

## Test technique MeilleursAgents - Geo data engineer

Le labo de recherche en économie de Meilleursagents souhaite étudier le parc locatif social de la commune de Cannes (06400). Nous disposons déjà d'une modélisation administrative du territoire à différents niveaux de granularité (commune, quartier, iris, parcelle). Le gouvernement produit chaque année le "*Répertoire des logements locatifs des bailleurs sociaux (RPLS)*". Ce fichier semble répondre à nos besoins et est disponible en open data.

L'objectif du test est de préparer et de présenter ce jeu de données pour faciliter son exploitation par les membres de l'équipe de recherche. Bon courage !

### Partie 1: nettoyage et géocodage du fichier brut

Chaque ligne du fichier représente un logement social sur le territoire observé. Un logement est localisé par son adresse. Pour effectuer des traitements géographiques sur ces données, nous souhaitons modéliser chaque logement par un point. Ce premier exercice consiste donc à géocoder chaque logement de sorte à obtenir une coordonnée précise.

Pour simplifier l'exercice, vous utiliserez une version allégée du RPLS 2018:

[https://storage.googleapis.com/ma-static-dev/geodude/rpls\\_cannes.csv](https://storage.googleapis.com/ma-static-dev/geodude/rpls_cannes.csv)

Avec son dictionnaire de variables:

[https://storage.googleapis.com/ma-static-dev/geodude/dictionnaire\\_rpls.xls](https://storage.googleapis.com/ma-static-dev/geodude/dictionnaire_rpls.xls)

Et sa note explicative originale:

[https://storage.googleapis.com/ma-static-dev/geodude/note\\_explicative.pdf](https://storage.googleapis.com/ma-static-dev/geodude/note_explicative.pdf)

Nous vous conseillons d'utiliser l'api de géocodage d'Etalab:

<https://geo.api.gouv.fr/adresse>

Nous attendons à la fin de l'exercice un nouveau fichier **rpls\_cannes\_geocoded.csv** contenant deux nouveaux champs "X" et "Y". L'exercice doit être exécuté **en Python, avec les librairies et outils de votre choix**.

Si vous rencontrez des difficultés, vous pouvez passer à l'exercice suivant en utilisant notre version du fichier géocodé. Nous n'avons pas utilisé la même api mais les résultats devraient être suffisamment proches:

[https://storage.googleapis.com/ma-static-dev/geodude/rpls\\_cannes\\_geocoded.csv](https://storage.googleapis.com/ma-static-dev/geodude/rpls_cannes_geocoded.csv)

## Partie 2: intégration au référentiel

Votre fichier **rpls\_cannes\_geocoded.csv** doit ensuite être stocké en base de données dans une ou plusieurs tables qu'il vous faudra définir au préalable.

Pour effectuer des analyses spatiales plus avancées il est souhaitable d'enrichir l'information du RPLS. Plus concrètement, on veut obtenir:

- Un identifiant unique pour chaque logement
- Une géométrie Postgis (extension GIS de PostgreSQL)
- Un identifiant de parcelle pour chaque logement
- *Bonus*: la distance physique (en mètres) d'un logement au centroid de la commune

Pour associer un identifiant parcellaire à chaque logement il faudra vous aider du référentiel géographique existant (cf: *Annexe: description du référentiel géographique*).

Pour accéder à la base de données, vous référer aux identifiants de connexion présents dans le mail.

Vous êtes libre d'utiliser les langages de programmation de votre choix. Nous vous encourageons cependant à tirer parti **de Postgis** avec le langage SQL. Pour faciliter sa mise en production, votre code doit être exécuté dans **un container Docker**.

## Partie 3: affichage et analyse des données

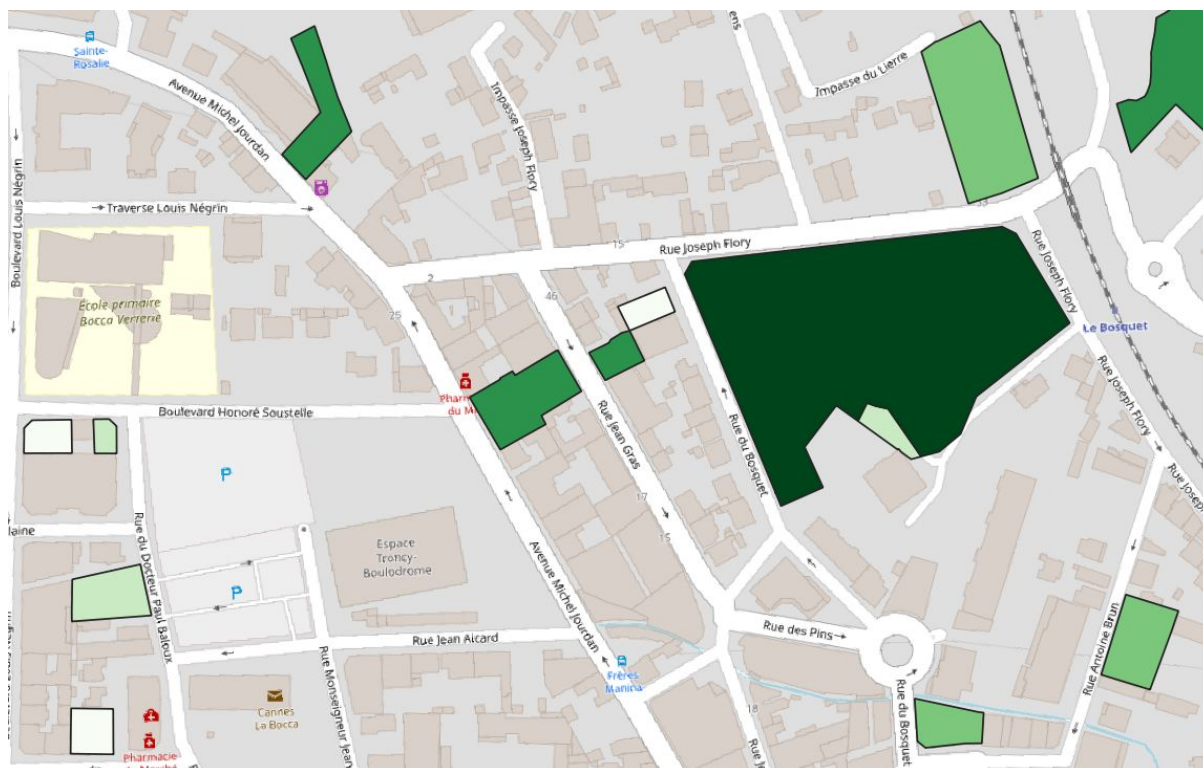
Pour cet exercice, l'esthétique importe peu. Le choix **des outils est libre**. Vous serez jugé sur votre capacité à visualiser **simplement** et **convenablement** vos données géographiques.

Si vous ne l'avez pas déjà fait, affichez le RPLS en tant qu'objets vectoriels ponctuels (un point par logement). Que dire de l'organisation spatiale des logements sociaux à Cannes ? Où peut on trouver des logements sociaux de plus de 5 pièces à Cannes ?

Pour vérifier nos volumes de logements sociaux, nous voulons cartographier la distribution du nombre de logements au sein de la commune.

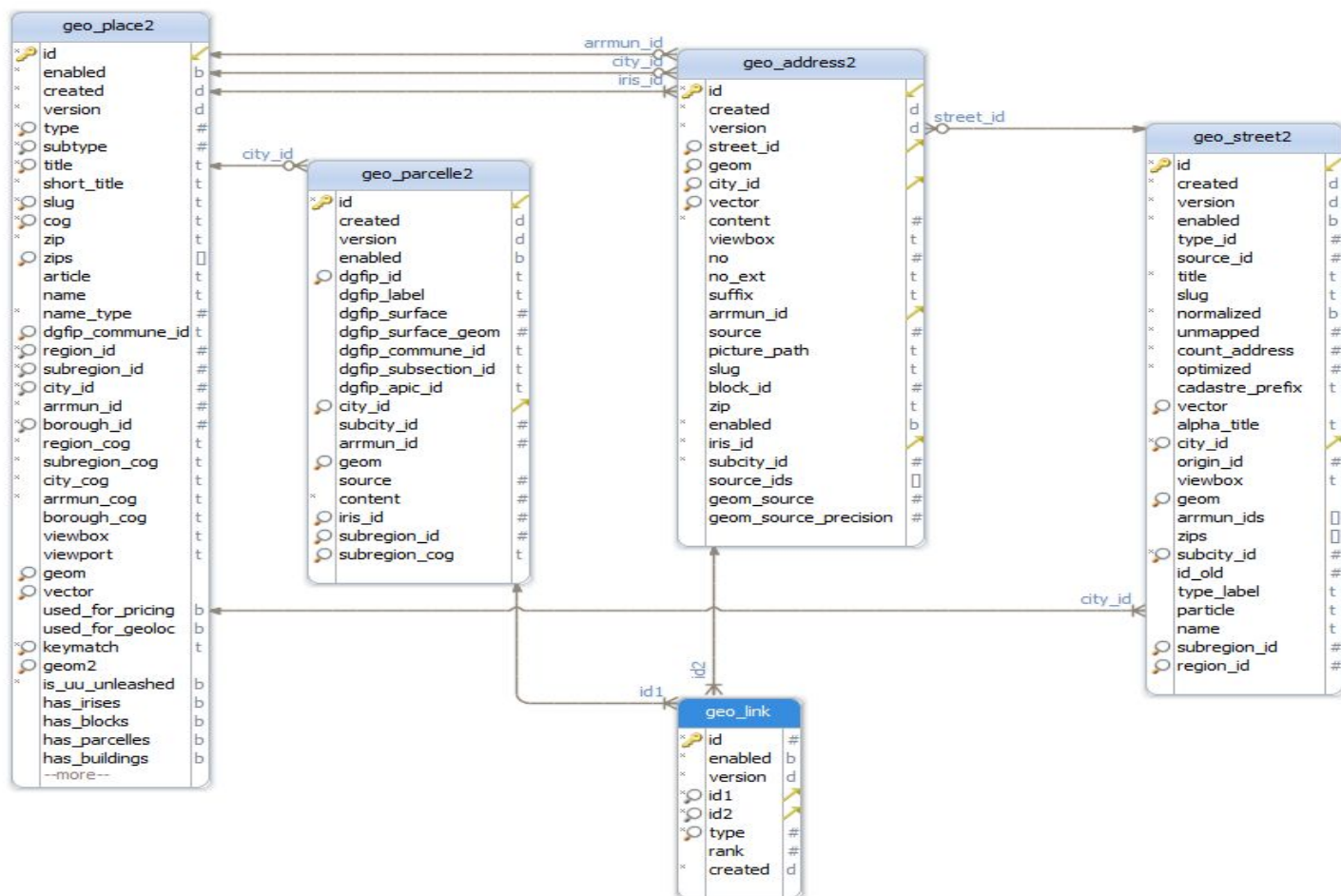
Pour effectuer cette carte vous allez devoir:

- Créer une nouvelle table agrégée au niveau de la parcelle. Pour un identifiant de parcelle (**geo\_parcelle2.id**) nous voulons connaître le nombre de logement sociaux.
- Afficher cette couche en utilisant la géométrie de la parcelle (**geo\_parcelle2.geom**) en utilisant une couleur différente pour chaque groupements de population.



*résultat attendu sur le quartier de la bocca*

## Annexe: description du référentiel géographique



Toutes les tables du référentiel se trouvent dans le schéma **geo**.

La table **geo\_place2** contient les différents niveaux géographiques pour la commune de Cannes. Le niveau géographique est déterminé par le **type** de chaque tuple de cette table (le niveau iris est 65 et le niveau commune est 40, l'id de la commune de Cannes dans cette table est 2199).

La table **geo\_address2** représente une adresse de la commune, par exemple: 22 Avenue de l'Amiral Wester Weymiss, 06400 Cannes. Les informations relatives à la rue sont disponibles dans la table **geo\_street2**.

La table **geo\_parcelle2** contient les parcelles de la commune. Le lien entre parcelles et adresses est de type *n-n*, de ce fait, ils sont modélisés via la table **geo\_link** avec le **type 1602**.

Chacune de ces tables (à l'exception de la table **geo\_link**) possède sa propre représentation géométrique dans la colonne **geom**.