

1. PRÉSENTATION

Nous sommes des employés de l'entreprise Orchy 3D, spécialisée en intelligence artificielle. En tant que membres de l'équipe du pôle DATA, nous avons pour devoir d'aider Jeon Su Kyung, une personne de la mairie de Tokyo pour rendre compte de la tendance d'utilisation des trottinettes électriques. Pour cela nous allons mener une étude statistique sur ces dernières avec comme variable le contexte météorologique et temporel, l'objectif final étant la réalisation d'un tableau de bord interactif s'actualisant à chaque nouvel ajout de données et une analyse statistique.

Dans ce rapport vous trouverez les différentes avancées du projet.

2. OBJECTIFS

- Extraction des données d'un serveur externe de façon automatisée ;
- Nettoyage et formatage des données pour faciliter leur traitement et lecture ;
- Analyse des données avec des représentations graphiques.

3. PROCÉDURE

3.1. Lancement

Dans un premier temps nous avons créé un programme pour récupérer les données brutes du serveur en format csv, avec interrogation sur plusieurs dates en une fois (utilisation de Python) : récupération de 3 fichiers à partir des 3 serveurs, compilant chacun la période choisie.

Nous avons fusionné les 3 fichiers csv en un et renommé des colonnes pour une meilleure compréhension.

L'automatisation reste encore à mettre en place. (quelques bases déjà mises en place sur Talend pour automatiser le processus Python).

Le fichier Github contient une partie des codes.

En parallèle de ce travail, nous avons également cherché, avec RStudio et notre échantillon de données, les premiers liens entre les différentes conditions et le nombre de trottinettes louées, dans un objectif de réalisation de graphique explicite.

Nous prêtons attention à chaque ajout de données pour améliorer la justesse des graphiques existant et également dans le but de concevoir de nouveaux graphiques.

Notre travail permettra ainsi d'apporter une réponse à la problématique suivante:

-> Le nombre de trottinettes louées augmente-t-il selon les conditions météorologiques ?

3.2. Présentation du logiciel

Il s'agit d'un fichier python.

Utilise les modules requests et time, importés en début de code.

En début d'utilisation le programme demande à l'opérateur de rentrer l'identifiant de départ (x) et l'identifiant de fin d'extraction (y), permettant d'extraire les données désirées. Les valeurs d'identification sont convertis dans le code pour que cela corresponde à un nombre de jour (ex : rentrer « 1 » en identifiant de départ et de fin extrait toutes les données du 01/12/2017 contenus sur les serveurs).

Le programme intègre une variable « eh » servant à énumérer le nombre de réponses négatives à une extraction.

Le token d'accès donné à notre groupe pour accéder au serveur est intégré au programme (gp_id), mais comme nous avons été informés de la possibilité que le token puisse changer, nous avons intégré la demande de confirmation du token actuel, avec la possibilité de rentrer le nouveau token à utiliser (celui-ci ne sera pas conservé dans le programme par la suite cependant).

Après toutes ces confirmations, le programme débute :

La première étape consiste en la création de la base (locations_data, meteo_data, calendrier_data) qui servira pour l'écriture des 3 fichiers csv obtenus en sortie, c'est à dire l'ajout des en-tête pour chacun d'entre eux dans leur propre variable (met_line, cal_line, loc_line). Ensuite vient la première série de boucle, pour le port 8100, correspondant aux données sur le jour. Le serveur est accédé en rentrant son adresse qui comprend l'id et le token, un minuteur est intégré dans cette section pour ne pas dépasser le nombre de demandes par secondes du serveur et empêcher qu'une extraction échoue.

Un print est employé pour indiquer à l'opérateur si le serveur est accédé avec succès (code 200) ou non, si la connexion a été établie alors on passe à l'extraction elle-même : chaque ligne de données est récupérée et placée dans sa liste dédiée grâce à l'utilisation de splitlines et strip. Afin d'avoir un premier nettoyage des données les en-têtes des données sont ignorées et les « NA » correspondant à des données absentes sont remplacées par une valeur vide.

La boucle se répète jusqu'à ce que tous les id demandés à être traités le sont.

Le début de la deuxième étape est la conversion de x et y pour les ports 8080 et 8090, contrairement au port 8100 ces données correspondent à 1 identifiant = 1 heure d'une journée. Cela correspond à une multiplication par 24 du nombre de serveurs à consulter pour ces deux ports.

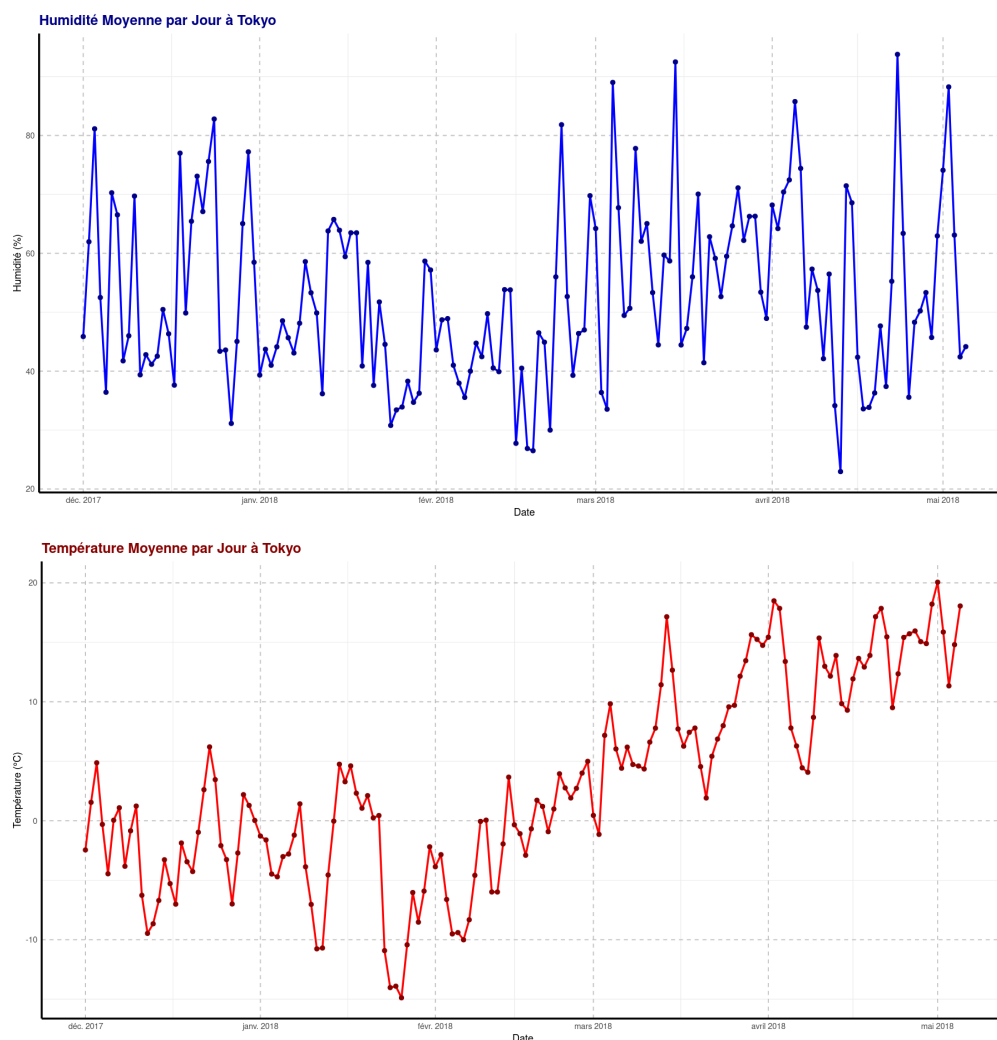
Après cela on rentre dans la deuxième boucle dédiée aux ports 8080 et 8090 qui suit les mêmes étapes que pour la première boucle.

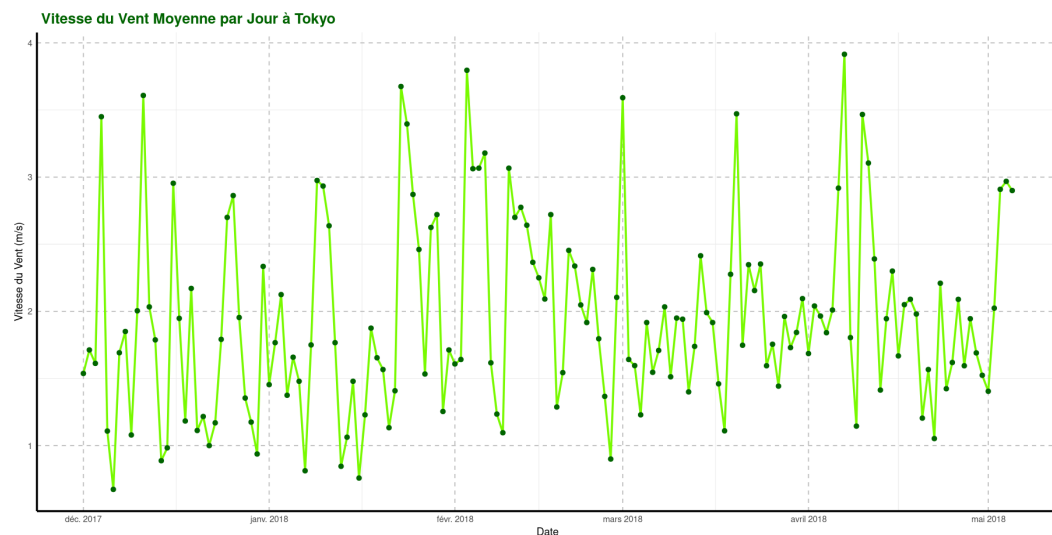
La dernière étape de traitement consiste en l'écriture des 3 fichiers csv grâce aux listes dédiés, par le biais des commande open et write.

L'opérateur est informé de la fin du processus et du nombre d'échec d'exportation, si ce nombre est supérieur à 1 une petite aide apparaît pour indiquer les erreurs les plus communes.

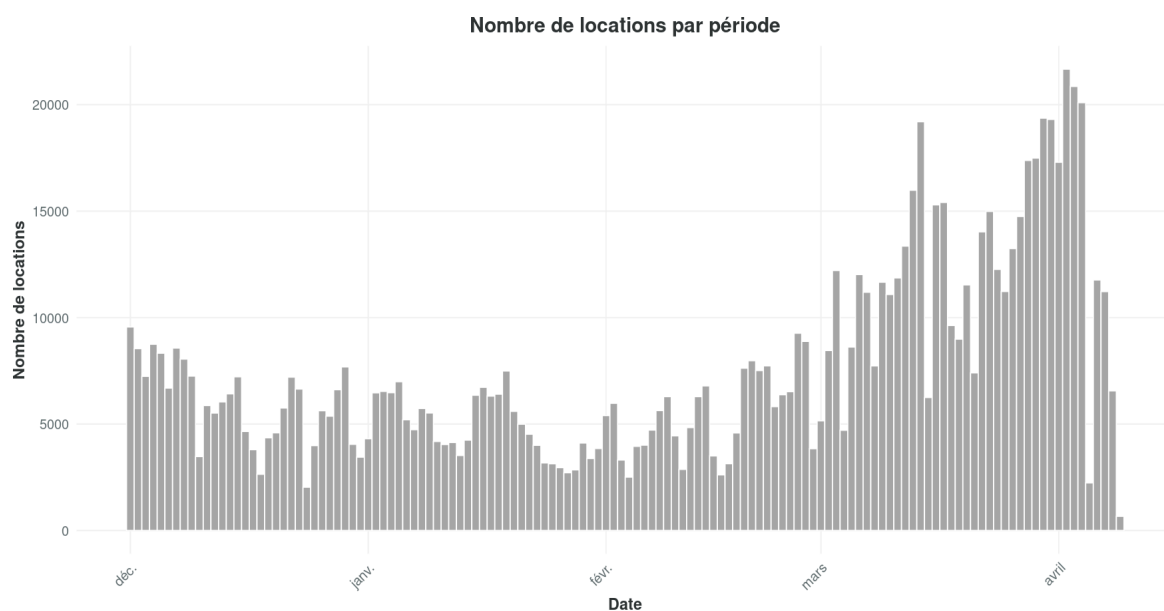
3.3. Présentation de l'échantillon

Pour mieux s'appropriier avec l'échantillon de données, ci-dessous des graphiques permettant d'observer les différentes moyennes de température, humidité et vent :





Et voici un histogramme du nombre de locations par période :



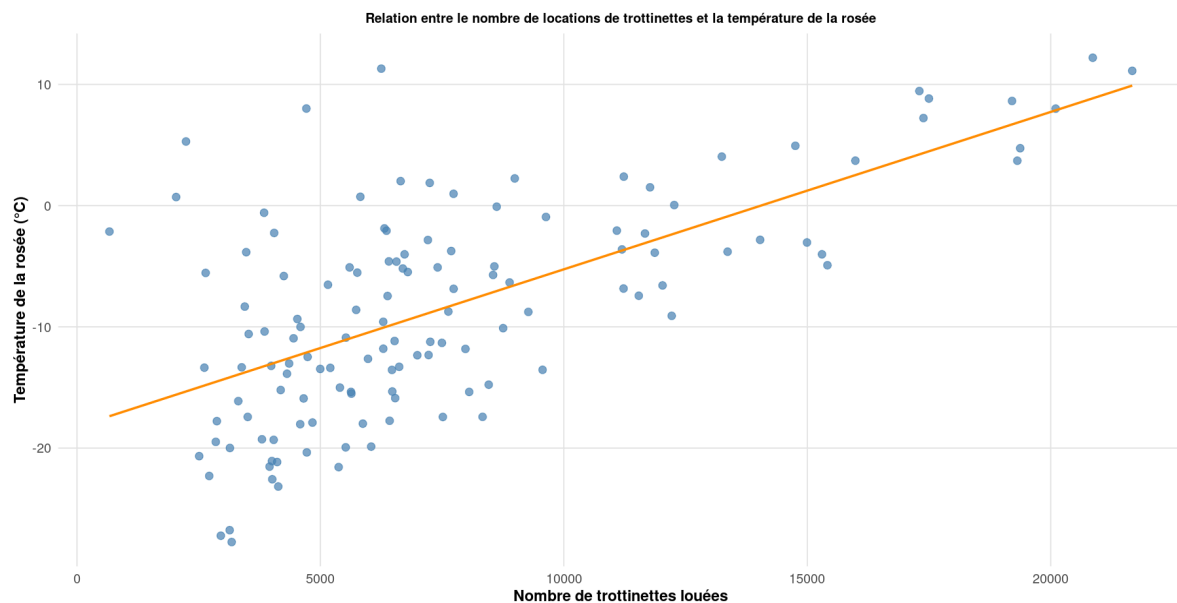
On remarque une baisse significative du nombre de trottinettes louées pour le jour et le réveillon de Noël, mais également pour le jour de l'an. A l'effet inverse d'autres événements comme la Saint-Valentin le 14 février ou encore le White Day le 14 mars, font considérablement augmenter le nombre de locations.

4. ANALYSE GRAPHIQUE :

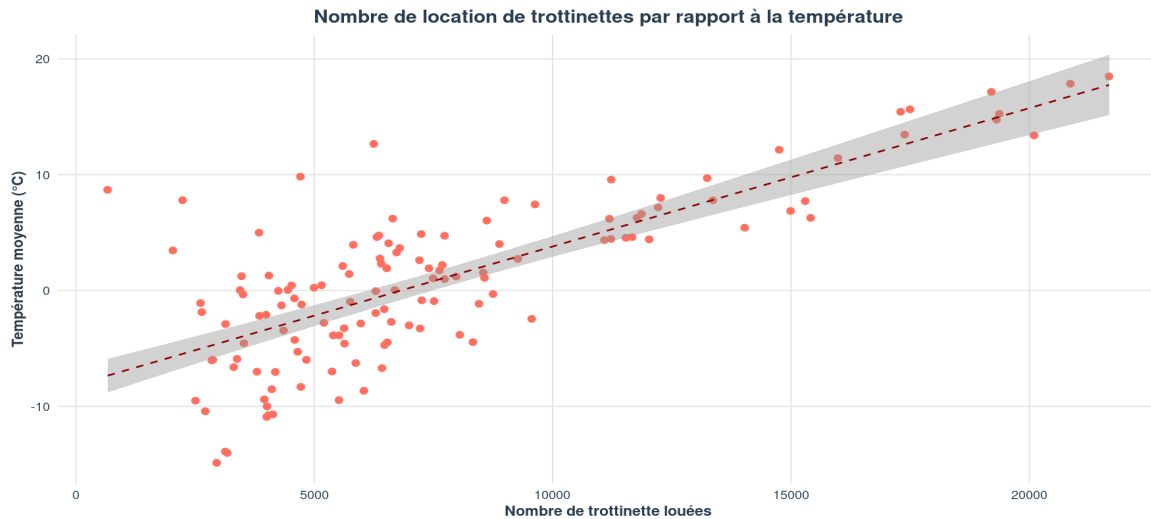
L'analyse graphique nous permettra de déterminer les différents contextes météorologiques pouvant influencer les variations du nombre de trottinettes louées par jour. Pour cela, nous allons nous appuyer sur la température de rosée, la température, la visibilité, l'humidité, la vitesse du vent, la pluie et la neige.

4.1 La Température

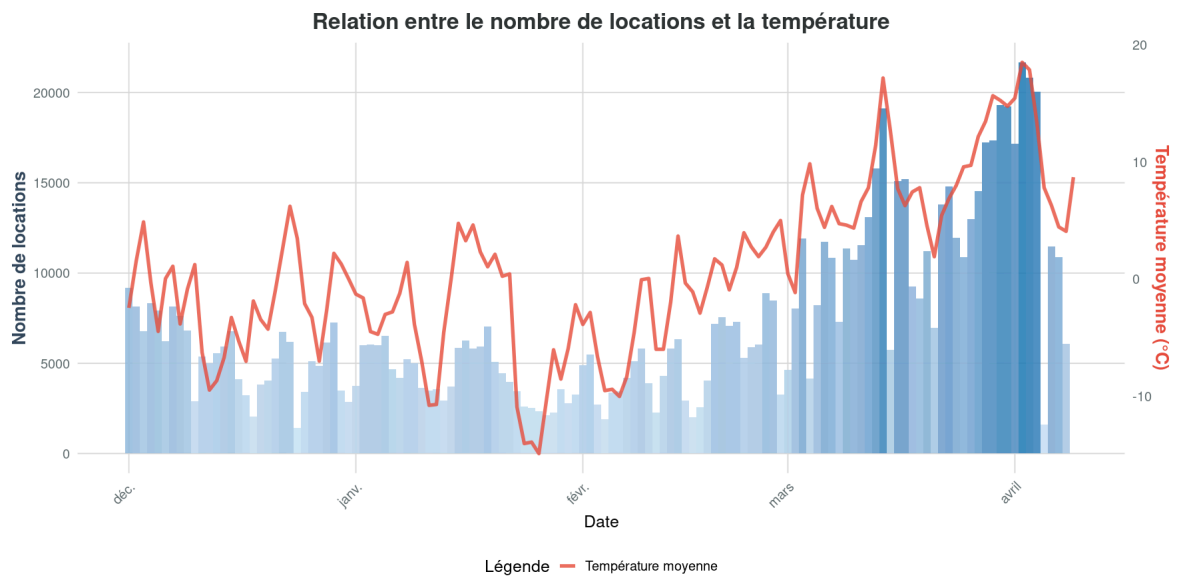
Dans un premier temps, nous allons nous intéresser à la température. Nous pensons que cette dernière a une influence sur le nombre de trottinettes. En effet, étant un moyen de transport ouvert, une température suffisante peut vite devenir un critère pour grand nombre d'utilisateurs. Nous allons vérifier cela avec les graphiques suivants :



Ce graphique met en relation la température de la rosée et le nombre de trottinettes louées. Nous avons pu observer une corrélation entre la température de la rosée et le nombre de trottinettes louées. Soit, plus la température de la rosée est élevée, plus le nombre de trottinettes louées à Tokyo est élevé, validant ainsi notre hypothèse.



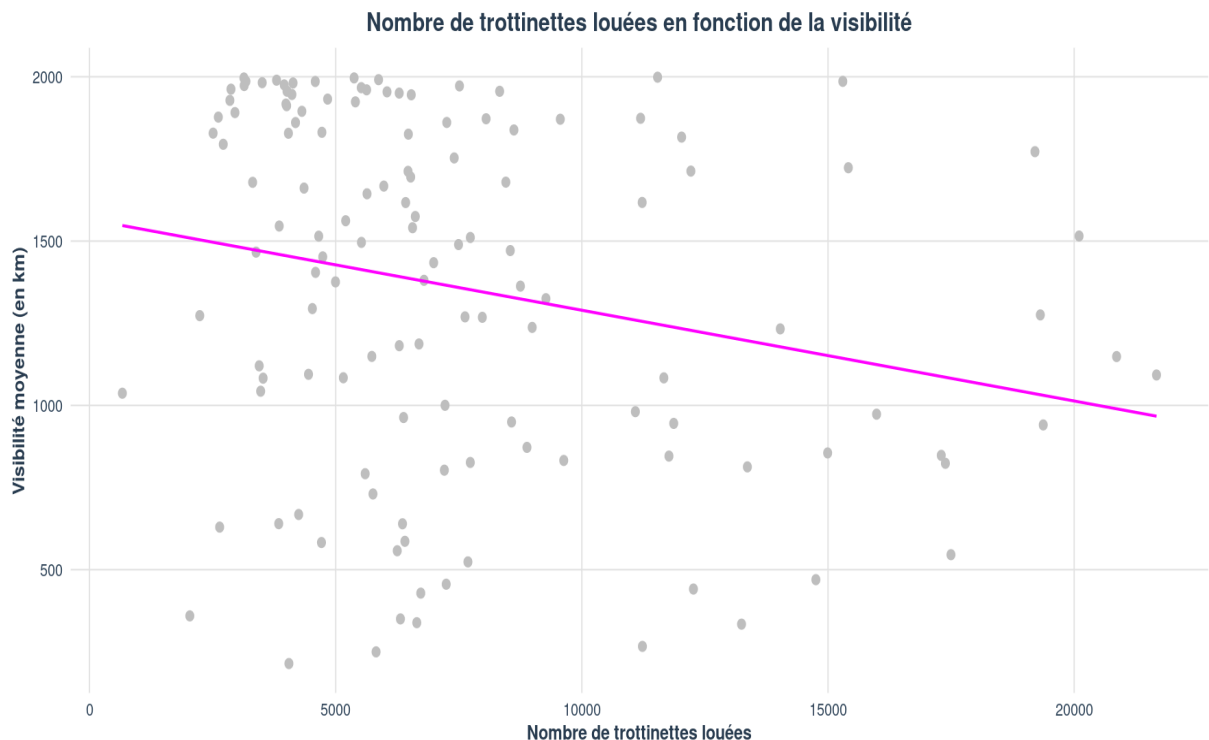
Le graphique ci-dessus met en relation le nombre de trottinettes louées et la température recensée à Tokyo. On observe que si la température à Tokyo augmente alors le nombre de trottinettes louées augmente lui aussi. Cela confirme bien notre idée. L'été sera sûrement prolifique sur la location de trottinettes à Tokyo !



Ici, la superposition de deux visualisations permet de se rendre compte de la corrélation entre la température et le nombre de trottinettes louées. On peut voir une grande similitude au niveau des tendances...

4.2 La Visibilité

La visibilité peut aussi faire baisser le nombre de locations de trottinettes. Beaucoup d'utilisateurs peuvent prendre des précautions face à un environnement imprévisible et brouillé par le manque de visibilité. Nous allons donc voir si cette possibilité est marquée par suffisamment d'emprise de la visibilité sur le nombre de locations.



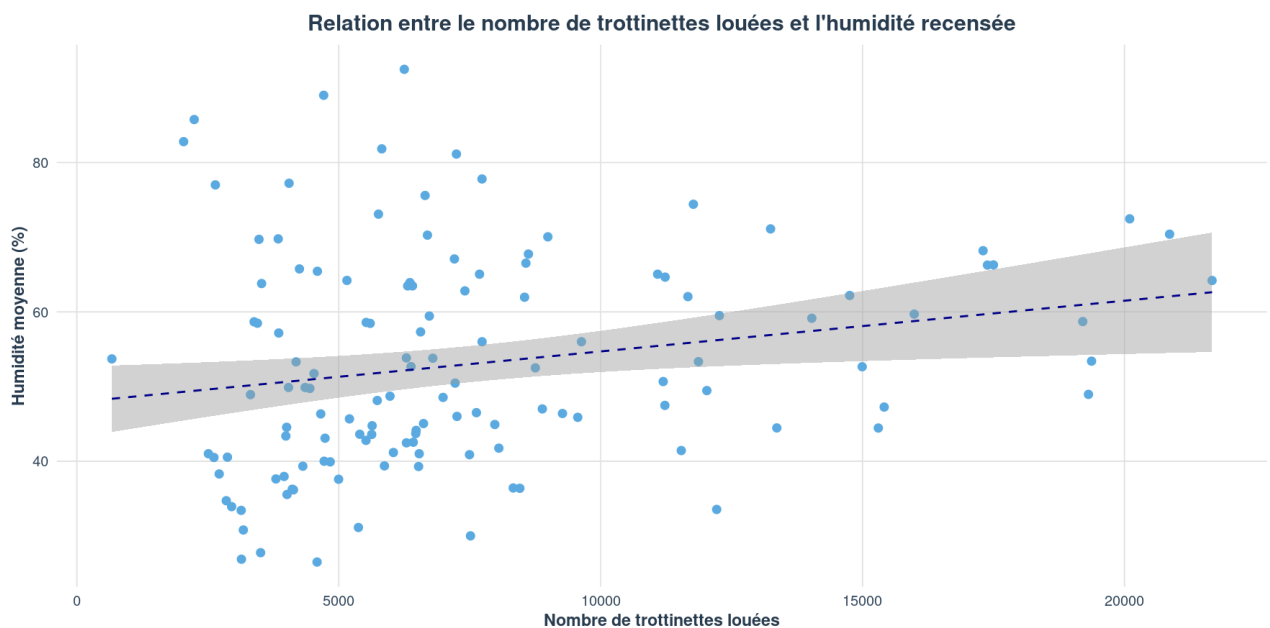
Ce graphique met en lien la visibilité en fonction du nombre de locations de trottinettes.

On constate que plus la visibilité diminue, plus le nombre de locations diminue, ce qui paraît pourtant incohérent.

Ce phénomène peut être dû au fait que les journées au Japon sont courtes et/ou que les citoyens louent des trottinettes qu'en cas d'extrême besoin c'est-à-dire lorsque les transports en commun ne sont plus disponibles. Donc, la nuit.

4.3 L'Humidité

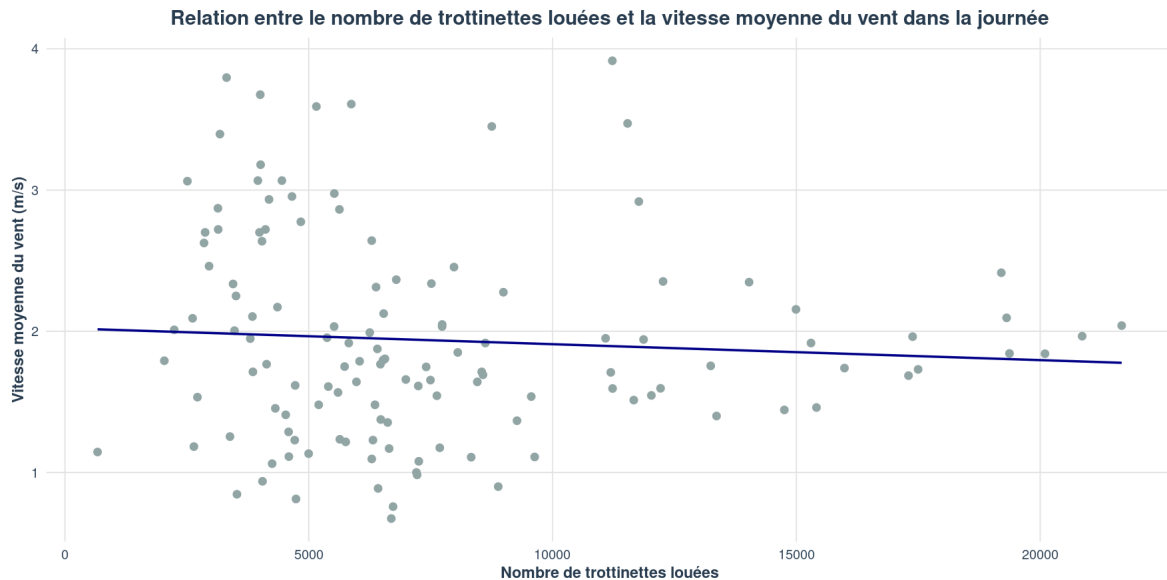
Bien que la température et la visibilité nous paraissaient des facteurs d'influences évidents sur le nombre de locations de trottinettes, l'humidité quant à elle semble moins pertinente car souvent moins explicite aux yeux de chacun. Il est cependant important de vérifier cette hypothèse avec le graphique suivant:



Le graphique ci-dessus porte sur la relation entre le nombre de locations de trottinettes et l'humidité. Malheureusement, on observe pas de corrélation claire entre l'humidité et le nombre de locations. Or, on peut émettre une hypothèse que l'humidité dépend de la température selon une corrélation positive. Plus la température augmente, plus l'humidité augmente. Donc le nombre de trottinettes louées augmenterait selon la température.

4.4 Le Vent

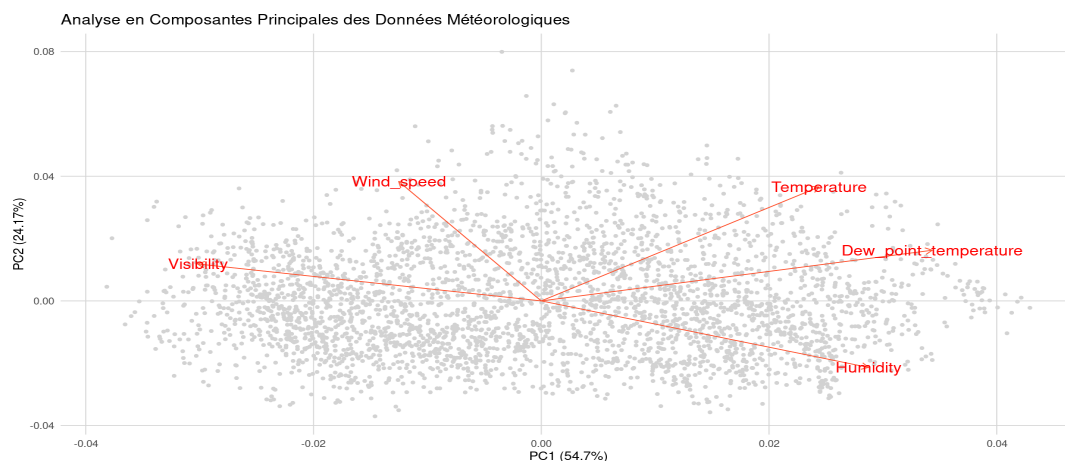
Cette fois-ci, nous pensons que le vent tient un rôle important sur le nombre de trottinettes louées. Par son caractère dangereux lors de forte rafale, il se peut que les utilisateurs de trottinettes fassent le choix de renoncer à ce type de transport. Nous allons donc vérifier cela avec le graphique suivant :



On constate que moins la vitesse du vent est grande, plus le nombre de trottinettes louées est élevé et inversement, plus la vitesse du vent est grande, moins le nombre de trottinettes louées est élevé.

Enfin, nous avons fait une visualisation intuitive des données météo soit une ACP (Analyse en Composantes Principales).

On voit un lien de corrélation entre la température, l'humidité et la température de la rosée. Mais aussi entre la visibilité et la vitesse du vent.



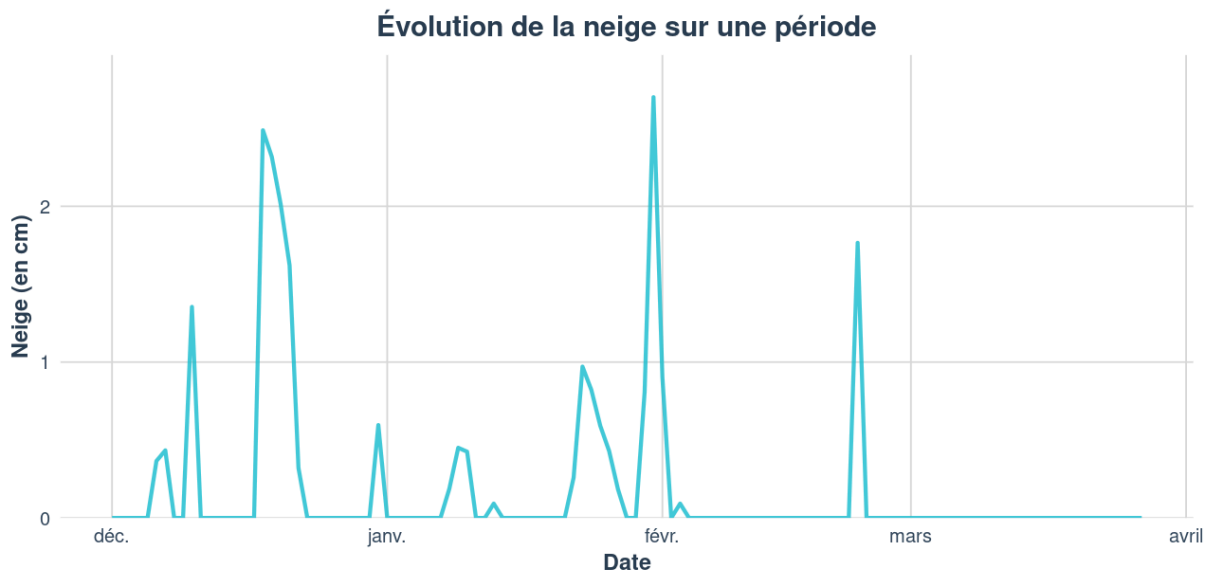
Cette visualisation permet de mieux se rendre compte des données météorologiques.

4.5. La Pluie et la Neige

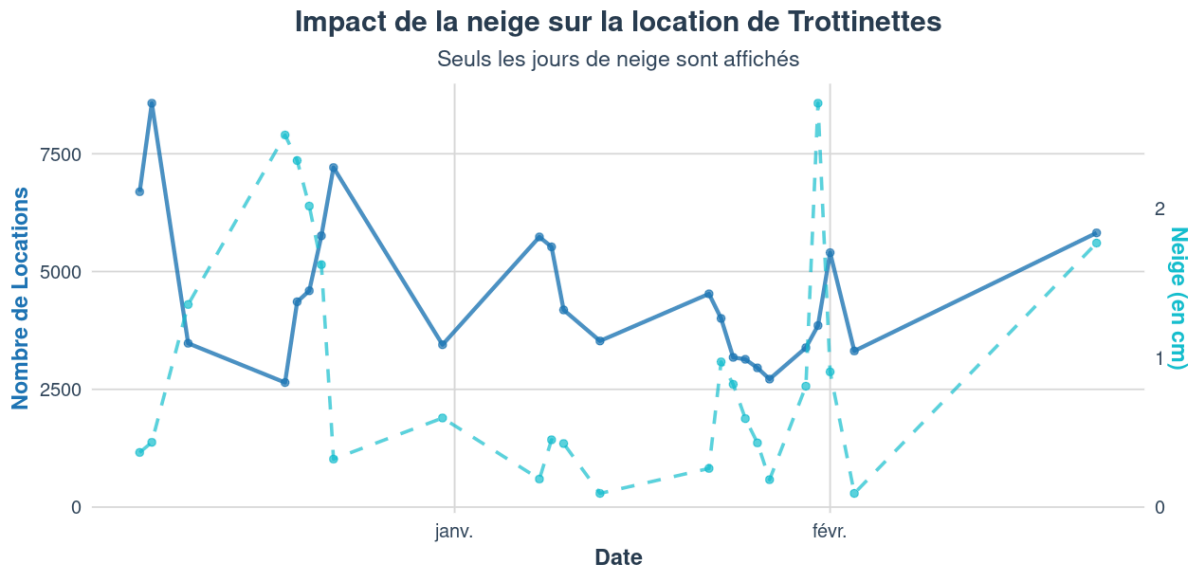
- Cas de la neige:

Nous allons désormais nous intéresser à un autre contexte météorologique, celui de la neige. On peut penser à une certaine influence de la part de cette dernière sur le nombre de locations de trottinettes, par son caractère dangereux et imprévisible. Cependant, cela doit être vérifié. Nous allons donc regarder dans un premier temps la répartition des jours de neige dans le temps pour tenir compte de son caractère fréquent ou non.

Nous chercherons ensuite son éventuelle influence sur la location de trottinette (en prenant compte également de la quantité de neige tombée).



On remarque que la neige semble assez fréquente en décembre et dans les deux premiers mois de l'année. Bien que la quantité ne soit pas très grande, cela est suffisant pour représenter un risque plus important de glissade pouvant repousser un certain nombre d'utilisateurs à ne pas utiliser de trottinette ce jour-là. On s'attend d'autant plus à une baisse du nombre de locations car la neige est liée à des températures assez faibles, qui, nous l'avons montré précédemment, joue un rôle sur l'utilisation moins importante des trottinettes.



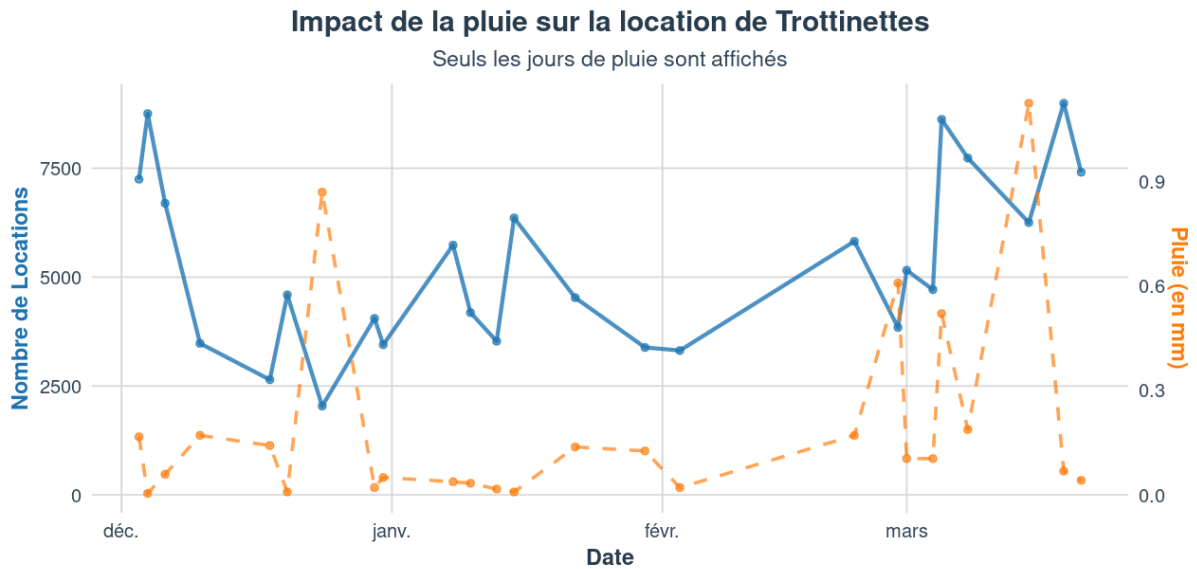
A première vue, on constate que l'évolution des deux courbes se fait de manière opposée, plus la neige est importante, plus le nombre de locations est faible et inversement. On peut donc bien affirmer que la neige influence le nombre de locations à la baisse, les habitants de Tokyo empruntent peu les trottinettes par temps de neige.

- Cas de la pluie :

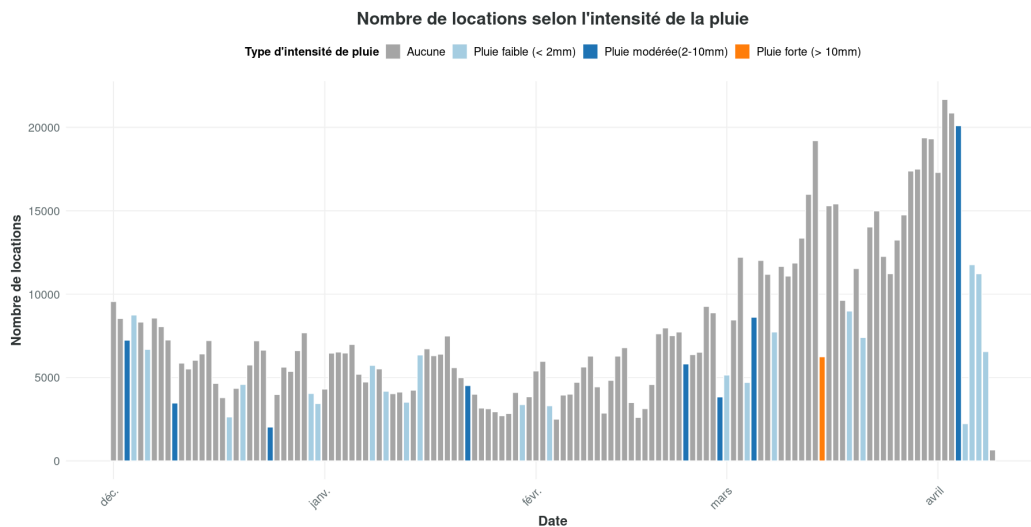
Dans un second temps nous nous intéresserons à la pluie. Très contraignante face à ce moyen de transport non couvert, on peut penser comme pour le cas de la neige, qu'elle influence le nombre de locations de trottinettes. Une nouvelle fois, cela reste à vérifier. Nous procéderons de la même manière que pour la neige pour démontrer cela.



La pluie semble assez répartie de décembre à avril mais avec des quantités différentes. Nous allons donc voir si on constate bien un impact de la pluie sur le nombre de location.

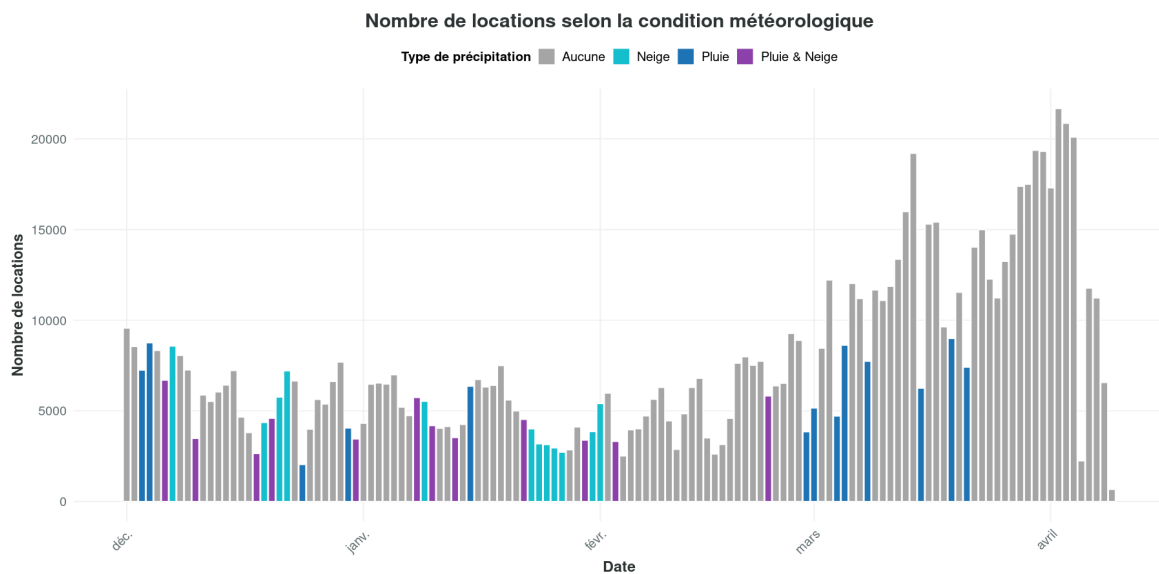


Ici, on constate une corrélation négative, plus la pluie est importante, plus le nombre de locations est faible et inversement. On remarque également que le nombre maximum de locations en un jour par temps sec se situe autour de 22 000, bien plus élevé que le nombre maximum de locations en un jour par temps de pluie qui est d'environ 9000. On peut donc bien affirmer que la pluie influence également le nombre de locations à la baisse.



On peut également ajouter grâce au graphique ci-dessus, présentant le nombre de trottinettes louées avec la représentation des jours de pluies de différentes intensités, que ces dernières amplifient notre hypothèse : plus une pluie est dense, moins le nombre de trottinettes est élevé et inversement.

- Bilan général pluie et neige :



Ce graphique représente bien l'influence de la pluie et de la neige sur la baisse générale du nombre de locations de trottinettes sur les jours concernés.

5. CONCLUSION :

La taille de notre échantillon était de 983648 données lors de la dernière actualisation, et se situe entre le 1er décembre 2017 jusqu'au 5 mai 2018.

Nous pouvons conclure que les paramètres météorologique influent sur la location de trottinettes