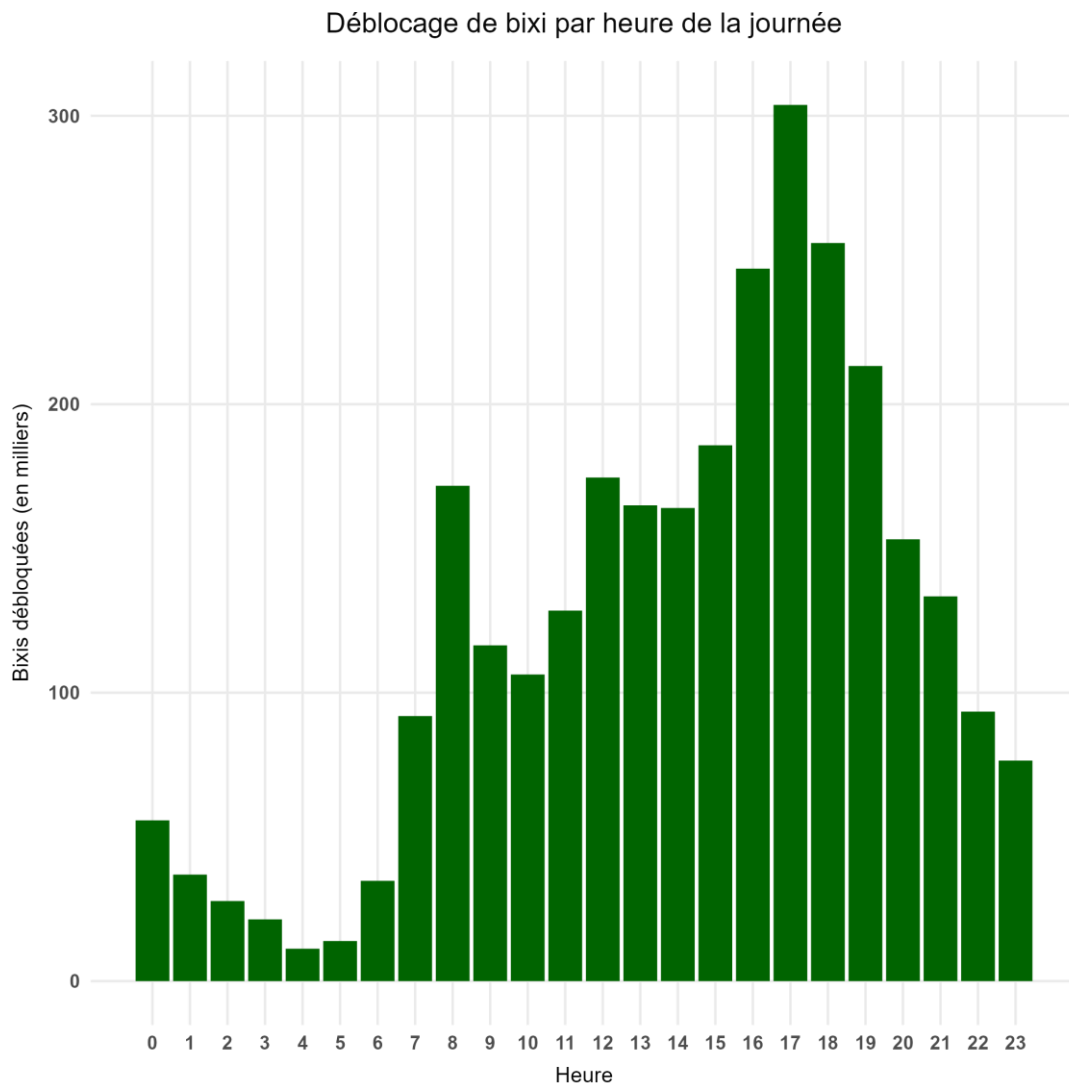
La disponibilité jouerait peut-être un rôle dans le nombre d'utilisateurs Bixi du Plateau-Mont-Royal. L'abondance de stations Bixi et de vélos contribuerait peut-être à motiver les gens à se servir davantage de ce mode de transport pour circuler. Le graphe ci-dessous nous montre

toutefois que l'arrondissement dans lequel le plus de stations sont installées est Ville-Marie, qui n'est que deuxième en termes d'utilisations.



L'utilisation de Bixi selon les heures de la journée

Le diagramme qui suit illustre l'utilisation des vélos Bixi par heure. On voit qu'il y a un pic à 17h. Plusieurs explications s'imposent ici. La première (et selon moi la plus plausible) serait que les individus prennent le vélo pour rentrer chez eux à la fin de leur journée de travail. Cela expliquerait également le petit pic vers 8h, alors que les gens se dirigent à leur emploi. On peut aussi supposer que ces utilisateurs préfèrent pédaler après leurs emplois plutôt qu'avant, pour éviter d'y arriver en sueur.



Pistes de recherches futures

Les résultats de ce travail, quoi que peu concluant, frayent la voie vers de futures recherches plus avancées. Je me suis moi-même posée plusieurs questions durant ce travail auxquelles je n'ai pas pu répondre par contrainte de temps, d'espace et de disponibilité de données. On pourrait par exemple se demander s'il existe un lien entre le revenu et l'utilisation des Bixi. Le Plateau-Mont-Royal serait-il un arrondissement riche? On pourrait également se demander si l'idéologie

politique joue un rôle dans cet enjeu. Le vélo est un mode de transport écologique. Serait-il juste d'assumer que les arrondissements qui votent plus à gauche font plus de vélo? Finalement, quelles seraient les déplacements principaux effectués en vélo par ses utilisateurs? Aller au travail, faire du sport, rencontrer des amis?