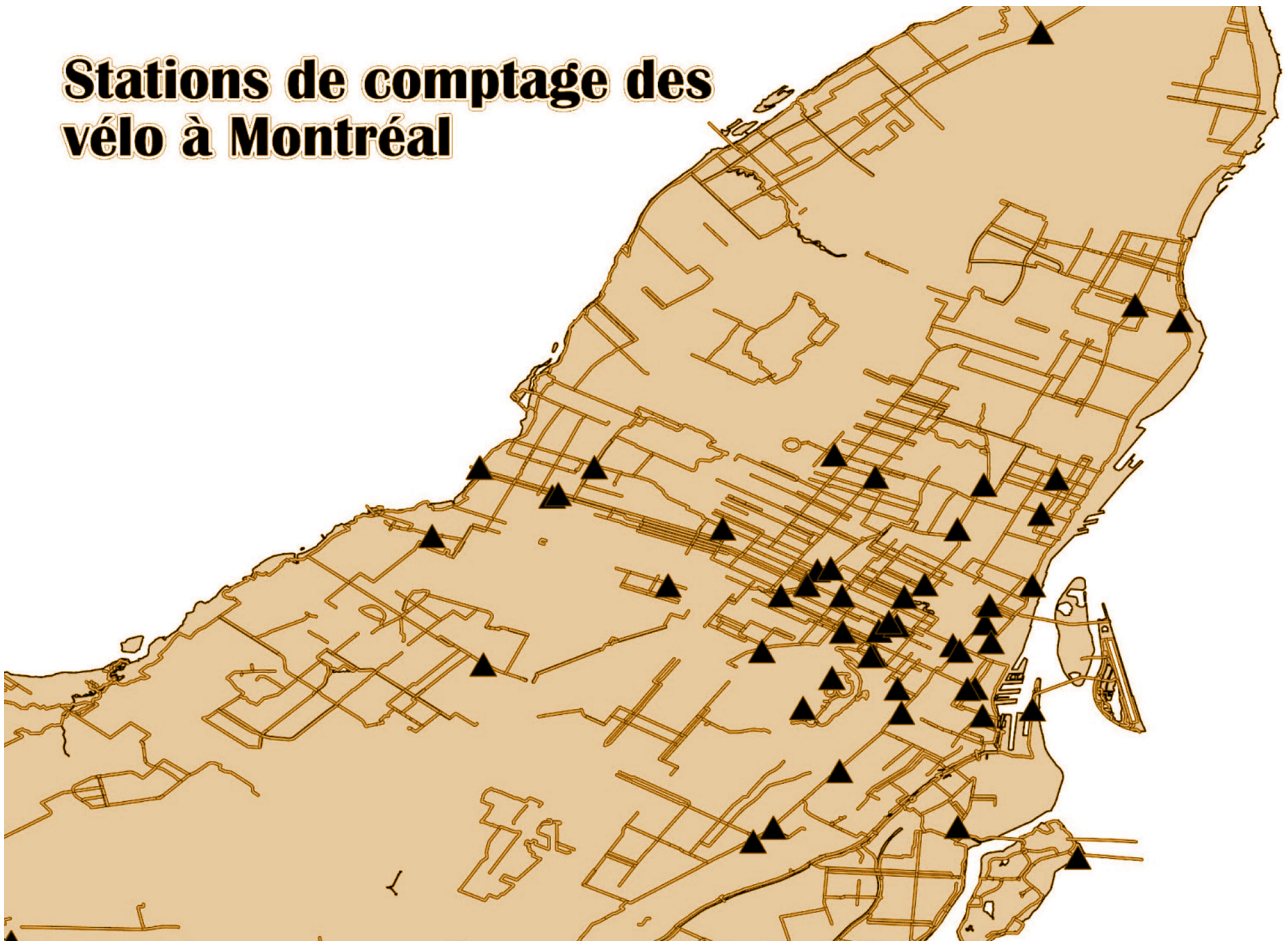


Projet Portfolio

Automatisation du processus de création d'histogrammes

Analyse de l'utilisation des pistes cyclables à Montréal

**Stations de comptage des
vélo à Montréal**



Christophe Dumoulin

[Code disponible sur Github](#)

Introduction

Avec l'augmentation significative du nombre de cyclistes au Québec ces dernières années, il est important de comprendre leur répartition géographique afin de mieux répondre à leurs besoins en matière d'infrastructure et de services. Heureusement, la Ville de Montréal met à disposition des données récoltées par les stations de comptage via le site web de Données Québec. Cet exercice vise à trouver une solution pour automatiser le traitement des données brutes et à produire des visualisations pour une meilleure compréhension de ces données.

Extraction des données

Évidemment, la première étape de ce projet est de récupérer les données. Il est important de s'assurer de pouvoir accéder à l'ensemble des données disponibles.

Données et ressources (16)

Données

CSV	Localisation des compteurs Dernière modification : 2023-01-06 Emplacement des compteurs avec la correspondance pour les noms dans les fichiers de comptage.	Explorer
CSV	Vélos - comptage sur les pistes cyclables, 2022 Dernière modification : 2023-01-06	Explorer
CSV	Vélos - comptage sur les pistes cyclables, 2021 Dernière modification : 2022-05-13	Explorer
CSV	Vélos - comptage sur les pistes cyclables, 2020 Dernière modification : 2021-10-04	Explorer
CSV	Vélos - comptage sur les pistes cyclables, 2018 Dernière modification : 2019-03-06 Informations disponibles : Date Site de comptage	Explorer
CSV	Vélos - comptage sur les pistes cyclables, 2017 Dernière modification : 2018-06-05 Informations disponibles : Date Site de comptage	Explorer

La page <https://www.donneesquebec.ca/recherche/dataset/vmtl-velos-comptage> contient les liens qui permettent de télécharger les fichiers CSV.

Il est possible de récupérer chaque lien sur cette page et télécharger le fichier CSV correspondant. Si la structure du site reste inchangée, le même script pourra être utilisé pour extraire les données des années suivantes sans modification. Les données sont alors sauvegardées et prêtes à être traitées.

Nettoyage des données

En examinant les données téléchargées, il est clair qu'elles sont difficiles à comprendre en raison de la variabilité des fréquences d'échantillonnage et des méthodes de référencement des stations de comptage. Certaines années présentent un compte de cyclistes toutes les 15 minutes, tandis que d'autres ne compilent ces données que quotidiennement. De plus, certains fichiers font référence aux stations de comptage par leur nom, tandis que d'autres utilisent un identifiant au format compteur_xxxxxxxxxxx. Ces incohérences rendent nécessaire un nettoyage et une normalisation des données pour en faciliter l'analyse et l'interprétation.

Date		Berri1	Boyer	BrÃ©beuf	CSC (CÃ'te S	Maisonneuve
01/01/2015	0:00	58	12	4	17	49
02/01/2015	0:00	75	7	5	15	113
03/01/2015	0:00	79	7	3	7	107
04/01/2015	0:00	10	1	21	0	35
05/01/2015	0:00	42	0	2	0	90
06/01/2015	0:00	101	0	1	1	153
07/01/2015	0:00	121	15	5	3	144
08/01/2015	0:00	70	0	36	0	139
09/01/2015	0:00	88	3	5	0	135
10/01/2015	0:00	29	7	6	0	70
11/01/2015	0:00	66	4	2	4	89
12/01/2015	0:00	183	4	4	16	323
13/01/2015	0:00	183	2	9	17	250
14/01/2015	0:00	183	2	6	33	276
15/01/2015	0:00	183	1	7	32	297
16/01/2015	0:00	177	5	2	7	280
17/01/2015	0:00	48	2	0	0	75

Données de 2015. Elles sont quotidiennes et utilisent le nom pour faire référence aux stations de comptage

Date	compteur_100054073	compteur_100003040	compteur_100052606
2022-01-01 0:00	0	0	0
2022-01-01 0:15	0	0	0
2022-01-01 0:30	0	0	0
2022-01-01 0:45	0	0	0
2022-01-01 1:00	0	0	0
2022-01-01 1:15	0	0	0
2022-01-01 1:30	0	0	0
2022-01-01 1:45	0	0	0
2022-01-01 2:00	0	0	1
2022-01-01 2:15	0	0	0
2022-01-01 2:30	0	0	0
2022-01-01 2:45	0	0	0
2022-01-01 3:00	0	0	0
2022-01-01 3:15	0	0	0
2022-01-01 3:30	0	0	0
2022-01-01 3:45	0	0	0
2022-01-01 4:00	0	0	0
2022-01-01 4:15	0	0	0

Données de 2022. Elles sont regroupées en sections de 15 minutes et utilisent un identifiant pour faire référence aux stations de comptage.

Pour standardiser et nettoyer les données de comptage de cyclistes, il est nécessaire d'utiliser des outils de traitement de données tels que Python. Le script `clean.py` peut facilement intégrer de nouveaux compteurs ou des données compilées à différentes fréquences. Il est également flexible, permettant à l'utilisateur de spécifier la période d'étude. Le script effectue un nettoyage et une normalisation des données pour en faciliter l'analyse et l'interprétation. Enfin, il crée un nouveau fichier incluant les données de toute la période étudiée compilées mensuellement, avec une référence géographique pour chaque station pour une intégration facile dans un système d'informations géographiques.

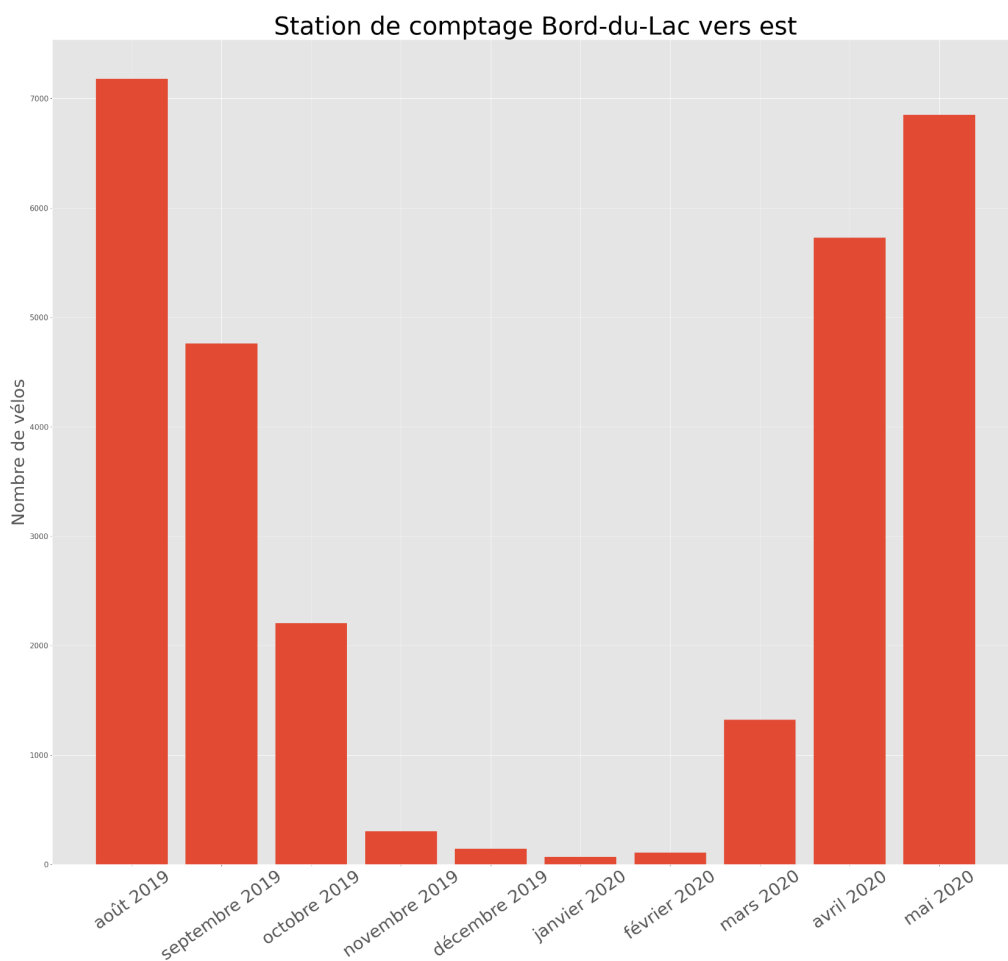
novembre 2022	décembre 2022	Latitude	Longitude	Nom	
4984	2082	45.5607132	-73.590636	16e Avenue - Bélanger	
4166	1007	45.5550844	-73.538819	Bennett	
57247	16820	45.516216	-73.56297	Berri	
		45.4469224	-73.782791	Bord-du-Lac vers est	
		45.4469936	-73.782642	Bord-du-Lac vers ouest	
21878	6585	45.53365	-73.59459	Boyer - Rosemont	
18385	5898	45.5434608	-73.616683	Boyer - Everett	
16988	64	45.527496	-73.57432	Brébeuf - Rachel	
10007	2603	45.55773	-73.64662	Christophe-Colomb-Lou	
21964	7009	45.5149	-73.607506	Côte Sainte-Catherine -	
42115	14104	45.5277722	-73.588832	Eco-Display - Métro Lau	
10270	1543	45.557593	-73.673222	Eco-Display Parc Stanley	
		45.54168	-73.68448	Edmond Valade	

Les données sont maintenant organisées et prêtes à être visualisées.

Cette structure permet d'intégrer facilement des statistiques descriptives, selon les besoins, qui pourront plus facilement être cartographiées. La structure actuelle est idéale pour produire des histogrammes pour visualiser le compte mensuel de vélos à chacune des stations.

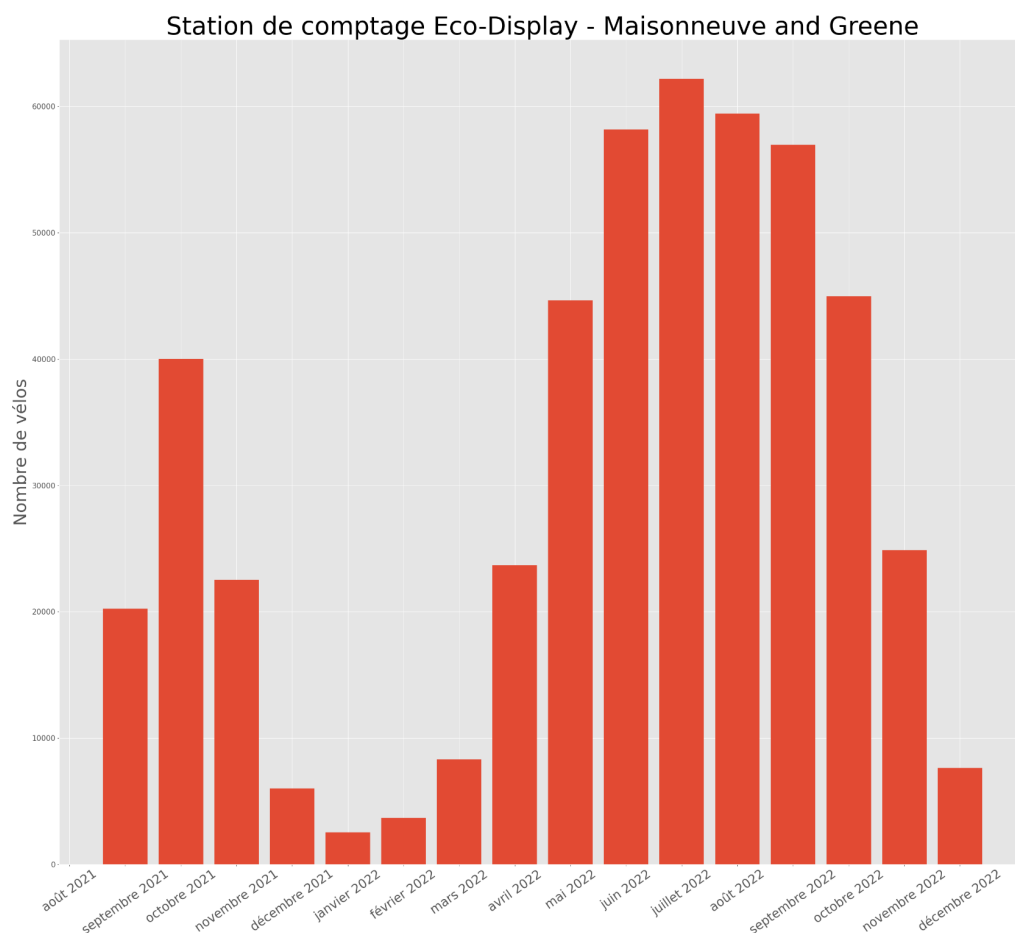
Production des histogrammes

La bibliothèque Matplotlib pour Python offre un puissant outil pour produire des histogrammes. En utilisant cette fonctionnalité, il est possible de visualiser rapidement le nombre de cyclistes comptés par mois à chaque station de comptage. Cela permet de mieux comprendre les tendances de circulation et de prendre des décisions éclairées en matière de planification des infrastructures cyclables.



Les histogrammes sont automatiquement produits seulement pour les mois
durant lesquels la station était en opération

Pour la période de 2013 à 2022, un total de 63 histogrammes sont produits. Une telle
visualisation des données nous permet rapidement de constater la variabilité saisonnière
du nombre de passage de cyclistes



Les histogrammes produits ont tous un style commun

Conclusion

En conclusion, il est clair que l'automatisation de la préparation de données est un élément clé pour améliorer l'efficacité d'organisations qui utilisent de grands jeux de données pour guider la prise de décision. Le projet présenté dans ce rapport démontre comment l'utilisation de technologies avancées peut faciliter la collecte et l'analyse de données, permettant ainsi une meilleure prise de décision. Les images incluses dans ce rapport illustrent les résultats obtenus grâce à cette automatisation, et il est évident que les

organisations qui intègrent de telles méthodes dans leur fonctionnement quotidien bénéficieront d'une amélioration significative de leur performance. Il est donc fortement recommandé aux employeurs de considérer ces résultats et d'envisager d'intégrer des processus d'automatisation similaires dans leur propre organisation.