

Case study 1 : Analyse avec R de Cyclistic, entreprise de location de vélo



Phase 1 : Poser les questions nécessaires à la bonne compréhension de l'étude

Problématique : Développer la clientèle de l'entreprise, en transformant des utilisateurs occasionnels en régulier. Question : Etude des comportements d'utilisateurs d'une plateforme de location de vélo.

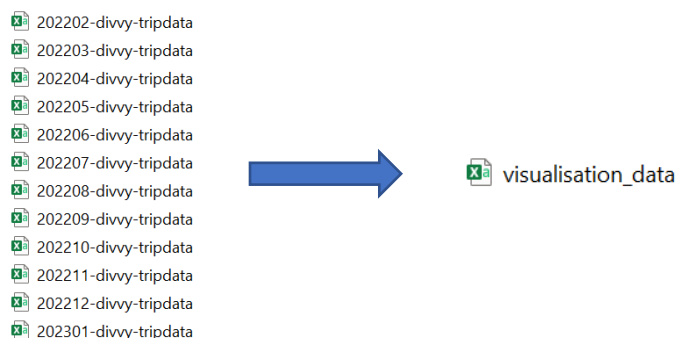
Quels sont les habitudes des cyclistes occasionnels comparées à celle des abonnés ? Comment convertir ces utilisateurs occasionnels en abonnés ?

Phase 2 : Préparer les données

Nous avons accès à l'historique de toutes les **locations réalisées au cours des 12 derniers mois**. A noter que nous avons accès à plus de données mais dans un souci de déterminer les comportements récents et de se **détacher de l'effet COVID** notre étude se limite seulement à la dernière année écoulée. Ces fichiers sont des csv et

Il y a énormément de données dans chacun des fichiers, aussi les données ne sont **pas échantillonnées**. Ces deux informations permettent d'avoir **une bonne qualité de données** car elles ne sont pas biaisées.

Tous les fichiers de données sont des .csv dont **les informations sont collectées par l'entreprise** elle-même, ce qui renforce la fiabilité de ces dernières. Les données sont alors téléchargées dans le même dossier en prenant soin d'utiliser une nomenclature unique afin de favoriser la compréhension et la manipulation de ceux-ci, le but étant de les **synthétiser afin de visualiser** les informations qu'ils contiennent.



Phase 3 : Nettoyage et manipulation des données

Après avoir **ouvert les 12 fichiers dans R**, et vérifié qu'il y avait le **même nombre de colonnes** dans chaque fichier et qu'elles portaient **toutes le même nom**, j'ai pu **combinaisonner les fichiers en un seul tableau** afin de manipuler les données plus facilement avec la fonction `bind_rows`.

Il s'agit désormais de **rendre cohérentes les informations** au sein de ce tableau, par exemple pour désigner des abonnés au service il y avait les dénominations « members » et « subscribers ».

Ensuite pour mener une analyse poussée des comportements quotidiens, des **colonnes supplémentaires sont nécessaires**, notamment ajouter le **jour de la semaine** correspondant à la location, ainsi que le **temps de chaque location** dont le calcul se fait en soustrayant les dates et heure de début et de fin de location.

Enfin il a fallu **nettoyer les données** qui peuvent gêner notre analyse :

- Les trajets correspondant au contrôle qualité par l'entreprise
- Les trajets dont la durée de location est négative
- Plus globalement les lignes comportant des données manquantes

Phase 4 : analyse des données

Durant cette phase, il a été question **d'extraire les informations nécessaires** pour répondre à la question que l'on s'est posée durant la phase 1.

Tout d'abord essayer d'avoir un **ordre de grandeur de la durée d'un trajet**, en calculant la durée minimale, moyenne, médiane et maximale. Puis en **catégorisant** ensuite par **type de consommateur (abonné ou occasionnel)**, puis en catégorisant encore plus en ajoutant **les jours de la semaine**.

On peut aussi **compter le nombre de location** réalisé de la même manière catégorisé par **type de consommateur** et par **jour de la semaine**.

Phase 5 : Partage des informations via des graphiques adaptés

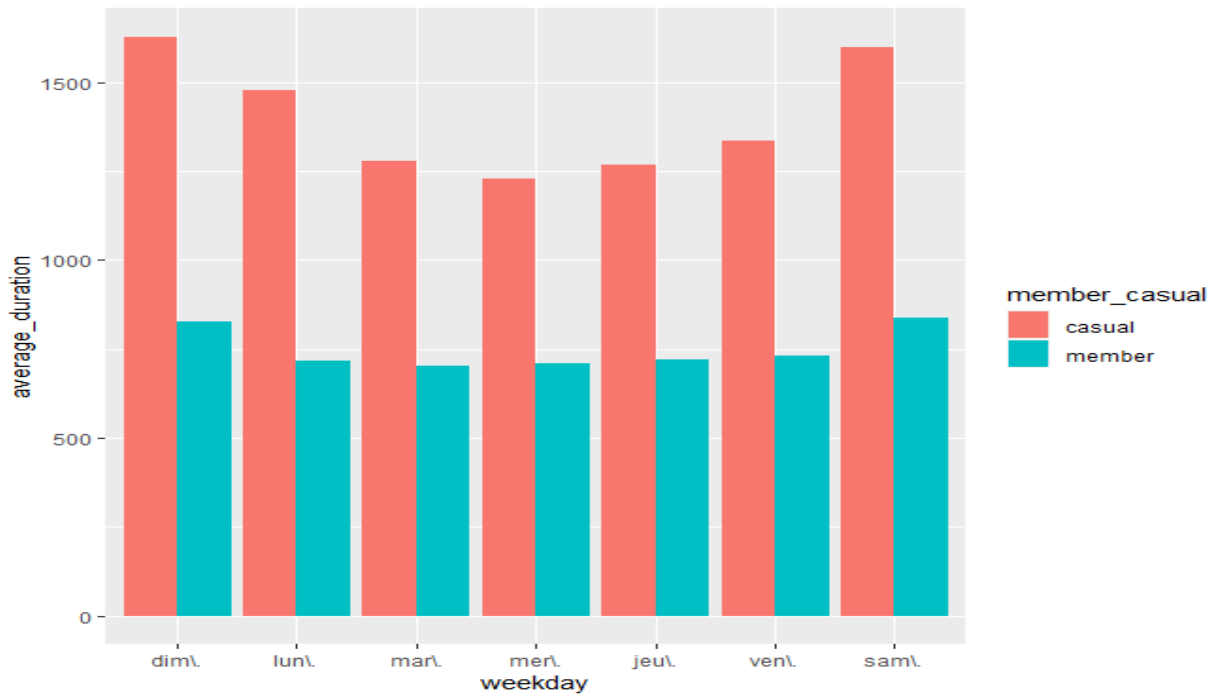


Figure 2 : Durée moyenne des locations en minutes par jour de la semaine

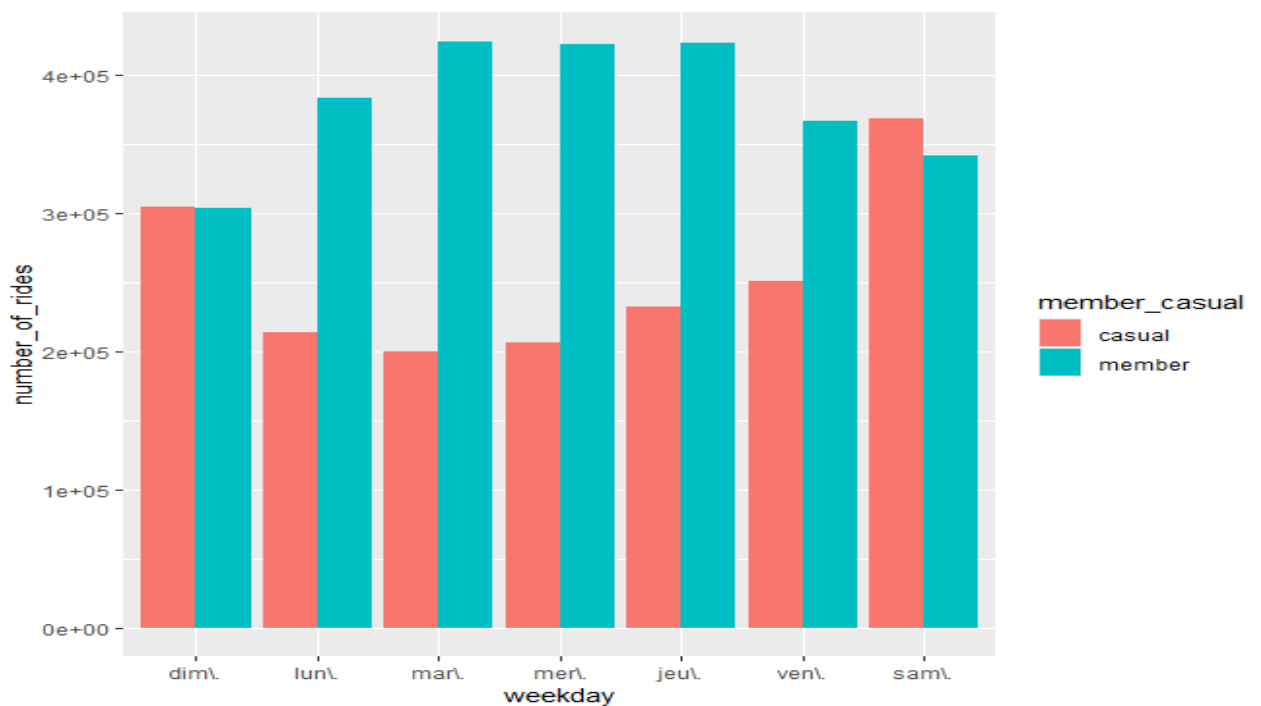


Figure 1: nombre de location de vélos par jour de la semaine

On peut observer plusieurs choses de ces graphiques :

- Les abonnés utilisent les vélos majoritairement la semaine
- Les utilisateurs occasionnels utilisent les vélos majoritairement le weekend
- Les utilisateurs occasionnels utilisent plus longtemps les vélos que les abonnés

En somme on peut en déduire les profils suivants :

- Les **abonnés** utilisent les vélos pour **aller au travail** c'est pour cela que leur **trajet sont court** et concentré sur les **jours de semaine**. On note une légère baisse les lundi et vendredi, surement des jours où ces personnes télétravaillent ou sont en WE prolongé.
- Les **utilisateurs occasionnels** utilisent les vélos de **manière récréative**, pour faire du tourisme au sein de la ville, se balader dans les parcs alentours. C'est pour cela que les locations ce segments là se concentrent le **weekend** et **durent plus longtemps**.

Phase 6 : Recommandation pour répondre à l'enjeux

Il faut donc viser les touristes, et les habitants qui souhaitent faire des balades à vélos.

On peut imaginer :

- **Créer des abonnements** séparés pour le weekend et pour la semaine
- **Faire des partenariats** avec l'office de tourisme, musée, parc, hôtel etc. pour augmenter la visibilité auprès de ce segment là
- Proposer des **itinéraires de balade touristiques** auprès des stations ou sur l'applications
- Proposer des **visites guidées** pour les balades à vélo pour les personnes abonnés