

Mini projet — Analyse de données avec Pandas

Ce mini-projet vous permet d’aller un peu plus loin en Python, en vous initiant à l’analyse et l’exploitation de fichiers CSV. La manipulation de fichiers CSV est facilitée par le module [Pandas](#).

Nous allons nous intéresser aux résultats du dernier baromètre vélo, dont les résultats ont été publiés en septembre 2025.

1 Introduction et contexte

La [fédération française des usagers et usagères de la bicyclette](#) réalise à intervalles réguliers une enquête, appelée « Baromètre vélo » de grande envergure sur l’utilisation du vélo au quotidien comme moyen de transport.

Le questionnaire porte sur les usages, les infrastructures et les relations entre les usagers et usagères de la route. Il vise à accompagner les pouvoirs publics dans ses réflexions sur l’aménagement de l’espace public, en vue de faciliter l’utilisation des mobilités actives.

Les résultats de l’année 2025 ont été publiés le 18 septembre 2025 et sont accessibles en *opendata*, via le site <https://opendata.parlons-velo.fr/>. Le volume de données est considérable. Dans ce travail, nous allons nous intéresser aux résultats de la région Île-de-France.

L’objectif est d’utiliser les résultats bruts de l’étude, et de les analyser pour produire quelques indicateurs statistiques pertinents, visant à identifier les points forts et les faiblesses du territoire en termes d’aménagements cyclables.

2 Récupération des données et mise en place du projet

- 1) Depuis le site Internet <https://opendata.parlons-velo.fr/>, télécharger l’archive « Données du Baromètre Vélo 2025 pour la région Île-de-France », extraire les fichiers `notice.pdf`, `reponses-region-11.csv` et `resultats-communes-region-11.csv`.
- 2) Créez dans votre répertoire de travail un dossier `MiniProjet` et le fichier `barometre_velo.py`.

Nous allons utiliser le module Python `pandas`, donc vous pouvez déjà ajouter la ligne ci-dessous à votre code :

```
1 import pandas as pd # Importe le module pandas et le renomme pd
2                     # pour faciliter l'écriture
```

- 3) Documentez-vous sur le fonctionnement de Pandas, en consultant <https://pandas.pydata.org/docs/>, rubrique « Getting started » pour une introduction, et rubrique « User guide » pour la documentation complète.

Méthodologie de projet

Ce travail est beaucoup moins guidé que les précédents TD. Pensez à travailler de manière méthodique et modulaire, en découpant convenablement votre code.

N'hésitez pas à prendre des initiatives et à faire des essais, voire à analyser les données en produisant des indicateurs qui vous semblent pertinents, même s'ils ne sont pas demandés dans l'énoncé.

Prenez le temps de vous former sans précipitation au fonctionnement de Pandas, avant de commencer à programmer.

3 Exploitation des fichiers CSV

- 1) Ouvrir le fichier `communes-region-11.csv` avec Python et construire une DataFrame à partir de ce fichier avec comme clé la colonne `uuid`.
- 2) Déterminer le nombre de réponses données pour la ville de Paris, puis pour la ville de Vitry-Sur-Seine, puis pour le département du Val-de-Marne.
- 3) Construire la liste contenant les identifiant des réponses pour lesquelles la ville évaluée est la même que la ville habitée.
- 4) Écrire un fichier CSV contenant uniquement les résultats associés à la ville de Paris.
- 5) Écrire un fichier CSV contenant, pour chaque ville, le nombre de personnes ayant répondu au baromètre.
- 6) Construire un fichier CSV contenant, pour chaque indicateur, la moyenne par ville.
- 7) Trouver l'indicateur le mieux noté, en moyenne, dans Paris.
- 8) Trouver l'indicateur le moins bien noté, en moyenne, dans le département du Val-de-Marne, pour l'indicateur de la question 16.
- 9) Calculer, pour chaque ville, des scores obtenus pour chaque thématique de question.
- 10) Trouver les 10 villes de la région parisienne pour lesquelles les personnes ayant répondu ont donné des avis différents les uns des autres sur l'ensemble du questionnaire. On pourra calculer l'écart-type.
- 11) N'hésitez pas à effectuer d'autres analyse ou à extraire d'autres informations, données, classements qui vous semblent pertinents.

4 Tracé de graphiques

On utilisera [Plotly](#).

- 12) Tracer le graphe indiquant le nombre de réponses par communes, ordonnées par ordre croissant de code INSEE. On ne listera que les villes ayant obtenu plus de 50 réponses.
- 13) Faire de même département par département.
- 14) Tracer le graphe avec en abscisses les villes du Val-de-Marne, et en ordonnées le score moyen pour la question 14.

- 15)** Tracer le graphe, pour le département du Val-de-Marne, mettant en corrélation la fréquence d'usage du vélo (question 6) avec la question 10. Pour ce faire, calculer, pour chaque réponse, le quotient Q_6/Q_{10} et tracer le nuage de points correspondant. Comment interpréter ce résultat ?