

RENDU PROJET FRAMEWORK BIG DATA

RABHI Yasser

Introduction :

Chaque année, l'Insee (Institut **national de la statistique et des études économiques**) estime la population des régions et des départements (France métropolitaine et DOM) à la date du 1er janvier. Ces estimations annuelles de population sont déclinées par sexe et par âge (quinquennal, classes d'âge).

Le niveau de population et la pyramide des âges sont évalués annuellement à partir des statistiques d'État civil et d'une estimation du solde migratoire.

Au cours d'une année donnée, l'évolution de la population d'une zone géographique résulte en effet de deux facteurs : le solde naturel et le solde migratoire.

Objectifs du projet :

Dans ce contexte, l'objectif de ce projet est d'exploiter ces données publiées sur le site de l'INSEE afin de pouvoir visualiser les différents changements en fonction de population, d'âge, de sexe ... [Source](#)

Sources de données :

- Un fichier .csv contenant une estimation de la population française au 1er janvier, par département, sexe et grande classe d'âge de 1975 à 2022.
- Un fichier .geojson contenant la définition des bordures des régions françaises
- Un fichier .geojson contenant la définition des bordures des départements françaises
- Un fichier .csv regroupant des informations sur les départements tel que le code et la superficie ainsi que les coordonnées géographiques.

Nettoyage et transformation des données :

Afin de pouvoir visualiser ces informations, les données sont formatées et les valeurs nulles sont enlevées. Voici les étapes de traitement de données :

- Télécharger le fichier de l'estimation de la population au mois de Janvier 2022
- Transformer les données en 3 fichiers csv :
 - o Ensemble de la population
 - o Population masculine
 - o Population féminine
- Mettre les fichiers sur QlikView et visualiser l'ensemble des valeurs
- Création des applications Analytics pour chaque fichier
- Charger les données sous structure Data Frame avec python
- Traitement des valeurs nulles et de colonnes vides
- Corriger l'encodage des départements pour afficher les caractères accentués

- Fusionner les fichiers de population au fichier de départements par la clé code département
- Joindre les départements à leurs régions et puis regrouper les données pour avoir les sommes par région.
- Ajouter à chaque région sa géométrie pour l'afficher sur la carte.
- Ajouter des colonnes tel que le pourcentage de population par sexe et la densité par région.

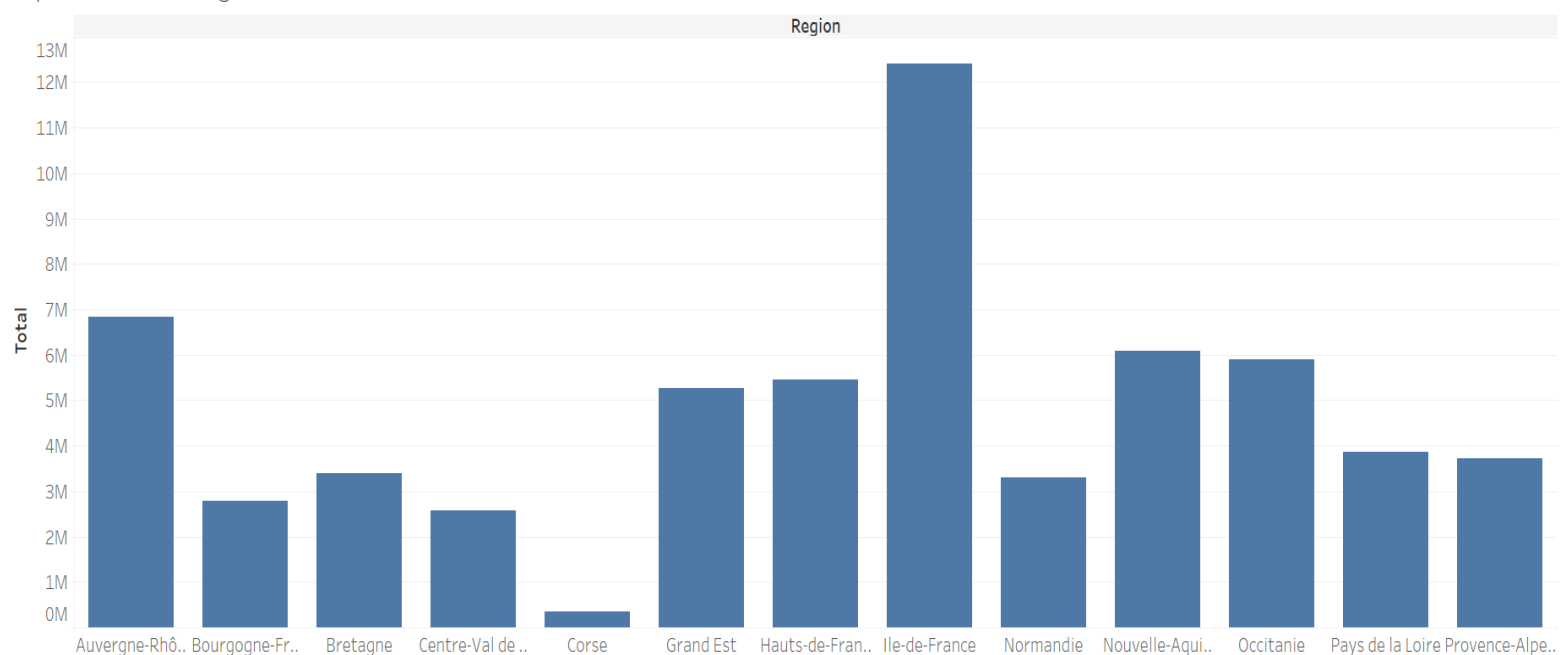
Visualisations :

1) Tableau :

Tableau est un outil de visualisation de BI. Les 3 fichiers csv regroupant l'ensemble de la population et les deux sous-ensembles masculin et féminin sont transmis sous formes de bases de données. Les dimensions utilisées sont les valeurs de régions avec plusieurs mesures comme population par tranche d'âge ou par densité.

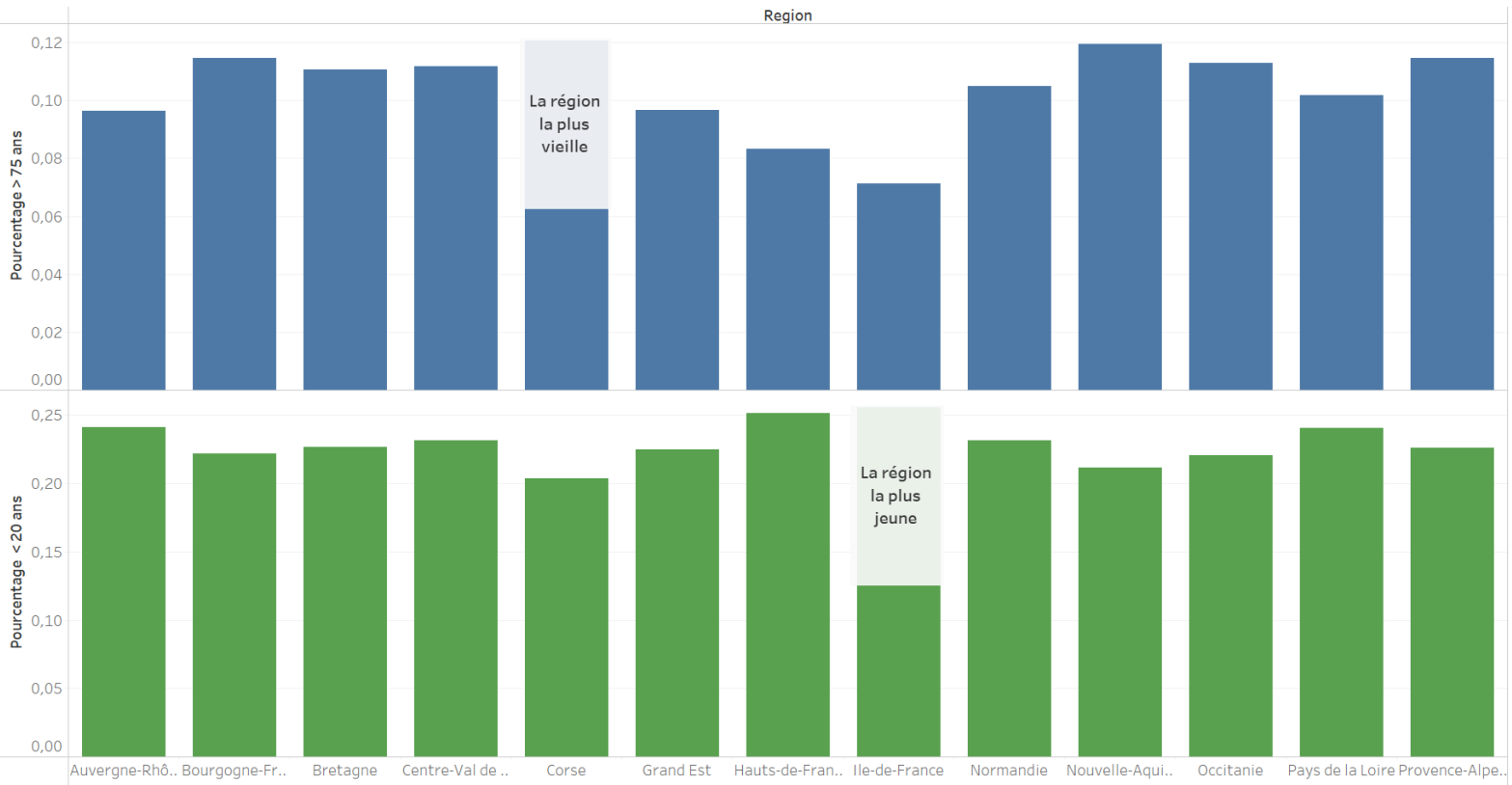
Population par région :

Population Par Région



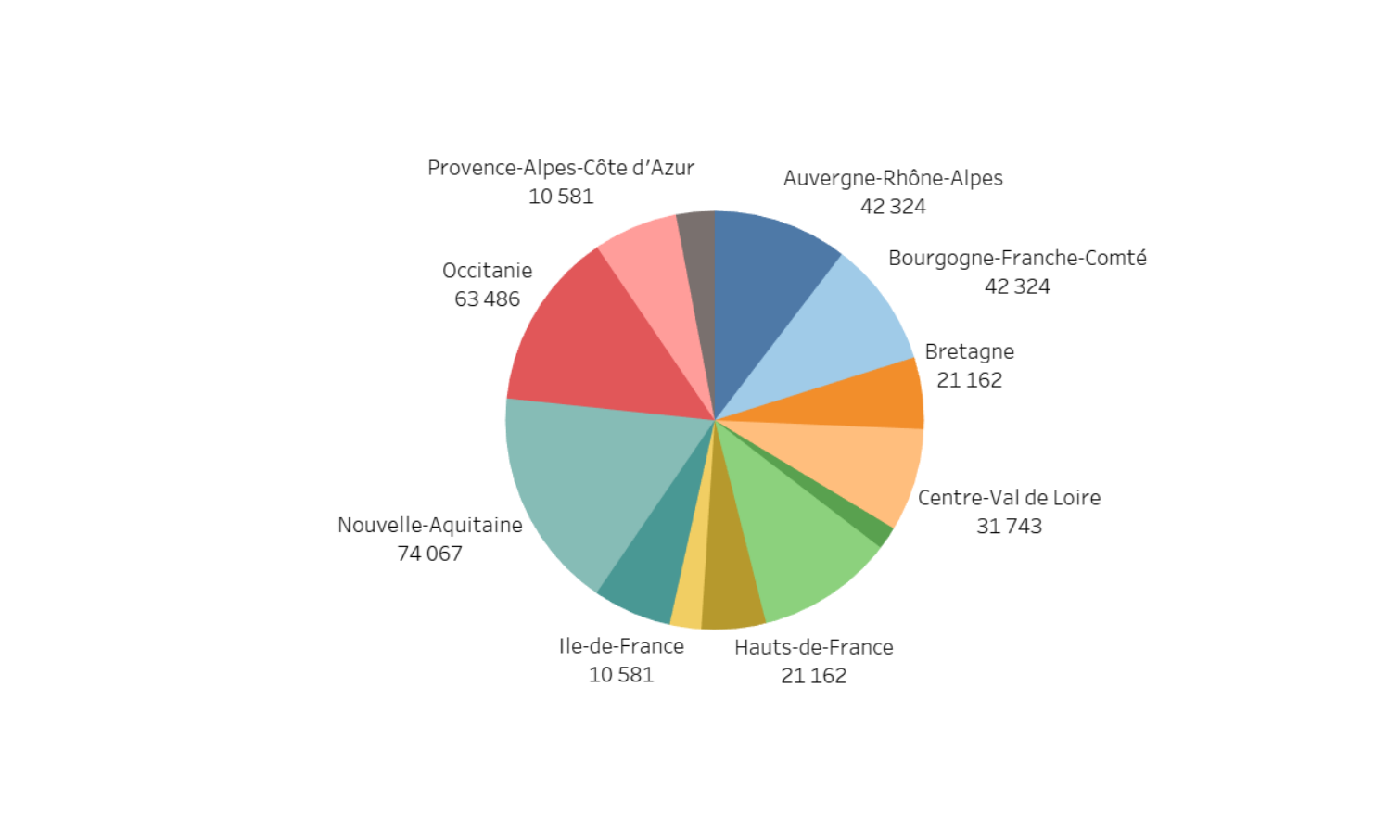
Pourcentage de personnes jeunes (0-19 ans) et âgées (>75ans) :

Pourcentage de personnes **jeunes** et **agées**



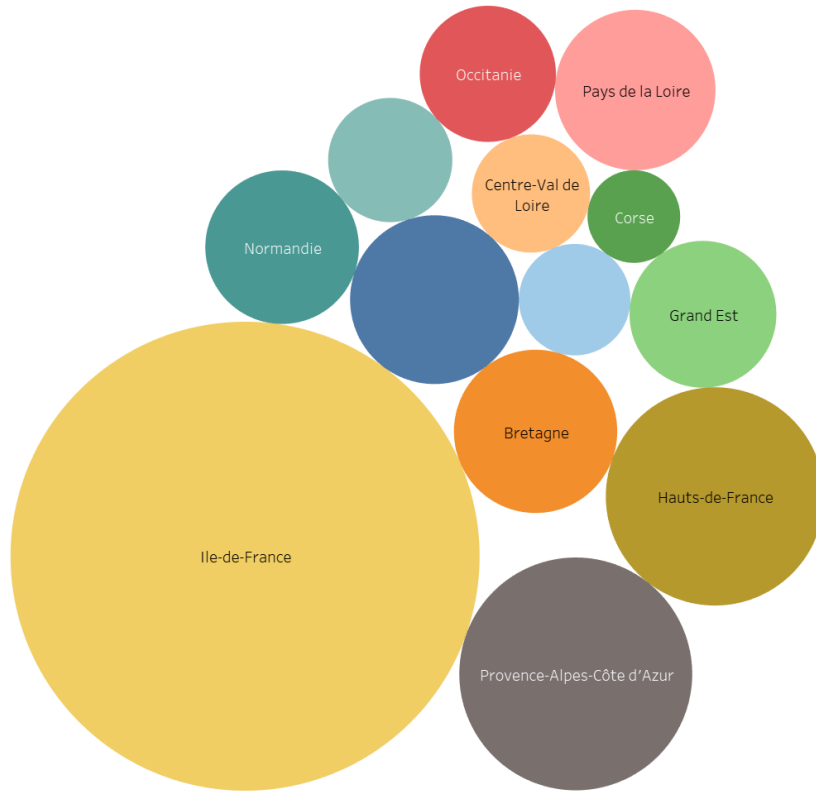
Graphique à secteur de régions par surface (km²) :

Graphique à secteurs de régions par surface



Bulles de régions par densité (Population / Surface) :

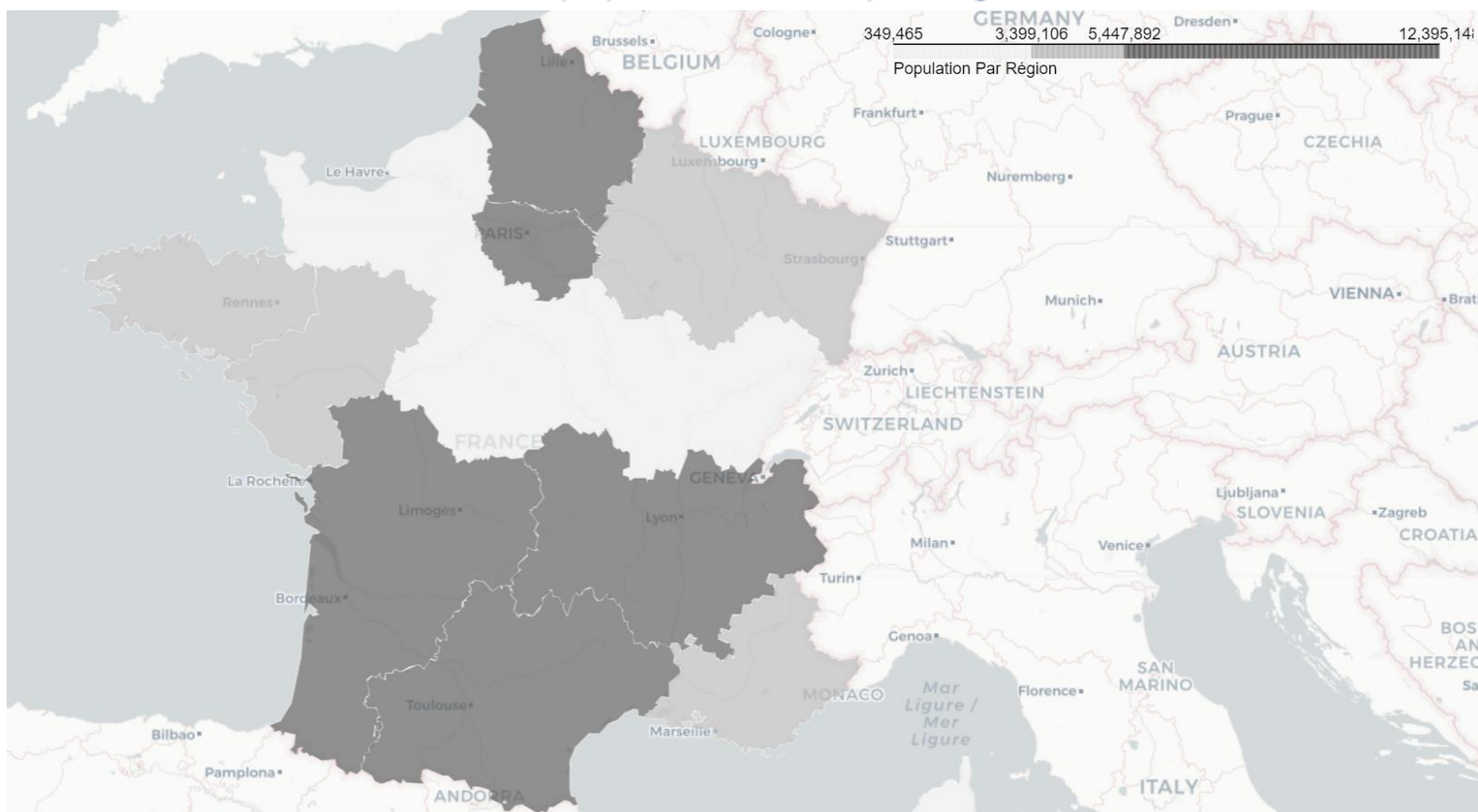
Bulles de régions par densité



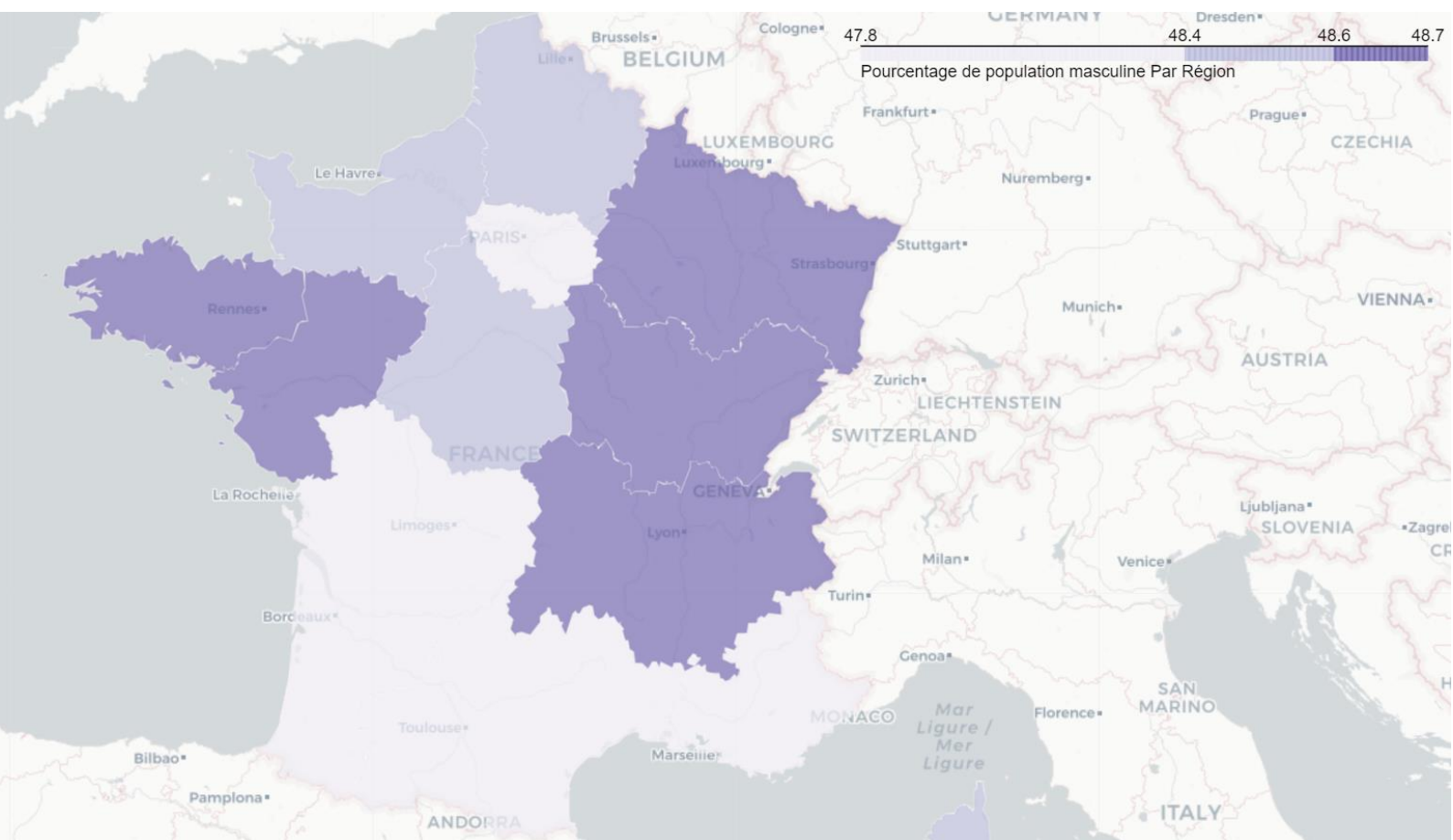
2) Folium :

La librairie Folium permet de créer des cartes interactives sur python. Les données sont transformées en Géo Data Frame contenant une colonne de géométrie. Les données sont aussi regroupées par région.

Carte de la population totale par région en 2022 :



Carte de pourcentage d'hommes par région en 2022 :



Carte de pourcentage de femmes par région en 2022 :

