

1 Présentation du service étudié

1.1 Principe général de Communauto

Communauto est un service basé dans différentes villes, principalement au Canada mais également à Paris, qui propose de la location de voitures sous deux formules :

free-floating : les utilisateurs peuvent utiliser une voiture garée dans l'espace public afin d'effectuer un trajet d'un point A à un point B, tout en restant dans la zone de service de Communauto. Cette formule est principalement destinée aux trajets quotidiens, par exemple entre le domicile et le lieu de travail.

basé-station : les utilisateurs peuvent utiliser une voiture garée dans l'une des nombreuses stations réparties dans la ville afin d'effectuer un trajet puis de ramener la voiture à la station de départ. Cette formule est plus destinée aux plus long trajets, planifiés à l'avance, par exemple dans le cadre de voyages sur le week-end.

Il est en plus possible pour les usagers de réserver un véhicule à l'avance pour les deux systèmes : le délai est de 30 minutes pour les voitures en free-floating et de 1 mois pour les voitures basées-stations.

1.2 Fonctionnement du système et lexique utilisé

Dans la pratique, les usagers se connectent un certain instant afin de réserver une voiture dans le futur ; les voitures basées-stations peuvent être réservées par créneaux jusqu'à un mois à l'avance tandis que les voitures en free-floating disponibles à cet instant peuvent être réservé jusqu'à 30 minutes à l'avance.

Dans la suite, on parlera de *temps de pré-réservation* pour le temps s'écoulant entre l'arrivée d'un usager et le début de sa réservation et de *temps de réservation* (et non de temps de trajet) pour la durée de réservation elle-même (uniquement pour les voitures basées-station).

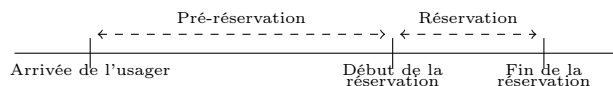


FIGURE 1 – Schéma d'une réservation

Pour les voitures basées-station, il est également possible pour un usager d'annuler une réservation effectuée ou encore de rendre la voiture avant la

fin annoncée de la réservation. Dans ce dernier cas, on parlera de *fin de réservation anticipée* par opposition aux *réservations menées à leur termes*

2 Analyse de données

2.1 Présentation des données

Grâce à un partenariat entre Communauto et Polytechnique Montréal, j'ai eu accès aux données correspondants aux réservations de voitures dans la ville de Montréal. J'ai ainsi pu étudier les données correspondant à l'année 2021, comprenant aussi bien les réservations effectuées sur les voitures en free-floating que sur celles en basées-stations, cela correspondant à respectivement à 973 819 et 441 871 réservations effectués soit 1 415 690 réservations au total.

Je disposais pour ces données d'un grand nombre de champ dont ceux que j'ai principalement utilisés sont les suivant :

- la date et l'heure de début et de fin de la réservation
- la distance effectuée durant la réservation
- les points de départ et d'arrivé de la réservation, correspondant à la localisation de la station pour les basées-station
- d'autres informations relatives à l'utilisateur du service (genre, langage, ancienneté et adresse approximative notamment)
- lorsque la réservation est annulée, la date d'annulation
- le statut de la réservation, parmi active, finie prématurément, annulée ou modifiée entre autres

On retrouve dans les données principalement des réservations actives (menées à leur terme), annulées ou finies prématurément ainsi qu'un très petit nombre de réservations au statut spécial car effectuées par Communauto pour maintenance. Leur répartition est présentée figure 2.1.

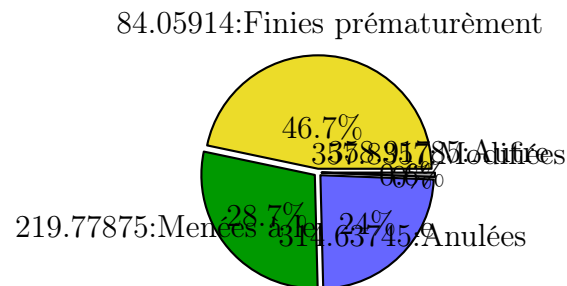


FIGURE 2 – Répartition des statuts de réservations dans les données

2.2 Nettoyage préalable des données

Pour étudier les données et tenter de déterminer les lois de probabilités par lesquelles on peut approximer les différentes distribution, il faut commencer par nettoyer et filtrer les données :

Dans un premier temps, il s'agit ainsi d'isoler les réservations réelles, c'est à dire les réservations effectuées par les usagers, et d'éliminer les autres, par exemple les réservations pour maintenance ou pour changer la voiture de station de rattachement. Ce filtrage se fait très simplement en considérant uniquement les réservation dites actives dans les données.

Suivant le paramètre étudié, j'ai pu avoir à filtrer selon le statut de la réservation, afin d'éliminer les trajets annulés pour mesurer le temps de réservation effectif par exemple.

2.3 Fluctuations au cours du temps de la capacité utilisée

Je me suis intéressé aux comportement des usagers, pour voir notamment quelles périodes sont les plus prisées par les usagers. La figure 2.3 présente ainsi la capacité utilisée typique durant une semaine, obtenue en moyennant chaque jour de la semaine sur les 53 semaines de l'année 2021.

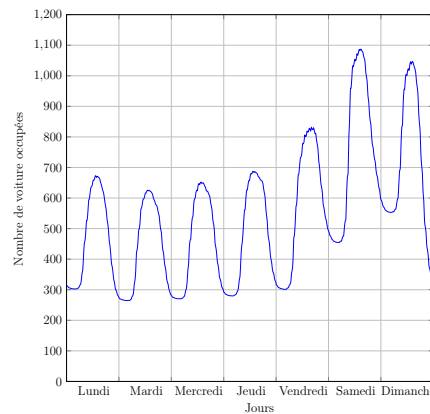


FIGURE 3 – Occupation du système au cours d'une semaine type

On y observe de grandes fluctuations de l'occupation au cours de la semaine, principalement à l'échelle de la journée avec des pics chaque jour aux alentours de 13h. De plus un effet du week-end est très clairement présent, tout d'abord avec des pics en journée plus haut que le reste de la semaine mais aussi avec des creux plus haut, représentant les usagers qui réservent une voiture afin de partir en week-end avec. Par ailleurs, on remarque que le

vendredi fait en quelque sorte la transition entre la semaine et le week-end, avec un pic intermédiaire en milieu de journée et une nuit entre vendredi et samedi plus chargée que les nuits de semaine, ce qui montre que les départs en week-ends commencent ici dès le vendredi soir.

2.4 Étude du temps de trajet des usagers

2.4.1 Premières analyses

On mesure un temps de trajet moyen de 7.8h alors que le maximum de la distribution est situé à 2h et le temps médian est lui de 3.5. La moyenne du temps de réservation est en effet bien augmentée par la présence de très long trajets, certains faisant plus de 3 semaines.

De plus le kurtosis de cette distribution vaut 160, ce qui indique que l'on a affaire à une distribution à queue lourde, c'est à dire qu'elle possède une queue plus épaisse qu'une loi normale et donc que les événements extrêmes y sont plus probables.

2.4.2 Détermination et lissage de la distribution

Afin de déterminer des informations plus fines, il faut étudier l'allure de la distribution : Pour cela, dans le cadre du temps de trajet des usagers, le champ *res_trip_length_hr* présent dans les données correspond précisément au temps de réservation de chaque élément.

Il ne comprend supposément que des temps définies au quart d'heure près, les réservations pour les voitures basés-station se faisant sur des créneaux de la taille d'un quart d'heure. Cependant, on retrouve parmi les données certaines valeurs aberrantes, telles des réservations durant 1 ou 40 minutes par exemple. Cela correspondant à seulement 13 réservations, j'ai décidé de les enlever de l'étude du temps de trajet.

On peut alors obtenir la répartition empirique des temps de trajet (courbe bleue sur la figure 2.4.2) qui reste néanmoins très irrégulière ; on peut en effet observer une certaine périodicité d'ordre 4, certainement due aux utilisateurs, qui auront une plus grande tendance à réserver un trajet d'une durée entière en heures. Pour la même raison, on observe un pic de trajets d'une durée de 10 heures ainsi que des pics successifs présents pour tout les jours entiers.

Afin de pouvoir exploiter la courbe, il est nécessaire de lisser la courbe, par exemple avec un estimateur à noyau de fenêtre assez large pour enlever ces périodicités. J'ai préféré un estimateur à taille de noyau variable afin de ne pas trop écraser les données présentes à l'origine et ainsi de garder l'allure générale de la courbe, ce qui est réalisé sur la courbe rouge en figure 2.4.2.

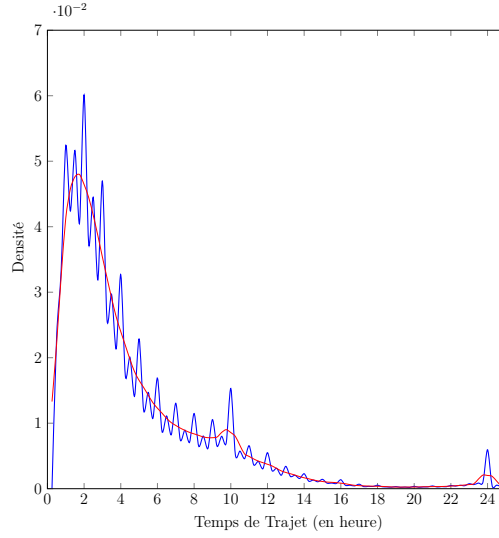


FIGURE 4 – Densité du Temps de trajet lissée

2.4.3 Détermination de la loi la mieux adapté aux données

A partir de cette courbe lissée, j'ai cherché à déterminer par quelle loi l'approcher pour correspondre le mieux aux données et ainsi adapter le mieux le modèle. Pour cela, je me suis concentré sur des lois à queues lourdes, à savoir notamment loi gamma, loi log-normale, loi de Weibull ou encore loi de Burr.

Je les aies comparées à la distribution empirique du temps de trajet en utilisant la distance en variation totale :

Définition 1. Distance en variation totale

Soient μ, ν deux mesures de probabilité de densités respectives f et g .

On note $d_{VT}(\mu, \nu) = \int_{\mathbb{R}} |f(x) - g(x)| dx$ la distance en variation totale entre μ et ν .

Après optimisation des différents paramètres pour les différentes lois, j'ai obtenu le tableau 2.4.3 présentant la distance minimale entre chaque loi et la distribution du temps de réservation.

Grâce à cela, il s'avère que c'est une loi log-normale de paramètres $\mu = 2.6$ et $\sigma^2 = 0.9$ qui permet la meilleur approximation du temps de réservation.

Le choix d'une loi log-normale est confirmé par l'étude des graphes quantile-quantile des différentes distributions théoriques présents figure 2.4.3 seule la loi log-normale sélectionnée permet une bonne approximation de la queue de distribution.

Loi approximant	Distance
Loi gamma	0.23
Loi Log-normale	0.17
Loi de Weibull	0.24

FIGURE 5 – Distance entre différentes lois et la distribution du temps de réservation

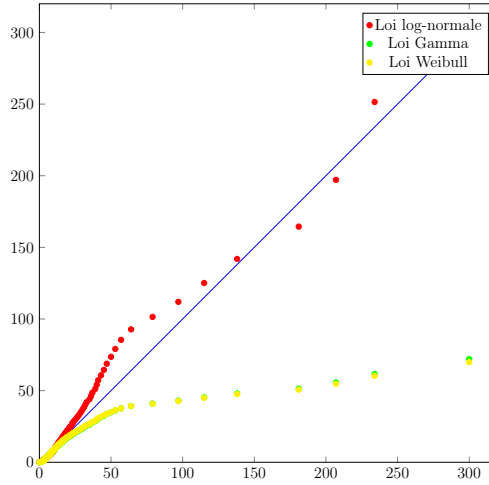


FIGURE 6 – QQPlot

Definition 2. Loi log-normale

Une variable aléatoire X suit une loi log normale de paramètres μ et σ^2 lorsque $Y = \ln(X)$ suit une loi normale de paramètres μ et σ^2 .

2.5 Capacité des stations

Dans le cadre de cette étude, il est intéressant de regarder la capacité des stations dans lesquelles sont stationnées les voitures basées-stations. Pour cela, j'ai regardé sur une semaine le nombre de voitures étant associé à chaque stations afin d'en faire un histogramme. Je me suis limité à une semaine afin de passer outre les voitures qui changent occasionnellement de station au grès des ajustements réalisés par Communauto tout en ayant assez de données pour que quasiment toutes les voitures aient été utilisées sur cet intervalle et ainsi avoir un histogramme assez fiable.

En traçant l'histogramme (figure 2.5), on observe une majorité de très petites stations : 66% des stations ayant 1 ou 2 voitures seulement. C'est explicable par la politique de Communauto d'avoir des station au plus près

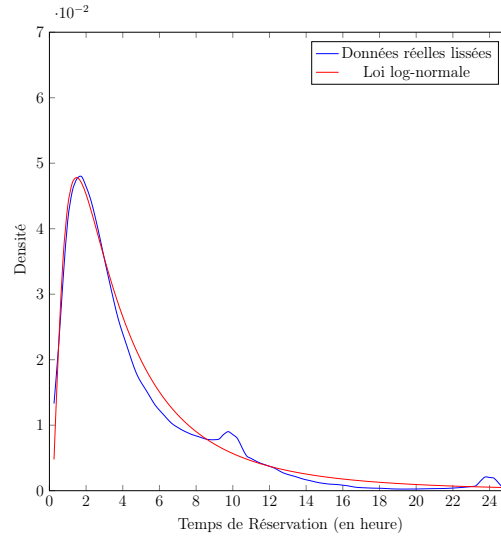


FIGURE 7 – Ajustement du temps de réservation par loi log-normale

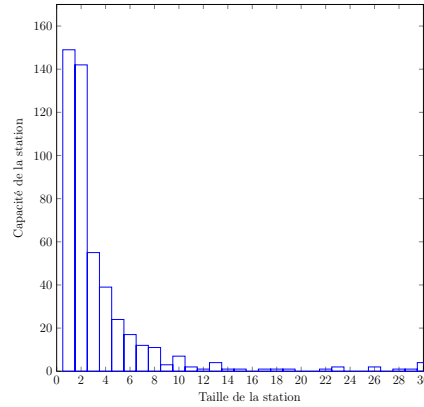


FIGURE 8 – Distribution du nombre de voitures par station

des usagers, multipliant ainsi la quantité de stations déployées, ainsi que par la difficulté d'obtenir de grandes zones afin d'y réaliser des stations, favorisant des petites stations avec quelques grandes stations pour quelques points d'intérêts.

2.6 Étude de la distance de trajet

Afin d'étudier la loi de la distance parcourue, il faut dans un premier temps retirer des données les réservations ayant une distance parcourue nulle, correspondant la majorité du temps à des réservations "à vide", sans bouger la voiture. (Cela pourrait également correspondre à de très petits trajets,

la distance relevée étant arrondie au kilomètre près, mais cela ne concerne qu'un faible nombre de trajets).

L'étude de la distance est après cela tout à fait similaire à celle de du temps de trajet, les distributions lissées étant similaires.

Dans ce cas, la distribution qui l'approche le mieux est encore une fois une loi log-normale, dont les paramètres sont $\mu =$, $\sigma^2 =$. La figure 9 présente la loi expérimentale et la loi approchée en superposition.

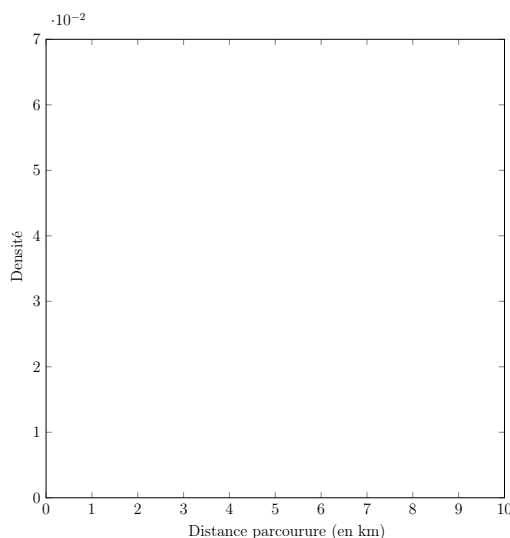


FIGURE 9 – Ajustement de la distance parcourue par loi log-normale

2.7 Corrélation entre le temps de réservation et la distance parcourue pour obtenir une voiture

2.8 Étude du temps éventuel d'annulation

Une autre donnée intéressante concerne l'annulation des réservations. En effet, comme la figure 2.1 cela correspond à une part importante des données : 24% des réservations initiales étant annulées.

Il est alors intéressant de regarder quand cette annulation a lieu dans le cycle de vie de la réservation. Pour cela, on peut s'intéresser à la loi du temps entre l'annulation de la réservation et son début théorique. Une loi empirique de ceci est tracé figure 10

La première chose que l'on remarque est la présence d'un petit nombre de réservations annulées après leur temps de début de réservation, correspondant à environ 0.5% des trajets annulés. Ceci n'est pas une erreur des données, la réservation étant payée par les usagers pour le temps

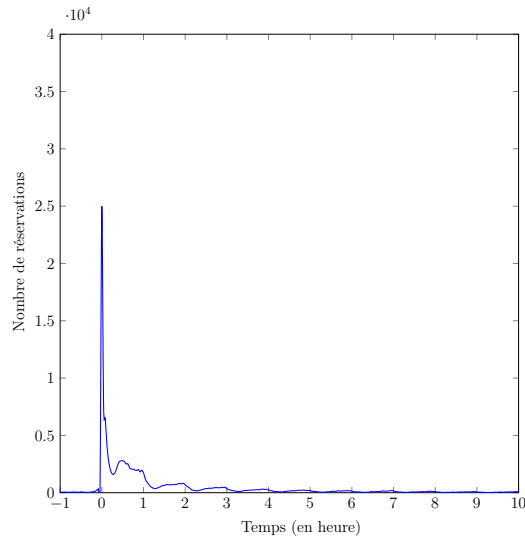


FIGURE 10 – Loi du temps entre annulation et début de la réservation

Ensuite la majorité des annulations ont lieu juste avant le début du trajet, comme montré par le pic observées sure les données en 0. Ainsi, les réservations annulées le sont en moyenne moins de 3 jours avant le début de la réservation : 62% étant annulées moins de 24h à l'avance.

On observe cependant pour la suite une tendance géométrique qui apparaît bien sur la courbe lissée, encore une fois à l'aide d'un estimateur à noyau.

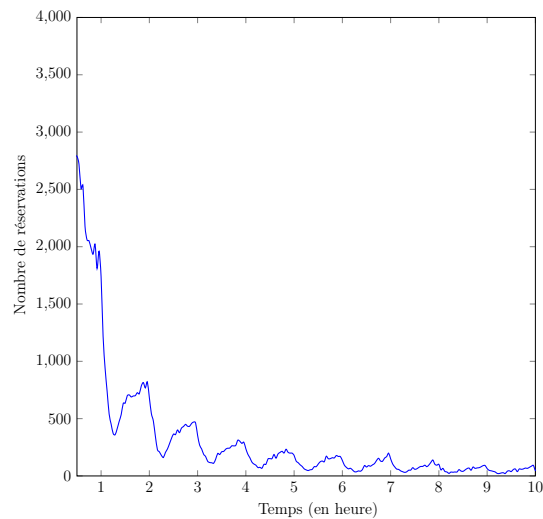


FIGURE 11 – Loi du temps entre annulation et début de la réservation

2.9 Étude du temps de pré-réservation

Les données disposent d'un champ *date d'inscription* correspondant en théorie à la date à laquelle l'utilisateur a réservé un créneau. Cependant, cette date ne correspond généralement pas à la date d'inscription réelle mais à la date de dernière action sur la réservation, que ce soit son annulation, sa modification ou bien un retour anticipé de la voiture. Ainsi les données ne sont pas forcément fiables et sont donc difficiles à exploiter.

3 Simulations numériques