

Découverte et qualité Open Data

Exercice 1.1 - Mise en œuvre des Données Ouvertes

- **La source de données choisi** : un fichier csv depuis le site <https://www.data.gouv.fr/fr/> qui présentent les loyers observés par les observatoires des loyers sur l'année 2022 selon le type de logement, l'époque de construction, le nombre de pièces ou l'ancienneté d'occupation. Les données sont mise à disposition selon deux échelles géographiques : l'agglomération et la ville centre. Elles correspondent aux résultats présentés sur www.observatoires-des-loyers.org.
- **Un scénario d'application** : la source de données peut être utilisée pour développer une application qui permet d'estimer le cout d'un loyer en fonction de plusieurs paramètres.
- Les données sont sous format csv, donc afin de pouvoir mieux les exploiter et faciliter les requêtes, il faudra les transformer en tables SQL. Aussi elles présentent beaucoup de valeurs nulles qu'il faut traiter.

Exercice 1.2 – Qualité des données

a. Les dimensions de qualité et les problèmes observés :

- 1- Complétude : le jeu de données compte beaucoup de valeurs nulles sur des colonnes dont l'information est primordiale dans notre cas d'utilisation « estimer le cout d'un loyer » tels que le type de l'habitat.
- 2- Précision : le jeu de données présente un problème de précision sur la colonne « zone complémentaire » qui définit l'échelles géographiques, en effet il n'y a que deux valeurs possible « ville centre » et « périphérie » chose qui peut avoir un impact sur l'exactitude des données, car dans le même centre-ville le loyer peut être différent suivant le quartier et l'emplacement de l'habitat.
- 3- Volatilité : nos données ont un degré élevé de volatilité car les couts des loyers varient très beaucoup dans le temps.

- 4- Ponctualité : les données présentes les couts des loyers sur l'années 2022, donc il ne sera plus ponctuel sur le long terme.
- b. Les données ne sont pas accompagnées d'un catalogue qui renseigne sur des dimensions de qualité.

Quelques outils de gestion de la qualité :

- Talend.
- OpenRefine.
- ZoomInfo OperationsOS.
- Ataccama.
- Dataedo.