Analyse comparative de la Base Adresse Nationale (BAN) et d'OpenStreetMap (OSM) dans le cadre du géocodage de magasins en France métropolitaine

Jérémy Rousseau

KAPP-RETAIL
Villeurbanne

Mémoire de Fin d'études

Année Universitaire 2022 - 2023

Analyse comparative de la Base Adresse Nationale (BAN) et d'OpenStreetMap (OSM) dans le cadre du géocodage de magasins en France métropolitaine

ROUSSEAU Jérémy

Bing Image Creator

Tuteurs :	Évaluateurs
Pédagogique : FIORINO Humbert	Président du Jury FIORINO Humbert
Entreprise : HIPPOLYTE Jérôme	Paule - Annick DAVOINE

Remerciements

Je tiens à remercier tout particulièrement M. Jérôme HIPPOLYTE, mon maître d'apprentissage, ainsi que les autres membres du Pôle IT pour m'avoir offert leurs aides, leurs conseils avisés et pour leurs écoutes.

Je remercie également M. Mickael BENARROUCH, Directeur et Fondateur de KAPP RETAIL de m'avoir accueilli au sein de sa structure.

Merci à tout le personnel de KAPP RETAIL et en particulier ceux du Pôle IT et Marketing pour l'accueil chaleureux.

Merci également à M. Humbert FIORINO mon tuteur pédagogique, pour sa disponibilité, ses conseils et son suivi durant mes périodes en apprentissage.

Enfin, un dernier mot pour remercier Mme. Amélie FIRMIN de la Direction des Ressources Humaines pour sa contribution au bon fonctionnement de mon stage en apprentissage au sein de KAPP RETAIL.

Résumé du rapport :

Ce mémoire présente une réflexion sur le choix le plus adapté entre deux sources de données, la Base Adresse Nationale (BAN) et OpenStreetMap (OSM), pour le géocodage des magasins en France métropolitaine. L'auteur propose de comparer ces sources selon plusieurs critères tels que la couverture géographique, la précision des données, la qualité des données, la fréquence de mise à jour et la disponibilité de données supplémentaires pour les magasins. Pour mener cette analyse à bien, ce travail utilise le géocodeur Nominatim pour OSM et l'API Adresse d'Etalab pour la BAN sur un échantillon de 385 magasins. Les résultats sont évalués en fonction des critères précédemment énoncés. Enfin, ce travail décrit également les étapes de la réflexion, qui comprennent un rappel des enjeux et du contexte et la méthodologie mise en place pour cette analyse comparative.

Mots-clés libres :

Géocodage ; Magasin ; BAN ; OSM ; Comparaison ; France

Abstract :

This thesis presents a reflection on the most suitable choice between two data sources, the National Address Base (BAN) and OpenStreetMap (OSM), for the geocoding of stores in metropolitan France. The author proposes to compare these sources according to several criteria such as geographical coverage, data accuracy, data quality, update frequency and the availability of additional data for stores. To carry out this analysis, this work uses the Nominatim geocoder for OSM and the Etalab Address API for BAN on a sample of 385 stores. The results are evaluated according to the previously stated criteria. Finally, this work also describes the stages of reflection, which include a reminder of the issues and the context and the methodology put in place for this comparative analysis.

Keywords :

Geocoding; Store ; BAN; OSM; Comparison ; France

Tables des matières

Table des illustrations

Introduction

Implanter un magasin[1] n'est pas un acte anodin, car cela suppose un investissement financier, souvent important. Plusieurs critères entrent donc en jeu pour éviter un échec commercial, le choix de l'emplacement en fait partie.
En effet, l'emplacement du magasin est crucial pour son succès. Choisir le mauvais emplacement peut entraîner un faible trafic de clientèle lié à une mauvaise visibilité, une difficulté d'accès ou encore une concurrence féroce, ce qui peut fortement impacter les ventes et les bénéfices du point de vente.

Il est donc important de mener une recherche approfondie pour trouver ces emplacements attractifs. Une première étape, peut-être tout simplement de générer toutes les zones commerciales, afin de détecter les zones à développer et donc intéressantes pour l'implantation d'un nouveau magasin, des zones saturées.
Cette génération se fait à l'aide de la position des magasins déjà existants. Effectivement, avoir une concentration de magasin proche les uns à côté des autres est souvent synonyme de zones commerciales. Ensuite, il nous reste plus qu'à délimiter plus ou moins facilement les contours de ces zones avant de les analyser pour détecter les zones à développer et donc intéressantes.

Tout ce cheminement s'appuie sur la connaissance de la position des magasins existants. Une position incorrecte d'un ou plusieurs magasins induit donc inévitablement une zone commerciale biaisée et par conséquent une analyse final non-viable. Pour remédier à cet effet « boule de neige », nous devons faire appel à du géocodage. C'est le processus qui consiste à affecter des coordonnées géographiques (longitude/latitude) à une adresse postale. Cette action va permettre de repositionner correctement les magasins et de corriger les zones commerciales.

Il existe une multitude de géocodeur pour faire ce repositionnement. Ces outils utilisés pour faire du géocodage s'appuient généralement sur plusieurs sources de données pour effectuer leur travail. Chaque source a ses spécificités, ses avantages et ses inconvénients. Elles peuvent ne pas donner le même résultat pour une adresse postale identique. Il est donc important d'utiliser la source de données la plus adaptée aux géocodages de magasins en France métropolitaine.

Deux sources de données sortent du lot du fait de leur popularité, de leur libre accessibilité et de leur libre réutilisation (à condition de citer la source). La Base Adresse Nationale (BAN) qui est une base de données ayant pour vocation à réunir l'ensemble des adresses géolocalisées du territoire national. OpenStreetMap (OSM) qui est quant à elle une base de données collaborative qui est constamment mise à jour par des contributeurs.
Nous allons donc comparer la BAN et OSM dans le cadre du géocodage de magasins en France métropolitaine et analyser quelle source de données est la plus adaptée à ce sujet.

Pour y répondre et les différencier, nous nous baserons sur plusieurs critères :

- **La couverture géographique** : on compare la précision et l'exhaustivité de la couverture géographique des deux sources de données, en examinant leur capacité à couvrir l'ensemble du territoire français métropolitain

[1] Établissement commercial où des marchandises sont exposées et vendues

    - **La couverture des zones rurales** : on évalue la capacité de chaque source de données à couvrir les zones rurales, en comparant la disponibilité et la précision des données de géolocalisation pour les magasins situés dans les zones rurales
- **Précision des données** : on évalue la précision de la position géographique des magasins dans chaque source, en comparant les écarts entre la position réelle du magasin et la position géographique fournie par chaque source
- **Qualité des données** : on évalue la qualité des données de géolocalisation dans chaque source, en examinant la quantité de données manquantes ou erronées dans les deux sources.
- **Fréquence de mise à jour** : on compare la fréquence de mise à jour des données de géolocalisation dans chaque source, en examinant la régularité et la rapidité des mises à jour de données dans les deux sources.
- **La disponibilité de données supplémentaires** : on évalue la disponibilité de données supplémentaires dans chaque source, telles que par exemple les horaires d'ouverture des magasins ou les évaluations et commentaires des clients.

Pour mener à bien cette analyse comparative, j'utiliserais le géocodeur Nominatim pour la source de données OSM et l'API[2] BAN d'Etalab sur un échantillon représentatif de 385 magasins. J'analyserais ensuite les résultats et les évaluerais en fonction des critères cités précédemment pour déterminer quelle source de données est la plus adaptée pour le repositionnement des magasins. Plus précisément, nous nous baserons sur des scripts Python et la librairie Geopy pour exécuter les géocodeurs sur cet échantillon des magasins disponibles en base de données. Nominatim et l'API BAN d'Etalab ont été utilisés dans le cadre de mon travail de génération des zones commerciales, il s'agit donc d'un choix arbitraire. Cette analyse s'intéresse aux sources de données (BAN et OSM) et n'évalue pas les géocodeurs et leurs traitements.

Ainsi, cette réflexion est construite en quatre étapes : nous verrons dans un premier temps un rappel plus détaillé des enjeux, du contexte et des critères d'évaluation. Ce rappel sera accompagné aussi d'une présentation des données et du cahier des charges. Dans un second temps, on parlera de la revue de littérature. Dans un troisième temps, nous nous intéresserons à la BAN et à l'évaluation des résultats du géocodeur d'Etalab vis-à-vis des critères d'évaluation retenus. Enfin, dans un dernier temps, nous découvrirons le cas d'OSM avec Nominatim et ses résultats comparés à ceux de la BAN.

[2] API est un acronyme qui signifie (en anglais) Application Programming Interface. Une API, permet à un ordinateur de demander une information à un autre ordinateur, par internet. (Source : api.gouv.fr)

1 Contexte, démarche et critères d'évaluation

Dans cette première partie, nous allons tenter de comprendre plus en détail le contexte et les enjeux de cette analyse comparative de la Base Adresse Nationale et d'OpenStreetMap dans le cadre du géocodage de magasins en France métropolitaine. Pour clarifier ce sujet, nous allons dans un premier temps nous replacer dans le contexte et les enjeux de cette analyse comparative puis, nous retracerons les démarches mise en place pour y répondre avant de définir les critères d'évaluation.

1.1 Contexte et enjeux de l'analyse comparative

Pour aider les enseignes dans la recherche du meilleur emplacement pour de nouveaux magasins, nous nous basons sur les zones commerciales. Il s'agit des espaces où se concentre plusieurs activités commerciales. Elles sont réparties en 3 grandes familles :

- Les zones commerciales de type centre-ville

Une zone commerciale de type centre-ville est un quartier d'une ville où il y a une concentration élevée de magasins, de restaurants et d'autres entreprises commerciales. C'est souvent le cœur de la ville, où se trouvent les principales attractions touristiques, les bureaux administratifs, les hôtels et les espaces publics tels que les places et les parcs.
Les centres-villes sont généralement très animés, surtout pendant les heures de pointe. Ils attirent une grande variété de personnes, des résidents locaux aux touristes en passant par les travailleurs du centre-ville. C'est un lieu de rencontre et de socialisation, où l'on peut se promener, faire du shopping, manger et boire, et profiter de la vie urbaine.
Les centres-villes peuvent être très attractifs pour les entreprises en raison de leur emplacement central et de leur visibilité. Cependant, ils peuvent également être confrontés à des défis tels que la congestion automobile, le manque de places de stationnement et des loyers commerciaux élevés.

jean-louis Zimmermann - flickr

Figure 1 : Centre-ville de Montélimar (26)

- Les zones commerciales de type retail-park

Une zone commerciale de type retail-park est un ensemble de magasins et de boutiques qui sont regroupés en un seul lieu, en dehors du centre-ville, souvent le long d'une route principale ou à proximité d'une autoroute. Les retail-parks sont généralement construits sur des terrains

vastes et plats, et comprennent souvent des zones de stationnement importantes pour les voitures.
Les magasins dans un retail-park sont souvent de grande taille et appartiennent à des chaînes de distribution, tels que les supermarchés, les magasins de bricolage, les magasins d'électronique ou encore les magasins de sport. Les retail-parks peuvent également inclure des restaurants, des cafés, des cinémas et d'autres équipements de loisirs.
Contrairement aux centres-villes, les retail-parks sont généralement conçus pour être accessibles en voiture, avec des parkings spacieux et des accès faciles à partir des routes principales. Ils sont souvent situés à la périphérie des villes ou dans des zones rurales, où les loyers sont moins élevés et où il y a plus d'espace pour construire des bâtiments de grande taille.
Les retail-parks sont populaires auprès des consommateurs qui recherchent des magasins de grande surface et des produits à prix abordables. Ils sont également pratiques pour les achats de gros ou pour les achats de dernière minute. Cependant, ils peuvent être moins attractifs pour les commerçants qui cherchent à être situés dans des endroits avec une forte concentration de population, ou pour les clients qui préfèrent une expérience de shopping plus intimiste et centrée sur les piétons.

commons.wikimedia.org

Figure 2 : Retail-park à Cardiff (Royaume-Uni)

- Les zones commerciales de type centre commerciale

Une zone commerciale de type centre commercial est un complexe commercial qui regroupe de nombreux magasins, boutiques et restaurants sous un même toit. Les centres commerciaux sont généralement situés dans des zones urbaines ou près d'un retail-park.
Les centres commerciaux sont souvent construits avec une grande variété de magasins et de boutiques, allant des grandes enseignes nationales aux petites boutiques spécialisées. Ils peuvent également inclure des cinémas, des restaurants et d'autres équipements de loisirs.
Les centres commerciaux sont généralement conçus pour offrir une expérience de shopping pratique et confortable pour les clients. Ils disposent souvent de grands parkings, d'une climatisation, de larges couloirs, d'ascenseurs et d'escalators pour faciliter la circulation des visiteurs.
Les centres commerciaux peuvent être très attractifs pour les commerçants, car ils offrent une grande visibilité, une forte concentration de clients et une sécurité pour les biens et les personnes. Ils peuvent également être une destination populaire pour les achats de vacances ou pour les achats en gros.

Cependant, les centres commerciaux peuvent également présenter des inconvénients, tels que des loyers commerciaux élevés, une concurrence féroce entre les magasins et une certaine redondance dans les offres de produits. De plus, ils peuvent être perçus comme une menace pour les centres-villes et les petits commerces locaux, qui peuvent avoir du mal à rivaliser avec leur offre et leur taille.

jean-louis Zimmermann - flickr

Figure 3 : Centre commercial Cap Sud (AVIGNON, 84)

Les enseignes vont privilégier ces zones pour implanter leurs nouveaux points de vente, car elles attirent un flux important de clientèle dû aux nombreuses activités commerciales proches les unes des autres. Ainsi, avoir accès à la localisation précise de ces zones est donc une forte valeur ajoutée dans l'aide que l'on peut apporter aux enseignes.

Il n'existe pas de carte ou de source de données exhaustive pour obtenir ces zones commerciales avec les besoins que nous avons. La seule solution est par conséquent de les générer nous-même. Pour ce faire, on peut s'appuyer sur la définition de ce qu'est une zone commerciale, un espace où se concentre plusieurs activités économiques, une zone peut donc se définir simplement comme plusieurs magasins assez proches de ces voisins sur un espace restreint. Ainsi, en connaissant la localisation des magasins sur le territoire métropolitain et en calculant leur distance au plus proche voisin, on a la possibilité de créer des nuages de point proche qui mettent en avant les surfaces susceptibles d'accueillir des zones commerciales. On peut ensuite rajouter les zones d'occupation du sol de CORINE Land Cover en les fusionnant avec les surfaces obtenues pour rendre les résultats encore plus détaillés.

Tout ce processus pour générer les zones commerciales est basé sur le principe que les informations disponibles sur les magasins sont correctes et fiables dans nos bases de données. Plus précisément, il faut que la position des points de vente soit la plus conforme à la réalité ce qui veut dire, soit au niveau de la « boite à lettre » ou sur le magasin lui-même, car dans le cas inverse le résultat obtenu serait biaisé. Le cas particulier des magasins des centres commerciaux est qu'il est préférable d'avoir ce point de positionnement à l'intérieur de la surface du magasin (sur le magasin lui-même), car la « boite au lettre » pointe souvent sur celle du centre commercial en générale et n'est donc pas très précise.

Voici un exemple de zone commerciale générée mais incorrecte (en rouge), les magasins étaient mal positionnés (ici sur la route) et par conséquent la zone s'est construite autour sans réelle cohérence. La véritable zone se trouve sur la droite et est délimité en vert. Il s'agit ici d'une zone de type centre commerciale avec son parking.

Figure 4 : Zone commerciale incorrecte (en rouge) et correcte (en vert)

En effet, les zones générées vont par la suite servir d'aide à la décision pour les enseignes et si l'on se fonde sur des positions de magasins fausse, nos zones seront incorrectes et non interprétable. Cela montre l'importance d'avoir des données de qualité en amont de nos traitements.

Par conséquent, pour corriger les zones, il nous faut simplement repositionner les magasins à l'aide du géocodage. Cela va nous permettre de replacer les points de vente en affectant les bonnes coordonnées géographiques (longitude/latitude) aux adresses postales des magasins. Il existe une multitude de géocodeur pour effectuer ce travail. Chaque géocodeurs a ses propres spécificités, ses avantages et ses inconvénients vis-à-vis de ses concurrents.

Voici les principales différences entre les géocodeurs :

- **La Source de données :** Les géocodeurs utilisent différentes sources de données pour géolocaliser les adresses. Certaines utilisent des données géographiques propriétaires (par exemple Google Maps), tandis que d'autres utilisent des données géographiques open-source telles qu'OpenStreetMap. Les données de qualité et leur couverture peuvent varier selon les référentiels utilisés.
- **Le Coût :** Les géocodeurs peuvent être gratuits ou payants. Les géocodeurs gratuits peuvent avoir une précision et une couverture plus limitées que les géocodeurs payants qui ont souvent accès à des données de meilleure qualité.
- **Facilité d'utilisation :** Certains géocodeurs peuvent être plus faciles à utiliser que d'autres, offrant par exemple des interfaces plus conviviales

La différence en termes de source de données est la plus impactante dans notre travail de repositionnement des magasins. En effet, les sources de données étant quelque fois différente entre elles (couverture géographique, qualité et exhaustivité des données), il n'est pas rare de ne pas avoir le même résultat pour une adresse postale identique et cela peut poser des problèmes s'il l'on recherche la vraisemblance. Par conséquent, nous nous intéresserons uniquement au contenu de ces sources de données sans chercher à étudier et à évaluer les géocodeurs en tant que telle.

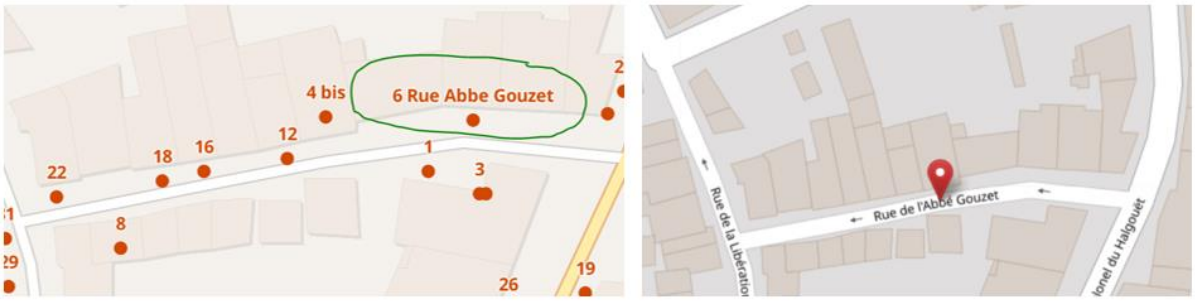

Figure 5 : "6 Rue Abbe Gouzet, Renac 35237" - BAN (à gauche) et OSM (à droite)

Voilà le cœur de la réflexion, trouver le référentiel d'adresse le plus adapté pour le repositionnement de nos magasins. Pour restreindre notre champ d'études, on va s'intéresser aux deux référentiels les plus populaires que sont la Base d'adresse nationale (BAN) et OpenStreetMap (OSM). Ils sont largement connus et utilisés, car étant librement accessibles et réutilisables. Nous détaillerons ces sources de données et les deux géocodeurs utilisés (un pour chaque référentiel) sur notre échantillon représentatif dans leurs parties respectives.

1.2 Les démarches mise en place

Cette partie va présenter dans un premier temps, les informations sur les magasins à disposition et la méthodologie utilisée pour obtenir un échantillon représentatif. Enfin, dans un second temps, on détaillera le cahier des charges utilisé pour répondre à cette analyse comparative.

1.2.1 Les données et l'échantillon représentatif

Dans cette sous-partie, nous allons prendre connaissance des données des magasins à disposition dans les bases de données, puis de la méthodologie appliquée pour obtenir un échantillon représentatif de cette population.

1.2.1.1 Les données

Nous avons actuellement environ 100 000 points de ventes en base de données. Il s'agit de magasin (en succursale ou en franchise) d'enseigne de multiples secteurs d'activités répartie sur le territoire métropolitain. Nous n'avons pas les enseignes indépendantes comme « Chez Stéphane optique » mais plutôt les enseignes « Alain Afflelou » ou encore « Atol les Opticiens ».

Pour chaque magasin, nous connaissons :

- La position XY (longitude latitude)
- Sa dénomination
- Son adresse postale
- Sa surface de vente

L'adresse postale[3] se compose de plusieurs variables. La particularité est que le champ adresse (numéro + nom de la rue ou nom de la zone commerciale) est aussi composée de plusieurs variables. Il y a trois variables : « adresse1 », « adresse2 » et « adresse3 ». En effet, la variable « adresse3 » est la plus précise et ensuite en allant en décroissant, on arrive jusqu'à la variable « adresse1 » la moins détaillée en termes d'information sur l'adresse.

[3] En France, l'adresse postale est composée du nom d'une rue et d'un numéro, du code postal et de la ville. Elle permet la localisation complète du destinataire d'un courier.

nom	adresse1	adresse2	adresse3	code_postal	ville	surface
CORA CAFETERIA	CC CORA CORMONTREUIL	ROUTE DE LOUVOIS		51350	CORMONTREUIL	moyenne
LA FEE MARABOUTEE	3 RUE DU 170 ÉME RI			88000	ÉPINAL	petite
OPTICAL CENTER	PAC GRAND PARC	ZONE DE LA CROISETTE		8000	LA FRANCHEVILLE	petite

Figure 6 : Extrait de l'échantillon

Les adresses que nous avons en base de données, sont généralement incomplètes et de mauvaise qualité. Au-delà de l'absence de données, il y a aussi énormément d'adresses d'usage. Ce sont des adresses que le facteur va connaître, car il saura les interpréter. Mais pour un géocodeur, cette adresse sera généralement mal traduite. Les adresses d'usage sont la particularité des magasins présents en Retail-park et dans les Centre commerciaux comme l'adresse « CC CORA CORMONTREUIL » pour le magasin « CORA CAFETERIA » ou encore «PAC GRAND PARC » pour le magasin « OPTICAL CENTER ». Ces adresses manquent aussi de précision, car elles pointent soit sur le centre commercial en lui-même ou sur la zone du retail-park. Ainsi, nous verrons quels référentiels est plus même de résoudre ces différents problèmes.

La surface de vente a été catégorisée pour répondre aux trois formats de commerces les plus répandus. Les petites surfaces avec une superficie située entre 0 à 400 m². Ensuite, les moyennes surfaces avec des superficies se trouvant entre 400 et 2 500 m². Enfin, nous avons les grandes surfaces avec des superficies qui dépassent 2 500 m².

1.2.1.2 La méthodologie de l'échantillon

Lorsque l'on effectue une analyse comparative, l'objectif est souvent de tirer des conclusions sur une population entière en se basant sur un échantillon de cette population pour gagner en efficacité. Il est donc important que cet échantillon soit représentatif pour que les conclusions que l'on en tire soient également représentatives de la population dans son ensemble.

Cet échantillon est considéré comme représentatif lorsque chaque individu de la population a une chance égale d'être sélectionné pour faire partie de l'échantillon. Si l'échantillon n'est pas représentatif, il peut y avoir des biais dans les résultats de l'analyse comparative qui peuvent conduire à des conclusions erronées.

Par exemple, si l'on veut tirer des conclusions sur un attribut de la population d'une ville, mais que l'on ne sélectionne que des individus d'un seul quartier de cette même ville, les résultats ne seront pas représentatifs de la population dans son ensemble. Si l'on veut que les conclusions soient représentatives, il est important de sélectionner des individus de différents quartiers de la ville. En résumé, avoir un échantillon représentatif permet de garantir que les conclusions que l'on tire de l'analyse comparative sont applicables à la population dans son ensemble.

Dans notre cas, il s'agit d'avoir un échantillon de magasin réparti homogènement sur le territoire métropolitain et distribué en termes de surface de vente de manière équitable. Nous allons maintenant voir comment procéder pour récupérer cet échantillon stratifié à deux niveaux.

Ainsi, pour sélectionner un échantillon représentatif des magasins en France métropolitaine, il est important de suivre une méthodologie rigoureuse pour garantir la représentativité et la qualité de l'échantillon, voici la méthode utilisée pour obtenir un échantillon représentatif en quelques étapes :

1. **Définir la population** : La population est l'ensemble de nos 100 000 magasins de France métropolitaine disponible dans nos bases de données. Il est important de définir clairement la population pour pouvoir ensuite définir l'échantillon.

2. **Déterminer la taille de l'échantillon** : La taille de l'échantillon dépend de plusieurs facteurs, notamment de notre taille de population (100 000), du niveau de confiance souhaité (ici 95%) et la marge d'erreur tolérée (5%). Ces valeurs clés sont les plus fréquemment utilisées lorsque l'on cherche la taille de l'échantillon.

$$Taille\ de\ l'échantillon = \frac{\frac{z^2 * p(1-p)}{e^2}}{1 + \frac{z^2 * p(1-p)}{e^2 * N}} = \frac{\frac{1{,}96^2 * 0{,}5(1-0{,}5)}{0{,}05^2}}{1 + \frac{1{,}96^2 * 0{,}5(1-0{,}5)}{0{,}05^2 * 100\,000}} \approx 383$$

3. **Sélectionner les strates** : Les strates sont les sous-groupes de la population qui ont des caractéristiques similaires. Dans notre cas, on sélectionne les régions pour répondre au critère de localisation géographique puis les surfaces de ventes en respectant les proportions. En d'autres termes, on cherche le nombre de magasins qu'il nous faut dans chaque région pour respecter les proportions de la population. Ensuite, dans chaque région, on affiche le nombre de magasins qu'il nous faut pour chacun des trois types de surfaces de ventes en fonction du nombre de magasins souhaité pour cette région. Vous trouverez un tableau récapitulatif en Annexe 1.

4. **Randomiser la sélection** : Pour éviter les biais de sélection, il est important de randomiser la sélection des magasins dans chaque strate. Dans QGIS, nous utilisons le traitement « Sélection aléatoire » pour sélectionner un certain nombre de magasins aléatoirement dans chaque strate. Un script Python que vous trouverez en Annexe 2 a permis cette automatisation.

5. **Vérifier la qualité de l'échantillon** : Il est important de vérifier la qualité de l'échantillon pour garantir sa représentativité. Pour cela, il est recommandé de comparer les caractéristiques des magasins sélectionnés avec celles de la population de référence. Si l'échantillon est représentatif, les proportions des magasins sélectionnés devraient être similaires à celles de la population.

Le résultat de cette méthodologie est un échantillon de 385 magasins dispatchés homogènement sur le territoire et avec une distribution équitable des surfaces de ventes. Vous trouverez en Annexe 3, un extrait de la liste des enseignes présentes dans cet échantillon.

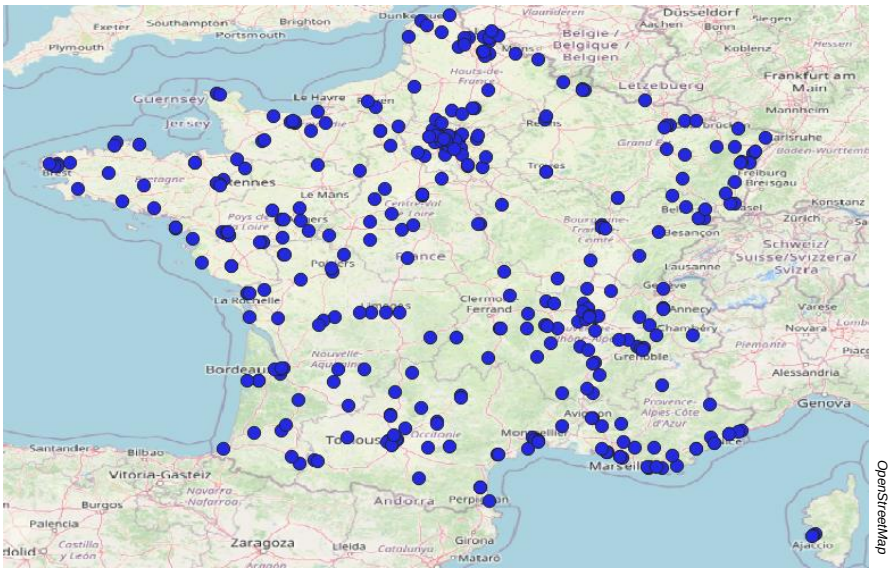

Figure 7 : Carte des 385 magasins de l'échantillon

1.2.2 Le cahier des charges

Maintenant, que nous avons un échantillon représentatif de notre population de magasins, nous allons visualiser le processus mis en place en langage Python pour géocoder nos 385 points de ventes à l'aide des deux géocodeurs choisis, celui d'Etalab et Nominatim. En effet, le géocodage d'une adresse peut sembler simple, vous entrez une adresse et le géocodeur vous renvoie des coordonnées X et Y, mais, en réalité, le processus est plus complexe qu'il n'y paraît pour obtenir un résultat le plus correct possible. Il y a tout un travail de nettoyage / normalisation en amont puis de vérification en aval à exécuter et donc des scripts Python ont été élaboré pour automatiser toutes ces tâches. On va regarder en détail les étapes qui composent ses scripts Python.

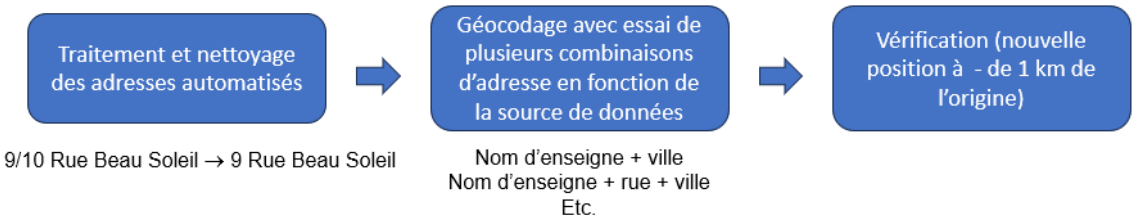

Figure 8 : Schéma général des étapes des scripts Python

Tout d'abord, nous avons voir l'étape de nettoyage et de complétion des adresses présentes dans notre échantillon. Comme dit précédemment, l'adresse postale d'un magasin est répartie sur plusieurs variables que nous devons fusionner pour la reformer. Le cas des variables « adresse1 », « adresse2 » et « adresse3 » se posent, car pour rappel il s'agit du champ adresse avec plus ou moins de précision.
Pour ce faire, on va chercher à détecter et à extraire (si présent) dans chaque variable (en privilégiant la plus détaillée en premier : « adresse3 » puis « adresse2 » et enfin « adresse1 ») :

- La voie de communication : on parle des rues, boulevard, etc.
- Le numéro de la voie de communication
- L'aire : on parle ici des noms de zone commerciale commençant souvent par ZAC, CC, ZONE, etc.
- Le nom du magasin

Ce travail d'extraction est accompagné d'un formatage des champs comme avec le numéro de voie de communication et ses espaces au millier (« 1 234 avenue de … ») ou encore les barre oblique mal gérées par les géocodeurs (« 184/192 rue de … »). Une fois terminé, on peut fusionner les variables pour recréer l'adresse postale. On va ensuite, dans l'optique d'optimiser nos chances d'avoir un résultat positif au géocodage, former plusieurs « modèles » d'adresse postale adaptés à chaque géocodeur.

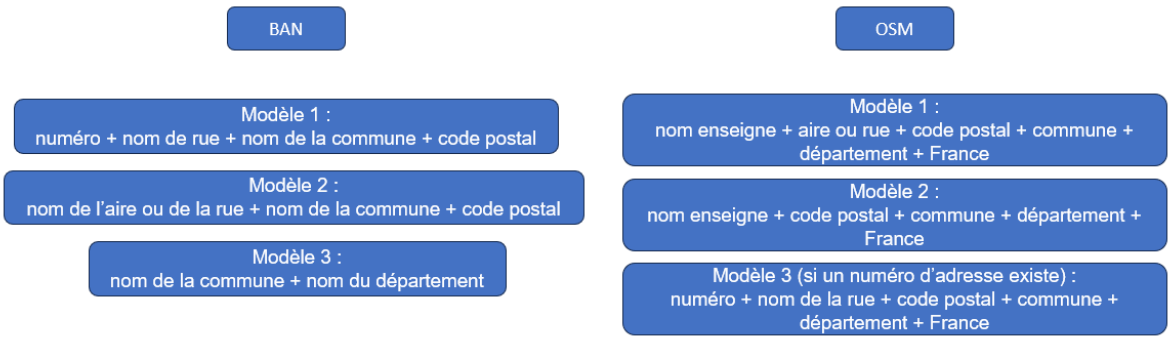

Figure 9 : Les différents modèles d'adresses pour la BAN et OSM

Ces différents modèles d'adresse postale vont nous permettre de voir l'étape suivante du géocodage. Cette étape permet d'automatiser l'envoi des adresses postales auprès des deux géocodeurs. Nous sommes aidés de la librairie Geopy qui va nous servir d'intermédiaire avec les géocodeurs. Geopy permet d'utiliser une multitude de géocodeur en Python très facilement. Ainsi, nous utilisons les classes Geopy, Nominatim et BANFrance qui font appel au géocodeur Nominatim et à l'API BAN fourni par Etalab.
Chaque adresse postale est envoyée au géocodeur pour qu'on puisse en retour obtenir ses coordonnées géographiques X et Y. Les différents modèles d'adresses nous permettent de tenter plusieurs essais pour maximiser nos chances d'avoir une réponse positive. Afin de respecter les conditions d'utilisation des géocodeurs, nous avons mise en place des pauses entre chaque requête auprès des géocodeurs. Ces pauses retardent l'exécution du processus pour éviter de dépasser la limite de requête et éviter les abus et la surcharge du service.

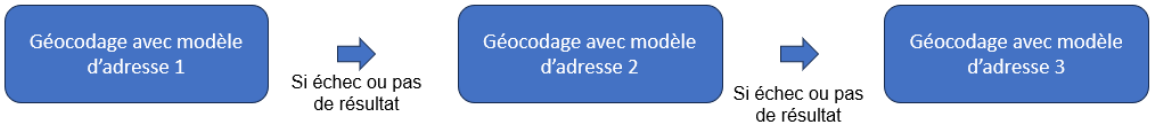

Figure 10 : Les différents essais de géocodage avec les modèles d'adresses

Afin de vérifier la cohérence de nos résultats, une dernière étape a été ajoutée. Cette étape de vérification mesure la distance à vol d'oiseau entre le point d'origine du magasin et la nouvelle position obtenue à l'aide du géocodage. Nous avons choisi une distance de sécurité de 1 km. Si la distance entre nos points est supérieure à la distance de sécurité, nous ne prenons pas de risque et nous transformons ce résultat en échec. Cela évite d'avoir un repositionnement trop éloigné de la position d'origine du magasin.

1.3 Les critères d'évaluation

Pour mener à bien cette analyse comparative de la BAN et d'OpenStreetMap, nous avons besoin de critères d'évaluation. Ils sont essentiels dans notre analyse, car ils permettent de mesurer et de comparer objectivement les performances, les caractéristiques et les qualités des deux référentiels. Nous allons donc voir ensemble en détail ces critères un par un.

1.3.1 La couverture géographique

La couverture géographique des référentiels est un critère important dans notre analyse comparative. En effet, nous allons examiner leur capacité à couvrir l'ensemble du territoire français métropolitain, notre périmètre d'étude. Notre échantillon proposant des magasins répartis sur l'ensemble de ce territoire, il sera facile de détecter si des zones administratives ne sont pas couvertes par les référentiels en ne proposant pas de réponse à nos adresses postales de magasins.
La couverture des zones rurales sera aussi étudiée. Sur cet aspect sous-jacent à la couverture géographique, il sera l'objet d'évaluer la capacité de chaque source de données à couvrir les zones rurales. Ces zones à faible densité de population seront détectées à l'aide des aires d'attraction des villes. Il s'agira de retenir les communes rurales hors influence d'un pôle (qui sont des communes n'appartenant pas à une aire d'attraction). Enfin, nous détecterons la disponibilité des données de géolocalisation pour les magasins situés dans ces zones et nous comparerons les résultats des deux référentiels.

1.3.2 La Précision des données

Le critère de précision des données va mettre en avant la fiabilité de la position géographique des magasins dans chaque source. En partant du principe que la position actuelle des magasins de notre échantillon est correcte. Nous allons évaluer cette précision en comparant les écarts entre la position réelle du magasin et la position géographique fournie en réponse par chaque référentiel. Plus la moyenne des écarts sera faible, plus le référentiel sera précis.
Nous distinguerons la précision des données en fonction du modèle d'adresse retenue par le géocodeur. Pour rappel, les modèles ont été rapidement présentés dans la partie 1.2.2 et seront plus détaillés par la suite.

1.3.3 La qualité des données

Le critère de qualité des données permettra de vérifier la justesse des données renvoyées par les référentiels. On quantifiera le nombre de données manquantes ou erronées renvoyées par les sources de données en se concentrant particulièrement sur les adresses d'usage des magasins des retails-parks et des centres commerciaux. Ensuite, nous chercherons des explications pour comprendre ces erreurs et ces manquements.

1.3.4 La fréquence de mise à jour

Ce critère compara la fréquence de mise à jour des données de géolocalisation dans chaque source grâce à leur documentation. On examinera la régularité et la rapidité des mises à jour de données. En effet, les données de géocodage sont sujettes à des changements constants, tels que de nouvelles constructions, des modifications de l'infrastructure routière, des changements de noms de rue, etc. Une fréquence de mise à jour élevée des données permet de garantir que les résultats de géocodage sont basés sur des informations récentes et précises.

1.3.5 La disponibilité de données supplémentaires

Ce dernier critère permet de vérifier si les référentiels proposent des données supplémentaires pour les magasins. Par défaut, les seules données misent à disposition sont les coordonnées géographiques et les adresses postales. Des informations supplémentaires telles que les horaires d'ouverture des magasins ou les évaluations et commentaires des clients seraient une forte valeur ajoutée.

1.4 Conclusion de la première partie

Nous avons présenté dans ce chapitre le contexte, les démarches mises en place pour répondre à notre problématique et enfin les critères d'évaluation utilisés pour départager nos deux référentiels. Le contexte nous a rappelé l'importance du positionnement des magasins dans la génération des zones commerciales. Des localisations incorrectes impliquent une génération des zones biaisée et non conforme à la réalité. La correction de la position des magasins se fait à l'aide du géocodage en transformant les adresses postales des points de vente en coordonnées géographiques X et Y. Les géocodeurs, les outils permettant de faire cette transformation sont nombreux et s'appuient sur des référentiels d'adresse différents. Chaque référentiel a ces particularités. La question est donc de savoir quels référentiels est le plus adapté au géocodage d'adresse de magasin. Nous nous concentrerons sur les deux plus populaires référentiels d'adresses libres OpenStreetMap et la Base Adresse Nationale. Pour mener à bien notre analyse comparative, nous avons mis en place un échantillon représentatif de nos données sur les magasins. Cet échantillon est par la suite utilisé dans des scripts Python permettant la normalisation automatique des adresses postales et leurs envoie au géocodeurs. Les résultats obtenus seront enfin évalués par cinq critères pour départager les deux référentiels. Le chapitre suivant dressera un état de l'art des travaux déjà effectués dans ce domaine.

2 La revue de littérature

Le géocodage est une technique utilisée pour associer des coordonnées géographiques à des adresses postales. Il existe une multitude d'outils de géocodage, appelés géocodeur. Ces géocodeurs ont des spécificités et des sources de données différentes. Nous allons nous intéresser à deux sources de données en particulier : la Base d'adresse Nationale et OpenStreetMap. Cette revue de littérature vise à explorer les ouvrages traitant d'une comparaison des géocodeurs ou des sources de données étudiées.

Comparaison des géocodeurs (Vandy Berten, 2015) :
Bien que cette source se concentre sur la Belgique, elle offre une comparaison générale et initiale des géocodeurs, ce qui peut fournir un aperçu général des critères à prendre en compte dans une analyse comparative plus approfondie.

Les logiciels et API pour géocoder (Grégory Gibelin, 2019) :
Cette source fournit une comparaison technique des géocodeurs, bien qu'elle date de 2019. Elle peut être utilisée comme point de référence pour comprendre les caractéristiques et les fonctionnalités à considérer lors de l'analyse comparative de la BAN et d'OSM.

OpenStreetMap (Jonathan Bennett) :
Cet ouvrage offre une introduction complète à OSM, explorant son histoire, ses principes, sa communauté et ses utilisations. Il met en évidence les avantages et les défis associés à cette plateforme de cartographie collaborative, qui peuvent être pertinents pour évaluer OSM.

Quality Assessment of the French OpenStreetMap Dataset (Jean-François Girres et Guillaume Touya) :
Cette étude évalue la qualité du jeu de données OSM pour la France. Elle souligne l'avantage de la réactivité et de la flexibilité d'OSM, mais aussi les problèmes liés à l'hétérogénéité des données et à la nécessité de suivre un cahier des charges bien défini. Ces résultats peuvent fournir des informations sur les limites potentielles d'OSM pour le géocodage en France métropolitaine.

Apports et limites d'OpenStreetMap pour l'analyse spatiale des équipements commerciaux en zone transfrontalière (Marianne Guérois et al.) :
Cet ouvrage examine l'utilisation d'OSM pour la cartographie des équipements commerciaux en zone transfrontalière. Elle met en évidence des aspects tels que la fiabilité spatiale, l'hétérogénéité régionale, les limites de couverture et la qualité des données. Ces informations peuvent être utiles pour comprendre comment OSM peut être appliqué au géocodage des magasins en France métropolitaine.

De l'adoption au rejet d'un commun numérique pour transformer la frontière entre État et citoyens (Sébastien Shulz) :
Cette source aborde la question de l'adoption de la forme sociotechnique du commun numérique, telle que la Base Adresse Nationale (BAN). Elle met en avant les raisons de son adoption, les changements dans la relation entre l'État et la société, ainsi que les résistances institutionnelles. Ces éléments permettre de comprendre l'origine de la BAN et la reconfiguration de la relation État/citoyen.

Géocodage / calcul de temps de parcours pour les communes de la base « mobilité professionnelles » (INSEE) (Matthieu Viry et al.) :
Cette source présente un cas d'usage du géocodage dans les communes étrangères, utilisant OSM, tandis qu'en France, la BAN est utilisée. Cet ouvrage montre l'importance de l'utilisation de la BAN et d'OSM dans un usage de géocodage.

La présente revue de littérature a examiné divers ouvrages et études portant sur les géocodeurs en générales, OpenStreetMap (OSM) et la Base Adresse Nationale (BAN). Malgré, le peu d'œuvre récente sur le sujet et donc la difficulté de voir l'état actuel de notre périmètre d'étude, les résultats et les conclusions des différentes sources permettent quand même de tirer quelques observations clés.

Tout d'abord, en ce qui concerne OSM, il a été souligné que cette plateforme de cartographie collaborative présente à la fois des avantages et des défis. La réactivité et la flexibilité d'OSM sont des atouts indéniables, mais l'hétérogénéité des données provenant de différentes sources et contributeurs, limite ses applications. L'établissement d'un cahier des charges bien défini peut contribuer à résoudre cette problématique, mais il reste encore des défis à relever, notamment en ce qui concerne la vérification automatique de la cohérence des contributions. De plus, une évaluation plus précise de la qualité des contributions d'OSM dans des contextes géographiques spécifiques est nécessaire.

En ce qui concerne la BAN, il est évident qu'il existe un manque d'ouvrages décrivant en détail cette base de données spécifique. Cependant, il a été mentionné que la BAN est privilégiée pour le géocodage en France métropolitaine, tandis qu'OSM est utilisé principalement pour le géocodage à l'étranger. Cette distinction met en évidence l'importance de disposer de bases de données nationales de référence pour garantir une couverture complète et fiable des informations géographiques.

En somme, une analyse comparative de la BAN et d'OSM dans le contexte du géocodage des magasins en France métropolitaine peuvent être une piste intéressante de recherche. Cette étude pourrait permettre de mieux comprendre les avantages et les limites de chaque système, en mettant par exemple l'accent sur la fiabilité spatiale, l'hétérogénéité régionale, la couverture des données, la qualité des informations et les perspectives d'amélioration.

En définitive, cette revue de littérature montre l'intérêt d'une évaluation approfondie des sources de données telles qu'OSM et la BAN dans le contexte spécifique du géocodage des magasins en France métropolitaine. Une meilleure compréhension des forces et des faiblesses de chaque système contribuera à améliorer la qualité de l'utilisation du géocodage et à soutenir les décisions dans le domaine des analyses commerciales et de la planification territoriale.

3 La Base Adresse Nationale

Dans cette troisième partie, nous allons nous intéresser à notre premier référentiel d'adresse, la Base Adresse Nationale (BAN). Pour clarifier ce sujet, nous verrons dans un premier temps une présentation générale de cette source de données. Nous retracerons les enjeux de sa création, ses auteurs et sa constitution. Dans un second temps, nous parlerons rapidement de l'API BAN d'Etalab aussi appelé API Adresse. Enfin, nous détaillerons les résultats obtenus en les évaluant à partir des critères mise en place.

3.1 Présentation générale de la BAN

La Base Adresse Nationale est une base de données ayant vocation à réunir l'ensemble des adresses géolocalisées du territoire national. Elle fait partie du Service Public des Données de référence. Il vise à mettre à disposition, en vue de faciliter leur réutilisation, les jeux de données de référence qui présentent le plus fort impact économique et social comme la Base SIRENE (fournisseuse des données d'identité des entreprises et des établissements).

Elle a été mise en ligne le 15 avril 2015. La création de cette base résulte d'un modèle innovant de collaboration entre pouvoirs publics, acteurs publics et société civile dans l'objectif de contribuer à la modernisation de l'action publique gouvernementale et territoriale en matière d'open data. Les quatre fondateurs sont :

- Etalab qui est une administration publique visant à améliorer le service public et l'action publique grâce aux données. Cette administration développe et maintient notamment la plateforme nationale des données ouvertes data.gouv.fr. Etalab est un département de la direction interministérielle du numérique (DINUM).
- L'Institut national de l'information géographique et forestière (IGN) est un établissement public ayant pour mission d'assurer la production, l'entretien et la diffusion de l'information géographique de référence en France.
- Le Groupe La Poste est une société anonyme[4] française principalement présente en tant qu'opérateur de services postaux.
- L'association OpenStreetMap France est une association à but non-lucratif, dont l'objectif est de promouvoir le projet OpenStreetMap et notamment la collecte, la diffusion et l'utilisation de données cartographiques sous licence libre.

Placé sous le co-pilotage de la DINUM, de l'Agence nationale de la cohésion des territoires (ANCT) et de IGN, la mise à disposition de la BAN constitue désormais un référentiel cartographique clé pour l'économie, la société et l'ensemble des services publics. En effet, cette base répond à de nombreux enjeux en garantissant notamment que les citoyens bénéficieront des meilleures conditions en terme :

- De secours aux personnes en assurant par exemple aux services d'urgence d'arriver au bon endroit
- De déploiement des réseaux en permettant aux opérateurs publics et privés de mieux coordonner leurs chantiers
- De livraison du courrier et des colis en améliorant la qualité de l'adressage
- D'évolution des services publics de proximité (carte scolaire, santé…)
- De partage de la donnée en vue de réutilisation dans le cadre de la réalisation d'une analyse cartographique

Autant d'éléments qui font de ce référentiel un véritable enjeu de société pour la France, aussi bien d'un point de vue de souveraineté, que d'un point de vue économique. Ce premier

[4] La société anonyme est une société de capitaux. Elle réunit des actionnaires qui investissent dans le capital de l'entreprise. (source : economie.gouv.fr)

géocodeur français propose une nouvelle solution aux besoins de géolocalisation comparable aux dispositifs des géants du web comme Google ou Microsoft.

Ainsi, cette base consiste à associer aux 25 millions d'adresses recensées sur le territoire français des coordonnées géographiques et gérer les 200 à 300 000 adresses créées chaque année. Elle ne contient donc aucune donnée nominative, mais renseignent seulement la position géographique des adresses et des lieux-dits. Elle a été construite à partir de données provenant d'acteurs historiques de l'adresse comme La Poste, l'IGN, la Direction Générale des Finances Publiques (DGFIP) avec le cadastre par exemple. Elles sont issues aussi des Bases Adresses Locales (BAL) qui sont l'inventaire des adresses créé par les communes. À terme, ces dernières devraient devenir la seule source.

En effet, les communes, par l'intermédiaire de leur Conseil municipal, sont les seules autorités compétentes dans la création des voies et des adresses. Il est courant pour ces communes de déléguer la compétence de gestion des adresses à des EPCI ou aux départements. La commune doit ensuite certifier ces adresses dans la BAL, c'est-à-dire valider que les adresses saisies sont justes. Une adresse certifiée est déclarée authentique par la mairie, ce qui renforce la qualité de la Base Adresse Locale et donc celle de la Base Adresse Nationale. Une des particularités de la BAL est que n'importe qui peut commencer sa BAL au format brouillon, mais la commune doit à la fin toujours valider cette BAL rendant difficile la contribution citoyenne.

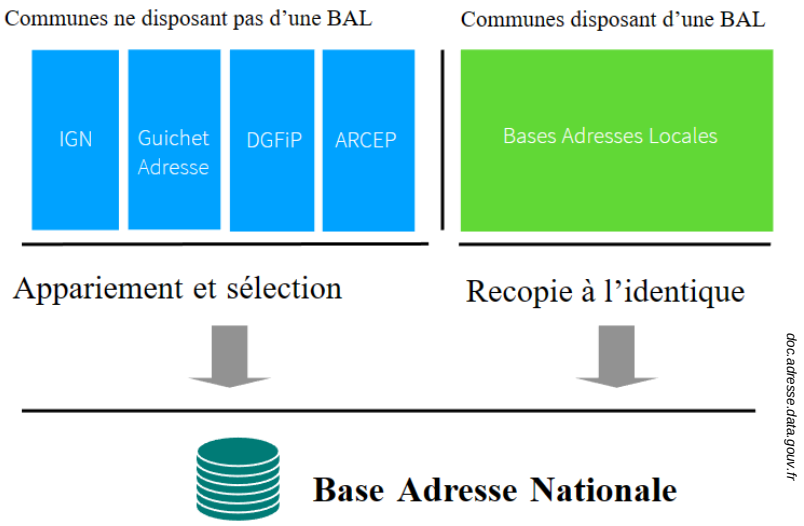

Figure 11 : Schéma sur la provenance des données de la BAN

La Base Adresse Nationale est accessible sous forme de fichiers et d'API sur le portail unique de référence adresse.data.gouv.fr. Ce référentiel étant la seule base de données d'adresses officiellement reconnue par l'administration, elle remplace par exemple la BD Adresse de l'IGN. Ainsi, cette plateforme est le point d'accès vers :

- La Base Adresse Nationale
- Les outils et les services permettant aux communes de mettre à jour leurs adresses et de les publier dans leur Base Adresse Locale

Dans le cadre de notre analyse comparative, nous utilisons l'API pour accéder à la BAN. Il s'agit d'un service de géocodage gratuit mis à disposition par Etalab.

Le 1er janvier 2020, les données Base Adresse Nationale sont sous licence ouverte. Cela a fortement aidé dans la réutilisation des adresses. Plus précisément, il s'agit de la licence ouverte Etalab 2.0. Cette licence permet d'utiliser librement et gratuitement les données de la BAN. Elle autorise la reproduction, la redistribution, l'adaptation et l'exploitation commerciale des données. Dans un souci de transparence de la donnée et de qualité des sources, il est

obligatoire de mentionner la paternité. Enfin, cette licence s'inscrit dans un contexte international en étant compatible avec les standards des licences Open Data développées à l'étranger et particulièrement celles du gouvernement britannique (Open Government Licence) ou encore les autres standards internationaux comme ODC-BY et CC-BY 2.0. Pour simplifier, la BAN est gratuite et librement réutilisable, à condition de citer la source. La licence ouverte Etalab 2.0 met en place un standard réutilisable par les collectivités territoriales qui souhaiteraient se lancer dans l'ouverture des données publiques.

Pour finir, la BAN est toujours en développement. Une page GitHub récapitule la feuille de route la BAN et explique comment contribuer à son évolution.

3.2 Présentation de l'API BAN d'Etalab

L'API BAN d'Etalab (aussi appelé API Adresse) est un service public ouvert à tous permettant d'interroger la base de données de l'intégralité des adresses du territoire français gratuitement, la Base Adresse Nationale. Accessible depuis l'url https://api-adresse.data.gouv.fr/search/, cette API couvre l'ensemble du territoire métropolitain, ainsi que les départements et régions d'outremer. En avril 2021, l'API Adresse enregistrait en moyenne 1 milliard d'appels pour un coût de 35 centimes d'euros par million d'adresses géocodées

Les usages de l'API Adresse ont principalement deux objectifs :

- Trouver par une requête une adresse pour la corriger et/ou récupérer ces coordonnées géographiques XY
- Fournir un fichier tabulaire pour obtenir en retour une version enrichie des coordonnées géographiques XY et d'autres informations (géocodage à partir d'un fichier CSV).

Pour notre analyse comparative, nous utilisons les requêtes pour obtenir les coordonnées géographique XY des adresses questionnées via l'intermédiaire de la librairie Python Geopy. Pour rappel, cette librairie est un client Python pour plusieurs services Web de géocodage populaires comme celui du logiciel SIG Arcgis ou encore du moteur de recherche Bing.

Ce service de géocodage a des limites, car contraint par les données à disposition dans la BAN, des adresses. Il est donc possible de géocoder seulement des adresses de type postal de maison, d'entreprise, etc. Une adresse comme «14 bis avenue Marie Reynoard 38000 Grenoble » fonctionne dans le géocodeur et renverra les coordonnées géographiques suivante [X : 5,728475 ;Y : 45.166164]. Cependant, s'il l'on cherche « Géant Casino, Saint-Martin-d'Hères », l'API renverra en réponse les coordonnées des « Rue du Casino » présentent sur le territoire (car ayant le mot-clé « Casino »). Le service n'est pas capable de récupérer des points d'intérêts (POI en anglais) comme notre « Géant Casino ». Il ne cherche que des adresses de type postales.

Cela explique les différents modèles d'adresses présentés dans la sous-partie 1.2.2 pour l'API Adresse. Ces modèles sont adaptés à cette spécificité du service en privilégiant les adresses de type postales. Nous verrons par la suite qu'OpenStreetMap permet de son côté d'obtenir des coordonnées pour les POI d'où l'intégration du nom d'enseigne dans ces modèles d'adresse.

Pour maximiser les chances d'avoir une réponse même légèrement imprécise, nous avons donc établi trois modèles d'adresse de type postales :

- Modèle 1 : numéro + nom de rue + nom de la commune + code postal
  - Exemple : « 14 BIS, AVENUE MARIE REYNOARD GRENOBLE 38000 »
- Modèle 2 : nom de l'aire ou de la rue + nom de la commune + code postal
  - Exemple : « AVENUE MARIE REYNOARD GRENOBLE 38000 » ou encore « ZAC DE BEAULIEU, PUILBOREAU 17138 »
- Modèle 3 : nom de la commune + nom du département
  - Exemple : « GRENOBLE, ISERE »

Ils fonctionnent par ordre de priorité, si le modèle 1 renvoie rien, on passe au suivant et ainsi de suite. Par conséquent, nous perdons en précision à chaque fois que nous passons au modèle suivant. Nous aurons un géocodage précis au numéro de rue, puis à la rue et enfin à la ville.

L'API renvoie en réponses les coordonnées exprimées en WGS-84 (EPSG 4326) et des attributs dans le format GeoJSON. Les attributs permettent d'avoir des informations supplémentaires sur ce qui est renvoyées et en voici la liste :

- id : identifiant de l'adresse
- type : type de résultat trouvé
  - housenumber : numéro « à la plaque » → il s'agit du résultat attendu pour le modèle 1
  - street : position « à la voie », placé approximativement au centre de celle-ci → il s'agit du résultat attendu pour le modèle 2
  - locality : lieu-dit
  - municipality : numéro « à la commune » position « à la voie », placé approximativement au centre de celle-ci → il s'agit du résultat attendu pour le modèle 3
- score : valeur de 0 à 1 indiquant la pertinence du résultat → nous n'utilisons pas ce score dans notre analyse, car nous avons déjà les positions correctes des magasins pour comparer avec le résultat
- housenumber : numéro avec indice de répétition éventuel (bis, ter, A, B)
- street : nom de la voie
- name : numéro éventuel et nom de voie ou lieu-dit
- postcode : code postal
- citycode : code INSEE de la commune
- city : nom de la commune
- district : nom de l'arrondissement (Paris/Lyon/Marseille)
- oldcitycode : code INSEE de la commune ancienne (le cas échéant)
- oldcity : nom de la commune ancienne (le cas échéant)
- context : n° de département, nom de département et de région
- label : libellé complet de l'adresse
- x : coordonnées géographique en projection Lambert 93 (EPSG : 2154)
- y : coordonnées géographique en projection Lambert 93
- importance : indicateur d'importance (champ technique)

Les attributs ci-dessous sont donc retournés au format GeoJSON. Avec la library Geopy, nous récupérons seulement les coordonnées géographiques en WGS-84 de la « première » adresse. En effet, l'API renvoie plusieurs réponses dans le GeoJSON classé par pertinence et avec le paramètre exactly_one = true, on ne garde la première (la plus pertinente). Vous trouverez en Annexe 4, la réponse au format GeoJSON renvoyé par l'API pour le « 14 bis avenue Marie Reynoard 38000 Grenoble ».

Derrière l'API Adresse, il y a le logiciel open source Addok. Il s'agit du moteur de géocodage développé initialement par Etalab pour la BAN capable d'atteindre environ 2000 recherches par seconde. Ce moteur de recherche a été paramétré pour ne travailler qu'avec des adresses de type postal d'où le fait qu'avec la BAN, il est impossible de chercher des points d'intérêts comme un centre commercial ou une enseigne avec leur simple dénomination, mais seulement avec leur adresse postale.

3.3 Les résultats du géocodage

Lors de notre première session de géocodage, nous avons donc utilisé l'API Adresse pour obtenir les coordonnées géographiques de nos magasins à l'aide des différents modèles

d'adresse postale mise en place. Sur un total de 385 recherches effectuées à partir de notre échantillon, nous avons réussi à obtenir des résultats pour tous les magasins, ce qui représente un taux de réussite de 100%.

Parmi ces 385 magasins traités, nous avons vérifié le respect de la distance maximale de sécurité de 1000 mètres entre les coordonnées obtenues et celle d'origine pour ce résultat. Sur ce critère, nous avons obtenu un total de 330 succès, c'est-à-dire que 330 magasins se trouvaient effectivement à une distance inférieure ou égale à 1000 mètres de leur emplacement d'origine.

Cependant, nous avons également rencontré 55 échecs lors de cette vérification. Ces échecs peuvent être dus à différents facteurs, tels que des adresses non-valides ou des erreurs de géocodage que nous tenterons d'expliquer par la suite. Heureusement, aucun des 55 échecs n'était lié à des adresses non géocodées, ce qui signifie que toutes les adresses ont pu être localisées sur la carte (même s'il s'agit d'une localisation erronée).

En résumé, grâce à l'API Adresse, nous avons obtenu des résultats complets pour l'ensemble des 385 magasins recherchés, avec 330 succès et 55 échecs.

3.4 Evaluation des critères

Dans cette partie, nous allons analyser les différents critères mise en place pour évaluer l'efficacité et la fiabilité de la Base Adresse Nationale. Nous nous pencherons sur les cinq aspects clés : la couverture géographique, la précision des données, la qualité des données, la fréquence de mise à jour et la disponibilité de données supplémentaires.

Tout d'abord, nous examinerons la couverture géographique offerte par l'API Adresse. Il est essentiel de déterminer dans quelle mesure cette API est capable de fournir des résultats dans différentes régions géographiques. Nous évaluerons si elle peut couvrir un large éventail de zones en nous intéressants particulièrement aux zones rurales.

Ensuite, nous nous pencherons sur la précision des données fournies par l'API Adresse. Il est crucial que les résultats obtenus correspondent fidèlement aux emplacements réels. Nous évaluerons la précision des coordonnées géographiques fournies par l'API et vérifierons si elles correspondent aux coordonnées correctes actuellement en base de données.

Nous aborderons ensuite la qualité des données. Cela comprendra l'évaluation de la cohérence et de l'intégrité des informations fournies par l'API. Nous analyserons les échecs en la présence d'erreurs telles que des localisations incorrectes et des informations manquantes dans les données et tenterons de les expliquer. Nous ferons un point sur le cas particulier des adresses d'usages des magasins présents en retail-park et en centre-commercial.

La fréquence de mise à jour sera également un critère essentiel à considérer. Nous évaluerons à quelle fréquence les données de la BAN sont mises à jour. Une mise à jour régulière est essentielle pour garantir la pertinence et l'actualité des résultats obtenus.

Enfin, nous examinerons la disponibilité de données supplémentaires. Il est important de déterminer si la Base Adresse nationale peut fournir des informations complémentaires telles que par exemple les horaires d'ouvertures des magasins qui seraient une valeur ajoutée.

En analysant ces cinq aspects clés sur les résultats de notre échantillon représentatif, nous serons en mesure de fournir une évaluation complète des performances et de sa pertinence pour le géocodage de magasins

3.4.1 La couverture géographique

Lors de l'évaluation de la couverture géographique de la Base Adresse Nationale, plusieurs résultats intéressants ont été observés. Tout d'abord, en ce qui concerne les régions avec le plus fort pourcentage d'échec au géocodage par rapport à leur population, la Bourgogne-Franche-Comté présente un taux d'échec de 38,89%. Cela signifie que près de 39% des adresses dans cette région n'ont pas pu être géocodées avec succès ou la distance de sécurité d'un kilomètre n'a pas été respecté. De même, la Corse affiche un taux d'échec de 33,33%, ce qui représente un pourcentage non négligeable de magasin non repositionné.

En revanche, certaines régions ont affiché un fort pourcentage de succès au géocodage par rapport à leur population totale. La Normandie et le Centre-Val de Loire se distinguent en obtenant un taux de succès de 100%. Cela signifie que toutes les adresses dans ces régions ont été géocodées avec succès et ont passé le test de la distance de sécurité, ce qui est un résultat très positif. De plus, on observe un taux de succès de 88,24% pour les magasins situés en zone rurale et de 85,60% pour les magasins en dehors de zone rurale. Ces chiffres indiquent une bonne capacité de la BAN à localiser les adresses, quelle que soit leur localisation géographique.

En se basant sur la documentation de la BAN et les résultats obtenus, il convient de noter que toutes les régions de France métropolitaine sont couvertes par la Base Adresse Nationale, ce qui signifie que l'API Adresse peut fournir des résultats dans l'ensemble du pays. De plus, il n'y a pas de différence significative entre le géocodage d'un magasin en zone rurale ou non. Cela indique une cohérence et une fiabilité globales de la BAN dans la localisation des adresses, quel que soit le type de zone.

En résumé, bien que certaines régions aient affiché des taux d'échec plus élevés, la base officiel des adresses en France a démontré une couverture géographique globalement solide en localisant avec succès la majorité des adresses à travers le territoire métropolitaine. Ces résultats sont un bon point pour la BAN dans notre comparaison avec OpenStreetMap.

Région	Echec	Succès
Auvergne Rhône-Alpes	13,73%	86,27%
Bourgogne-Franche-Comté	38,89%	61,11%
Bretagne	14,29%	85,71%
Centre-Val de Loire	0,00%	100,00%
Corse	33,33%	66,67%
Grand Est	9,38%	90,63%
Hauts-de-France	6,45%	93,55%
Ile-de-France	11,86%	88,14%
Normandie	0,00%	100,00%
Nouvelle-Aquitaine	12,20%	87,80%
Occitanie	18,42%	81,58%
Pays de la Loire	21,74%	78,26%
Provence-Alpes-Côte d'Azur	24,24%	75,76%
Total général	**14,29%**	**85,71%**

Zone rurale	Echec	Succès
faux	14,40%	85,60%
vrai	11,76%	88,24%

Figure 12 : Tableaux « couverture géographique » pour la BAN

3.4.2 La précision des données

Lors de l'évaluation de la précision des données fournies par la BAN, des résultats significatifs ont été observés. Pour les échecs, la moyenne des distances entre la position géocodée et la position réelle du magasin était de 2348 mètres, après avoir exclu les distances extrêmes du calcul. Cela indique une certaine imprécision dans la localisation des magasins, qui sont mis en échec principalement par le test de la distance de sécurité d'un kilomètre. En revanche, pour les succès, la moyenne des distances entre la position géocodée et la position réelle du magasin était de 141 mètres. Cela démontre une précision plus élevée lorsque la localisation est réussie.

Si on s'intéresse plus précisément aux modèles d'adresse utilisés pour permettre ces succès. On constate que le modèle 1, qui comprend le numéro et le nom de la rue, la commune et enfin le code postal, a obtenu une distance moyenne de 75 mètres. Cela indique une précision plus élevée lorsque nous avons le numéro de rue.

Le modèle 2, qui utilise le nom de l'aire ou de la rue, le nom de la commune et le code postal, a obtenu une distance moyenne de 354 mètres pour les succès. Bien que légèrement moins précis que le modèle 1, montre que le géocodeur parvient à trouver la rue même si nous n'avons pas le numéro de rue exacte avec la BAN.

Parmi les 330 magasins géocodés, 252 ont été localisés avec succès en utilisant le modèle 1, tandis que 78 ont été localisés avec succès en utilisant le modèle 2. Il est intéressant de noter que le modèle 1 a été le plus utilisé et s'est révélé le plus précis dans les réponses fournies par le géocodeur.

En ce qui concerne les échecs, le modèle 2 a été le plus utilisé, et il n'a pas été nécessaire de recourir au modèle 3 (précision à la ville). Le modèle 2 parvient toujours à trouver les rues même s'il ne s'agit pas de la bonne rue.

En résumé, bien que la moyenne des distances pour les échecs soit importante, les succès ont démontré une précision relativement élevée, en particulier lorsque le modèle 1 est utilisé. Ces résultats soulignent l'importance de choisir de bien normaliser l'adresse postale et de former plusieurs modèles appropriés pour obtenir des résultats plus précis et fiables lors de l'utilisation de l'API Adresse.

	Moyenne des distances
Echec	2348
Succès	141

Succès	Moyenne des distances
modele 1	75
modele 2	354
Total général	**141**

Nombres de magasins	Echec	Succès	Total général
modele 1	10	252	**262**
modele 2	45	78	**123**
Total général	**55**	**330**	**385**

Figure 13 : Tableaux « précision des données » de la BAN

3.4.3 La qualité des données

Lors de l'évaluation de la qualité des données fournies par l'API Adresse, plusieurs résultats importants ont été observés. Les 55 échecs provenaient de l'échec au test de distance de sécurité d'un kilomètre. Cela signifie que la localisation obtenue avec le géocodeur étaient trop éloignés de la position réelle du magasin.

En essayant de comprendre ces résultats trop éloignés de la réalité, il a été constaté qu'un bon nombre d'adresses (24 au total) étaient localisées sur des routes départementales ou nationales. Le géocodeur n'a pas réussi à trouver la correspondance dans le référentiel de la Base Adresse National.

D'autre part, il a été observé que l'échec était aussi dû à des géocodages à la rue trop imprécis. En effet, certaines adresse (5 au total) pointaient sur de grandes artères telles que de grand boulevard. Et comme un géocodage à la rue positionne les coordonnées au centre de cette même rue. Il était trop loin de la position réelle du magasin.

Certains échecs ont également été observés en raison d'adresses incomplètes ou non reconnues. Par exemple, une adresse incomplète telle que "1 D 568, CHATEAUNEUF-LES-MARTIGUES 13220" ou une adresse non connue comme "CHEMIN DE HALAGE, MERVILLE 59660". Ces adresses ne sont pas correctement géocodées en raison de leur formulation ou de leur non-reconnaissance dans la base de données.

En ce qui concerne les adresses d'usage des magasins dans les retail-parks et les centres commerciaux en général, il a été constaté que leur géocodage était souvent incorrect ou imprécis. Ces adresses étaient soit complètement erronées, soit mal placées dans la zone commerciale, manquant de précision. La raison la plus probable de cette problématique est que ces adresses étant des adresses d'usage plutôt que des adresses postales standard, cela rend difficile à géocodage de manière précise.

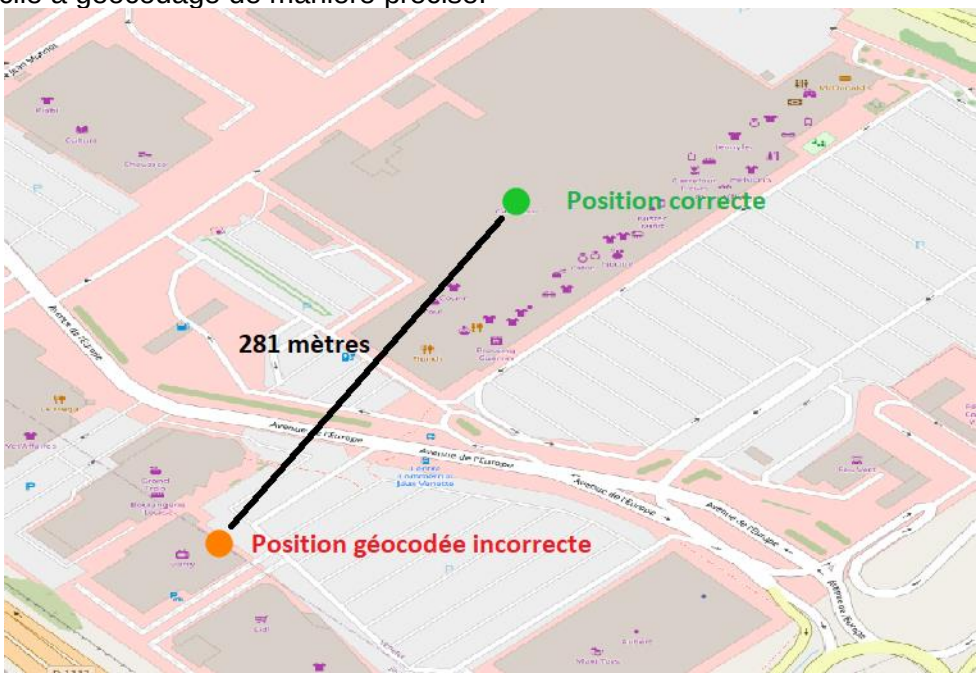

Figure 14 : Exemple d'adresse d'usage avec « CC VENETTE,VENETTE 60280 »

En résumé, les résultats de l'évaluation de la qualité des données de la Base Adresse Nationale ont révélé plusieurs problèmes. Certains échecs étaient dus à des adresses localisées sur des routes départementales ou nationales et des adresses incomplètes ou non

reconnues. De plus, les adresses d'usage des magasins dans les retail-parks et les centres commerciaux ont été mal interprétés, avec des résultats souvent incorrects ou imprécis. Ces résultats soulignent la difficulté de la BAN à trouver les magasins présents dans les retail-park et les centres commerciaux du fait de leur adresse d'usage.

3.4.4 La fréquence de mise à jour

Il a été constaté que les Bases Adresses Locales sont collectées et intégrées dans la Base Adresse Nationale chaque jour. Cela garantit que les données sont régulièrement mises à jour pour inclure les nouvelles adresses ou les modifications apportées aux adresses existantes.

En ce qui concerne la disponibilité des fichiers à télécharger, il a été observé qu'ils sont produits deux fois par semaine. Cela signifie que les utilisateurs de la BAN peuvent accéder à des fichiers contenant les données actualisées au moins deux fois par semaine, ce qui permet de bénéficier de données relativement récentes.

De plus, il a été mentionné qu'une version des fichiers est archivée chaque mercredi. Cela peut être particulièrement utile pour les utilisateurs qui souhaitent accéder à des versions antérieures des données ou effectuer des comparaisons entre différentes versions de la Base Adresse Nationale.

En résumé, les résultats de l'évaluation de la fréquence de mise à jour indiquent que les Bases Adresses Locales sont collectées quotidiennement et intégrées dans la Base Adresse Nationale, assurant ainsi la régularité de la mise à jour des données. De plus, les fichiers à télécharger sont produits deux fois par semaine, avec une version archivée chaque mercredi. Ces résultats démontrent l'engagement envers la mise à jour fréquente des données pour offrir aux utilisateurs de la BAN des informations actuelles et fiables.

3.4.5 La disponibilité de données supplémentaires

En termes de données supplémentaires, le code INSEE de la commune est disponible pour chaque adresse. Le code INSEE est un identifiant unique attribué à chaque commune en France, ce qui permet de faire des jointures avec ce code pour obtenir par exemple des indicateurs socio-économiques à la commune

Ensuite, les coordonnées géographiques sont aussi fournies en projection Lambert 93 en plus du WGS-84. La projection Lambert 93 est le système de coordonnées officiel utilisé en France pour représenter les positions géographiques de manière précise et cohérente.

De plus, pour les magasins situés dans les villes de Paris, Lyon et Marseille, l'API Adresse fournit également le nom de l'arrondissement correspondant. Ces informations sont particulièrement utiles pour les magasins situés dans ces grandes villes, car elles permettent de préciser davantage leur emplacement géographique au sein de la ville.

Cependant, il est important de noter qu'en-dehors de ces informations supplémentaires spécifiques à la localisation du magasin (code INSEE, coordonnées géographiques en projection Lambert 93 et arrondissement pour certaines villes), il n'y a pas d'autres données disponibles propre aux magasins. Les informations complémentaires se limitent aux détails liés à l'adresse et non au magasin en lui-même.

3.5 Conclusion de la troisième partie

En conclusion, ce chapitre a présenté de façon générale la Base Adresse Nationale (BAN) et l'API Adresse, ainsi que les résultats des critères d'évaluation.

Dans l'ensemble, les résultats ont été très positifs. La BAN et l'API Adresse ont réussi à géocoder 100% des adresses analysées, avec un taux de succès de 85%. Cela démontre une bonne performance dans la conversion des adresses en coordonnées géographiques.
Une des forces de la BAN est sa couverture géographique globale, qui inclut également les zones rurales. Cela garantit que les adresses dans toutes les régions de France métropolitaine peuvent être géocodées avec succès même dans des zones faiblement peuplées.

En ce qui concerne la précision des données, le modèle 1 (géocodage au numéro de rue) et le modèle 2 (géocodage à la rue) ont montré une certaine précision, avec des distances moyennes, relativement faibles lors des succès au géocodage. Cependant, il convient de noter que certaines difficultés ont été rencontrées lors de l'interprétation des adresses d'usage des magasins en retail-parks et en centres commerciaux, ainsi que pour les adresses situées sur des routes nationales et départementales.
La fréquence de mise à jour quotidienne de la BAN est un aspect positif, garantissant que les données sont régulièrement mises à jour et reflètent les changements récents dans les adresses.

En ce qui concerne les données supplémentaires, il est important de noter que les informations complémentaires fournies se limitent à des détails liés à l'adresse, tels que le code INSEE de la commune, les coordonnées géographiques en projection Lambert 93 et le nom de l'arrondissement pour certaines villes. Cependant, aucune donnée spécifique sur le magasin lui-même n'est disponible en plus.

En conclusion, la BAN et l'API Adresse offrent une solution globalement bonne pour le géocodage des adresses de magasin en France. La couverture géographique étendue, la précision des données au niveau de la rue ou du numéro de rue, la fréquence de mise à jour régulière sont des points forts. Cependant, il est important de prendre en compte les limites liées à l'interprétation des adresses d'usage des magasins en retail-parks et en centres commerciaux, ainsi que les adresses situées sur les routes nationales et départementales.

Ainsi, la véritable limite de la BAN est sa vocation à ne pas prendre en compte les adresses d'usages ou les Points d'Intérêt (POI) et d'être seulement un référentiel d'adresses purement postales. Le chapitre suivant parlera du second référentiel analysé, OpenStreetMap

4 OpenStreetMap

Dans cette quatrième partie, nous allons nous intéresser à notre dernier référentiel d'adresse, OpenStreetMap (OSM). Pour clarifier ce sujet, nous verrons dans un premier temps une présentation générale de cette source de données. Nous retracerons les enjeux de sa création, ses auteurs et sa constitution. Dans un second temps, nous parlerons rapidement de l'API Nominatim, le géocodeur d'OSM. Enfin, nous détaillerons les résultats obtenus en les évaluant à partir des critères mise en place.

4.1 Présentation générale d'OpenStreetMap

OpenStreetMap est un projet collaboratif ayant vocation à créer une base de données géographique libre du monde entier. On peut voir ce projet comme l'équivalent Open Source et collaboratif de Google Maps ou encore d'un Wikipédia de la cartographie libre. Le point fort de cette base de données est sa communauté. En effet, OpenStreetMap est un projet collaboratif. Il repose donc sur les contributeurs pouvant aider à la création, la modification et la mise à jour des données.

Ce projet a été initié par Steve Coast en 2004 et est hébergé à l'University College de Londres depuis sa mise en route en juillet 2004. Afin de soutenir le projet, la Fondation OpenStreetMap est créée en août 2006. Cette organisation à but non-lucratif a pour principaux objectifs de gérer l'infrastructure matérielle nécessaire au fonctionnement d'OpenStreetMap et de protéger juridiquement le projet. La fondation ne contrôle pas le projet OSM et n'est pas non plus propriétaire des données. Elle supporte simplement le projet. Au niveau national, l'association OpenStreetMap France est représentante de la Fondation OpenStreetMap.

Le succès de Wikipédia et la prédominance des données cartographiques propriétaires au Royaume-Uni, ainsi que dans d'autres pays ont été les deux facteurs à l'initiative de la naissance du projet OpenStreetMap. En effet, son fondateur Steve Coast est parti du constat que les états conservent le plus souvent les droits de reproduction des données cartographiques qu'ils possèdent, alors même qu'ils sont financés par leurs principaux utilisateurs, leurs administrés. Avec le développement des logiciels libres, le créateur d'OpenStreetMap a souhaité suivre ce modèle pour l'accès aux données cartographiques, afin de permettre à tous de consulter et d'utiliser les données sans entrave.

Ainsi, cette base consiste à créer une base de données géographiques mondiale. Cette base va donc devoir stocker et donner accès à des données vectorielles brutes qui représente différents types d'informations cartographiques telles que des routes, des bâtiments, des boulangeries, etc. Les données vectorielles sont accompagnées de métadonnées (aussi appelées « Tags » ou propriétés) permettant de les qualifier.

Nos informations cartographiques sont modélisées dans OpenStreetMap par trois éléments fondamentaux :

- Les nœuds : Ils sont les éléments principaux de la base de données. On les caractérise par une position géographique (latitude, longitude, hauteur en option). Ils servent principalement à définir des points d'intérêts (POI) comme des fontaines, des commerces ou encore des arrêts de bus. Pour ce faire, ils doivent être accompagnés d'attributs (que nous détaillerons par la suite) pour décrire ces points. Ils permettent également de définir des "chemin" et peuvent faire partie d'une relation. Il y a actuellement (au 05/2023) plus 466 millions de nœuds dans la base de données en France métropolitaine (source : taginfo.openstreetmap.fr).
- Les chemins : Ils sont une interconnexion entre au moins deux nœuds. Ils peuvent modéliser des lignes caractérisant (avec les attributs) des autoroutes, des voies ferrées

ou encore des rivières par exemple. L'orientation des chemins a une importance lors de leur création pour définir un côté gauche ou droite ou un sens de circulation d'une route par exemple. Lorsque dans un chemin, le dernier nœud est le même que le premier, ce chemin fermé forme une surface. Cette surface peut représenter par la suite, des champs, des lacs, etc. En France métropolitaine, il y a (au 05/2023) environ 66 millions de chemin, dont 54 millions de chemin fermé présent dans la base de données (source : taginfo.openstreetmap.fr).

- Les relations : il s'agit d'une collection d'objets permettant de regrouper des éléments ayant des caractéristiques communes. Par exemple, avec les différents nœuds définissant les arrêts, on peut modéliser la ligne de bus. Il y a (au 05/2023) environ 850 000 relations en France métropolitaine (source : taginfo.openstreetmap.fr).

Comme nous l'avons vu, les éléments qui composent la base OpenStreetMap sont accompagnés d'attribut (tag en anglais) permettant de les préciser de manière plus approfondi. Par exemple, un élément de type "chemin" peut modéliser grâce aux attributs, une autoroute, ainsi que sa limitation de vitesse. Un attribut est donc formé d'une clé (il est aussi possible d'avoir une clé secondaire) et d'une valeur, et il est possible d'en attribuer plusieurs par éléments. Pour notre autoroute, il s'agirait de l'attribut « highway=motorway » pour la définir et pour limitation de vitesse à 90 km/h de l'attribut « maxspeed = 90». Pour nous aider, un référentiel des attributs et des conventions en vigueur est à disposition sur le Wiki d'OpenStreetMap et il est régulièrement enrichi.

L'attribut qui nous intéresse dans notre comparaison avec la BAN est principalement « addr:housenumber ». Cet attribut composé de la clé principale « addr » et de la clé secondaire « housenumber » permet de définir le numéro de rue de l'adresse postale. Il y avait 7 894 194 attributs « addr:housenumber » de disponible en France métropolitaine en mai 2023 (source : taginfo.openstreetmap.fr). Nous sommes bien loin des plus de 25 millions d'adresses référencés dans la BAN. La répartition sur le territoire des numéros adresse d'OSM est aussi très hétérogène en variant d'une région à l'autre comme le montre la carte ci-dessous.

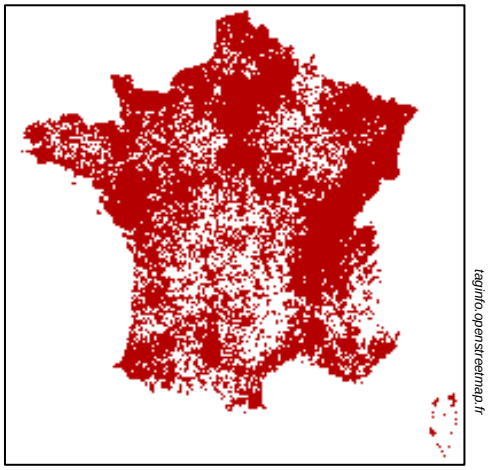

Figure 15 : Répartition géographique de la clé addr:housenumber

Si les adresses postales d'OpenStreetMap ne sont pas encore très exhaustives, on va essayer de s'appuyer sur d'autres attributs comme « shop » ou encore « amenity » qui décrivent les magasins et les restaurants. Il y a plus de 300 000 magasins de disponible sur le territoire en mai 2023 (source : taginfo.openstreetmap.fr) dont plus de la moitié sont attribué à des éléments de type nœud. Cela nous montre l'intérêt d'utiliser les POI dans notre processus de géocodage des magasins.

En effet, pour repositionner nos magasins, nous ne sommes pas obligés de passer par une adresse de type postale. Ce qui nous intéresse vraiment n'est pas de savoir où est cette adresse, mais de connaître la localisation du magasin. Ainsi, les points d'intérêts nous sont utiles pour remplacer l'utilisation des adresses postales. Dans OpenStreetMap, les POI prennent la forme d'un point isolé possédant au moins un attribut. Par exemple, un nœud avec l'attribut « shop=clothes », définit un POI représentant un magasin de vêtements.

Nous allons donc adapter nos différents modèles d'adresses en y ajoutant le nom de l'enseigne. Avec le nom du magasin dans nos modèles, on va essayer de trouver une correspondance avec les POI d'OSM. Ces points d'intérêts vont nous être d'une grande aide pour interpréter les adresses d'usage.

La base de données OpenStreetMap est un projet collaboratif à l'échelle mondiale. Les sources des données sont nombreuses et donc on va s'intéresser cas celle alimentant le territoire métropolitain. Ces sources de données doivent être sous licence libre et compatible avec celle d'OpenStreetMap (que nous présenterons par la suite).

Ainsi, les données peuvent venir de grand acteur de l'information géographique comme l'IGN avec sa BD Ortho accessible aux contributeurs pour vectoriser les éléments géographiques. Les données de la plateforme officielle de l'Open Data, data.gouv.fr, sont aussi utilisées comme source de données. Le fond d'occupation des sols Corine Land Cover France 2006 a été partiellement importé en octobre 2009. Pour les limites administratives des communes et leur population, le projet utilise principalement les références de l'INSEE. Des données des ministères de l'Écologie et du ministère des Sports viennent enrichir OpenStreetMap en apportant des informations sur les zones côtières (GéoLittoral) et sur les équipements sportifs (Recensement des Équipements Sportifs). Le réseau hydrographique provient de la BD Topage et de la couche hydrographique de la BD Topo de l'IGN.

Cette liste n'est pas encore exhaustive, mais nous allons maintenant nous intéresser aux sources de données des adresses postales. Les adresses présentes dans OpenStreetMap proviennent en grande majorité du cadastre de la Direction Générale des Finances Publiques (DGFiP). L'importation des données est semi-automatique, car elles contiennent des erreurs qu'il faut corriger manuellement et vérifier leurs cohérences avec les données existantes dans OSM avant de valider le tout. Les adresses ne sont donc pas encore intégrées partout. Cependant, il y a plus de chance d'avoir des données sur les adresses dans les grandes villes, car en plus du cadastre, les contributeurs (plus nombreux en ville) peuvent s'appuyer sur l'Open data des métropoles pour ajouter les adresses. Enfin, La Poste fournit les codes postaux.

Toutes des données obtenues via cette multitude de sources sont intégrées manuellement par les contributeurs dans la base. Au 24 mai 2023, la plateforme accueille plus de 10 millions d'utilisateurs (source : planet.openstreetmap.org). N'importe quelle personne ou entité (université, association, entreprise, etc.) peut contribuer en ajoutant ses propres connaissances au projet OpenStreetMap et rejoindre la communauté. Il suffit simplement de créer un compte.

Il existe plusieurs manières de participer au projet. En passant par des éditeurs comme JSOM ou iD, on peut ajouter des données géographiques pérennes dans le projet OpenStreetMap. Les contributions sont mises en ligne immédiatement et il n'y a pas de vérification des modifications apportées. Il est donc aussi possible d'aider à la correction des erreurs et des outils comme Osmose qui référence les erreurs de géométrie ou d'attributs présents au sein d'OSM. Ce dernier type de contribution est important, car comme le projet reposant sur de nombreux contributeurs et que les modifications ne sont pas tous vérifiées, la qualité des données est un enjeu pour OSM. Un autre point à aborder est la répartition des collaborateurs

sur le territoire. En effet, nos contributeurs ne sont pas forcément répartis de façon homogène. Cela implique que certaines zones vont être plus détaillées que d'autres.

Il existe aussi plusieurs moyens de récupérer les données d'OSM. Par exemple, l'outil GéoDataMine permet d'extraire des données par thèmes et territoires sous format CSV, Excel, GeoJSON et Shapefile. Overpass Turbo permet aussi de récupérer les données à l'aide de l'API Overpass via des requêtes et ensuite, le résultat est visualisé sur une carte interactive. Enfin, à partir d'une adresse ou du nom d'un lieu, le géocodeur d'OpenStreetMap Nominatim, que nous détaillerons dans la partie suivante, nous renvoi un certain nombre d'informations, dont des coordonnées géographiques et une version standardisée de l'adresse.

OpenStreetMap est sous licence Open Database License (ODbL). Nous sommes libres de copier, distribuer, transmettre et adapter nos données, à condition que nous créditions OpenStreetMap et ses contributeurs. La différence avec la licence ouverte Etalab 2.0 de la BAN, est que si nous modifions ou utilisons les données dans d'autres œuvres dérivées, celle-ci doivent être distribué sous la même licence ODbL.

Pour finir, la plateforme OpenStreetMap est constamment mise à jour et un répertoire Github est accessible à cette URL.

4.2 Présentation du géocodeur Nominatim

L'API Nominatim (du latin « par le nom ») est l'outil de recherche du site web d'OpenStreetMap qui permet d'interroger la base de données. Accessible depuis l'URL https://nominatim.openstreetmap.org/search, Nominatim couvre l'ensemble de l'échelle mondiale et traite jusqu'à 30 millions de requêtes par jour sur un seul serveur.

Les Usages de l'API sont divers, car on peut :

- Rechercher les objets OSM à partir d'une description textuelle ou d'une adresse (les requêtes de recherche peuvent être structurées ou de forme libre)
- Rechercher les objets OSM par leurs coordonnées géographiques. L'API fonctionne en trouvant l'objet OSM le plus proche et en renvoyant ses informations d'adresse.
- Lister les objets qui ont été supprimés dans OSM, mais qui sont retenus dans Nominatim au cas où la suppression serait accidentelle
- Liste des polygones non valide détecté par Nominatim

Pour notre analyse comparative, nous traiterons le premier cas d'usage avec l'aide de la librairie Python Geopy. Nous respecterons la politique d'usage de l'API en ne faisant par exemple qu'une requête par seconde.

Ce service de géocodage a des limites, car comme nous l'avons vu précédent, les adresses postales ne sont pas le point fort d'OpenStreetMap. Par exemple, nous recherchons le magasin de l'enseigne « King jouet » à Redon (35). Son adresse est « 2 rue de la Vieille Ville REDON ». La recherche sur Nominatim renvoi un résultat pointant sur la rue, car les numéros de rue (attribut « addr:housenumber ») n'ont pas encore été complétement référencés dans cette ville et donc ne sont pas connu par Nominatim.

Figure 16 : Résultat avec "2 rue de la Vieille Ville REDON"

Cependant, avec « King Jouet, rue Vieille Ville REDON », Nominatim récupère les coordonnées du magasin. En effet, il existe un POI avec l'attribut « name=King Jouet » pour représenter le magasin.

Figure 17 : Résultat avec " King Jouet, rue Vieille Ville REDON"

Il est donc préférable pour nos modèles d'adresse d'être composé du nom d'enseigne et l'adresse postale pour répondre à la présence ou non d'information dans OpenStreetMap. Les POI aident aussi dans l'interprétation des adresses d'usage, car « King Jouet de Redon » renvoi aussi un résultat positif.

Ainsi, pour maximiser les chances d'avoir une réponse même légèrement imprécise, nous avons donc établi trois modèles d'adresse mélangeant le nom de l'enseigne et son adresse postal :

- Modèle 1 : nom enseigne + aire ou rue + code postal + commune + département + France
  - Exemple : «AUDITION SANTE,ROUTE DE SAVERNE, 67205, OBERHAUSBERGEN, BAS-RHIN, France»
- Modèle 2 : enseigne + code postal + commune + département + France
  - Exemple : «AUDITION SANTE, 67205, OBERHAUSBERGEN, BAS-RHIN, France »
- Modèle 3 (si un numéro d'adresse existe) : numéro + nom de la rue + code postal + commune + département + France
- Exemple : «75 ROUTE DE SAVERNE, 67205, OBERHAUSBERGEN, BAS-RHIN, France»

Ils fonctionnent par ordre de priorité, si le modèle 1 renvoi rien, on passe au suivant et ainsi de suite. Nous privilégions les POI dans le modèle 1 et 2, pour ensuite tester l'adresse postale dans le modèle 3, si un numéro d'adresse existe.

L'API renvoi en réponses les coordonnées exprimées en WGS-84 (EPSG 4326) et des attributs dans le format GeoJSON ou XML ou JSON (au choix). Les attributs permettent d'avoir des informations supplémentaires sur ce qui est renvoyées et en voici la liste :

- place_id : identifiant de l'objet interne à Nominatim
- osm_type, osm_id : type de l'élement et son identifiant OSM
- boundingbox : les coordonnées des coins de la surface de l'objet
- lat, lon : les coordonnées géographique du centroïdes de l'objet
- display_name : nom complet de l'objet avec son adresse
- class, type : clé et valeur de l'attribut principale
- importance : indicateur d'importance (champ technique)
- icon : lien URL vers l'icône de l'objet
- address : adresse détaillé (format dictionnaire)
- extratags : les attributs de l'objet (format dictionnaire)
- namedetails : liste des noms disponibles pour qualifier l'objet (format dictionnaire)

Les attributs ci-dessous sont donc retournés au format demandé. Avec la library Geopy, nous récupérons seulement les coordonnées géographiques en WGS-84 du premier résultat. En effet, Nominatim renvoie une liste de résultats triés par pertinence. Vous trouverez en Annexe 5, la réponse brute de Nominatim pour « E. Leclerc Drive Relais Nantes - Bouffay, Allée du Port Maillard ».

Pour comprendre ce qu'il y a derrière l'API Nominatim, il faut d'abord savoir que le géocodeur extrait toute les données OSM utile pour son géocodage, afin de créer un index géographique à l'aide de l'utilitaire osm2pgsql. Les informations recueillies sont ensuite entreposées dans une base de données PostgreSQL. L'API est une combinaison de C, PL/pgSQL et PHP. Le code source est disponible sur GitHub et est sous licence GPLv2.

4.3 Les résultats du géocodage

Pour cette deuxième session de géocodage avec OpenStreetMap, nous avons donc utilisé le géocodeur Nominatim et obtenu un total de 317 résultats corrects sur les 385 magasins que nous avions à traiter, ce qui représente un taux de réussite de 82,34 %.

Sur les 385 magasins de notre échantillon, nous avons obtenu 289 magasins qui ont reçu une réponse positive et qui ont passé le test de la distance de sécurité de 1000 mètres. Cela signifie que pour ces magasins, le géocodeur Nominatim a pu trouver une correspondance dans un rayon de moins 1000 mètres autour de la position réelle du magasin.

Cependant, nous avons également rencontré 96 échecs, dont 68 magasins n'ont pas pu être géocodés du tout. Ces échecs peuvent être dus à diverses raisons et nous tenterons de les comprendre par la suite.

En résumé, notre session de géocodage de l'échantillon de magasin avec le géocodeur Nominatim a été relativement réussi, avec un taux de réussite de 82,34 %. Le géocodeur a permis d'obtenir des résultats pour 289 magasins. Cependant, il y a eu quelques échecs de géocodage, notamment 68 magasins qui n'ont pas pu être géocodés du tout.

4.4 Evaluation des critères

Dans cette partie, nous allons analyser les différents critères mise en place pour évaluer l'efficacité et la fiabilité d'OpenStreetMap. Nous nous pencherons sur les cinq aspects clés : la

couverture géographique, la précision des données, la qualité des données, la fréquence de mise à jour et la disponibilité de données supplémentaires.

4.4.1 La couverture géographique

Lors de l'évaluation de la couverture géographique d'OpenStreetMap, nous avons obtenu des résultats variables selon les régions. En Corse, nous avons rencontré un taux d'échec assez important de 66,67%. Cette même région avait déjà été mise en évidence lors de l'évaluation de la BAN pour son taux échec assez significatif de 33,33% et par la répartition géographique hétérogène de la clé addr:housenumber sur la carte. De même, les régions de Nouvelle-Aquitaine, Auvergne Rhône-Alpes, Pays de la Loire et Provence-Alpes-Côte d'Azur ont présenté un taux d'échec assez conséquent de 30%. Cela signifie que près de 30% des adresses dans ces régions n'ont pas pu être géocodées avec succès ou la distance de sécurité d'un kilomètre n'a pas été respecté. En revanche, la Normandie a affiché un taux de succès élevé de 94,74%. Pour les régions d'Occitanie, Grand Est et Centre-Val de Loire, nous avons obtenu un taux de succès de 80%.

Ces résultats montrent une certaine hétérogénéité par rapport à ceux de la Base Adresse Nationale (BAN). Selon la documentation, OpenStreetMap a pour vocation de couvrir l'échelle mondiale, cependant, les résultats varient d'une région à l'autre peuvent laisser penser que cette couverture n'est pas parfaite et est donc à relativiser.

En ce qui concerne la distinction entre les zones rurales et les zones non-rurales, nous avons constaté que Nominatim avait plus de difficultés à trouver des correspondances pour les magasins en zone rurale par rapport à l'API Adresse de la BAN. Les magasins présents en zone rurale ont affiché un taux de succès de 70,59%, comparé à 88,24% pour la BAN. Quant aux magasins en dehors des zones rurales, ils ont enregistré un taux de succès de 75,27%, comparé à 85,60% pour la BAN.

En résumé, les résultats de notre géocodage de magasins avec Nominatim mettent en évidence des différences régionales significatives, avec des taux d'échec plus élevés en Corse et dans certaines régions de France. De plus, Nominatim semble avoir plus de difficultés à trouver des correspondances pour les magasins en zone rurale par rapport à l'API Adresse de la BAN. Sur ce point, la Base Adresse Nationale est plus performante qu'OpenStreetMap.

Région	Echec	Succès
Auvergne Rhône-Alpes	33,33%	66,67%
Bourgogne-Franche-Comté	22,22%	77,78%
Bretagne	23,81%	76,19%
Centre-Val de Loire	12,50%	87,50%
Corse	66,67%	33,33%
Grand Est	15,63%	84,38%
Hauts-de-France	22,58%	77,42%
Ile-de-France	25,42%	74,58%
Normandie	5,26%	94,74%
Nouvelle-Aquitaine	31,71%	68,29%
Occitanie	18,42%	81,58%
Pays de la Loire	34,78%	65,22%
Provence-Alpes-Côte d'Azur	30,30%	69,70%
Total général	**24,94%**	**75,06%**

Zone rurale	Echec	Succès
faux	24,73%	75,27%
vrai	29,41%	70,59%

Figure 18 : Tableaux « couverture géographique » pour OSM

4.4.2 La précision des données

Pour évaluer la précision des données, nous avons calculé les distances moyennes entre la localisation obtenue par géocodage et la position réelle du magasin pour les échecs et les succès. Après avoir exclu les valeurs extrêmes, nous avons obtenu une distance moyenne de 2615 mètres pour les échecs, tandis que la distance moyenne pour les succès était de 76 mètres. Comparativement, la Base Adresse Nationale (BAN) a affiché une distance moyenne de 2348 mètres pour les échecs et 141 mètres pour les succès. OpenStreetMap est donc plus précise en termes de distance avec la position réelle de magasin que la Base Adresse nationale.

En utilisant différents modèles d'adresses pour le géocodage, nous avons constaté des variations dans les distances moyennes pour les succès. Le modèle 1, qui utilise le nom de l'enseigne, l'aire ou la rue, le code postal, la commune et le département dans sa composition, a enregistré une distance moyenne de 11,57 mètres pour les succès. Pour la BAN, cette distance moyenne était de 75 mètres.
Ce modèle se distingue par l'utilisation des Points d'Intérêt (POI) et permet à OpenStreetMap d'être plus précise que la BAN par rapport à son modèle 1 en les plaçant directement sur les magasins au lieu de les placer sur les boîtes aux lettres (qui peuvent être loin de l'emplacement réel du magasin).

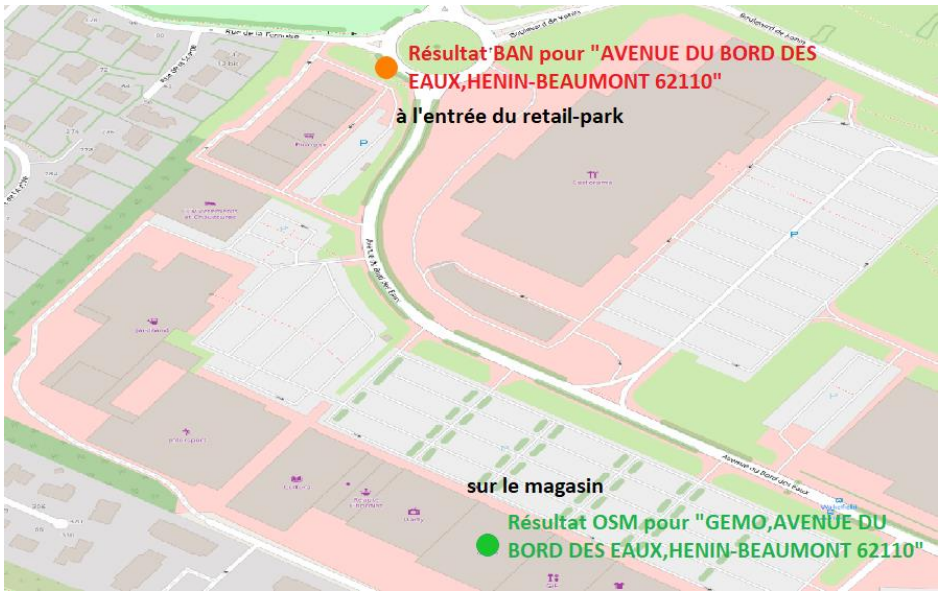

Figure 19 : Avantage de l'utilisation des POI pour gagner en précision

Le modèle 2, qui se compose du nom de l'enseigne, du code postal, de la commune et du nom du département, a affiché une distance moyenne de 75,52 mètres pour les succès. En comparaison, la distance moyenne pour la BAN était de 354 mètres. Comme pour le modèle 1, ce modèle utilise également les POI pour le géocodage.

Enfin, le modèle 3, qui effectue le géocodage au niveau du numéro de rue (en utilisant le numéro et le nom de la rue, le code postal, la commune et le département) a enregistré une distance moyenne de 130,44 mètres pour les succès. Ce modèle ne fait pas appel aux POI, mais se concentre sur le numéro de rue comme la BAN. Il pourrait être comparé au modèle 1 de la BAN qui lui aussi fait usage d'un géocodage au numéro de rue. Dans ce cas, il serait bon de constater que la BAN est plus précise avec son géocodage au numéro de rue avec seulement 75 mètres.

En analysant les résultats, il est intéressant de noter que le modèle le plus utilisé pour les succès était le modèle 3, avec 106 occurrences, suivi du modèle 2 avec 97 occurrences. Pour la BAN, le modèle le plus utilisé était le modèle 1. Cela suggère que lorsque les POI représentant le magasin ne sont pas disponibles dans OpenStreetMap, Nominatim se rabat sur le numéro de rue pour effectuer son géocodage. En ce qui concerne les échecs, le modèle 1 a été le plus utilisé, avec 69 occurrences. Nous essayerons de comprendre ce phénomène dans la partie suivante.

En résumé, l'utilisation des POI par Nominatim a permis à OpenStreetMap d'être plus précis que la Base Adresse Nationale. Par exemple, en étant positionner directement sur l'enseigne dans les grands centre-commerciaux au lieu d'être en dehors au niveau de la boite aux lettres. Cependant, le géocodage au numéro de rue reste l'apanage de la BAN en étant plus précis que la Base OpenStreetMap.

	Moyenne des distances
Echec	2615
Succès	76

Succès	Moyenne des distances
modele 1	11,57
modele 2	75,52
modele 3	130,44

Nombre de magasins	Echec	Succès	Total général
modele 1	69	86	**155**
modele 2	20	97	**117**
modele 3	7	106	**113**
Total général	**96**	**289**	**385**

Figure 20 : Tableaux "précision des données" pour OSM

4.4.3 La qualité des données

Avec cette deuxième session de géocodage, nous avons rencontré 96 échecs (contre 55 avec la BAN), parmi ces échecs, 68 n'ont pas été géocodés et 28 ont échoué au test de la distance de sécurité d'un kilomètre. Parmi ces 68 échecs de géocodage, la quasi-totalité provenait du modèle 1. Après avoir mené notre enquête, nous avons découvert que le modèle 1 de ces 68 enseignes ne renvoyait aucune réponse. Dans ce cas, nos scripts Python sautent le modèle 2 pour tester directement l'adresse postale du modèle 3. Malheureusement, ces magasins n'avaient pas de modèle 3 à disposition et par conséquent, ils sont mis en échec en gardant le modèle 1 comme dernier essai.

Pour les 28 magasins qui n'ont pas réussi le test de la distance de sécurité, certaines erreurs reviennent fréquemment. Par exemple, le géocodage à la rue se fait dans la mauvaise ville, comme "RUE DES JONCS, 79300, CERIZAY, DEUX-SEVRES, France" au lieu de "13 RUE DES JONCS, 79300, BRESSUIRE, DEUX-SEVRES, France". Dans d'autres cas, Nominatim propose une adresse telle que "Rue Plein Ciel, Orgères, Ille-et-Vilaine, France" au lieu du quartier "PLEIN CIEL, 35000, RENNES, ILLE-ET-VILAINE, France".

Une autre situation se présente lorsque le POI pour l'enseigne n'est pas présent dans la ville, et Nominatim va chercher le POI dans une autre ville, par exemple avec l'enseigne KRYS de la ville de GUEBWILLER dans le HAUT-RHIN, où Nominatim renvoie la position de l'enseigne KRYS de la commune voisine.

De plus, lorsque la même enseigne se trouve plusieurs fois dans la même ville avec une adresses « ouverte », comme "WELCOM', 03000, MOULINS, ALLIER, France", il est difficile pour Nominatim de déterminer lequel est le bon magasin. Pour remédier à ce problème, il serait intéressant de mettre en place un comptage d'enseignes par ville en utilisant les magasins présents dans notre base de données, et si une seule enseigne est présente en ville, il n'y aurait pas de problème et nous pourrons tester le modèle 1 ou 2, sinon nous passerions directement au modèle 3 (si possible). Cela demanderait une vérification approfondie, car avec les POI, le risque de confusion peut être important avec les adresses demandées.

En ce qui concerne les adresses d'usage des magasins dans les retail-parks et les centres commerciaux, en général, nous avons constaté que leur géocodage était plutôt correct et précis, avec des coordonnées situées directement sur le magasin lui-même. Cette précision

est rendue possible par la présence d'un POI correspondant à l'enseigne, qui a été ajouté manuellement par les contributeurs.

Un autre aspect à prendre en compte est la répartition des contributeurs sur le territoire métropolitain. Comme nous l'avons vu, OpenStreetMap s'appuie sur les connaissances des contributeurs pour ajouter de nouvelles données. Cependant, il est possible que ces contributeurs ne soient pas répartis de manière homogène sur le territoire. Cela peut expliquer les résultats hétérogènes en termes de couverture géographique.

En résumé, nous avons constaté plus d'échecs de géocodage avec OpenStreetMap qu'avec la BAN, mais dans l'ensemble, les adresses d'usage des magasins sont interprétées de manière plus précise par OpenStreetMap grâce à la présence de POI. Cependant, cela n'est pas homogène sur l'ensemble du territoire et dépend de la présence de contributeurs pour ajouter les informations nécessaires.

Raison Echec	Nombre
modele 1	69
distance > 1000 mètres	2
Pas de géocodage	67
modele 2	20
distance > 1000 mètres	19
Pas de géocodage	1
modele 3	7
distance > 1000 mètres	7
Total général	**96**

Figure 21 : Tableau "qualité des données" pour OSM

4.4.4 La fréquence de mise à jour

Chaque jour, les données OpenStreetMap sont constamment améliorées par des milliers de contributeurs à travers le monde. C'est l'une des forces de cette plateforme collaborative : la capacité d'actualiser et de corriger les informations en temps réel. Comme nous l'avons vu, les contributeurs d'OSM peuvent apporter des modifications à la base de données, ajouter de nouveaux éléments, corriger des erreurs ou mettre à jour des informations existantes. Cela permet de maintenir les données géographiques aussi précises et à jour que possible.

Dans le cas du géocodeur Nominatim, qui utilise les données d'OpenStreetMap, la détection des mises à jour continue de fonctionner de manière ininterrompue. Elle surveille les nouveaux changements et mises à jour publiés sur les serveurs d'OSM et les applique au fur et à mesure de leur disponibilité. Cela garantit que les données utilisées par Nominatim sont régulièrement actualisées et reflètent les dernières modifications apportées par la communauté des contributeurs.

En comparaison, la Base Adresse Nationale (BAN) est mise à jour quotidiennement. Cependant, la différence réside dans le mode de mise à jour. Alors que la BAN est gérée de manière centralisée par des entités spécifiques, les mises à jour en temps réel des données OSM sont le fruit du travail collectif de milliers de contributeur répartis dans le monde entier.

Il est important de noter que la fréquence de mise à jour des données peut varier en fonction des régions. Certaines zones peuvent bénéficier d'une mise à jour plus fréquente en raison d'une activité plus intense des contributeurs dans ces régions aux dépens d'autres territoire disposant de moins de contributeurs.

En conclusion, les données OpenStreetMap sont continuellement améliorées grâce à l'engagement des contributeurs qui mettent à jour la base de données, ce qui permet à Nominatim d'utiliser des données constamment actualisées. Bien que la BAN soit également mise à jour quotidiennement, la nature collaborative d'OSM permet une mise à jour en temps réel et une réactivité plus rapide aux changements géographiques, mais que sur certains territoires abondants de contributeur.

4.4.5 La disponibilité de données supplémentaires

Lorsque nous effectuons un géocodage avec l'API Nominatim, nous avons la possibilité de récupérer des informations supplémentaires sur l'adresse géocodée. En plus des informations de base telles que le numéro de rue, la ville ou encore le code postal, Nominatim peut également fournir les attributs spécifiques à l'objet géocodé. Cela comprend des détails tels que le numéro de téléphone du magasin, son site web, les heures d'ouverture, l'accessibilité en fauteuil roulant, et bien d'autres.

Ce paramètre de la requête nous permet d'obtenir des informations complémentaires sur le magasin géocodé, ce qui peut être utile en nous apportant des précisions sur les magasins obtenus.

En comparaison avec la Base Adresse Nationale (BAN), l'avantage de Nominatim réside dans sa capacité à fournir ces informations supplémentaires spécifiques à la localisation du magasin et au point de vente lui-même. Alors que la BAN peut fournir des informations de base sur l'adresse, telles que le nom de la rue et le code postal par exemple, les données supplémentaires sur le magasin telles que le numéro de téléphone ou les heures d'ouverture ne sont pas incluses dans les résultats de géocodage de la BAN.

Il est important de noter que la disponibilité de ces informations supplémentaires peut varier en fonction des données fournies par les contributeurs d'OpenStreetMap. La présence et la précision de ces données dépendent de la qualité et de la complétude des informations ajoutées par la communauté d'éditeurs d'OSM.

En résumé, lors d'un géocodage avec l'API Nominatim, vous pouvez bénéficier de données supplémentaires spécifiques au magasin, telles que le numéro de téléphone, le site web et l'accessibilité en fauteuil roulant. Cette fonctionnalité offre des possibilités d'obtenir des informations complémentaires et détaillées sur les lieux géocodés, ce qui nous apporte de la valeur ajoutée à ce géocodage. Comparé à la BAN, Nominatim présente cet avantage en fournissant des données supplémentaires spécifiques à chaque emplacement géocodé.

4.5 Conclusion de la quatrième partie

En conclusion, ce chapitre a présenté de façon générale OpenStreetMap et son géocodeur Nominatim, ainsi que les résultats des critères d'évaluation. L'utilisation du géocodeur Nominatim d'OSM a permis d'obtenir un taux de réussite global de 82,34%, avec des résultats pour 289 magasins sur un total de 385. Malheureusement, 68 magasins n'ont pas été géocodés.

Nous avons constaté que Nominatim, grâce à l'utilisation des Points d'Intérêt (POI), offre une interprétation plus précise des adresses d'usage des magasins par rapport à la BAN. Cela permet notamment une localisation plus précise des magasins dans les centres commerciaux et les retail-parks, en étant positionné directement sur l'enseigne plutôt qu'à l'extérieur au niveau de la boîte aux lettres. Nominatim offre également la possibilité de récupérer des informations supplémentaires sur les magasins, telles que les numéros de téléphone, les sites web et les heures d'ouverture contrairement à la BAN et son API Adresse.

Cependant, il convient de noter que les résultats de géocodage avec Nominatim sont hétérogènes d'une région à l'autre, avec des taux d'échec plus élevés en Corse et dans certaines régions de France à cause de la répartition inégale des contributeurs pour ajouter les informations nécessaires. De plus, Nominatim a rencontré plus de difficultés pour géocoder les magasins en zone rurale par rapport à l'API Adresse de la BAN.

La BAN, quant à elle, se distingue par sa précision dans le géocodage au niveau du numéro de rue, offrant des résultats plus précis que ceux d'OSM. De plus, la BAN est également mise à jour quotidiennement, garantissant la disponibilité de données à jour pour le géocodage des magasins, mais n'offrant pas d'actualisation en temps réel.

Pour résumer, OpenStreetMap et Nominatim offrent une solution bonne pour le géocodage des magasins en France. La précision des données grâce à la bonne interprétation des adresses d'usage lié aux POI, la fréquence de mise à jour en temps réel, la disponibilité en données supplémentaires sont des points forts. Cependant, par rapport à la BAN, la couverture géographique et la présence de données varient d'une région à l'autre. Les adresses postales sont plus précises avec la BAN.

Conclusion

Pour conclure ce mémoire, un rappel de son objectif initial est nécessaire. Le but était de mener à bien une analyse comparative de la Base Adresse Nationale (BAN) et le projet OpenStreetMap (OSM). Notre champ d'étude se basait sur le cas du géocodage des magasins en France métropolitaine.

Les critères d'évaluation ont permis d'analyser les résultats obtenus à l'aide de l'API Adresse et du géocodeur Nominatim. Il a été mis en avant que les Points d'intérêts que propose OSM dans sa base de données, permettre une bonne interprétation des adresses d'usages très répandus pour les magasins présent dans les retail-park et les centres commerciaux. Les actualisations en temps réel et la présence d'informations supplémentaires pour les magasins telle que les horaires d'ouvertures, font pencher la balance vers OSM pour être le référentiel le plus à même pour le géocodage des magasins en France métropolitaine. Cependant, les résultats positifs de notre base de données collaboratives ont montré qu'ils variaient en fonction des régions à cause de la répartition hétérogène de ces contributeurs. Il a été aussi été démontré que la BAN, en plus d'une bonne couverture géographique, avait accès à un plus grand référentiel d'adresse numérotés et par conséquent était plus précise qu'OSM avec des adresses postales normalisés (hormis avec les adresses d'usages).

Les deux référentiels ont donc chacun leurs points faibles et leurs points forts. Le choix entre la BAN et OSM pour le géocodage de magasin en France métropolitaine dépendra donc plus des besoins spécifiques de l'utilisateur liés à ses données. La BAN sera privilégiée dans des cas d'adresses postales de magasins bien normalisés et numérotés, car la précision du résultat sera plus importante. OpenStreetMap sera quant à elle mise en avant lors du géocodage d'adresses d'usages et le besoin de rechercher des informations supplémentaires pour les magasins en France métropolitaine

Les objectifs fixés pour cette analyse comparative ont été en grande partie atteints. Tous les critères d'évaluation ont été traités dans leurs ensembles. Cependant, la revue de littérature a montré ses limites en ne proposant que peu d'œuvres sur nos deux référentiels (surtout pour la BAN) dû à un sujet assez récent. Les sources d'information existante sur le web sont pour leur part souvent obsolète. Cette analyse comparative est donc inédite et apporte un regard axé sur la pertinence du contenu des référentiels d'adresses de la BAN et d'OSM dans le cadre du géocodage des magasins en France métropolitaine.

En effet, cette analyse comparative ne traite pas du fonctionnement des géocodeurs utilisés (API Adresse et Nominatim). Tout le processus d'indexage des adresses, tous les algorithmes de rapprochement entre les adresses à géocoder et celles de références, mais aussi la mise en place des résultats « pertinents » ne sont pas analysés et évaluer. Ces traitements peuvent pourtant avoir un impact sur les résultats finaux et notre analyse comparative. Ce premier biais peut aussi être accompagné par notre reformatage d'adresse et notre processus Python de géocodage automatique, car il existe sûrement d'autres méthodes pour appliquer ce reformatage, et ce géocodage automatique de manière plus juste. Cette analyse comparative se base donc sur l'intérêt général d'utiliser ces bases de données pour le géocodage des magasins, sans approfondir sur les traitements entre les données à géocodés et les géocodeurs. Une perspective utile pour une prochaine étude serait donc d'intégrer ce fonctionnement du géocodeur dans l'analyse comparative même si ce premier état des lieux a permis de bien dégrossir notre sujet en donnant des premiers résultats aidant à faire un choix en adéquation avec son besoin.

Plusieurs pistes intéressantes n'ont pas été développées dans cette analyse comparative. La Base Adresses Nationale Ouverte (BANO) est un projet initié par 2014 par OpenStreetMap

France. Ce projet a pour vocation de constituer une base libre des adresses à l'échelle de la France sous licence ODBL. Ainsi, poursuivant les mêmes ambitions que la BAN créé un an après elle, elle s'appuyait sur trois sources jusqu'en 2019 :

- Les adresses récupérées via l'Open Data
- Les adresses déjà présentes dans OpenStreetMap
- Les adresses fournies dans le Cadastre.

À partir de 2019, la mise à disposition des données ayant évolué, les sources de la BANO ont changé. Elle se fonde toujours sur les adresses d'OpenStreetMap, mais aussi maintenant sur celle des Bases Adresses Locales (BAL) et de la Base Adresses Nationale (BAN). OpenStreetMap reste la source prioritaire de la BANO, afin de prendre en compte rapidement les dernières corrections faites par les contributeurs. Cependant, il faut noter que la BANO n'a pas vocation à être une source pour OSM, mais simplement de créer une base d'adresse à partir de différentes sources libres tout en proposant la couverture la plus étendue possible et la plus homogène. La BANO est donc un référentiel d'adresse reposant sur la BAN et OpenStreetMap.

Une autre piste intéressante est celle de la base SIRENE. Cette base recense les données d'identité des entreprises et des établissements en France. Elle contient leur adresse postale, mais pas leur coordonnée géographique. Le porte-parole d'OpenStreetMap France, monsieur Christian Quest, a géocodé cette base à l'aide de la BAN, de la BANO et des POI d'OSM. Les résultats obtenus sont disponibles en accès et en réutilisation libre. Ces deux pistes seraient à approfondir pour notre cas de géocodage de magasin en France métropolitaine.

Bibliographie

Partie 1

Bouba-Olga, O. (2021, mai 25). *Qu'est-ce que le « rural » ? Analyse des zonages de l'Insee en vigueur depuis 2020—Géoconfluences*. http://geoconfluences.ens-lyon.fr/. http://geoconfluences.ens-lyon.fr/actualites/eclairage/grille-densite-zonage-aires-urbaines-definition-rural

Bouyrie, S. (2022, mai 16). Tout sur le géocodage (ou presque). *Infostat Marketing*. https://infostat-marketing.com/tout-sur-le-geocodage/

Géocodage pour le géomarketing : Géolocaliser les magasins, agences et clients. (2015, mai 21). *Parabellum Geographic Insight*. https://www.pginsight.com/geocodage-geomarketing-geolocalisation-magasins-et-clients/

LEBRUN, J. (2020, février 17). 6 considérations pour choisir un service de géocodage précis. *Korem*. https://www.korem.com/fr/6-considerations-pour-choisir-un-service-de-geocodage/

Van der Feer, J. (2020, août 19). *Moyennes et grandes surfaces : Quelle différence ?* Fiches pratiques.https://fiches-pratiques.chefdentreprise.com/Moyennes-et-grandes-surfaces-quelle-difference

Partie 2

Bennett, J. (2010). *OpenStreetMap*. Packt Publishing.
Girres, J.-F., & Touya, G. (2010). Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, *14*(4), 435-459. https://doi.org/10.1111/j.1467-9671.2010.01203.x

Grégory Gibelin. (2019, octobre 10). *Les logiciels et API pour géocoder*. Makina Corpus. https://makina-corpus.com/sig-webmapping/les-logiciels-et-api-pour-geocoder

Guérois, M., Ysebaert, R., Giraud, T., Maranget, B., Hamez, G., Boquet, M., & Dorkel, N. (2020). *Apports et limites d'OpenStreetMap pour l'analyse spatiale des équipements commerciaux en zone transfrontalière—Note de synthèse* (p. 5 p.) [Report, RIATE]. https://hal.science/hal-03587237

Shulz, S. (2021). De l'adoption au rejet d'un commun numérique pour transformer la frontière entre État et citoyens. La trajectoire de la Base Adresse Nationale entre contribution citoyenne, autogouvernement et État-plateforme. *Réseaux*, *225*(1), 151-186. https://doi.org/10.3917/res.225.0151

Vandy, B. (2015, novembre 30). *Comparer des géocodeurs*. Smals.be https://www.smals.be/fr/content/comparer-des-geocodeurs

Viry, M., Giraud, T., Guérois, M., Ysebaert, R., Lambert, N., & Feredj, A. (2016). *Géocodage / calcul de temps de parcours pour les communes de la base « mobilité professionnelles » (INSEE)* (p. 11 p.) [Report, RIATE - Réseau interdisciplinaire pour l'Aménagement et la Cohésion des Territoires de l'Europe et de ses voisinages CNRS - CGET - Université Paris Diderot]. https://hal.science/hal-03589037

Partie 3

Addok. (2020, octobre 30). addok.readthedocs.io. https://addok.readthedocs.io/en/latest/

API Adresse. (2022, mars). guides.etalab.gouv.fr. https://guides.etalab.gouv.fr/apis-geo/1-api-adresse.html#les-donnees-d-adresses

Couturier, S., & Waechter, C. (2022, mars 15). *La Base Adresse Nationale (BAN) franchit de nouvelles étapes en poursuivant son action au sein de l'IGN*. https://www.ign.fr/espace-presse/la-base-adresse-nationale-franchit-de-nouvelles-etapes

De Blomac, F. (2017, avril 12). *Un nouveau géocodeur pour les deux ans de la BAN*. DécryptaGéo, l'information géographique. https://decryptageo.fr/un-nouveau-geocodeur-pour-les-deux-ans-de-la-ban/

Documentation de la plateforme adresse.data.gouv.fr. (s. d.). doc.adresse.data.gouv.fr. Consulté 30 mai 2023, à l'adresse https://doc.adresse.data.gouv.fr/

La France lance la première base adresse nationale collaborative. (2015, avril 22). numerique.gouv.fr. https://www.numerique.gouv.fr/actualites/la-france-lance-la-premiere-base-adresse-nationale-collaborative/

Les dessous de l'adresse. (2022, mars 15). ign.fr. https://www.ign.fr/institut/nos-domaines-dintervention/urbanisme/les-dessous-de-ladresse

Service Public de la Donnée. (s. d.). data.gouv.fr. Consulté 30 mai 2023, à l'adresse https://www.data.gouv.fr/fr/pages/spd/reference/

Partie 4

About OpenStreetMap. (s. d.). OpenStreetMap. Consulté 30 mai 2023, à l'adresse https://www.openstreetmap.org/about

Cabot, M. (s. d.). *OpenStreetMap*. igm.univ-mlv.fr. Consulté 30 mai 2023, à l'adresse http://igm.univ-mlv.fr/~dr/XPOSE2012/OpenStreetMap/index.html

Données. (s. d.). *OpenStreetMap France*. Consulté 30 mai 2023, à l'adresse https://www.openstreetmap.fr/donnees/

France/Cadastre/Import semi-automatique des adresses. (s. d.). OpenStreetMap Wiki. Consulté 30 mai 2023, à l'adresse https://wiki.openstreetmap.org/Cadastre/Import_semi-automatique_des_adresses

FR:Sources de données potentielles/France. (s. d.). OpenStreetMap Wiki. Consulté 30 mai 2023, à l'adresse https://wiki.openstreetmap.org/wiki/FR:Sources_de_donn%C3%A9es_potentielles/France

Huraux, T. (2020, octobre 16). Récupération des données OpenStreetMap (OSM). *Scalian*. https://medium.com/scalian

Nominatim. (s. d.). Consulté 30 mai 2023, à l'adresse https://nominatim.org/

Quest, C. (2017, juillet 27). Géocodage de Points d'Intérêt…. *Medium*. https://cq94.medium.com/

Vandy, B. (2020, février 17). *Géocodage : Contourner les lacunes d'OpenStreetMap (partie 1)*. smalsresearch.be. http://www.smalsresearch.be/geocodage-contourner-les-lacunes-dopenstreetmap-partie-1/

Conclusion

Chrzanowski, P. (2014, septembre 3). *BANO, la Base d'Adresses Nationale Ouverte*. Open Knowledge France. http://fr.okfn.org/2014/09/03/bano-la-base-dadresses-nationale-ouverte/

Le projet BANO - La Base Adresses Nationale Ouverte, par OpenStreetMap France. (s. d.). bano.openstreetmap.fr. Consulté 19 juin 2023, à l'adresse https://bano.openstreetmap.fr/

Quest, C. (2014, mai 19). *BANO? BANCO!* | *OpenStreetMap France*. http://prev.openstreetmap.fr/. http://prev.openstreetmap.fr/node/18587.html

Quest, C. (2017, septembre 15). Géocodage de la base SIRENE. *Medium*. https://cq94.medium.com/g%C3%A9ocodage-de-la-base-sirene-2f0e14e87a8d

Tables des Annexes

Annexe 1 : Tableau des pourcentages par Région puis surface de ventes

Région et type de surface	% de la population parente	Nombre
Grand Est	**8,29%**	**32**
grande	6,88%	2
moyenne	32,05%	10
petite	61,07%	20
Nouvelle-Aquitaine	**10,58%**	**41**
grande	6,32%	3
moyenne	29,50%	12
petite	64,18%	26
Auvergne Rhône-Alpes	**13,12%**	**50**
grande	5,17%	3
moyenne	27,18%	14
petite	67,66%	34
Normandie	**5,10%**	**20**
grande	7,01%	1
moyenne	30,80%	6
petite	62,19%	12
Bourgogne-Franche-Comté	**4,71%**	**18**
grande	6,20%	1
moyenne	34,49%	6
petite	59,31%	11
Bretagne	**5,58%**	**21**
grande	7,61%	2
moyenne	28,56%	6
petite	63,83%	13
Centre-Val de Loire	**4,08%**	**16**
grande	7,36%	1
moyenne	30,04%	5
petite	62,61%	10
Corse	**0,67%**	**3**
grande	4,91%	0
moyenne	26,56%	1
petite	68,53%	2
Ile-de-France	**15,53%**	**59**
grande	4,92%	3
moyenne	22,73%	13
petite	72,35%	43
Occitanie	**9,80%**	**38**
grande	5,71%	2
moyenne	29,16%	11
petite	65,13%	25
Hauts-de-France	**7,98%**	**31**
grande	6,03%	2
moyenne	32,40%	10
petite	61,57%	19
Pays de la Loire	**5,99%**	**23**
grande	7,75%	2
moyenne	26,71%	6
petite	65,54%	15
Provence-Alpes-Côte d'Azur	**8,57%**	**33**
grande	4,75%	2
moyenne	25,53%	8
petite	69,72%	23
Total général	**100%**	**385**

Annexe 2 : Script Python pour sélectionner les magasins aléatoirement

```
import csv
from qgis.core import (
    QgsVectorLayer,
    QgsProcessing
)
import processing
#####
_PATH_INPUT = 'C:/Users/jrous/OneDrive -
TRIBEKAI/Mémoire/TailleEchan.csv'#Le tableau récapitulant ce que
nous avons besoin pour chaque région et type de surface
_PATH_MAG = 'C:/Users/jrous/OneDrive -
TRIBEKAI/Mémoire/Data_pour_echantillon.gpkg'#Les magasins
#####
with open(_PATH_INPUT, mode='r', newline='', encoding="ANSI") as
input_file:
    input_data = csv.reader(input_file, delimiter=';')
    header = next(input_data)  # on enlève l'en-tête

    mag = QgsVectorLayer(_PATH_MAG, "Magasins", "ogr")

    liste_region = ['Alsace, Champagne-Ardenne et Lorraine',
'Aquitaine, Limousin et Poitou-Charentes', 'Auvergne et Rhône-
Alpes', 'Basse-Normandie et Haute-Normandie', 'Bourgogne et Franche-
Comté',
                    'Bretagne', 'Centre', 'Corse', 'Ile-de-France',
'Languedoc-Roussillon et Midi-Pyrénées', 'Nord - Pas-de-Calais et
Picardie', 'Pays de la Loire', 'PACA']

    liste_surface = ['grande', 'moyenne', 'petite']

    liste_layer = []

    for region in liste_region:#on filtre les magasins région par
région
        expression = '"Region" = \'{}\' '.format(region)
        alg_params = {
            'EXPRESSION': expression,
            'INPUT': mag,
            'OUTPUT': QgsProcessing.TEMPORARY_OUTPUT
        }
        filtre1 = processing.run('native:extractbyexpression',
alg_params)['OUTPUT']
        filtered_rows = []
        count = 0
        for ligne in input_data:#on garde les surfaces liés à cette
région
            if ligne[0] == region:
                filtered_rows.append(ligne)
                count += 1
                if count >= 3:
                    break
```

```
        for surface in liste_surface:#on sélectionne aléatoirement
les magasins pour chaque surface
            expression = '"SURFACE" = \'{}\' '.format(surface)
            alg_params = {
                'EXPRESSION': expression,
                'INPUT': filtre1,
                'OUTPUT': QgsProcessing.TEMPORARY_OUTPUT
            }
            filtre2 = processing.run('native:extractbyexpression',
alg_params)['OUTPUT']
            for filtered_row in filtered_rows:
                if filtered_row[1] == surface:
                    valeur = filtered_row[2]#combien de magasins il
nous faut pour cette surface dans cette région
                    break
            alg_params = {
                'INPUT': filtre2,
                'METHOD': 0,  # Nombre d'entités sélectionnées
                'NUMBER': valeur
            }
            processing.run('qgis:randomselection', alg_params)
            alg_params = {
                'INPUT': filtre2,
                'OUTPUT': QgsProcessing.TEMPORARY_OUTPUT
            }
            liste_layer.append(processing.run(
                'native:saveselectedfeatures',
alg_params)['OUTPUT'])#on sauvegarde la sélection dans une couche

    alg_params = {
        'CRS': QgsCoordinateReferenceSystem('EPSG:4326'),
        'LAYERS': liste_layer,
        'OUTPUT': QgsProcessing.TEMPORARY_OUTPUT
    }
    result = processing.run('native:mergevectorlayers',
alg_params)['OUTPUT']#on fusionne toutes les couches en une seul
    QgsProject.instance().addMapLayer(result)
```

Annexe 3 : Extrait de la liste des enseignes présentes dans l'échantillon

Nom des enseignes	CAVAVIN
AUDITION SANTE	COLUMBUS CAFE & CO
LA MAMMINA	CARREFOUR MARKET
INTERMARCHE CONTACT	LITERIE CONFORT
HAPPY CASH	GEMO VETEMENTS
CORA CAFETERIA	VIB'S
LA FEE MARABOUTEE	CASH CONVERTERS
GAMM VERT	ERAM
FLUNCH	SILIGOM
JD SPORTS	PLEIN CIEL
LIDL	H & M
PROMOCASH	MS MODE
LEROY MERLIN	MOBALPA
CELIO	SUD EXPRESS
AUDIKA	LA HALLE AU SOMMEIL
BESTDRIVE	PROFIL +
NOCIBE	LE DRIVE INTERMARCHE
MAXI ZOO	POINT.P
ARMOR LUX	ALDI
KRYS	LEVI'S STORE
ATOL LES OPTICIENS	BAGELSTEIN
MC DONALD'S	COPRA
PULSAT	BLANC BRUN
LA PATATERIE	BUT COSY
LA BOUCHERIE RESTAURANT	LA MIE CALINE
LE MANEGE A BIJOUX E.LECLERC	ESPACE FOOT
ACUITIS	MUY MUCHO
OPTICAL CENTER	BRUCE FIELD
LA FOIR'FOUILLE	MY BIGBANG
VINS SUR 20	OLD WILD WEST
FRENCH COFFEE SHOP	MICROMANIA-ZING
KIABI	ARMAND THIERY
CASINO	NATURALIA
CARREFOUR CONTACT	LA CROISSANTERIE
VM MATERIAUX	SITIS
GRAND FRAIS	EXTRA
SUPER U	MATSURI
MARC ORIAN	LE BONHOMME DE BOIS
TRESOR	CHRISTINE LAURE
THIRIET	SPEEDY
EDEN PARK	NICOLAS
LE MARCHE DE LEOPOLD	MONTRE SERVICE
M+ MATERIAUX	MAISONS DU MONDE
TBS	ADIDAS
IKKS	KING JOUET
KIKO	LOUIS PION
MAX PLUS	LA BRIOCHE DOREE
COURTEPAILLE	LACOSTE

Annexe 4 : Réponse de l'API Adresse pour "14 bis avenue Marie Reynoard 38000 Grenoble"


```
{
    "type": "FeatureCollection",
    "version": "draft",
    "features": [
        {
            "type": "Feature",
            "geometry": {
                "type": "Point",
                "coordinates": [
                    5.728475,
                    45.166164
                ]
            },
            "properties": {
                "label": "14bis Avenue Marie Reynoard 38100 Grenoble",
                "score": 0.7582936363636363,
                "housenumber": "14bis",
                "id": "38185_4853_00014_bis",
                "name": "14bis Avenue Marie Reynoard",
                "postcode": "38100",
                "citycode": "38185",
                "x": 914321.79,
                "y": 6455576.78,
                "city": "Grenoble",
                "context": "38, Isère, Auvergne-Rhône-Alpes",
                "type": "housenumber",
                "importance": 0.74123,
                "street": "Avenue Marie Reynoard"
            }
        },
        {
            "type": "Feature",
            "geometry": {
                "type": "Point",
                "coordinates": [
                    5.705631,
                    45.184806
                ]
            },
            "properties": {
                "label": "14bis Rue Ampère 38000 Grenoble",
                "score": 0.4787618181818182,
                […]
            }
        },
    ],
    "attribution": "BAN",
    "licence": "ETALAB-2.0",
    "query": "14 bis avenue Marie Reynoard 38000 Grenoble ",
    "limit": 5
}
```

Annexe 5 : Réponse de Nominatim pour "E. Leclerc Drive Relais Nantes - Bouffay, Allée du Port Maillard"


```
[
    {
        "place_id": 20403566,
        "licence": "Data © OpenStreetMap contributors, ODbL 1.0. https://osm.org/copyright",
        "osm_type": "node",
        "osm_id": 2320501677,
        "boundingbox": [
            "47.2147812",
            "47.2148812",
            "-1.552299",
            "-1.552199"
        ],
        "lat": "47.2148312",
        "lon": "-1.552249",
        "display_name": "E. Leclerc Drive Relais Nantes - Bouffay, Allée du Port Maillard, Bouffay, Decré - Cathédrale, Centre Ville, Nantes, Loire-Atlantique, Pays de la Loire, France métropolitaine, 44007, France",
        "class": "shop",
        "type": "supermarket",
        "importance": 1.02001,
        "icon": "https://nominatim.openstreetmap.org/ui/mapicons/shopping_supermarket.p.20.png",
        "address": {
            "shop": "E. Leclerc Drive Relais Nantes - Bouffay",
            "road": "Allée du Port Maillard",
            "neighbourhood": "Bouffay",
            "suburb": "Centre Ville",
            "city": "Nantes",
            "municipality": "Nantes",
            "county": "Loire-Atlantique",
            "ISO3166-2-lvl6": "FR-44",
            "state": "Pays de la Loire",
            "ISO3166-2-lvl4": "FR-PDL",
            "region": "France métropolitaine",
            "postcode": "44007",
            "country": "France",
            "country_code": "fr"
        },
        "extratags": {
            "website": "https://www.leclercdrive.fr/region-pays-de-la-loire/nantes/drive-relais-nantes---bouffay.aspx",
            "operator": "E. Leclerc",
            "check_date": "2023-03-25",
            "description": "Point de retrait E. Leclerc",
            "contact:city": "Nantes",
            "opening_hours": "Mo-Sa 10:00-21:00",
            "brand:wikidata": "Q1273376",
            "contact:street": "Allée du Port Maillard",
            "brand:wikipedia": "en:E.Leclerc",
            "contact:postcode": "44000",
            "contact:housenumber": "12"
        },
        "namedetails": {
            "name": "E. Leclerc Drive Relais Nantes - Bouffay",
            "brand": "E. Leclerc",
            "name:fr": "E. Leclerc Drive Relais Nantes - Bouffay"
        }
    }
]
```