



Challenge 2 - The agency



Note : Vous pouvez utiliser le langage Python pour répondre à l'exercice

Stéphane et Alex, deux agents immobiliers qui souhaitent offrir une nouvelle expérience à leurs clients en utilisant les données.

En tant que Data Engineer, vous allez accompagner ces agents dans leur quête.

Nous vous fournissons des données qui présentent les loyers observés par les observatoires des loyers du réseau selon le type de logement, l'époque de construction, le nombre de pièces ou l'ancienneté d'occupation.

Les données sont mise à disposition pour chaque année de collecte et selon deux échelles géographiques : l'agglomération et la ville centre.

Vous trouverez les données dans le fichier CSV accompagnant l'exercice.

1. Proposez un schéma plus significatif pour les données
2. En utilisant un format de stockage de votre choix, créez une nouvelle table de données adéquate à ce schéma.
3. Selon vous, quels nettoyages faut-il opérer sur ces données afin qu'elles soient prêtes à une analyse exploratoire ?
4. Effectuez une analyse exploratoire sur les données et créez des visualisations qui peuvent aider Stéphane et Alex à montrer facilement des statistiques sur les logements à leurs clients.
5. Supposant que les données viennent d'un système en temps réel, que seront vos propositions de design d'une pipeline qui va de l'ingestion de données aux visualisations. (Architecture technique + outils recommandés)