



Etude de cas

Sujet : Velib'

Dotsu Olympe, Mellak Maysaraa
UNIVERSITE BRETAGNE SUD

Table des matières

Introduction	3
Méthodologie	3
Description de la démarche	3
Outils et technologies utilisés.....	4
Justification des choix méthodologiques	4
Développement / Réalisation	5
Exploration du sujet et des données	5
Analyse exploratoire	5
Analyse des régulations existantes	6
Modélisation statistiques	6
Difficultés rencontrés	6
Solutions mises en place	7
Résultats et Analyse	7
Présentation des résultats obtenus	7
Analyse temporelle.....	7
Analyse des déséquilibres d'arrivées et de départs.....	8
Analyse de la durée moyenne des trajets par type de vélo	8
Analyse de la distance moyenne parcourue par type de vélo	8
Analyse du remplissage des stations	8
Classification	9
Analyse de la régulation réelle	9
Analyse de la flotte et des ateliers.....	9
Modélisation supervisée : prédiction de l'état des stations.....	9
Analyse critique	9
Comparaison avec les objectifs de départ	10
Conclusion et Perspectives	10
Limites du projet	11
Pistes d'amélioration ou suites possibles	11
Planification dynamique des tournées :	11
Anticipation des déséquilibres	11
Intégration géographique et zonage multi-indicateurs	11
En combinant les données de signalements, d'indisponibilités et de flux anormaux sur des cartes interactives, les régulations seraient optimisées, ce qui réduirait les inégalités d'accès au service.....	12
Annexes.....	12

Données utilisées	12
Extraits de code Python utilisés	12
Tableau des résultats	21

Introduction

La région parisienne compte plus de 12 millions d'habitants, ce qui explique pourquoi le transport parisien est l'un des plus importants au monde. Ces dernières années, un nouveau type de transport s'est imposé : le vélo en libre-service, dont Vélib' est un acteur majeur de ce milieu. Utilisé comme simple moyen de transport ou en combinaison avec d'autres, ce service a permis à de nombreux usagers de se déplacer plus rapidement et de manière écologique.

Cependant, cette popularisation tend à engendrer un manque de fluidité du fonctionnement, tel que l'encombrement des stations principales et la désertion d'autres. Dès lors, une question centrale se pose :

Comment mieux réguler la répartition des vélos dans les stations Vélib' de la région parisienne, afin de garantir une disponibilité équilibrée du service sur l'ensemble du territoire ?

Tout au long de cette étude, nous nous efforcerons de trouver des solutions afin d'optimiser la distribution des vélos à travers plusieurs objectifs.

Analyser les données mises à disposition (cours, disponibilité des stations, régulations, signalements...)

- Mettre en évidence les tendances et déséquilibres spatio-temporels les plus récurrents.
- Comprendre si les interventions actuelles permettent réellement de corriger les déséquilibres.
- Proposer des stratégies de régulation, fondées sur l'analyse des données :

Pour atteindre nos objectifs, nous commencerons par explorer en profondeur le sujet et les données associées. Nous réaliserons par la suite une analyse exploratoire permettant de mieux comprendre les habitudes d'utilisation du service Vélib'. Nous analyserons ensuite les régulations actuellement mises en place, afin d'évaluer leur efficacité. Ces analyses nous permettront d'identifier des points d'amélioration concrets. Nous pourrions alors les développer à l'aide de modèles statistiques en vue de proposer des stratégies de régulation optimisées.

Cette étude a donc pour but l'amélioration du fonctionnement d'un service de transport en métropole en mêlant analyse et compréhension des données. Grâce à notre démarche, nous explorerons les causes de ce déséquilibre et construirons des solutions basées sur un modèle statistique.

Méthodologie

Tout au long de cette étude, nous avons appliqué une démarche méthodologique, visant à analyser les données et proposer des solutions concrètes.

Description de la démarche

La première étape est la préparation de notre jeu de données. Les données prises sur deux périodes (février et juin 2023) seront fusionnées puis nettoyées. De plus, plusieurs variables

(vitesse moyenne, durée, heure de départ/arrivée...) ont été créées ou simplifiées pour plus de praticité.

Ensuite, une analyse exploratoire sera menée afin d'observer les différents comportements d'utilisation sur plusieurs échelles (temporelles et géographiques) à partir de différentes variables. Cette analyse nous permettra d'identifier les principaux déséquilibres associés aux pics d'utilisation entre certaines zones.

Dans un troisième temps, nous réaliserons une classification en nous appuyant sur des variables comme le nombre de départs/arrivées ou encore le taux de saturation afin de regrouper les stations par profil. Cela nous permettra d'identifier des typologies de stations (stations très sollicitées, sous-utilisées, structurellement déséquilibrées, etc.).

Enfin, selon les résultats obtenus, nous choisirons d'appliquer certains modèles statistiques. En effet, on cherchera à anticiper les déséquilibres futurs en fonction du temps et de l'espace.

Outils et technologies utilisés

L'ensemble de l'étude a été réalisé en langage Python, plus précisément dans un environnement Jupyter Notebook. Nous avons ainsi mobilisé plusieurs bibliothèques :

- **NumPy** : permet de manipuler des calculs numériques, des listes de nombres, mais aussi la création d'ensembles de données.
- **Pandas** : afin de trier, filtrer, agréger et transformer des données en quelques lignes de code, tout en intégrant des méthodes statistiques pour générer des insights.
- **Matplotlib, Seaborn** : pour les visualisations statistiques.
- **Scikit-learn** : pour le prétraitement des données et les modèles de machine learning (clustering, régression...).
- **Plotly** : pour des visualisations interactives et géographiques.
- **Datetime** : pour l'analyse temporelle.

Justification des choix méthodologiques

L'analyse exploratoire permet d'identifier les premiers déséquilibres et tendances d'utilisation des vélos en libre-service. Elle constitue donc la base de l'étude.

Les représentations spatio-temporelles (heatmaps, graphiques par heure/jour, cartes interactives) sont particulièrement adaptées au sujet étant directement liées à la problématique du déséquilibre. Elles offrent ainsi, pour une première approche, une facilité d'analyse.

L'utilisation de la classification K-Means permet de regrouper les stations selon leurs caractéristiques et de mettre en place une stratégie de régulation selon le profil de la station. La modélisation statistique vise à anticiper les déséquilibres futurs et à proposer des actions stratégiques afin d'améliorer les régulations mises en place.

Développement / Réalisation

Exploration du sujet et des données

Dans un premier temps, nous nous sommes intéressés au contenu de chaque fichier de données et à leur utilité :

- **Courses usagers** : ce fichier contient toutes les courses effectuées par les usagers sur deux périodes (février et juin 2023). Il permet de repérer les stations les plus sollicitées ainsi que les différents flux.
- **Disponibilité** : ce fichier présente la mise à jour en continu de l'état des stations (chaque minute ou à chaque changement). On peut ainsi identifier les périodes de saturation ou de vide.
- **Régulation mouvements entre stations** : ce fichier recense l'historique des interventions de régulation (transfert de vélos d'une station à une autre). Il aide à mesurer l'impact direct des actions de maintenance et à identifier les stations les plus sensibles au vandalisme et à la panne.
- **Signalements de vélos abandonnés** : ce fichier liste les vélos laissés hors station. Ces données sont alors précieuses et aident à mieux planifier les interventions.
- **Flotte de véhicules disponibles** : ce fichier indique la quantité de véhicules assurant la régulation et la maintenance du réseau. Ces informations sont importantes, car elles joueront un rôle majeur dans la mise en œuvre réelle des solutions de régulation.

À la suite de cela, nous avons préparé notre jeu de données. Issues de deux périodes (février et juin 2023), les données ont d'abord été fusionnées puis nettoyées. Certaines modifications de variables ont ensuite été effectuées. En effet, les colonnes de dates (`date_départ`, `date_arrivée`) ont été converties au format `datetime`. Puis, pour plus de précision, nous avons également ajouté de nouvelles variables telles que la durée en minutes, la vitesse moyenne, le jour de la semaine, l'heure de départ/arrivée, ou encore le statut de la station (`pleine`, `vide`, `normale`).

Analyse exploratoire

Dans un second temps, nous avons mené une analyse exploratoire afin de décrire les comportements d'utilisation du service Vélib'.

Nous nous sommes tout d'abord intéressés à la distribution de plusieurs variables, notamment la durée des trajets, la distance parcourue et la vitesse moyenne. Nous avons également observé les différents flux (trajets) sur différentes échelles temporelles, comme le nombre de départs et arrivées par heure tout au long de la semaine. Ces observations ont permis de révéler des pics d'utilisation.

Nous voulions aussi identifier les trajets les plus fréquents, c'est pourquoi l'utilisation de cartes interactives et de heatmaps s'est révélée particulièrement utile. Nous avons ainsi repéré les stations les plus sollicitées et les zones les plus actives.

Enfin, afin de vérifier les déséquilibres entre la périphérie et le centre de Paris, nous avons analysé le statut des stations (`pleine`, `vide` ou `normale`) lors des départs et arrivées des trajets. Cette étape a donc confirmé certaines hypothèses posées dans le sujet : le déséquilibre entre périphérie et centre-ville aux heures de pointe, et la saturation de certaines stations à forte demande.

Analyse des régulations existantes

Dans cette partie, nous avons étudié les données de régulation à travers deux axes : les mouvements de régulation au niveau des stations et au niveau des ateliers.

Nous avons tout d'abord analysé les mouvements de régulation des stations. On s'intéresse donc ici au rééquilibrage du nombre de vélos entre les stations. Grâce aux données sur les stations d'origine, les stations de dépôt et le nombre de vélos transportés, nous avons visualisé les mouvements de régulation les plus fréquents. Cela nous a permis de relever un déséquilibre marqué entre les stations les plus souvent réapprovisionnées ou vidées. Afin de vérifier si ce déséquilibre est stratégique, nous avons décidé de croiser les besoins réels (les stations les plus utilisées, régulièrement vides ou saturées) avec les interventions effectuées par les équipes de régulation.

Dans un second temps, nous avons étudié les régulations liées aux ateliers. On distingue deux actions principales : le retrait des vélos en panne et la réinjection des vélos réparés. Nous avons ainsi comparé les stations les plus souvent vidées vers les ateliers et celles recevant le plus de vélos réparés. Cela nous a permis de visualiser les zones les plus touchées par l'usure ou le vandalisme, mais aussi de repérer les stations bénéficiant le plus d'injections. Enfin, afin de mieux comprendre les besoins en maintenance, nous avons comparé le nombre de vélos mécaniques (VM) et de vélos à assistance électrique (VAE) envoyés en atelier. Cette phase a donc mis en évidence le rôle essentiel de la maintenance dans l'équilibre du réseau.

Modélisation statistiques

Pour mieux comprendre le fonctionnement des stations Vélib' et anticiper les déséquilibres, plusieurs modèles statistiques ont été utilisés. Tout d'abord, un modèle de classification supervisée appelé Random Forest a été mis en place. Ce modèle permet de prédire l'état d'une station (pleine, vide, normale) en fonction de plusieurs variables comme l'heure, la capacité, et le nombre de vélos disponibles. Cette prédiction aide à anticiper les besoins en régulation.

Ensuite, un modèle de clustering, appelé KMeans, a été utilisé pour regrouper les stations selon leur profil. Cela signifie que les stations ont été classées en plusieurs groupes selon le nombre de vélos mécaniques et électriques disponibles. Ce regroupement permet de repérer les stations qui ont un profil similaire d'utilisation.

Pour aller plus loin, un croisement entre les groupes (clusters) identifiés et le taux de remplissage des stations a été réalisé. Cette analyse a permis de détecter certaines stations qui, malgré un fort potentiel d'utilisation, restent mal régulées. Ces stations devraient être ciblées en priorité pour améliorer l'efficacité du service.

Difficultés rencontrés

Pendant le projet, plusieurs difficultés ont été rencontrées. Tout d'abord, les noms de certaines stations étaient écrits de différentes manières, ce qui compliquait les analyses. Un nettoyage des données a donc été nécessaire pour uniformiser les noms.

Ensuite, le volume de données à traiter était très important, ce qui rendait certaines opérations, comme les jointures ou les regroupements, très longues à exécuter. Des optimisations ont dû être mises en place pour accélérer ces traitements.

Enfin, certaines stations manquaient de données complètes, notamment sur les signalements ou les états de remplissage. Cela a nécessité une analyse plus prudente pour ne pas fausser les résultats. De plus, il manquait des informations sur la logique des tournées de régulation, ce qui compliquait l'interprétation des mouvements de vélos.

Solutions mises en place

Pour répondre à ces difficultés, plusieurs solutions ont été adoptées. Tout d'abord, de nouvelles colonnes appelées `station_clean` ont été créées pour fiabiliser les croisements de données. Ensuite, des visualisations ont été réalisées pour mieux cibler les données et repérer les anomalies.

L'analyse s'est concentrée sur les indicateurs les plus fiables, comme le taux de remplissage, les ratios entre les flux et le nombre de vélos, ainsi que les regroupements (clustering) pour identifier les profils de stations.

Résultats et Analyse

Présentation des résultats obtenus

Analyse temporelle

L'analyse temporelle des données, réalisée à partir des heatmaps des mois de février et juin (voir Annexe 5 et 6), met en évidence une tendance horaire. On observe deux pics d'utilisation quotidiens : un premier le matin entre 8h et 10h et un second en fin de journée, entre 17h et 19h correspondant aux heures de pointe. De plus on remarque qu'en février ces pics sont plus marqués. En revanche en juin (une période associée aux vacances scolaires) l'activité est plus répartie. Malgré cela, le taux d'utilisation est plus élevé en juin. Par exemple, à 18h, on compte environ 400 500 utilisations, contre un peu plus de 300 000 en février.

En ce qui concerne la répartition par jour de la semaine, les diagrammes (Annexe 1,2,3,4) montrent une tendance assez régulière durant les jours de semaine en février, avec une fréquentation stable du mardi au vendredi, et une baisse notable le lundi et le week-end. Ce qui peut être probablement liée à une réduction des déplacements domicile-travail. En juin les jeudis et vendredis deviennent les jours les plus fréquentés. Enfin, on note une hausse générale du nombre de trajets en juin, avec un pic atteignant près de 800 000 trajets, contre environ 500 000 en février, ce qui confirme l'impact des saisons sur la demande du service Vélib'.

L'analyse montre donc des pics d'utilisation aux heures de pointe (8h–10h, 17h–19h) et une fréquentation plus élevée les jeudis et vendredis de juin. On note aussi que l'usage de Velib' est plus intense en été.

Analyse des déséquilibres d'arrivées et de départs

On observe grâce aux graphiques (Annexe 7 et 8) que certaines stations sont régulièrement en surplus d'arrivé ou en surplus de départs. En effet en février, les surplus d'arrivées se concentrent autour de Gambetta ou encore Belleville, tandis que les départs sont plus vers Enghien et la Gare de Lyon. Cela s'explique donc par le fait que les utilisateurs quittent principalement les arrondissements périphériques et les communes limitrophes pour rejoindre le centre.

En juin, ces déséquilibres s'amplifient avec plus de 800 vélos de surplus à Gambetta et près de 4000 vélos à la Gare de Lyon. Cette augmentation est donc causée par la haute saison.

La typologie des stations (Annexe 13 et 14) renforce les conclusions de l'analyse des déséquilibres observés. Cette correspondance entre les profils de stations et les flux montre l'importance d'une régulation entre ces zones.

Analyse de la durée moyenne des trajets par type de vélo

L'analyse des graphiques de durée moyenne par heure pour les vélos mécaniques et électriques (Annexes 9 et 10) montre qu'en février, les vélos électriques présentent une durée moyenne plus élevée, en particulier pendant les premières heures de la journée (environ 35 minutes vers 1h du matin). Cette différence peut s'expliquer par la simplicité d'utilisation offerte par l'assistance électrique, ainsi que par les trajets de retour effectués en dehors des horaires des transports en commun. En juin, les durées moyennes sont plus homogènes entre les deux types de vélos. La météo peut ainsi jouer un rôle dans cette tendance.

Analyse de la distance moyenne parcourue par type de vélo

L'analyse de la distance moyenne parcourue par type de vélo (Annexes 11 et 12) montre également que les vélos électriques enregistrent régulièrement des distances plus longues (plus de 4 000 mètres) que les vélos mécaniques (environ 3 000 mètres). En juin, cette tendance reste inchangée. L'assistance électrique est donc privilégiée par les utilisateurs pour les trajets de plus longue distance.

Ces observations suggèrent qu'il serait pertinent d'augmenter la disponibilité des VAE dans les zones périphériques ou faiblement desservies, et celle des vélos mécaniques dans les zones centrales où les trajets courts sont majoritaires.

Analyse du remplissage des stations

L'analyse du remplissage des stations montre que en juin, 37 % des stations sont fréquemment vides, tandis que seulement 13 % sont souvent pleines, révélant une forte demande logistique pendant la période estivale. Ce déséquilibre indique que le système n'arrive pas à réguler efficacement les flux de vélos entre les zones à forte demande et celles à faible demande. De plus les stations critiques (systématiquement pleines ou vides) restent les mêmes entre février

et juin. Cette situation montre un besoin urgent de régulation plus efficace dans ces zones pour éviter les stations saturées ou vides, surtout lors des pics d'activité.

Classification

L'utilisation de la classification KMeans a permis de classer les stations en quatre profils distincts : petites stations peu utilisées, stations équilibrées, stations surchargées en vélos électriques et stations mécaniques à fort volume. Cette segmentation a révélé l'existence de 230 stations à fort potentiel mais mal desservies, représentant des cibles idéales pour une régulation prioritaire. En effet, ces stations concentrent une demande importante sans être suffisamment approvisionnées en vélos, ce qui génère un déséquilibre constant. En adaptant les stratégies de régulation à chaque profil, il serait possible de maximiser l'efficacité du service tout en réduisant les problèmes d'accès aux vélos dans les zones critiques.

Analyse de la régulation réelle

Concernant l'analyse de la régulation réelle, il apparaît que les tournées de réapprovisionnement sont souvent planifiées à l'avance. De plus, les VAE, bien que très sollicités, sont moins régulés que les vélos mécaniques : moins de 25 % des mouvements concernent ces vélos électriques. Cela crée une pression supplémentaire dans les zones où les VAE sont en forte demande. Pour remédier à cette situation, il serait nécessaire d'améliorer les algorithmes de planification pour intégrer des données en temps réel, afin de réguler les stations en fonction de l'évolution de la demande.

Analyse de la flotte et des ateliers

L'analyse de la flotte et des ateliers a mis en évidence une surcharge logistique en juin, où le volume de vélos déplacés dépasse largement les capacités prévues. Les véhicules de régulation, qui doivent à la fois gérer les transferts, les réparations et les interventions techniques, ne peuvent donc pas être efficaces. De plus, le nombre de stations critiques reste stable entre février et juin, dépassant les 1 400 stations. On en déduit alors que les déséquilibres sont durables et pas seulement saisonniers.

La distinction des véhicules dédiés à la régulation et de ceux affectés aux interventions techniques, permettrait de mieux répondre aux besoins des utilisateurs.

Modélisation supervisée : prédiction de l'état des stations

La modélisation supervisée par le modèle de Random Forest a permis de prédire l'état d'une station (pleine, vide, normale) en fonction de variables telles que l'heure, la capacité, et le nombre de vélos disponibles. Le modèle affiche un score F1 global supérieur à 0,90, ce qui prouve que les états des stations sont prévisibles. En intégrant ce modèle dans un système de gestion dynamique, il serait possible de réagir en temps réel pour redistribuer les vélos en fonction des flux prédits et ainsi réduire les déséquilibres.

Analyse critique

Le système Vélib' montre des signes de saturation en période de forte demande, particulièrement en juin avec des déséquilibres entre les stations de départ et celles de

destination. L'analyse menée montre clairement que les stations situées dans les zones périphériques sont souvent vides, tandis que celles du centre affichent un excès de départs.

L'étude met également en lumière un décalage important entre la demande réelle et les opérations logistiques. Les régulations actuelles ne ciblent pas toujours les stations les plus critiques, et les tournées de réapprovisionnement, planifiées à l'avance, manquent de réactivité face aux déséquilibres en temps réel.

La mise en place de la modélisation statistique et la classification offrent des solutions opérationnels simples et exploitables pour anticiper et corriger ces problèmes.

Comparaison avec les objectifs de départ

Objectif initial	État
Identifier les pics et flux critiques	Réalisé (pics horaires, typologies de stations)
Détecter les stations déséquilibrées	Réalisé via ratios, clustering, remplissage
Évaluer la régulation actuelle	Réalisé (constat de décalage avec l'usage)
Proposer une modélisation statistique	Random Forest + KMeans
Recommander des solutions opérationnelles	Formulées (ciblage dynamique, types de vélos, flotte)

Conclusion et Perspectives

Ce projet a permis d'analyser en profondeur les données d'utilisation du service Vélib', en comparant deux périodes marquantes : février, où l'affluence est modérée, et juin, où elle est bien plus importante. L'analyse a révélé des déséquilibres dans la répartition des vélos, avec des stations souvent vides en périphérie et saturées dans les zones centrales. Les opérations de régulation actuelles ne ciblent pas toujours les stations les plus critiques, ce qui aggrave ces déséquilibres. Cependant, les modèles prédictifs développés permettent d'anticiper l'état des stations, offrant de meilleures régulations. L'étude a également démontré que certaines

stations, pourtant stratégiques, restent sous-desservies, ce qui diminue l'efficacité globale du service.

Limites du projet

L'analyse a révélé plusieurs limites :

les données de régulation manquent de détails, notamment sur le nombre exact de vélos déplacés lors de chaque intervention. De plus, les modèles de prédiction ne prennent pas encore en compte des facteurs externes comme la météo, les événements urbains ou les grèves, qui peuvent influencer fortement l'utilisation des vélos. Enfin, les modèles utilisés, bien qu'interprétables, restent relativement simples et pourraient être optimisés pour plus de précision.

Pistes d'amélioration ou suites possibles

Pour améliorer la régulation du système Vélib', plusieurs axes d'optimisation sont envisageables :

Niveau d'importance par station :

Il serait pertinent de mettre en place un niveau d'importance pour chaque station, calculé à partir de plusieurs indicateurs comme le ratio entre les départs et les vélos disponibles, le nombre de signalements, les vélos indisponibles ou encore l'écart entre les collectes et les réinjections vers les ateliers. Cet indicateur permettrait de prioriser les stations à traiter en urgence.

Planification dynamique des tournées :

Actuellement, les tournées de régulation sont peu réactives. Chaque matin un classement des stations prioritaires serait généré en fonction de leur score. Cela permettrait d'optimiser les trajets des camions de régulation en tenant compte de leur capacité, de la proximité géographique des stations et du nombre de vélos nécessaires. Ce système améliorerait l'efficacité logistique en intervenant sur les stations avant qu'elles ne soient saturées ou vides.

Anticipation des déséquilibres

À partir des données historiques de flux, il serait possible de prédire les stations qui risquent d'être vides ou pleines selon l'heure. Il serait alors envisageable de déclencher des régulations préventives, ce qui éviterait les surcharges par exemple.

Intégration géographique et zonage multi-indicateurs

En combinant les données de signalements, d'indisponibilités et de flux anormaux sur des cartes interactives, les régulations seraient optimisées, ce qui réduirait les inégalités d'accès au service

Annexes

Données utilisées

Extraits de code Python utilisés

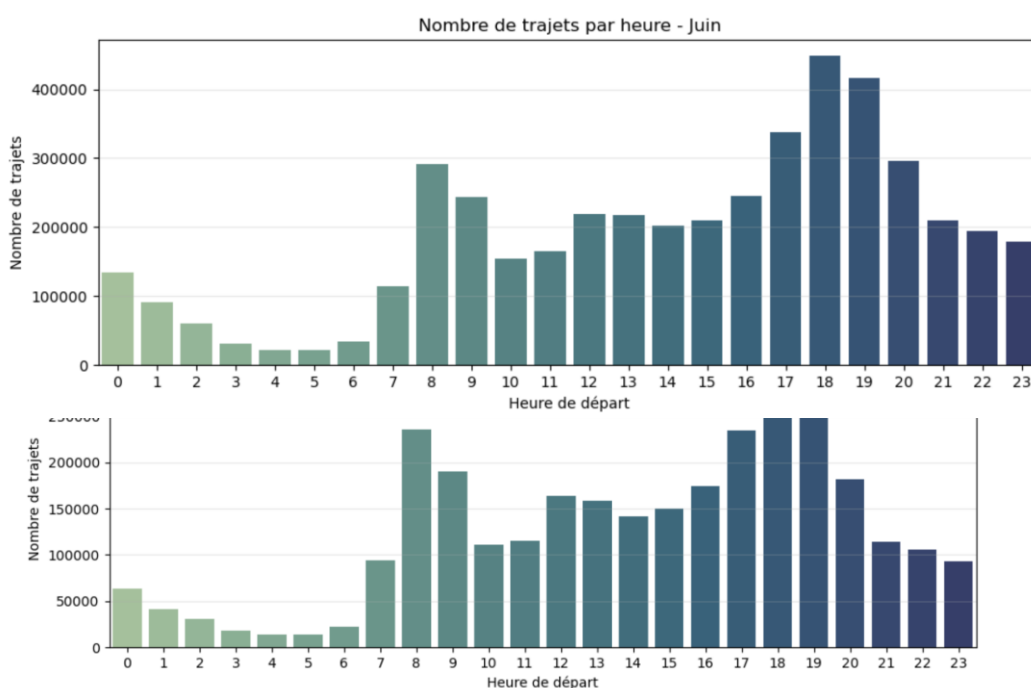
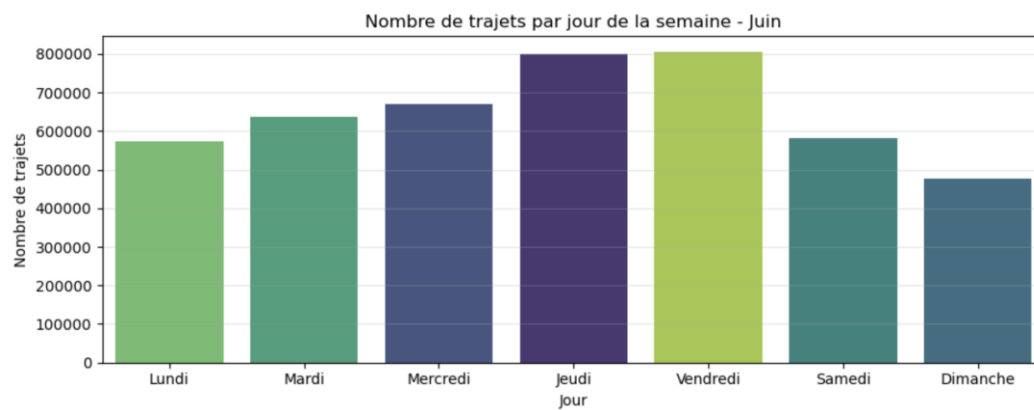
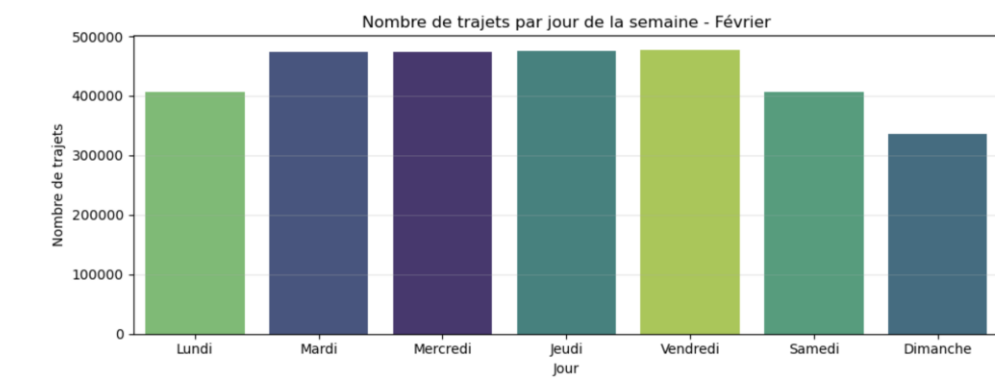
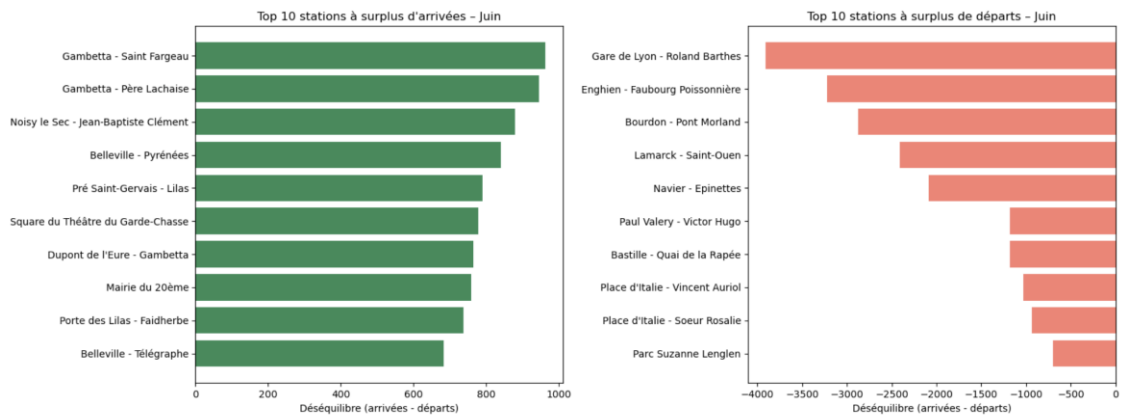
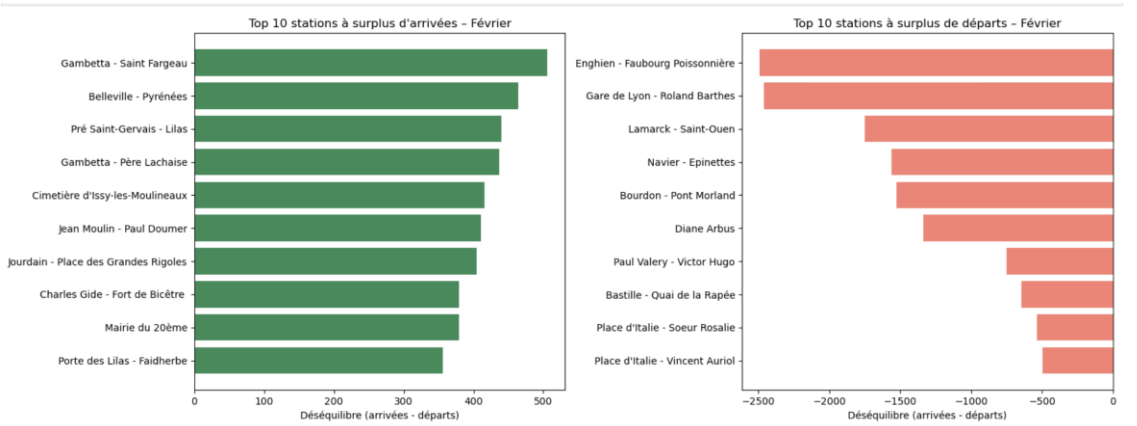
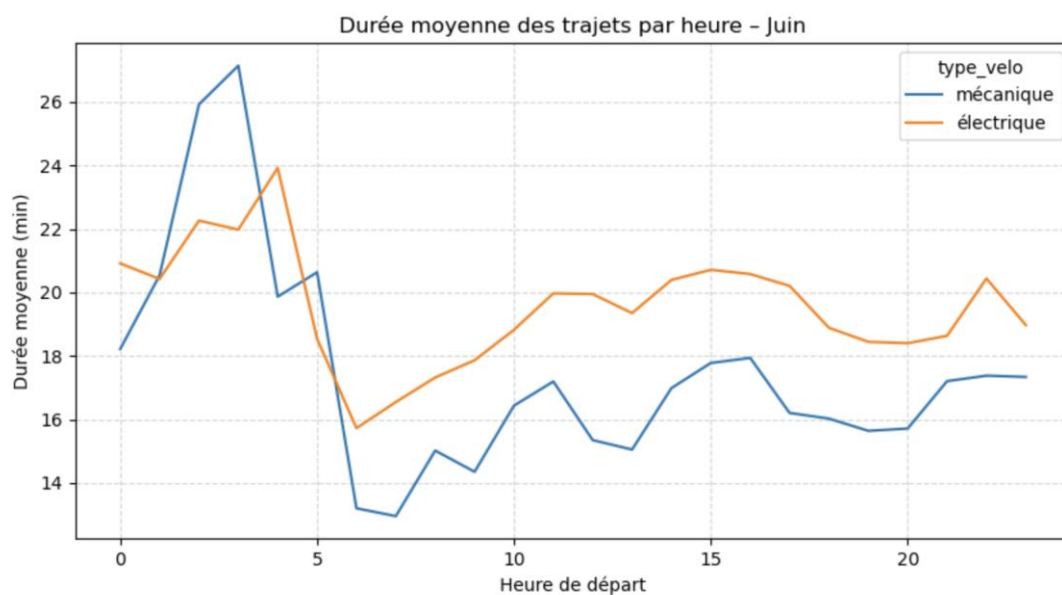
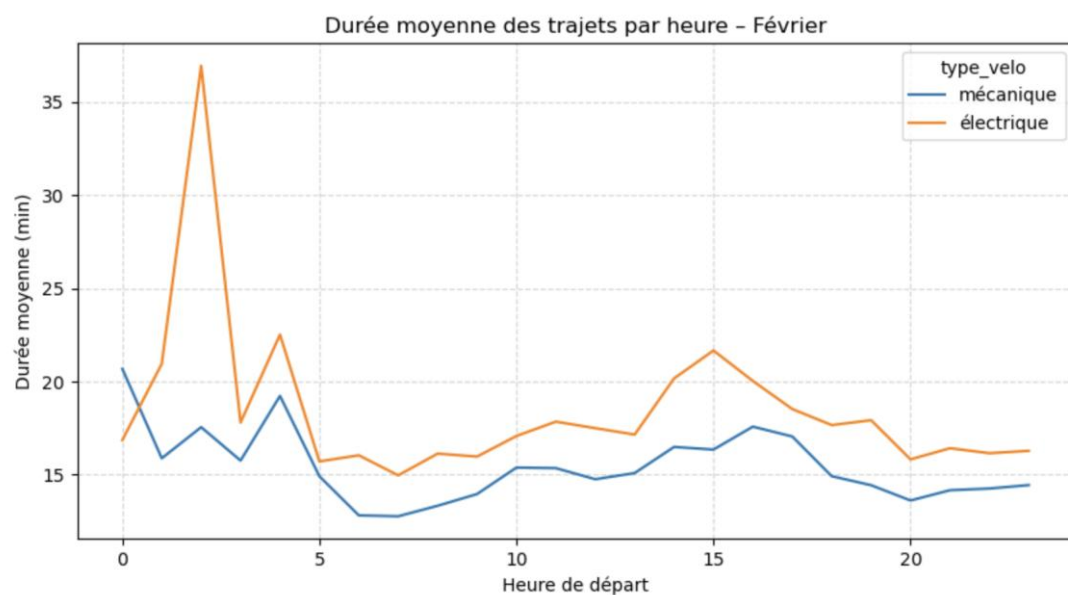
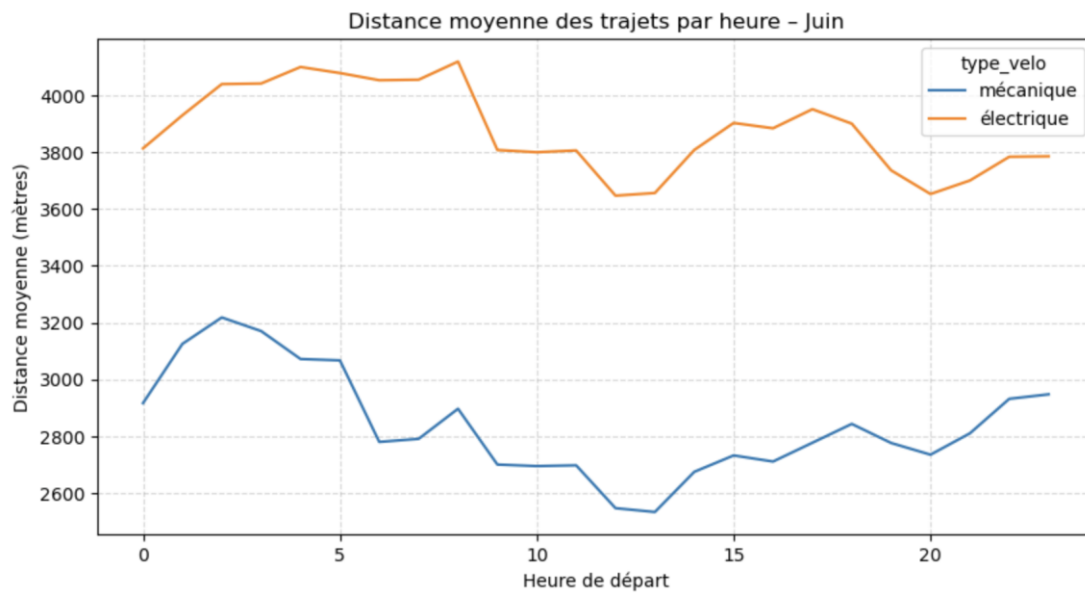
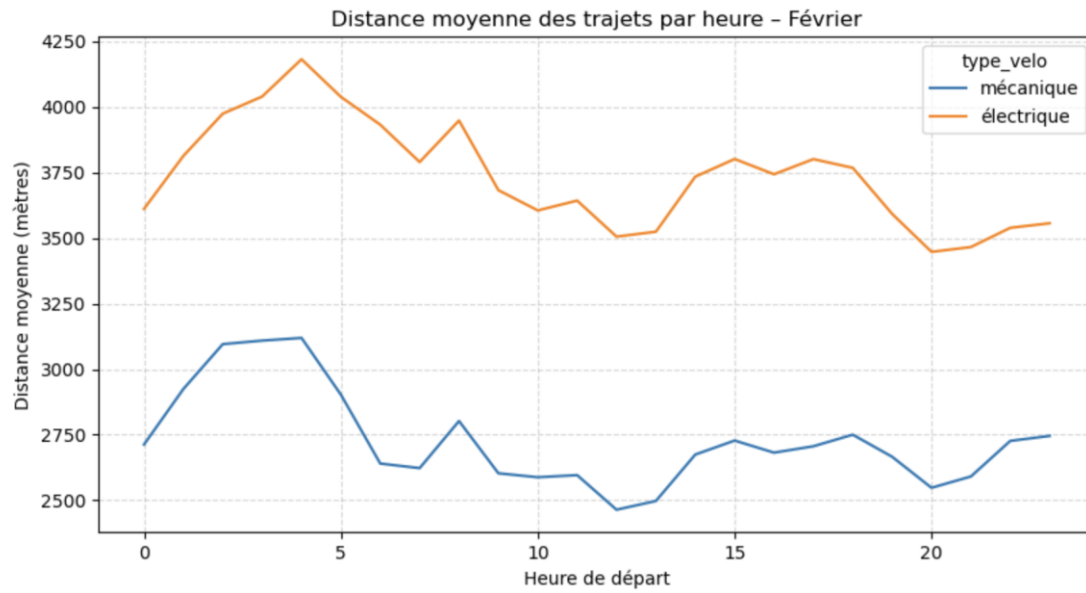


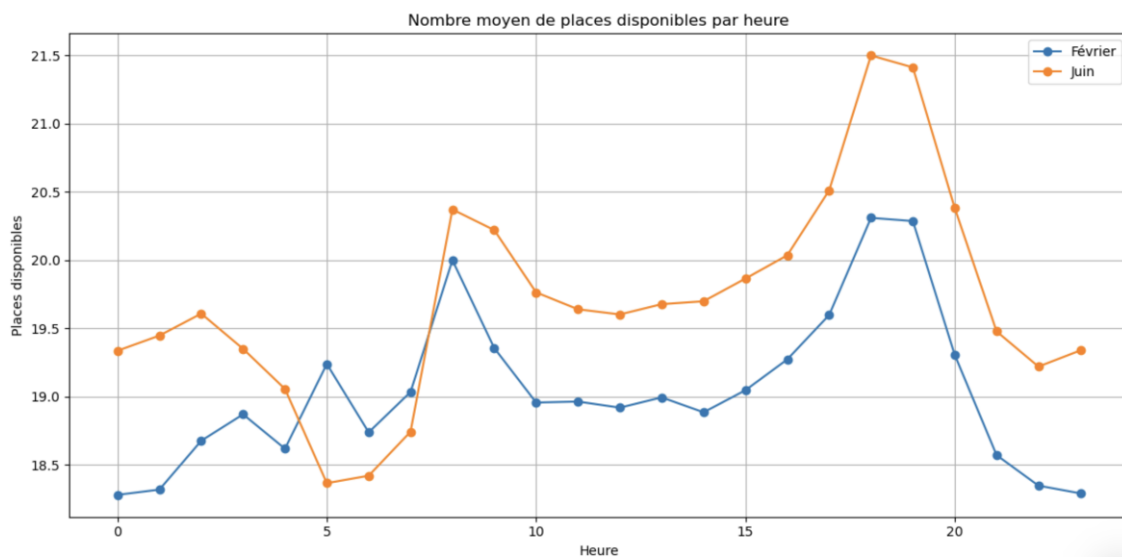
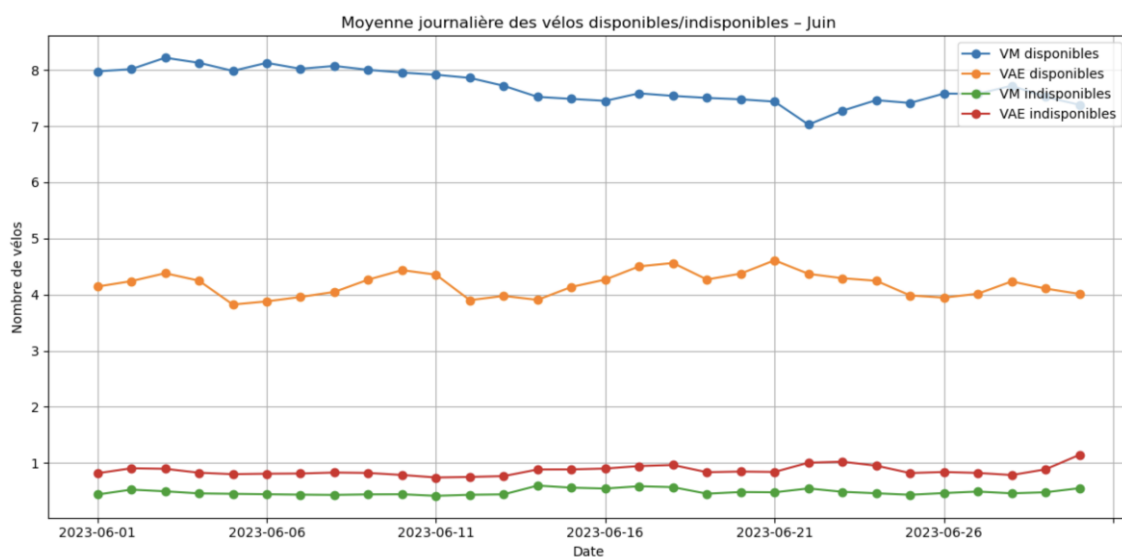
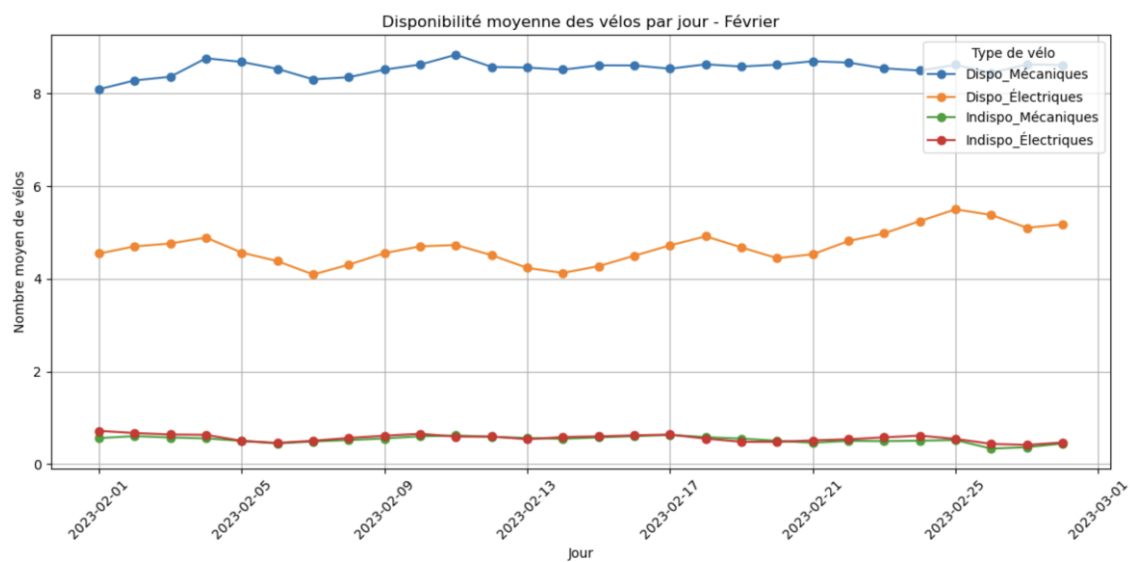
Figure 1 Nombre de trajets par heure -Fevrier

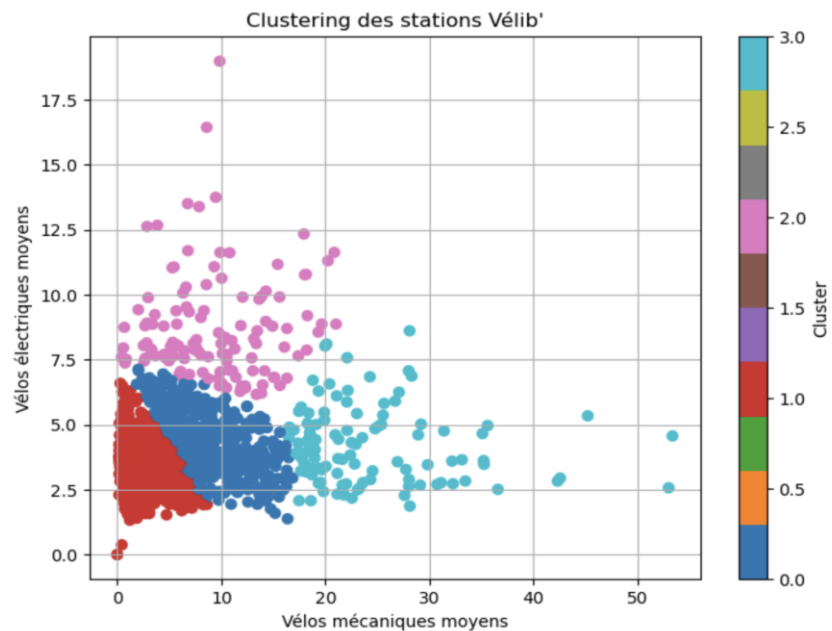
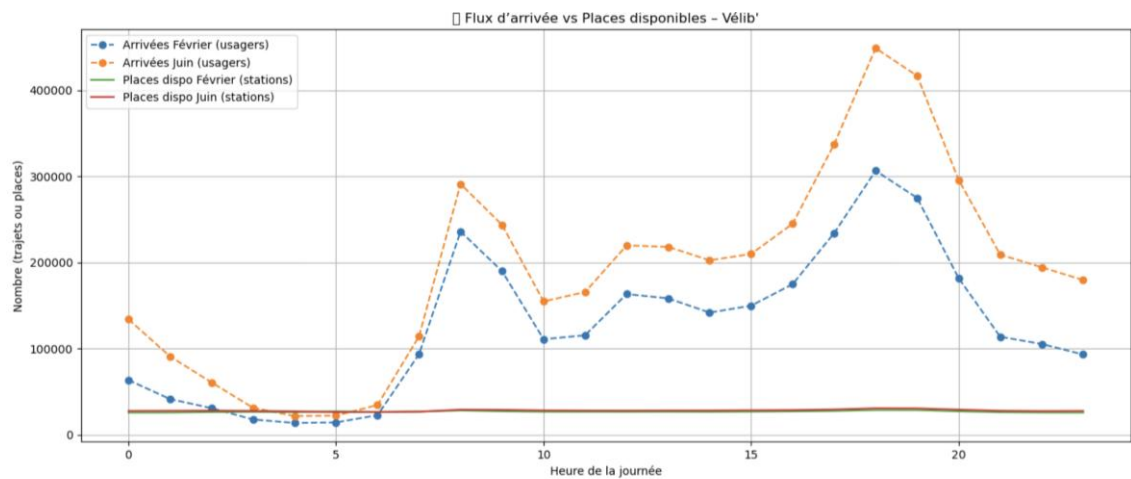
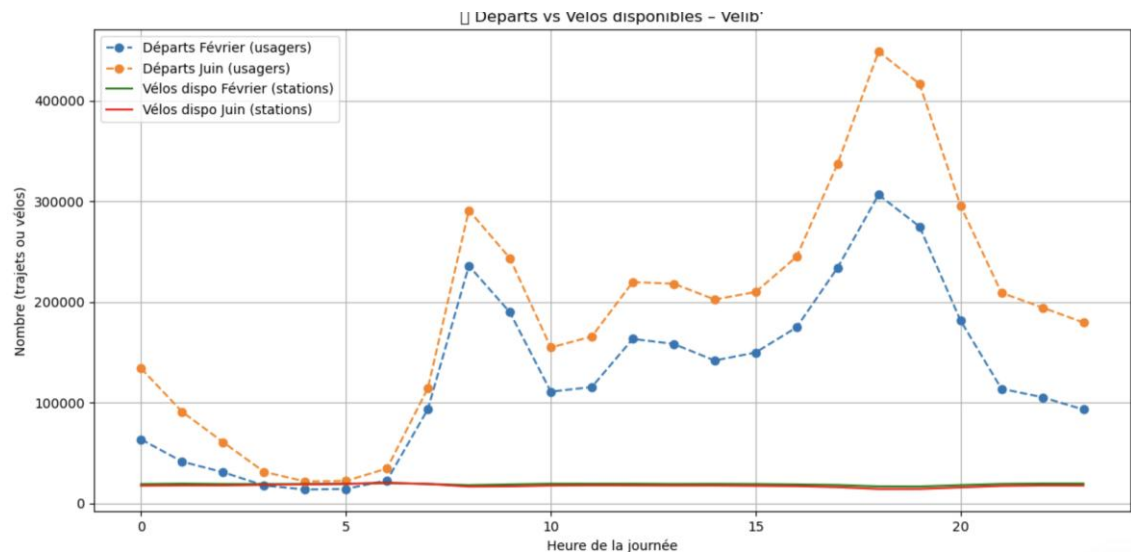




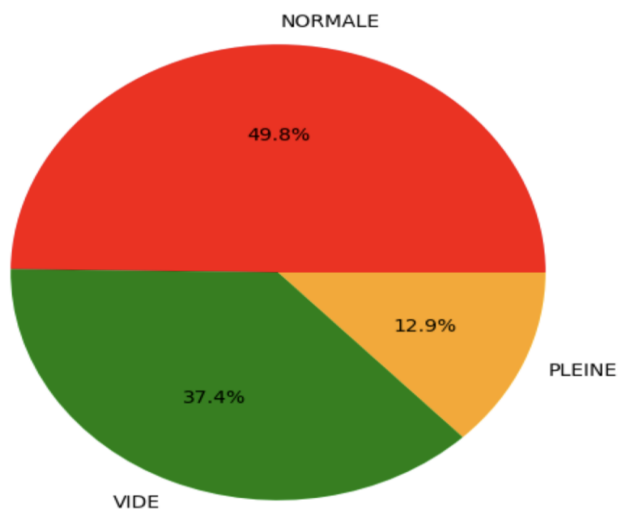




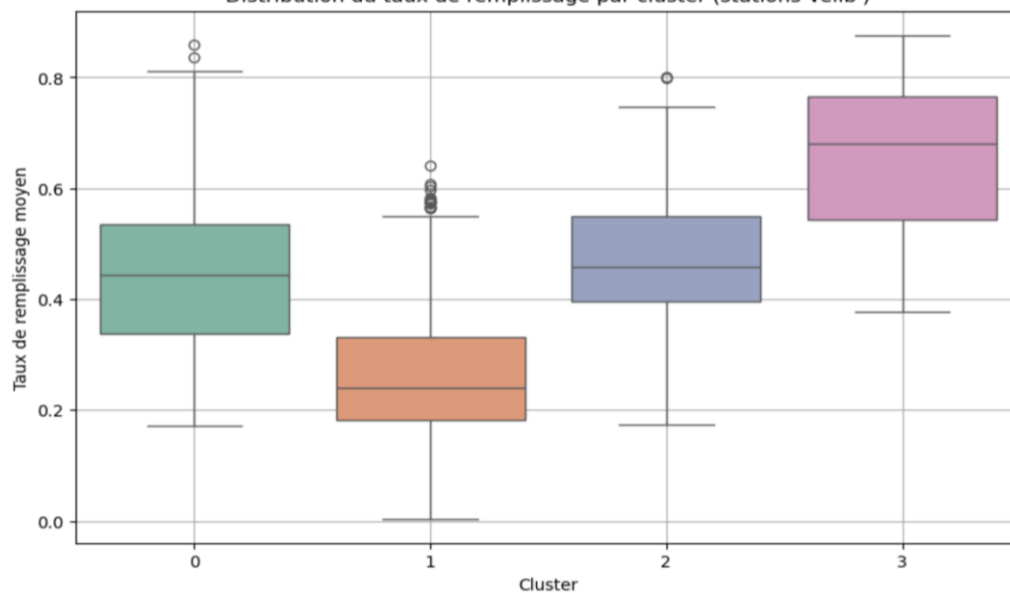




Répartition des statuts de stations juin



Distribution du taux de remplissage par cluster (stations Vélib')



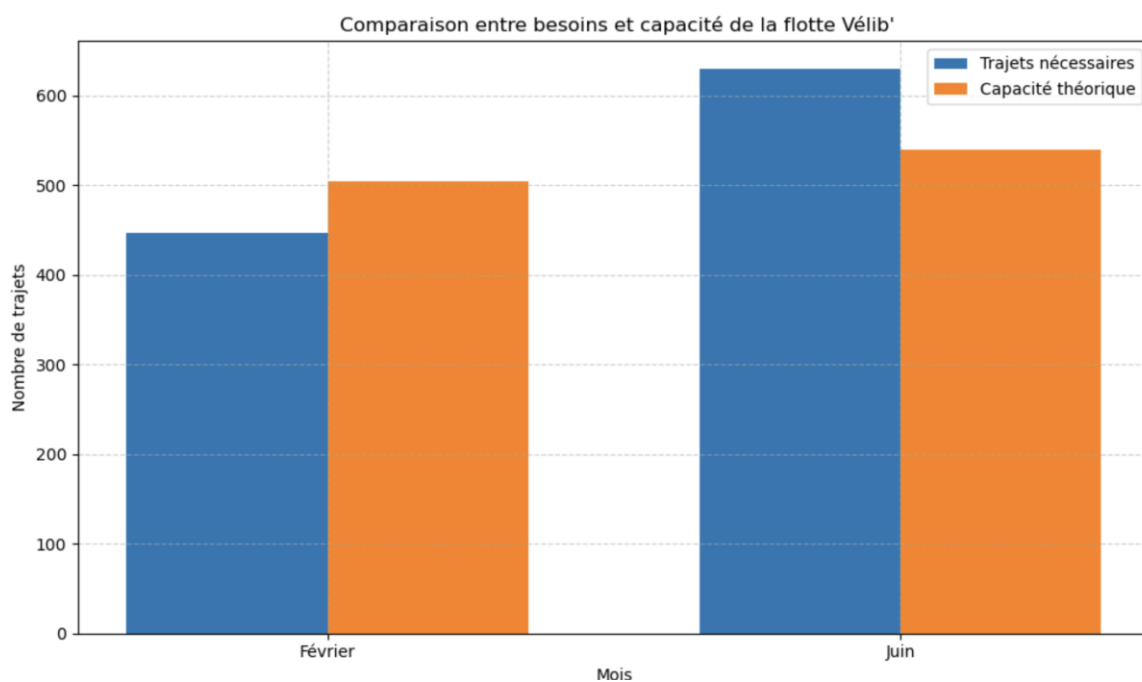


Tableau des résultats

	arrivees_moy	departs_moy	velos_disponibles	places_disponibles	ratio_arrivees_vs_places	ratio_departs_vs_velos
quai des celestins - henri iv	3824.0	3851.0	9.035868	3.638934	1050.857096	426.190358
place de l'hotel de ville	4204.0	4263.0	8.753920	4.168103	1008.612350	486.981830
strasbourg - saint-denis	4445.0	4494.0	9.951470	4.897346	907.634494	451.591565
traversiere - ledru-rollin	7872.0	7839.0	23.819544	8.716633	903.101042	329.099503
saint-antoine sevine	6999.0	6968.0	15.881663	7.954600	879.868205	438.744974
pont neuf - rivoli	6538.0	6639.0	16.498381	8.260679	791.460377	402.403116
centre georges pompidou	6631.0	6646.0	15.679806	8.705876	761.669451	423.857298
lacuee - lyon	3506.0	3486.0	10.258223	4.689576	747.615535	339.824943
chabrol - d'hauteville	3365.0	3353.0	5.084186	4.584681	733.965939	659.495909
turbigo - francaise	3482.0	3531.0	9.942349	4.755065	732.271789	355.147465
place de la republique - temple	5466.0	5525.0	9.295618	7.691909	710.616868	594.366101
petites ecuries - faubourg saint-denis	8089.0	8188.0	13.603444	11.388912	710.252203	601.906403
charenton - diderot	4510.0	4472.0	10.958049	6.393288	705.427329	408.101833
gare saint-lazare - cour du havre	10137.0	10057.0	26.549028	15.093764	671.601856	378.808588
saint-gilles - turenne	3488.0	3454.0	11.076997	5.252511	664.063371	311.817373

Stations souvent VIDÉES en février (prise) :

nom_station_prise	
Westermeyer – Paul Vaillant-Couturier	22
Malesherbes – Place de la Madeleine	21
Bercy – Villot	20
BNF – Bibliothèque Nationale de France	18
Bercy.	14
M.- A. Lagroua Weill-Hallé – F. Dolto	14
Grenelle – Dr Finlay	14
Gare RER – Maurice Grandcoing	13
Parc André Citroën	12
Terroirs de France – Lheureux	12

Name: count, dtype: int64

Stations souvent REMPLIES en février (dépose) :

nom_station_dépose	
Place d'Italie – Soeur Rosalie	34
Gare Montparnasse – René Mouchotte	25
Place d'Italie – Vincent Auriol	24
Porte d'Ivry	20
Pyrénées – Charles Renouvier	17
Maine – Jean Zay	17
Odessa – Départ	15
René Coty – Place Denfert-Rochereau	14
Bobillot – Paulin Mery	14
Edgar Quinet – Gaité	14

Name: count, dtype: int64

Stations VIDÉES qui sont critiques : []

Stations REMPLIES qui sont critiques : []

stations avec plus de régulations Février :

nom_station_prise	nom_station_dépose	
Laffitte – Italiens	Caulaincourt – Place Constantin Pecqueur	45
BNF – Bibliothèque Nationale de France	Porte d'Ivry	36
Westermeyer – Paul Vaillant-Couturier	Porte de Charenton	36
Terroirs de France – Lheureux	Place d'Italie – Soeur Rosalie	31
Jussieu – Guy de la Brosse	Epée de Bois – Mouffetard	30
Bercy – Villot	Place d'Italie – Soeur Rosalie	29
Rond-Point Rhin et Danube	Mozart – Jasmin	28
Bercy.	Pyrénées – Charles Renouvier	28
Malesherbes – Place de la Madeleine	Gérando – Rochechouart	27
Primo Levi – Frigos	Place d'Italie – Vincent Auriol	26

Name: total, dtype: int64

stations avec plus régulations Juin :

nom_station_prise	nom_station_dépose	
Hippodrome de Longchamp	Place de la Porte de Passy	55
Madeleine Vionnet	Cambrai – Benjamin Constant	55
Hippodrome de Longchamp	Route de Sèvres – Porte de Bagatelle	49
Westermeyer – Paul Vaillant-Couturier	Place d'Italie – Soeur Rosalie	38
	Place d'Italie – Vincent Auriol	36
Quai Jules Guesde – Saint-Simon	Place d'Italie – Vincent Auriol	34
Malesherbes – Place de la Madeleine	Naples – Rome	33
Porte Molitor	Versailles – Claude Terrasse	32
BNF – Bibliothèque Nationale de France	Bobillot – Paulin Mery	31
Victor Cresson – Telles de la Poterie	Bara – Henri Farman	30

Name: total, dtype: int64

	velos_injectes	velos_collectes	velos_indisponibles	solde_net
station_clean				
quai panhard et levassor	119.0	27.0	4.559952	92.0
porte de versailles	0.0	14.0	3.913880	-14.0
quai de la gare - pont de bercy	0.0	10.0	3.794628	-10.0
jardin modele - pierre boudou	0.0	7.0	3.626218	-7.0
place des marseillais	12.0	20.0	3.600431	-8.0
richard lenoir - place de la bastille	8.0	44.0	3.528251	-36.0
gare rer de vincennes	2.0	27.0	3.498802	-25.0
gare saint-lazare - cour du havre	15.0	25.0	3.479340	-10.0
maison de radio - france	0.0	23.0	3.256901	-23.0
froment - breguet	0.0	17.0	3.230581	-17.0
hotel de ville de saint-ouen	121.0	14.0	3.221413	107.0
transvaal - charles de gaulle	9.0	10.0	3.144025	-1.0
traversiere - ledru-rollin	0.0	23.0	3.134468	-23.0
constantine - universite	0.0	17.0	3.133661	-17.0
place de barcelone - mirabeau	0.0	26.0	3.117686	-26.0

	velos_injectes	velos_collectes	velos_indisponibles	solde_net
station_clean				
perrault - place du louvre	0.0	7.0	5.337497	-7.0
place du 8 mai 1945	0.0	0.0	4.815625	0.0
charles tillon - cimetiere communal	0.0	12.0	4.586385	-12.0
hippodrome de longchamp	0.0	2.0	4.570518	-2.0
stade pershing	0.0	2.0	4.414634	-2.0
route de la muette - porte de neuilly	0.0	4.0	3.904204	-4.0
place jean-baptiste clement	0.0	12.0	3.847586	-12.0
sainte-elisabeth - turbigo	0.0	7.0	3.709884	-7.0
theatre des amandiers - palais des sports	0.0	0.0	3.674374	0.0
hotel de ville de saint-ouen	6.0	11.0	3.603454	-5.0
bercy - villot	0.0	8.0	3.494732	-8.0
theophile gautier - francois millet	0.0	8.0	3.459967	-8.0
institut du monde arabe - saint-germain	0.0	16.0	3.439581	-16.0
gare saint-lazare - cour du havre	0.0	19.0	3.357132	-19.0
hotel de ville d'asnieres sur seine	12.0	4.0	3.243066	8.0

--

station_clean	velos_injectes	velos_collectes	velos_indisponibles	solde_net
quai panhard et levassor	119.0	27.0	4.559952	92.0
porte de versailles	0.0	14.0	3.913880	-14.0
quai de la gare - pont de bercy	0.0	10.0	3.794628	-10.0
jardin modele - pierre boudou	0.0	7.0	3.626218	-7.0
place des marseillais	12.0	20.0	3.600431	-8.0
richard lenoir - place de la bastille	8.0	44.0	3.528251	-36.0
gare rer de vincennes	2.0	27.0	3.498802	-25.0
gare saint-lazare - cour du havre	15.0	25.0	3.479340	-10.0
maison de radio - france	0.0	23.0	3.256901	-23.0
froment - breguet	0.0	17.0	3.230581	-17.0
hotel de ville de saint-ouen	121.0	14.0	3.221413	107.0
transvaal - charles de gaulle	9.0	10.0	3.144025	-1.0
traversiere - ledru-rollin	0.0	23.0	3.134468	-23.0
constantine - universite	0.0	17.0	3.133661	-17.0
place de barcelone - mirabeau	0.0	26.0	3.117686	-26.0

	arrivees_moy	departs_moy	velos_disponibles	places_disponibles	ratio_arrivees_vs_places	ratio_departs_vs_velos
quai des celestins - henri iv	7011.0	7015.0	7.011902	5.162782	1357.988762	1000.441847
saint-antoine sevigne	11703.0	11667.0	13.291420	9.632715	1214.922215	877.784326
westermeyer - paul vaillant-couturier	4825.0	4672.0	17.962800	4.205705	1147.251162	260.093087
place de l'hotel de ville	6045.0	6037.0	7.338764	5.329224	1134.311584	822.618072
place de la republique - temple	9540.0	9656.0	7.482542	8.656009	1102.124556	1290.470604
quai de la loire - jean jaures	9051.0	9051.0	5.218318	8.514282	1063.037317	1734.466982
pont neuf - rivoli	10945.0	10970.0	14.319860	11.065177	989.139179	766.068953
chabrol - d'hauteville	4652.0	4662.0	4.942670	4.765730	976.135911	943.214888
traversiere - ledru-rollin	11323.0	11452.0	19.207249	11.658282	971.240847	596.233211
lacuee - lyon	5277.0	5224.0	8.744824	5.794984	910.615118	597.381978
charenton - diderot	6226.0	6244.0	9.553296	7.064109	881.356708	653.596443
turbigo - francaise	5442.0	5480.0	8.441274	6.490489	838.457609	649.191094
strasbourg - saint-denis	6357.0	6449.0	5.887463	7.594321	837.072830	1095.378485
parvis corentin celton	5545.0	6152.0	45.491338	6.725958	824.417843	135.234536
bourse du travail	6130.0	6299.0	7.410077	7.583431	808.341267	850.058687

Stations critiques VIDÉES (régulation prise) :

	station_prise_clean \
478	madeleine vionnet
377	hippodrome de longchamp
381	hippodrome de longchamp
927	westermeyer - paul vaillant-couturier
928	westermeyer - paul vaillant-couturier
770	quai jules guesde - saint-simon
512	malesherbes - place de la madeleine
762	porte molitor
47	bnf - bibliotheque nationale de france
36	berthier - porte de clichy

	station_depose_clean	total	vae	vm
478	cambrai - benjamin constant	55	28	27
377	place de la porte de passy	55	25	30
381	route de sevrès - porte de bagatelle	49	29	20
927	place d'italie - soeur rosalie	38	12	26
928	place d'italie - vincent auriol	36	14	22
770	place d'italie - vincent auriol	34	2	32
512	naples - rome	33	5	28
762	versailles - claudette terrasse	32	4	28
47	bobillot - paulin mery	31	7	24
36	guy moquet - compoint	30	9	21

Stations critiques REMPLIES (régulation dépôt) :

	station_prise_clean \
478	madeleine vionnet
377	hippodrome de longchamp
381	hippodrome de longchamp
927	westermeyer - paul vaillant-couturier
928	westermeyer - paul vaillant-couturier
770	quai jules guesde - saint-simon
512	malesherbes - place de la madeleine
762	porte molitor
47	bnf - bibliotheque nationale de france
36	berthier - porte de clichy

	station_depose_clean	total	vae	vm
478	cambrai - benjamin constant	55	28	27
377	place de la porte de passy	55	25	30
381	route de sevrès - porte de bagatelle	49	29	20
927	place d'italie - soeur rosalie	38	12	26
928	place d'italie - vincent auriol	36	14	22
770	place d'italie - vincent auriol	34	2	32
512	naples - rome	33	5	28
762	versailles - claudette terrasse	32	4	28
47	bobillot - paulin mery	31	7	24
36	guy moquet - compoint	30	9	21

	velos_injectes	velos_collectes	velos_indisponibles	solde_net
station_clean				
perrault - place du louvre	0.0	7.0	5.337497	-7.0
place du 8 mai 1945	0.0	0.0	4.815625	0.0
charles tillon - cimetiere communal	0.0	12.0	4.586385	-12.0
hippodrome de longchamp	0.0	2.0	4.570518	-2.0
stade pershing	0.0	2.0	4.414634	-2.0
route de la muette - porte de neuilly	0.0	4.0	3.904204	-4.0
place jean-baptiste clement	0.0	12.0	3.847586	-12.0
sainte-elisabeth - turbigo	0.0	7.0	3.709884	-7.0
theatre des amandiers - palais des sports	0.0	0.0	3.674374	0.0
hotel de ville de saint-ouen	6.0	11.0	3.603454	-5.0
bercy - villot	0.0	8.0	3.494732	-8.0
theophile gautier - francois millet	0.0	8.0	3.459967	-8.0
institut du monde arabe - saint-germain	0.0	16.0	3.439581	-16.0
gare saint-lazare - cour du havre	0.0	19.0	3.357132	-19.0
hotel de ville d'asnieres sur seine	12.0	4.0	3.243066	8.0

Signalements par jour (juin) :

jour

2023-06-01 10

2023-06-02 21

2023-06-03 19

2023-06-04 22

2023-06-05 22

Name: count, dtype: int64

Signalements par jour (fevrier) :

jour

2023-02-01 22

2023-02-02 11

2023-02-03 23

2023-02-04 19

2023-02-05 22

Name: count, dtype: int64

Top localisations (juin) :

Localisation

11 Quai De Conti, 75001 Paris, France	12
8 Rue Morère, 75014 Paris, France	3
9 Rue Dulcie September, 93400 Saint-Ouen-sur-Seine, France	3
14 Place Carpeaux, 92800 Puteaux, France	3
Le Parvis 92800 Puteaux	3
43 Avenue Franklin Delano Roosevelt, 75008 Paris, France	3
14 Rue Claude Monet 92100 Boulogne-Billancourt	3
11 Boulevard De Bonne Nouvelle 75002 Paris	3
Route Saint-Louis 75012 Paris	3
7 Rue Des Grands Augustins, 75006 Paris, France	3
Allée Des Chênes, 75012 Paris, France	3
49 Cours De L'île Seguin, 92100 Boulogne-Billancourt, France	2
Route Des Lacs 75116 Paris	2
57 Rue Grétilat 94400 Vitry-sur-Seine	2
2 Rue Foucher Lepelletier 92130 Issy-les-Moulineaux	2

Name: count, dtype: int64

Top localisations (fevrier) :

Localisation

11 Quai De Conti, 75001 Paris, France	6
14 Rue Savier, 92240 Malakoff, France	5
3 Rue Louis Rousseau, 94200 Ivry-sur-Seine, France	5
D 7, 92400 Courbevoie, France	4
39 Rue Beauregard 75002 Paris	3
D 7 92400 Courbevoie	3
26 Avenue Du Général De Gaulle, 93170 Bagnolet, France	3
Gare de Bercy	3
Place Lucien Herr, 75005 Paris, France	3
274 Avenue Des Grésillons 92600 Asnières-sur-Seine	3
29 Rue De Paris 92110 Clichy	2
26 Rue Bartholdi 92100 Boulogne-Billancourt	2
123 Rue De Malabry, 92350 Le Plessis-Robinson, France	2
7 Rue Marie Et Louise 75010 Paris	2
14p Rue De La Grange Aux Belles, 75010 Paris, France	2

Name: count, dtype: int64

Nombre de stations à fort potentiel mais mal remplies : 230

	Nom station	cluster	total_velos_moyens \
0	11 Novembre 1918 - 8 Mai 1945	0	8.869565
7	Adolphe Lalyre - Armand Silvestre	0	12.046672
9	Adrien Lesesne - Docteur Bauer	0	10.533430
12	Alexandre Dumas - Voltaire	2	20.066709
15	Alfred de Vigny - Place du Général Brocard	0	10.855149

taux_replissage

0	0.248246
7	0.313438
9	0.384310
12	0.356578
15	0.360817