
Projet Génie Logiciel - M2 ISI 2021/2022

Sujet : Rentabilité d'investissement régional dans le service et le commerce
Parsa Farsi - Michele Grimaldi

1 Présentation du projet

Le but de ce projet est d'explorer et de traiter de données de base INSEE sur les communes de France afin de préparer un classement régional par rapport à la rentabilité d'investissement dans le service et le commerce. Ensuite, on représente le classement sur une carte des régions de France.

Le traitement se fait principalement en SQL et qui sera complété avec Python notamment pour la représentation du classement sur la carte de France.

2 Plan du rapport :

Ce rapport est constitué deux parties.

2.1 Partie 1 : SQL

- Création la base SQL "inse" et des tables ci-dessous à partir la bases de données INSEE sur les communes (MDB-INSEE-V2.csv) et la base officielle des codes postaux (Laposte_Hexasmal.csv) :
 - Activite_Economique
 - Fiscalite
 - Logement
 - Population_INSEE
 - Sante
 - Laposte_Hexasmal
- Insertion de données dans les tables
- Traitement des données des tables ci-dessus afin de créer et préparer des indicateurs régionaux sur la rentabilité d'investissement

2.2 Partie 2 : Python

- Connection à la base de donnée
- Recuperation des données via les requetes
- Creation des fichiers geoJson
- Creation des Map avec les GeoJson et le package Folium

3 Partie 1 : SQL

3.1 Création des tables et l'insertion de données

La première étape du projet consiste à la création de la base SQL "inse" et des tables mentionnés dans le plan du rapport à partir de la bases de données INSEE sur les communes récupéré du site "www.data.gouv.fr" et la base officielle des codes postaux du site "www.data.opendatasoft.com".

Après récupération des données, on les analyse au but de la catégorisation, cela consiste à étudier différents aspect de la table et classer les données de manière qui peuvent nous servir à la création des indicateurs régionaux sur la rentabilité d'investissement dans la partie de traitement. Cette analyse nous a amené à décomposer la table INSEE (inséré à partir de la bases de données INSEE sur les communes)

aux tables suivants :

- Activite_Economique -> Cette table contient des données liées au nombre des différents types d'entreprise existant ou en cours de création ainsi qu'au nombre de la dynamique entrepreneuriale[1].
- Fiscalite -> Dans cette table, on a généralement accès aux différents variables qui sont en rapport avec la fiscalité. On peut voir différents scores comme le score de PIB[2] et de VA[3] régional, le taux d'évasion de la clientèle[4] et les moyennes des salaires des différents secteurs d'activité.
- Logement -> La table logement comme son nom indique est remplie des variables liés aux immobiliers comme le taux de propriété, le nombre de logement etc.
- Population -> Elle représente les variables démographique.
- Santé -> En utilisant cette table, on accède aux données de Santé des communes de France comme le nombre des pharmacies et des parfumeries, le nombre des infirmiers, des dentistes etc.

3.2 Traitement des données

3.3 Identification des indicateurs

Suite à la création fait dans la partie précédente, on cible certains colonnes de chaque table afin de commencer la partie traitement et créer les indicateurs.

Pour cela, on a fait des recherches sur les paramètres qui peuvent donner l'importance à une région au sujet d'investissement dans le service et le commerce. Par exemple, Pour la table « Fiscalite », selon la figure 1, on constate que les investisseurs et les actionneurs sont des bénéficiaires de la valeur ajoutée. Donc le colonne du score VA régional est un paramètre à prendre en considération et qui est un des indicateurs pour l'investissement.

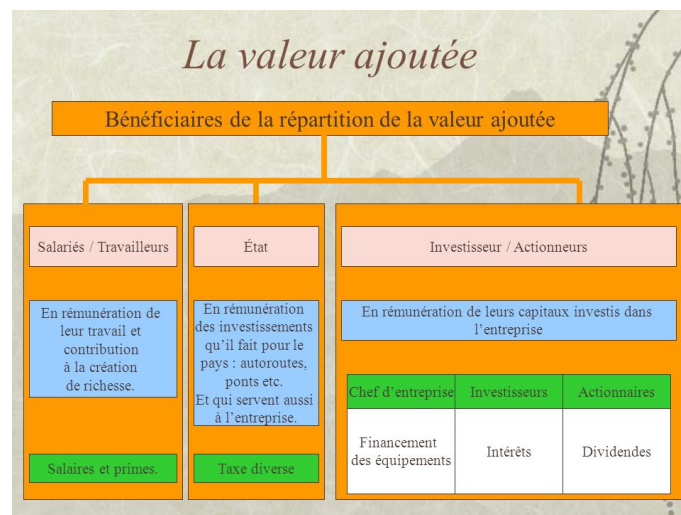


FIGURE 1 – VA , <https://slideplayer.fr/slide/1212990/>

De même, on a identifié les indicateurs ci-dessous dans les tables :

- Activite_Economique
 - dynamique_entrepreneuriale_service
 - nb_creation_services
 - nb_entreprises_secteur_service
 - nb_creation_commerces
 - nb_entreprises_secteur_commerce

- Fiscalite
 - score_pib
 - score_va_region
 - pib_regionnal
 - valeur_ajoutee_regionale
 - taux_evasion
- Logement
 - nb_logements
- Population_INSEE
 - evolution_pop_pourcent
- Sante
 - nb_pharmacie_et_parfumerie

3.4 Calcul des indicateurs

Afin de finaliser le traitement des données, on calcule les indicateurs pour chaque région. pour certains, on applique d'abord les opérations mathématiques en utilisant des coefficients définis. Par exemple pour la table « Activite_Economique », on a défini trois coefficients coef1, coef2 et coef3 qui sont multipliés par les indicateurs [dynamique_entrepreneuriale_service], [nb_creation_services] et [nb_entreprises_secteur_service] de manière à prendre en considération l'importance des unes vis_à_vis des autres.

Au final on normalise tous les indicateurs pour avoir leur valeurs en pourcentage. De ce fait, en additionnant tous les indicateurs, on déduit un indicateur final qui nous sert à classer les régions selon leur rentabilités d'investissement dans le service et le commerce.

4 Partie 2 : Python

4.1 Description et objectifs

Une fois on a obtenu tous les indicateurs en SQL, on se connecte à la base de donnée via Python et on récupère les indicateurs pour toutes les régions. Ces indicateurs sont ensuite insérée dans des fichiers "GeoJSON", un format de fichier qui encode les données géographiques et qui vont nous permettre de visualiser les donnée sur la carte géographique de la France.

Pour ce faire, on crée d'abord un fichier "GeoJson" pour chaque région en ajoutant les contours grâce aux coordonnées GPS disponible dans la table "Laposte_Hexasmal". On utilise les features qui caractérisent la région (le nom, le code régional, les coordonnées de contours) et au final on ajout le(s) indicateur(s) correspondant.

Ensuite pour créer la carte, on utilise le package *folium* de Python qui utilise la base "CartoDB"[5]. On modifie les couleurs, les légendes et aussi les paramètres de la carte, dans un template en "jQuery"[6] qui utilisera ensuite les données des fichiers "GeoJSON" pour afficher la carte des régions et leurs couleurs respectifs.

Les figures ci-dessous montre différents cartes pour les tables traités. Chaque région est donc coloré selon les valeurs des indicateurs. Par exemple, pour la figure 2 qui correspond à la table "Fiscalite", la région Île-de-France est frottement coloré ce qui signifie, selon les indicateurs de cette table, il y a plus de chance d'avoir un investissement rentable dans cette région en comparant des autres régions.

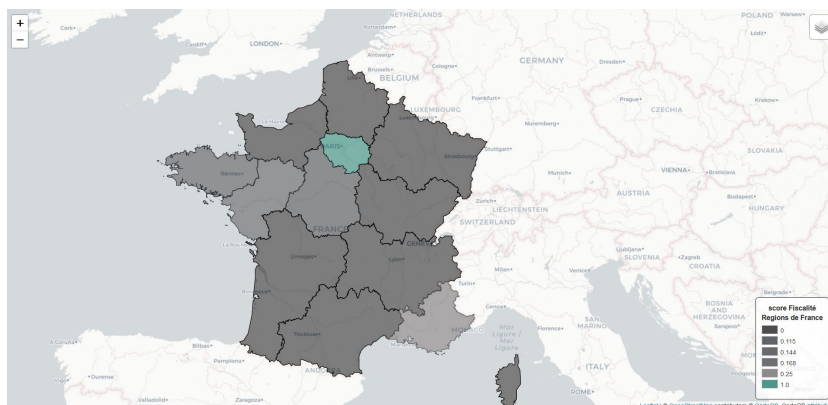


FIGURE 2 – Carte selon les indicateurs de Fiscalité

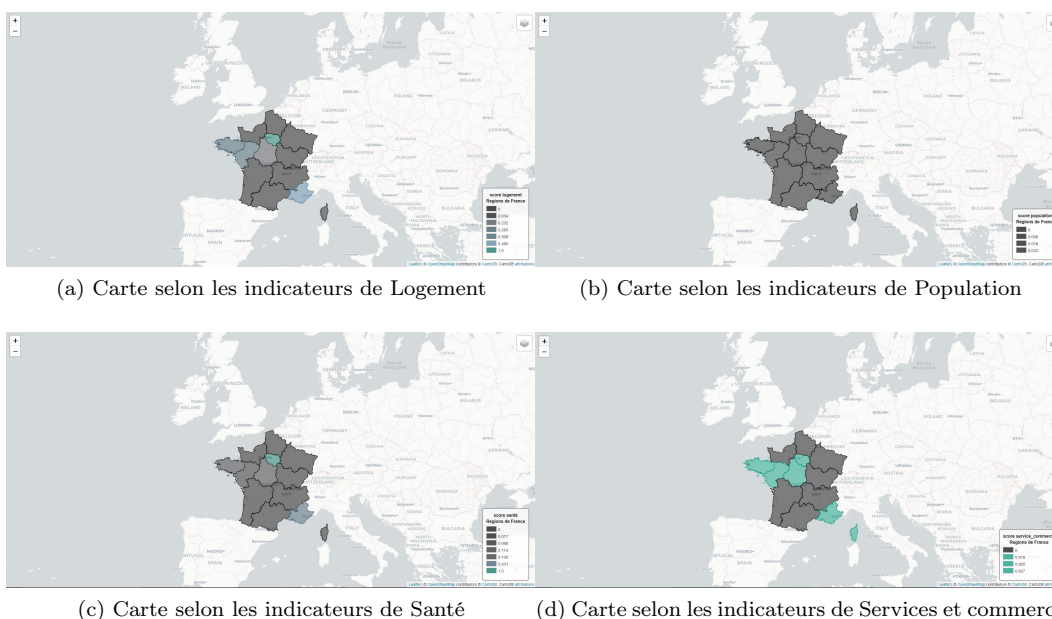


FIGURE 3 – Carte selon les indicateurs

5 Conclusion

Au terme de ce projet on a pu faire le traitement des données sur une base officielle de INSEE, ce qui nous a amené à travailler sur les différents aspects d'une analyse en utilisant différents logiciels et outils.

Pour la conclusion, on représente la carte de Figure 4 selon l'indicateur final qui s'est calculé en additionnant tous les indicateurs des différentes tables.

Cet indicateur nous montre le classement des régions selon leur potentiel pour un investissement rentable dans le service et le commerce.

Selon le résultat de cette analyse, on voit bien que la région de l'Île de France a le meilleur potentiel parmi toutes les régions. Ce potentiel diminue de plus en plus qu'on va vers le sud et l'est et qu'il devient important dans la région Île de France et son entourage.

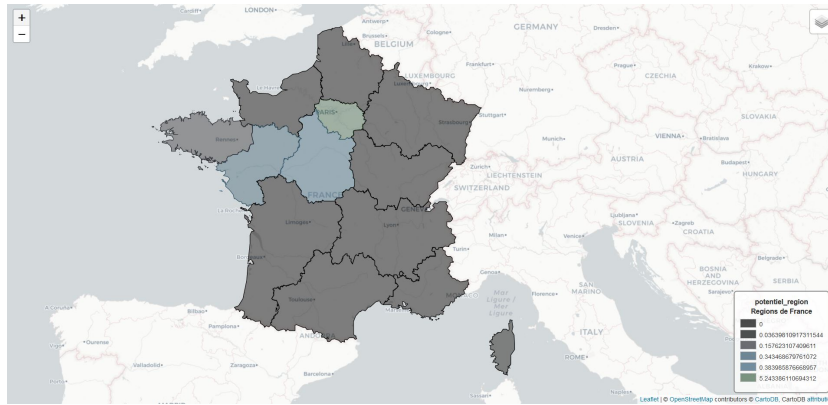


FIGURE 4 – Potentiel des régions

Annexes

[1] Dynamique Entrepreneuriale se représente par le volume de création de nouvelles entreprises et/ou de nouvelles idées. La dynamique entrepreneuriale est proportionnelle au nombre de nouvelles entreprises créées et/ou de nouvelles solutions commerciales développées, visant à résoudre des problèmes existants.

[2] Score de PIB (Produit Intérieur Brut) est un indicateur économique permettant de mesurer la production de richesses d'un pays, le produit intérieur brut mesure la valeur de tous les biens et services produits dans un pays sur une année.

[3] Score de VA (Valeur Ajoutée) représente la richesse nouvelle produite par l'entreprise lors du processus de production qui pourra être répartie sous forme de revenus. Elle permet de calculer la richesse brute créée par une entreprise, avant rémunération de ses salariés, de ses apporteurs de capitaux et des administrations.

[4] Taux d'évasion de la clientèle désigne généralement la part des dépenses commercialisables des clients d'une zone de chalandise qui est dépensée dans des commerces hors zone.

[5] CartoDB ou CARTO est une plate-forme de cloud computing qui fournit des outils de cartographie Web et de science des données spatiales.

[6] jQuery est une bibliothèque JavaScript

Bibliographie

1. <https://www.igi-global.com/dictionary/the-influence-of-the-local-ecosystem-on-entrepreneurial-intentions/84237>
2. <https://www.vie-publique.fr/fiches/270043-quest-ce-que-le-produit-interieur-brut-pib>
3. <https://www.lafinancepourtous.com/decryptages/entreprise/gestion-et-comptabilite/comptes-de-l-entreprise/partage-de-la-valeur-ajoutee-et-partage-des-profits/valeur-ajoutee-qu-est-ce-que-c-est/>
4. <https://www.definitions-marketing.com/definition/taux-d-evasion-commerciale/>
5. <https://slideplayer.fr/slide/1212990/>
6. <https://en.m.wikipedia.org/wiki/CartoDB>