

Projet BIXI - Partie 1: Analyse exploratoire

MATH60604 - Modélisation statistique

Abdoul Wassi Badirou, Alfred Assal, James Roy, Samuel Croteau

2023-09-22

Contents

1	Introduction	1
2	Exploration des données	2
2.1	Exploration univariée	2
2.1.1	Les variables qualitatives	2
2.1.2	Les variables quantitatives	3
2.2	Exploration multivariée	4
2.2.1	Exploration des relations avec la durée des trajets	4
2.2.2	Exploration des relations avec l'utilisation des bixis	8
2.2.3	Exploration des relations avec le revenu	12
3	Conclusion	13
4	Contribution des membres de l'équipe	13

1 Introduction

Dans le cadre du cours de modélisation statistique nous devons, à travers un projet, mettre en pratique les connaissances théoriques acquises. Le projet porte sur les activités de BIXI Montréal qui gère un système de vélopartage dans la grande région de Montréal. L'objectif du projet est d'identifier les facteurs qui influencent l'utilisation des Bixi, la durée des déplacements et les revenus qui en découlent. Le projet est divisé en 4 parties. La première porte sur l'analyse exploratoire et les autres consistent à explorer différents modèles à savoir: des modèles de régression linéaire, modèles linéaires généralisés et des modèles linéaires mixtes. Un jeu de données est mis à notre disposition. Il est important de rappeler que les données utilisées sont un sous-ensemble purement aléatoire des données complètes, coligées par l'entreprise.

La présente analyse porte sur la première partie du projet: faire une analyse exploratoire des données pour mieux comprendre le contexte et modèle d'affaires.

Pour y arriver, nous analysons d'abord les variables individuellement. Ensuite à travers une analyse mutivairée, nous explorons les relations entre les différentes variables par rapport aux variables cibles. Les faits saillants de nos résultats sont ensuite résumés dans la conclusion.

2 Exploration des données

La visualisation de la structure compacte des données ci-dessous, nous montre que le jeu de données est constitué de 10000 observations et 14 variables. Nous y retrouvons toutes les variables décrites dans l'énoncé du projet. Nous allons dans un premier temps explorer chacune des variables individuellement, puis analyser les relations entre elles et les variables cibles. Les variables cibles sont: la durée totale des déplacements partant de chaque station par jour (**dur**), le revenu total qui en découle (**rev**) et le nombre total de déplacements correspondant (**n_tot**).

```
## 'data.frame': 10000 obs. of 14 variables:
## $ station: int 10 10 10 10 10 10 10 10 10 10 ...
## $ mm : int 4 5 6 6 7 7 7 8 8 8 ...
## $ dd : int 24 24 3 18 21 22 28 18 20 22 ...
## $ wday : chr "Saturday" "Monday" "Thursday" "Friday" ...
## $ mem : int 1 1 1 0 1 1 0 0 1 1 ...
## $ holiday: int 0 1 0 0 0 0 0 0 0 0 ...
## $ dur : num 917.9 1123.4 115 56.7 285.7 ...
## $ avg : num 29.6 26.1 16.4 28.3 16.8 ...
## $ rev : num NA NA NA 11 NA ...
## $ n_AM : int 3 2 2 1 5 3 0 0 5 2 ...
## $ n_PM : int 28 39 3 1 10 15 3 2 6 10 ...
## $ n_tot : int 31 43 7 2 17 21 3 4 14 13 ...
## $ temp : num 14.1 14.4 17.5 21.2 20 19.6 19.2 25.4 25.9 28.2 ...
## $ rain : num 0 0 6.3 4.8 1.5 0 0 0 0 0 ...
```

2.1 Exploration univariée

2.1.1 Les variables qualitatives

Le tableau 1 présente un sommaire descriptif des variables qualitatives.

Station La variable station est une variable qualitative qui représente les différentes stations de départ des trajets en BIXI. Nous l'avons encodé sous forme de variable catégorielle nominale car il n'y a pas de relation d'ordre entre les différentes stations. Le jeu de données couvre 792 stations Bixi.

Sur la figure 1 nous voyons le nombre de stations en fonction du nombre de leurs observations dans les données. On constate qu'il n'y a pas le même nombre d'observations pour toutes les stations. Donc pour certaines stations, il y a des observations manquantes, soit pour une journée, soit pour un mois, soit pour une catégorie de membre. Cela est due au fait que le jeu de donnée que nous utilisons est un sous-ensemble aléatoire des données réelles.

Mois Il s'agit d'une variable catégorielle ordonnée car les mois arrivent dans un ordre donné. La saison Bixi s'étend d'avril (4) à novembre (11) comme on peut le voir à la figure 1. En dehors des mois d'avril et novembre, nous avons approximativement le même nombre d'observations par mois. Il y a environ 2 fois moins d'observations pour les mois d'avril et novembre. C'est cohérent avec le fait que la saison s'étend de la mi-avril à la mi-novembre. De ce point de vue, notre sous-ensemble de jeu de données est représentatif de la saison réelle.

Table 1: Sommaire descriptives des variables qualitatives

	station	mm	dd	wday	mem	holiday
	147 : 29	8 :1513	10 : 374	Monday :1443	0:4726	0:9755
	598 : 28	10 :1499	24 : 373	Tuesday :1410	1:5274	1: 245
	886 : 27	6 :1438	12 : 368	Wednesday:1416		
	385 : 26	9 :1430	15 : 363	Thursday :1411		
	580 : 26	7 :1383	9 : 351	Friday :1424		
	653 : 26	5 :1343	13 : 351	Saturday :1445		
	(Other):9838	(Other):1394	(Other):7820	Sunday :1451		

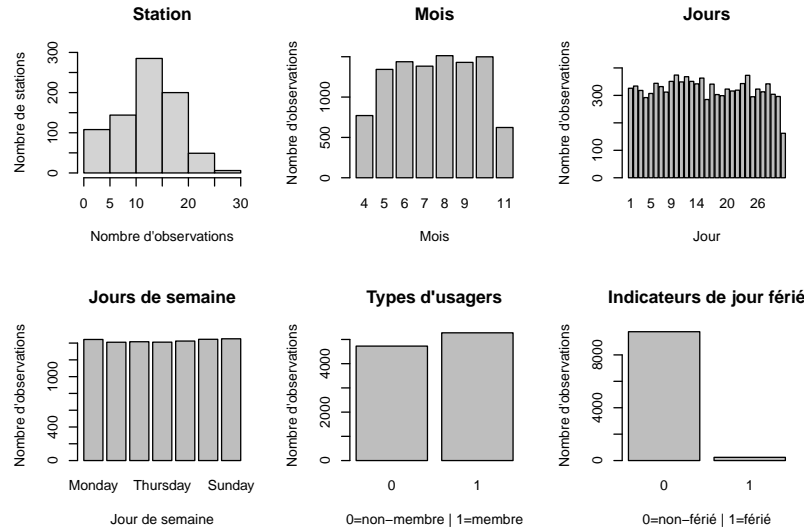


Figure 1: Diagrammes en bâton des variables qualitatives

Jour Il s'agit d'une variable catégorielle ordonnée. Les jours s'étendent du 1 au 31 en fonction des mois. Dans nos données, le premier jour de la saison est le 9 avril et le dernier jour de la saison est le 15 novembre. Or sur le site de Bixi, la saison commence le 15 avril. Il nous manque donc dans nos données la première semaine réelle de la saison.

Il y a environ 2 fois moins d'observations pour la journée du 31 par rapport aux autres mois. La distribution des autres jours semble approximativement uniforme.

Jour de semaine Il s'agit d'une variable catégorielle ordonnée. Mais le choix du premier jour est relatif. La figure 1 nous montre que la distribution des jours de semaine est approximativement uniforme car nous avons environ le même nombre d'observations pour les différents jours de la semaine.

Membres Il s'agit d'une variable catégorielle nominale. La figure 1 nous montre qu'il y a légèrement un peu plus d'observations pour la catégorie membre que de non-membre. Mais cela ne nous permet pas de savoir si le portefeuille de clientèle de Bixi est composé plus de membre que de non-membre.

Jour férié Les observations sur des jours fériés représentent environ 2%. Nous devons plustard y accorder une attention particulière dans la modélisation afin d'éviter que cette catégorie soit négligée, car le revenu engrené ces jours ne sont pas négligeable.

2.1.2 Les variables quantitatives

« « « < HEAD

Le tableau 2 présente un sommaire descriptif des variables quantitatives. En moyenne la durée totale des trajets qui partent d'une station par jour est d'environ 5h. Par ailleurs, il y a trois journées qui totalisent des déplacements

Table 2: Sommaire descriptives des variables quantitatives

dur	avg	rev	n_AM	n_PM	n_tot	temp	rain
Min. : 1.117	Min. : 1.117	Min. : 1.418	Min. : 0.000	Min. : 0.00	Min. : 1.00	Min. : 0.80	Min. : 0.000
1st Qu.: 66.092	1st Qu.:11.285	1st Qu.: 8.883	1st Qu.: 1.000	1st Qu.: 2.00	1st Qu.: 4.00	1st Qu.:11.90	1st Qu.: 0.000
Median : 169.875	Median :13.839	Median : 18.111	Median : 2.000	Median : 6.00	Median : 11.00	Median :17.90	Median : 0.000
Mean : 267.795	Mean :15.404	Mean : 27.337	Mean : 4.309	Mean : 11.42	Mean : 19.72	Mean :16.84	Mean : 1.934
3rd Qu.: 375.179	3rd Qu.:17.717	3rd Qu.: 34.491	3rd Qu.: 6.000	3rd Qu.: 16.00	3rd Qu.: 27.00	3rd Qu.:21.20	3rd Qu.: 0.500
Max. :3756.900	Max. :58.817	Max. :748.535	Max. :63.000	Max. :165.00	Max. :310.00	Max. :28.20	Max. :37.000
		NA's :5274					

supérieurs à 50h. Deux d’entre elles sont associées à la station 256, les jeudis 2 et 9 septembre. L’autre est associée à la station 470, le samedi 21 août. Ces durées extrêmes sont tout de même surprenantes. Celles associées à la station 256 paraissent moins aberrantes tout de même, puisque le nombre de déplacement de ces journées est aussi extrême tant en AM qu’en PM. Cela suggère qu’il y a probablement eu ces deux jeudis consécutifs des événements dans le secteur qui pourraient expliquer ces utilisations extrêmes. De plus, ils sont associés à des usagers membres. Cependant, la durée totale associée à la station 470 est la plus extrême et la plus surprenante. En effet, le nombre de déplacement associé n’est pas le plus extrême. En plus, c’est associé à des usagers non-membres qui ont un profil de consommation de courte durée. Ce déplacement peut s’expliquer si un utilisateur non-membre prend un bixi et le conserve chez lui plutôt que de le déposer à une station.

Toutes les variables quantitatives ont des valeurs extrêmes sauf la température moyenne quotidienne, comme on peut le voir sur la figure 3. Elle présente également une distribution quasiment symétrique (voir figure 2). Les autres ont une distribution qui s’étale à droite. Une transformation sera nécessaire éventuellement dans la modélisation pour les rendre symétrique.

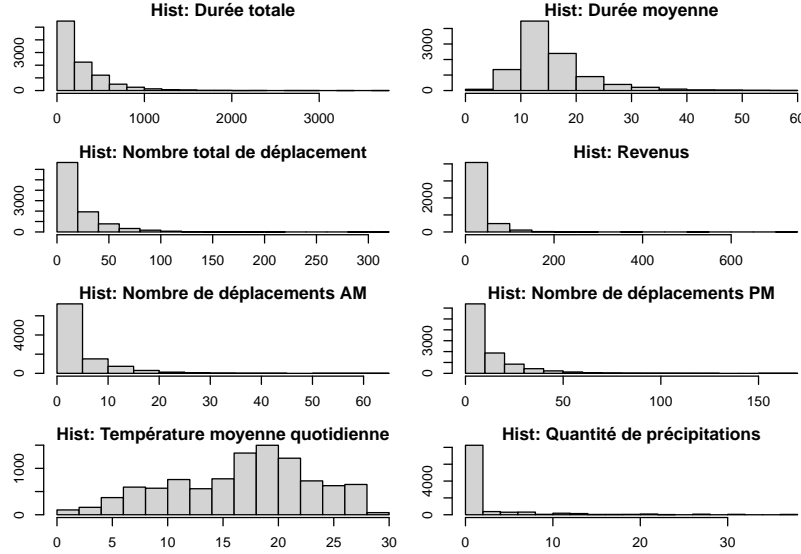


Figure 2: Histogramme des variables quantitatives

2.2 Exploration multivariée

Avant de débiter l’exploration multivariée des variables cibles, il est important d’avoir une vue d’ensemble des relations entre les variables. Pour ce faire, une matrice de corrélation comme illustrer dans la figure 4 est une représentation simple qui permet de visualiser les tendances entre chaque variable. Par exemple, il est possible de voir que la corrélation entre `dure` et `rev` est de 0.99. De plus, les relations de `dure` et `rev` vont se comporter de façon quasi identique avec les autres variables. Ainsi, il est possible d’en représenter qu’une seule des deux lors de l’analyse multivariée. Cette problématique sera abordée plus en détail dans la section 2.2.3.

En ce qui concerne la matrice de corrélation, deux variables ont été enlevées à des fins pratiques. La première est `wday` (les journées de la semaine) et la deuxième est `mem` (membre de bixi ou non). Par contre, il est logique d’affirmer par déduction que ces deux variables restent tout de même intéressantes à croiser avec nos variables d’intérêt.

2.2.1 Exploration des relations avec la durée des trajets

En croisant la variable `dur` avec la variable `wday` et `mem`, nous pouvons obtenir le figure 5. D’abord, on remarque que les membres utilisent beaucoup plus les Bixi (au niveau de la durée) que les non-membres, et ce, peu importe le jour de la semaine. Les types d’utilisation semblent aussi changer, par exemple, les non-membres utilisent relativement plus les Bixi les fin de semaine que les membres. En effet, les membres semblent avoir une utilisation croissante la semaine et décroissante la fin de semaine. Naturellement, on pourrait penser que c’est parce qu’ils utilisent le Bixi comme un moyen de transport régulier pour aller au travail ou à l’école par exemple.

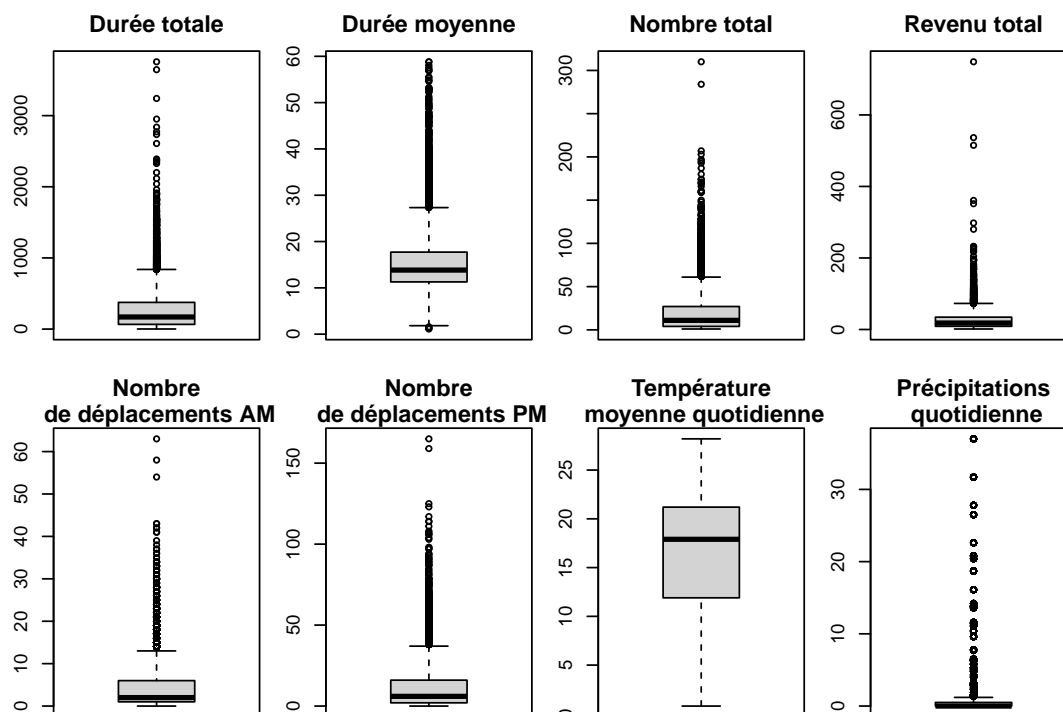


Figure 3: Boîte à moustache des variables quantitatives

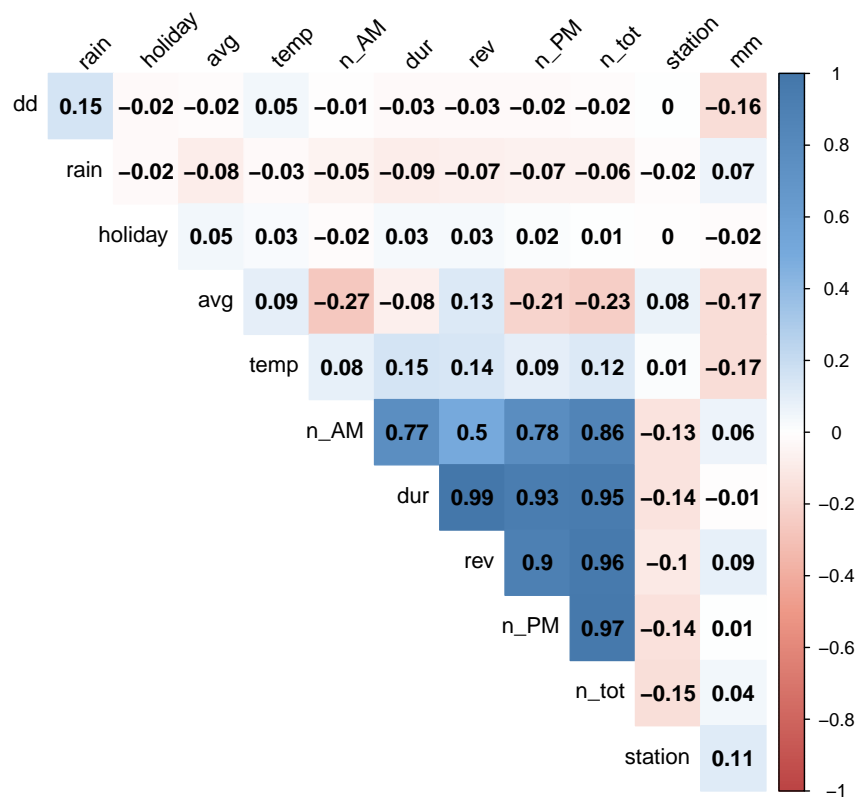


Figure 4: Matrice de corrélation

D'un point de vue d'affaires, il pourrait être intéressant pour l'entreprise Bixi de faire des accords avec les gros employeurs de Montréal pour augmenter ses clients et accélérer la transition vers des moyens de transport plus doux. Dans ce même ordre d'idée, la localisation des Bixi proche des milieux de travail et des milieux scolaires devient une avenue intéressante si l'entreprise a pour but d'augmenter son nombre d'abonnés mensuel.

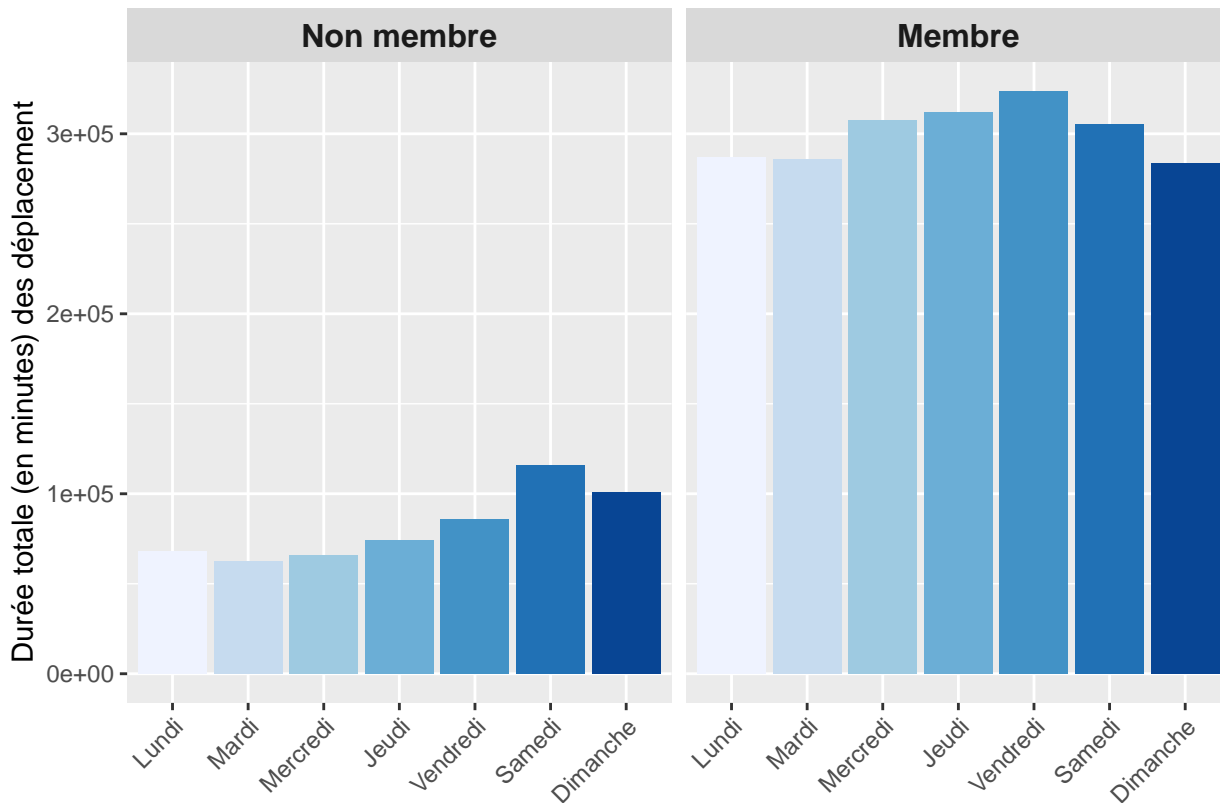


Figure 5: Durée totale des déplacement par jours selon les membres et les non membres

En croisant la variable **dur** avec la variable **mm** et **mem**, nous pouvons obtenir la figure 6. Plusieurs analyses sont pertinentes à partir de ce graphique. D'abord, les membres utilisent beaucoup plus les Bixi (au niveau de la durée) que les non-membres, et ce, peu importe le mois de l'année. D'une façon similaire, l'utilisation semble atteindre son pic vers août pour les deux types de clients.

En croisant la variable **dur** avec la variable **temp** et **mem**, nous pouvons obtenir la figure 7. Naturellement, on voit une relation positive entre la température totale moyenne et la durée totale de déplacement. La relation semble toutefois ralentir voir stagner vers les 20°C. Cela peut venir du fait que l'utilisation des Bixi est plus élevée en été (donc les mois avec une température plus élevée). Ainsi, ce graphique donne globalement les mêmes informations que le précédent étant donné que la température est fortement corrélée avec les mois.

Au niveau des déplacements moyens (avec la variable **avg**), il semble que la durée moyenne des déplacements diminue après les 15°C.

Au niveau de la corrélation, elle est de 0.53 pour les non-membres et de 0.67 pour les membres pour le déplacement moyen total et de 0.26 pour les non-membres et de 0.34 pour les membres pour le déplacement moyen.

En croisant la variable **dur** avec la variable **rain** et **mem**, nous pouvons obtenir la figure 8. Pour les non-membres l'effet de la pluie sur la durée totale moyenne n'est pas autant significatif que pour les membres. Cela est assez étonnant, car il semble que les membres devraient avoir un usage régulier et ainsi moins sensible aux facteurs exogènes comme la pluie.

Au niveau des déplacements moyens (avec la variable **avg**), l'effet est assez similaire entre les membres et les non-membres, les déplacements sont en moyennes plus courts avec la pluie.

Au niveau de la corrélation entre la pluie et la durée totale moyenne des déplacements, elle est de -0.36 pour les membres et de -0.24 les non-membres. Pour la durée moyenne, les corrélations sont de -0.38 pour les membres et de -0.18 pour les non-membres.

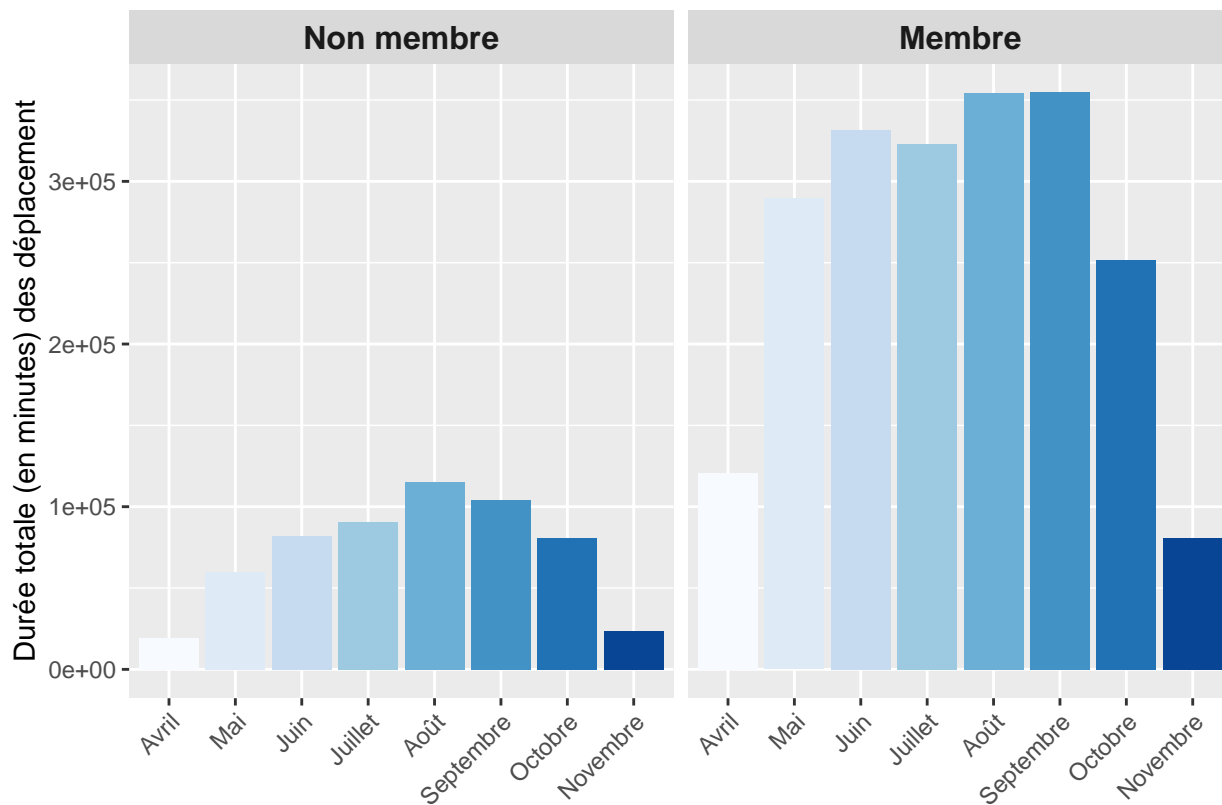


Figure 6: Durée totale des déplacement par jours selon les membres et les non membres

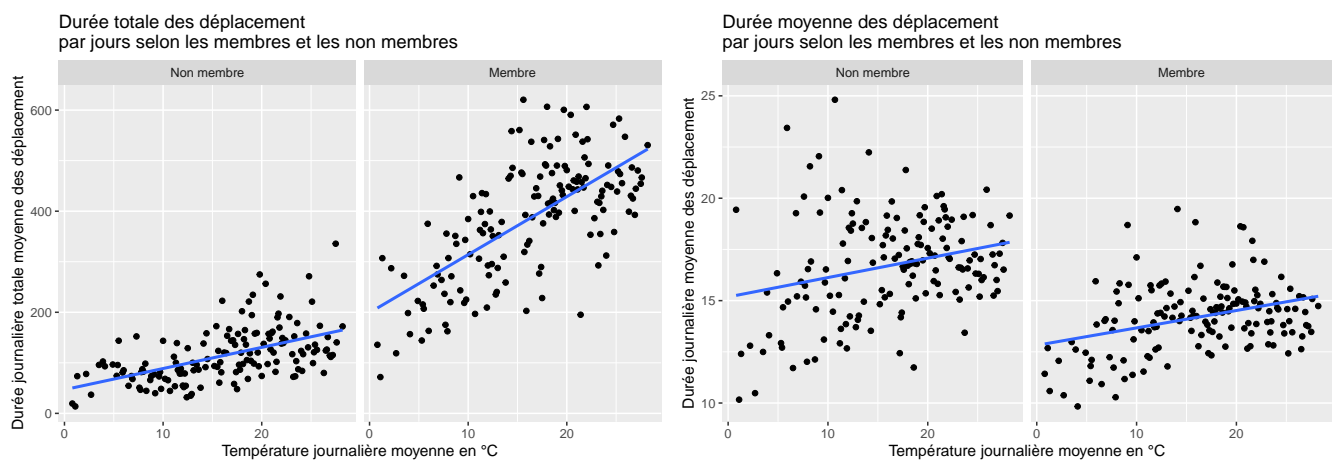


Figure 7:

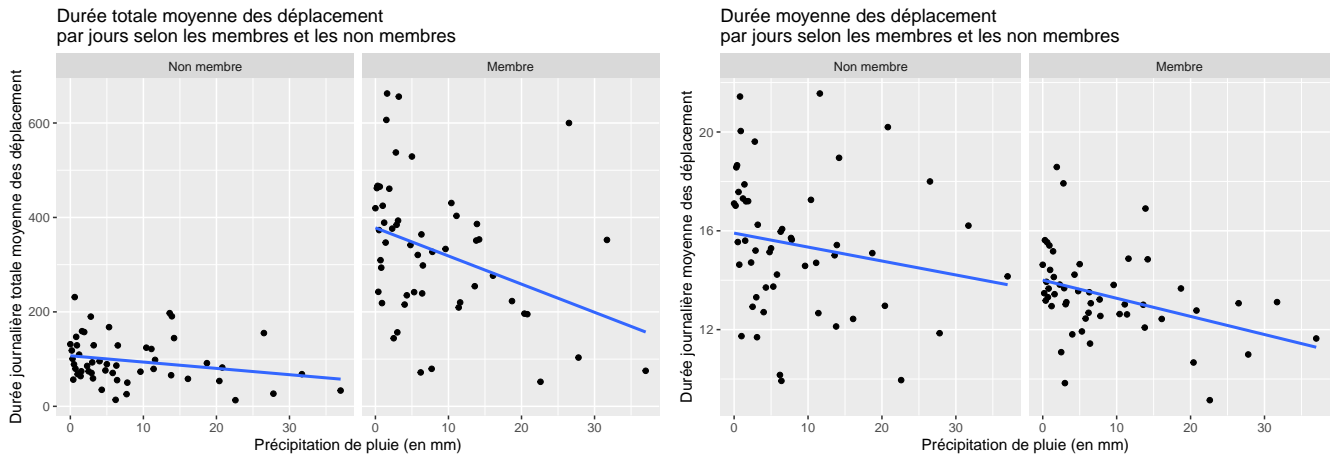


Figure 8:

En croisant la variable `dur` et `avg` avec `mem`, nous pouvons obtenir le tableau 3.

La durée moyenne des trajets est assez similaire entre les membres et les non-membres. En revanche, les membres ont une durée totale de trajets beaucoup plus haute que les non-membres, ce qui rejoint ce que nous avons observé précédemment. Il apparaît aussi que les membres sont responsables de la majeure partie des déplacements (en durée). Ainsi, selon la rentabilité du programme d'adhésion (donnée non-disponible), il pourrait être opportun d'attirer des nouveaux membres avec des promotions et des campagnes publicitaires.

En croisant la variable `dur` et `avg` avec `holiday`, nous pouvons obtenir le tableau 4.

La durée moyenne des trajets est légèrement supérieure pour les jours fériés. Aussi, la durée totale relative des trajets est supérieure pendant les jours fériés. Cela s'explique par le fait que les usagers ont plus de temps disponible pour utiliser les Bixi. Il serait aussi avantageux pour Bixi d'augmenter la disponibilité des vélos lors des jours fériés.

Selon le graphique 9, il est possible de modéliser les stations par leur durée totale de déplacements de la saison. On voit que la plupart des stations sont sous le seuil des 10 000 minutes.

En fonction de la durée totale durant la saison, le top 3 des stations sont :

- de la Commune / Place Jacques-Cartier (Vieux-Montréal)
- de la Commune / St-Sulpice (Vieux-Montréal)
- Métro Mont-Royal (Rivard / du Mont-Royal) (Plateau Mont-Royal)
- de Brébeuf / du Mont-Royal (Plateau Mont-Royal)
- BAnQ (Berri / de Maisonneuve) (Quartier des Spectacles)

Ainsi, on remarque que les deux stations les plus populaires en terme de durée sont situées dans le Vieux-Montréal. Il est possible d'attribuer cela aux touristes.

2.2.2 Exploration des relations avec l'utilisation des bixis

Dans la prochaine section, nous allons explorer les relations entre l'utilisation des Bixis, soit la variable `n_tot` qui représente le nombre total de déplacement, ainsi que `n_AM` (nombre de déplacement en avant-midi) et `n_PM` (nombre de déplacement en après-midi) dans certains cas et diverses variables de notre base de données.

La figure 10 montre la relation entre le nombre de déplacement total moyen par jour par station par rapport à la température pour les membres et les non-membres.

Table 3: Détail sur la durée selon les abonnements

Membre	Durée moyenne	Durée totale	Durée totale relative
Non-membre	16.73821	572884	121.2196
Membre	14.20860	2105066	399.1403

Table 4: Détail sur la durée selon les fériés

Jour	Durée moyenne	Durée totale	Durée totale relative
Non férié	15.35440	2597155.75	266.2384
Férié	17.38251	80794.15	329.7720

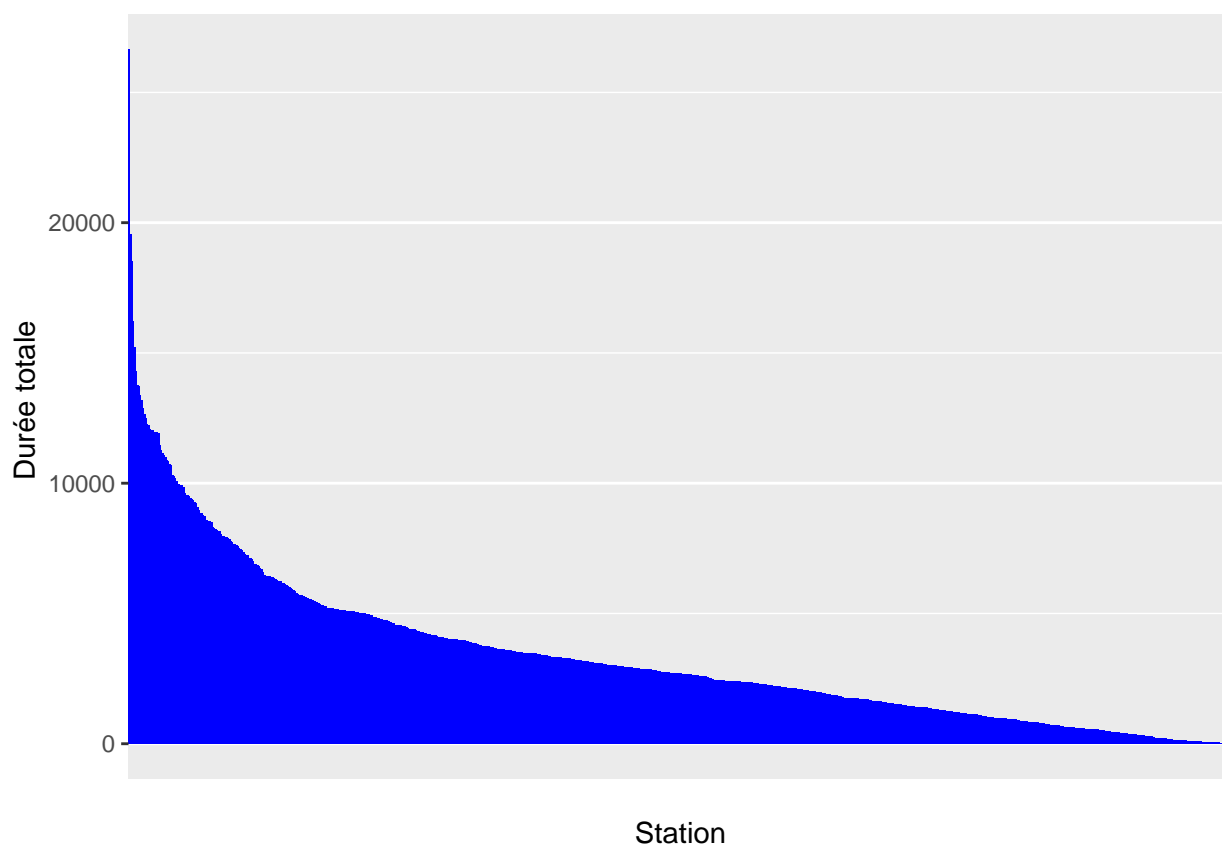


Figure 9: Répartition des stations selon la durée totale

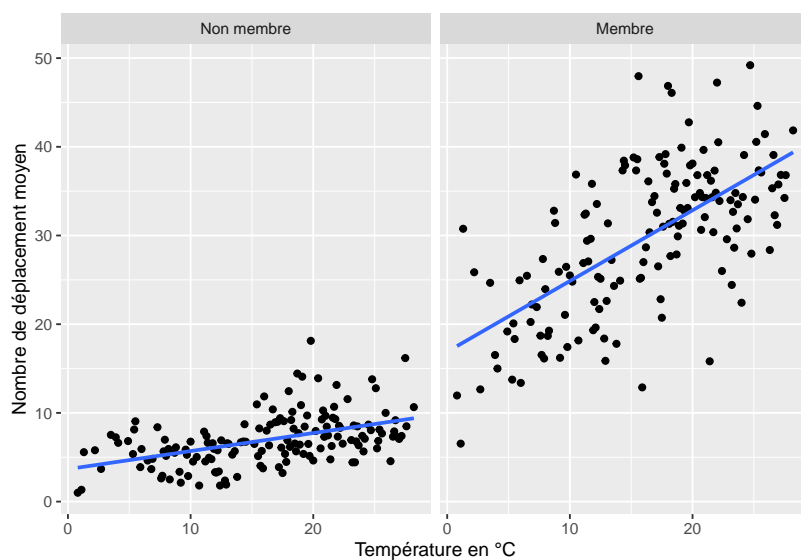


Figure 10: Nombre moyen de déplacements par rapport à la température

D'après ce graphique, on constate qu'il semble exister une relation linéaire entre le nombre de déplacement moyen effectué en bixi et la température pour les membres et non-membres. On observe que le nombre de déplacement moyen a tendance à augmenter avec la température. Une forte corrélation positive de 0.65 pour les membres et une corrélation moyenne de 0.48 pour les non-membres.

D'un point de vue d'affaires, si l'entreprise Bixi désirerait par exemple effectuer de la maintenance sur certains vélos et que ceci nécessiterait de récupérer une certaine quantité de bixi, il faudrait que l'entreprise entreprenne ces changements pendant les journées où il fait plus froid, ou durant les périodes plus froides soit durant les mois d'avril, mai, octobre ou novembre comme le montre le tableau qui suit afin d'éviter de causer des problèmes de manque de vélos pour les usagers.

D'un point de vue d'affaires, si l'entreprise Bixi désirerait par exemple effectuer une mise à jour logiciel ou effectuer de la maintenance sur certains vélos et que ceci nécessiterait de récupérer une certaine quantité de Bixi, il faudrait que l'entreprise entreprenne ces changements pendant les journées où il fait plus froid, ou durant les périodes plus froides soit durant les mois d'avril, mai, octobre ou novembre comme le montre le tableau afin d'éviter de causer des problèmes de manque de vélos pour les usagers.

Table 5: Température moyenne par mois

Mois	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre
Température (°C)	9.94	14.62	21.29	20.43	23.61	17.37	12.22	5.46

Par la suite, nous avons observé la relation entre le nombre de déplacements total moyen par jour par station d'après la quantité de précipitation tombée dans une journée. La figure 11 montre cette relation.

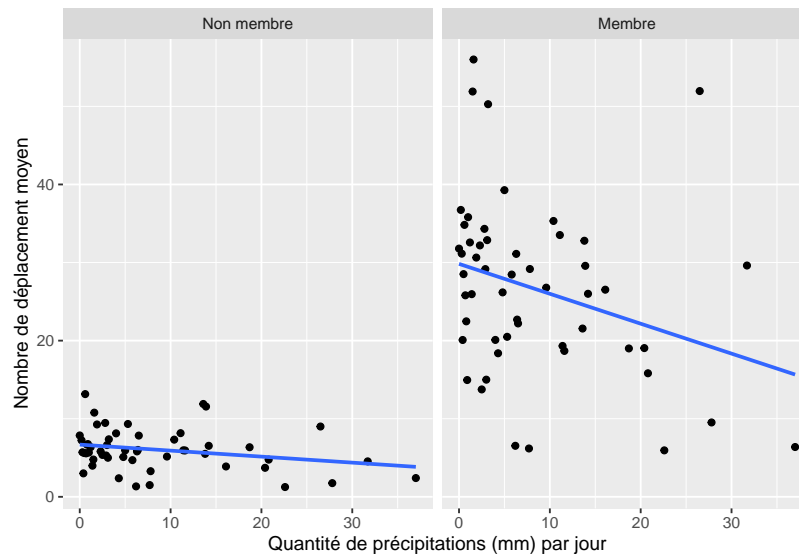


Figure 11: Nombre moyen de déplacements par rapport à la quantité de précipitations

D'après ce graphique, il semble y avoir une faible corrélation entre le nombre de déplacements et la quantité de précipitation dans une journée. En effet, la corrélation est de -0.26 pour les non-membres et -0.30 pour les membres.

La figure 12 montre la relation entre le nombre de déplacements moyen pour différents moments de la journée et le nombre total de déplacements d'après la journée par station.

Pour les membres et non-membres, on remarque qu'en moyenne, il y a plus de déplacements en après-midi qu'en avant-midi et durant la nuit. De façon générale, le nombre de déplacements en avant-midi s'approche de celui de la nuit. Pour les membres, le nombre de déplacements total est très semblable pour tous les jours de la semaine et est un peu plus faible pour les jours de la fin de semaine. Pour les non-membres, les jours de fin de semaine sont ceux avec le plus de déplacements comparativement aux jours de semaine.

Finalement, on peut tirer de cela que les membres utilisent plus les Bixis durant la semaine pour aller au travail par exemple et que plus de membres les utilisent l'après-midi. Pour les non-membres, on peut supposer qu'ils sont plus portés à utiliser les Bixis les fins de semaines pour des loisirs et moins pour le travail la semaine.

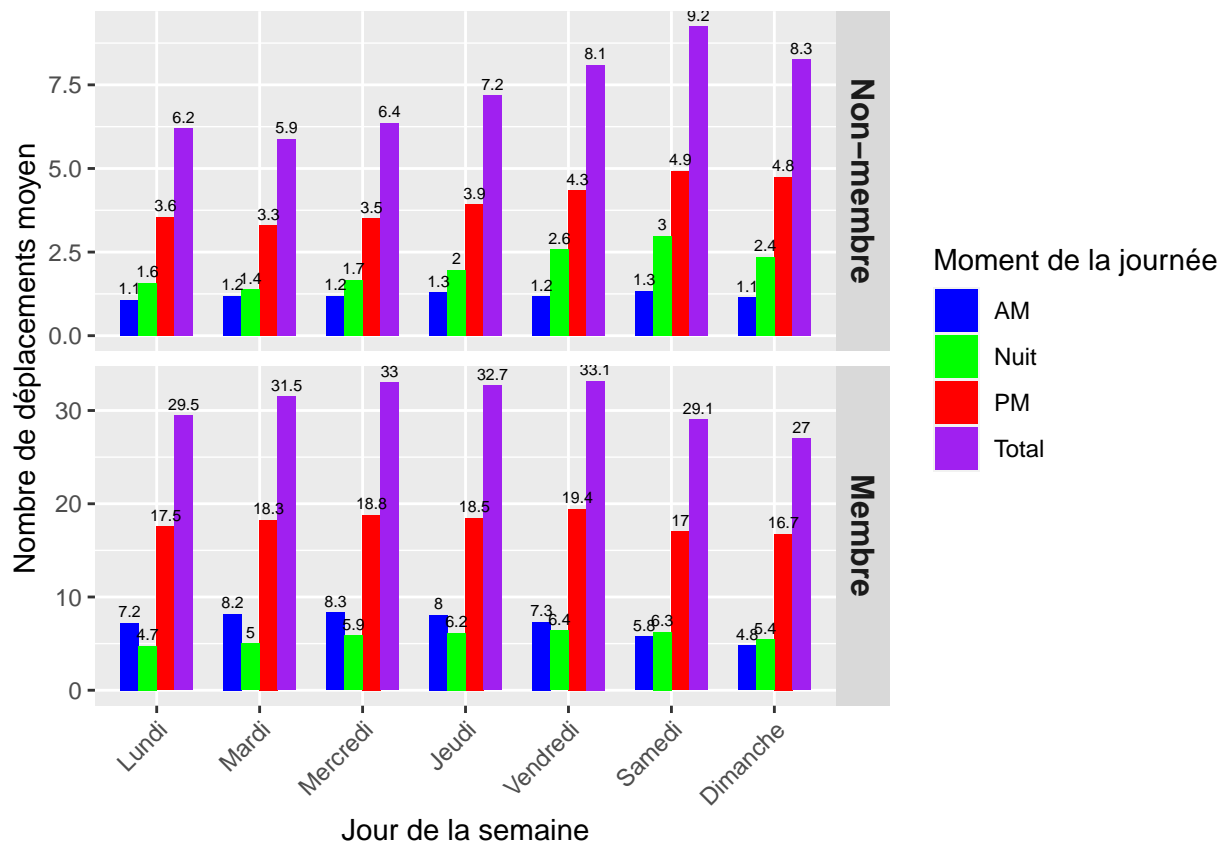


Figure 12: Nombre de déplacements moyen en bixi d'après le jour de la semaine

En moyenne, le nombre de déplacements total pour les membres est plus élevé à 30.8 par jour par station et à 7.32 par jour par station pour les non-membres.

D'un point de vue d'affaires, comme les membres utilisent plus les Bixis la semaine et qu'il y a plus de déplacements pour les membres que les non-membres. Il serait judicieux de toujours s'assurer d'une bonne disponibilité de bixi la semaine et ce surtout en après-midi.

Le tableau 6 montre la relation entre le nombre de déplacements par jour par station et si la journée est férié ou non.

Ce tableau montre qu'en moyenne il y a plus de déplacements durant les jours fériés.

La figure 13 représente la relation entre l'utilisation moyenne des bixis par jour par station et le mois de l'année.

On remarque que l'allure des deux graphiques est très semblable entre les membres et les non-membres. On parle ici d'une très forte corrélation de 0.86.

Par la suite, en comparant ces données avec les données du tableau 5, on obtient une très forte corrélation d'environ 0.85 pour les membres et une forte corrélation d'environ 0.59 pour les non-membres. Nous pouvons donc faire l'hypothèse que la relation entre le nombre de déplacements moyen par mois est dépendant de la température de ce mois. Les gens ont tendance à plus utiliser le bixi lors des mois plus chauds que les mois plus froids (avril, octobre et novembre).

On remarque que l'allure des deux graphiques est très semblable entre les membres et les non-membres. On parle ici d'une très forte corrélation de 0.86. On peut alors dire que le comportement des membres et des non-membres

Table 6: Détail sur le nombre de déplacement selon les fériés

Jour	Nb. déplacement moyen	Nb. déplacement total
Non férié	19.69103	192086
Férié	20.91020	5123

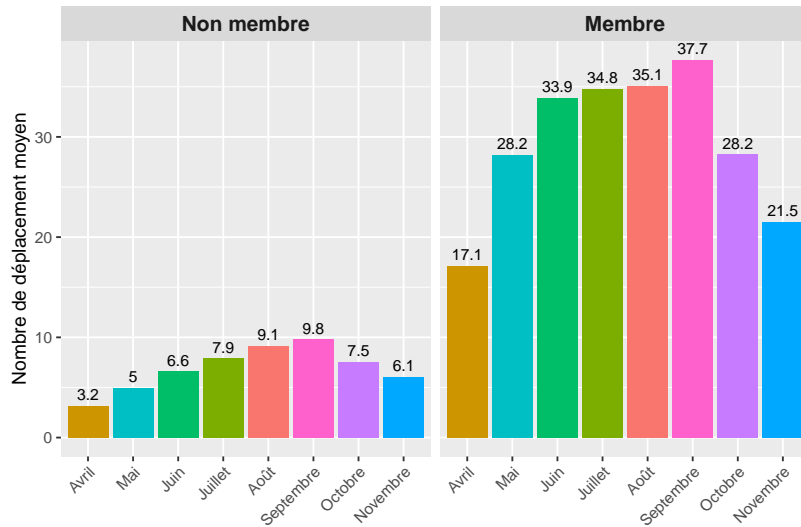


Figure 13: Nombre de déplacement moyen jours selon les membres et les non membres

est très semblable par rapport au mois.

Cependant, on remarque que dans les deux cas (membre et non-membre), le mois avec le plus de déplacement moyen est le mois de septembre qui pourtant n'est pas le mois le plus chaud. Notre hypothèse est qu'il s'agit de la période de la rentrée scolaire et du retour au travail pour plusieurs, alors plusieurs optent pour un transport plus écologique comme le bixi à ce moment de l'année avant de tomber dans la période plus froide de l'année avec les mois d'octobre et novembre.

D'un point de vue d'affaires, l'entreprise Bixi devrait prioriser et s'assurer de pouvoir mettre à disposition une grande quantité de bixi pour les périodes plus chaudes de l'année ainsi qu'en septembre.

Par la suite, nous avons identifié les stations qui généraient le plus de déplacements pour les membres et les non-membres. Les deux tableaux 7 et 8 présentent ces résultats.

On remarque que l'arrondissement avec le plus de stations dans le top 5 est le Plateau Mont-Royal et Le Vieux-Montréal. D'un point de vue d'affaires, il serait pertinent de prioriser ces arrondissements pour l'ajout de nouveaux bixis pour toujours en garder une grande quantité dans ces arrondissements. Puis, aussi d'effectuer plus de maintenance préventive des installations de ces arrondissements, car celles-ci génèrent beaucoup d'attraction et il ne faudrait pas qu'elles soient hors service.

2.2.3 Exploration des relations avec le revenu

Dans cette section, il sera principalement question de la variable revenu **rev**, puisque c'est la variable qui met en relation le revenu généré par les non-membres avec les autres variables. De ce fait, d'un point de vue entrepreneurial, c'est une variable d'intérêt capital, puisqu'elle participe au calcul de profit de BIXI. Cependant, puisque l'encodage des données du revenu n'est offert que pour les non-membres et que le revenu est calculé à partir de la durée du trajet, la corrélation entre ces deux dernières est relativement forte comme on peut le voir dans le tableau 4. Autrement dit, l'entreprise peut constater qu'une augmentation de la durée des trajets **dur** revient principalement à une augmentation du revenu **rev**. Conséquemment, visualiser la relation entre le revenu et les journées de la

Table 7: Détail sur les stations les plus utilisées pour les membres

Stations	Emplacement	Arrondissement	Nb. de déplacement total
256	Métro Mont-Royal (Rivard / du Mont-Royal)	Plateau Mont-Royal	1329
316	Prince-Arthur / du Parc	Plateau Mont-Royal	1189
169	De Bréboeuf / du Mont-Royal	Plateau Mont-Royal	1164
179	Messier / Du Mont-Royal	Plateau Mont-Royal	1136
167	Boyer / du Mont-Royal	Plateau Mont-Royal	1105

Table 8: Détail sur les stations les plus utilisées pour les non-membres

Stations	Emplacement	Arrondissement	Nb. de déplacement total
470	de la Commune / Place Jacques-Cartier	Le Vieux Montréal	652
336	Duluth / St-Laurent	Plateau Mont-Royal	373
256	Métro Mont-Royal (Rivard / du Mont-Royal)	Plateau Mont-Royal	320
762	St-Dominique / Napoléon	Plateau Mont-Royal	275
469	de la Commune / St-Sulpice	Le Vieux Montréal	270

semaine `wday`, le mois `mm`, la température `temp` et les précipitations `rain` reviennent à visualiser la même chose que dans la sections de l’exploration des relations avec la durée des trajets. Dès lors, il reste néanmoins intéressant d’observer comment le revenu se comporte lors des journées fériées `holiday`.

Dans le cas de la relation entre les journées fériées et le revenu, il est possible d’apercevoir qu’en moyenne les non-membres utilisent plus bixi lors des journées fériées, puisque le revenu moyen est plus élevé lors de ces journées (9). Cela étant dit, il y a un parallèle à faire avec la figure illustrée plus haut 5 où l’on peut voir que les non-membres ont plus tendance à se déplacer en bixi la fin de semaine. Par conséquent, on peut soutenir l’hypothèse que l’utilisation de bixi par les non-membres est liée à des fins plus récréatives.

Du point de vue de l’entreprise, la localisation des stations de bixi pour les non-membres devient alors cruciale. Autrement dit, placer des stations de bixi proches des lieux culturels et les lieux de divertissement pourraient ainsi favoriser le revenu de l’entreprise.

3 Conclusion

En conclusion, on peut constater que la durée d’utilisation est plus élevée pendant l’été pour les usagers de Bixi. Parallèlement, la durée relative d’utilisation est plus élevée la fin de semaine pour les non-membres que pour les membres. Il semblerait aussi que la pluie ait un effet défavorable sur la durée d’utilisation. Cela dit, les moments où l’utilisation des Bixis est la moins fréquente incluent les journées et les mois les plus froids (avril, octobre et novembre), ainsi que les fins de semaine et les matinées. Par ailleurs, l’inverse est aussi vrai, c’est-à-dire que les périodes de pointes de l’année sont entre juin et septembre, soit les mois les plus chauds. En complément, les journées de semaine et l’après-midi sont aussi les moments avec le plus d’achalandages. De plus, les secteurs avec les stations qui génèrent le plus de déplacements sont le Plateau Mont-Royal et le Vieux-Montréal. Ceux-ci ont également généré quelques valeurs extrêmes, mais il nous est impossible de dire pour le moment si ce n’est causé que par l’achalandage ou par des erreurs d’encodage (valeur aberrante). Finalement, l’entreprise devrait donc s’assurer de toujours avoir le plus de Bixis possibles disponibles pour les périodes fortes, procéder aux maintenances lors des périodes plus faibles et s’assurer du bon état des stations et des vélos dans les arrondissements avec les stations les plus en demande.

Bien sûr, ceci n’est que la première partie du projet. Dans la prochaine phase, nous devons explorer les modèles de régression. Par conséquent, il sera primordial d’appliquer une transformation afin d’obtenir une distribution symétrique pour les variables en fonction des modèles qui seront utilisés, étant donné qu’à l’exception de la température, toutes les autres variables montrent une distribution étirée vers la droite.

4 Contribution des membres de l’équipe

- Abdoul Wassi Badirou: Introduction, analyse univariée et conclusion
- Alfred Assal: Analyse multivariée (toutes les covariables et la variable cible `rev`) et conclusion

Table 9: Détails entre les journées fériées et les journées normales selon le revenu rapporté par les non-membres, par station, par jour

	Revenu moyen	Revenu total
journée normale	27.17594	125498.480
Journée fériée	34.20486	3694.125

- Samuel Croteau: Analyse multivariée (toutes les covariables et la variable cible **n_tot**) et conclusion
- James Roy: Analyse multivariée (toutes les covariables et la variable cible **dur**) et conclusion