Table des matières

La première partie de ce document est un bref résumé de travaux effectué lors de mon dernier stage. Pour la documentation concernant les scripts du GitHub rendez-vous à I : Présentation du Répertoire Stage-2023.

Repenser le secteur du nettoyage B2B :

Je suis chargé du poste de responsable stratégies et outils marketing au sein de RnPur une start-up souhaitant amener du vent frais sur le marché du nettoyage B2B.

Gilles Costa ingénieur des Mines était alors commerciale au sein de Capgemini, acteur majeur du conseil et de la transformation numérique.

Lorsqu'il s'est décidé en 2016 de révolutionner le secteur du nettoyage B2B

Le marché était alors caractérisé par les points suivants :

- Une offre sclérosée, rigide et dépassée

- Des environnements de travail en pleine transformation confrontant des clients à la recherche de solutions innovantes, flexibles et agiles à une offre fermée, opaque et non évolutive

- Relation clients et salariés rude

- Informatique très en retard et peu intégré dans la chaine de valeur (du salarié au client)

- Un potentiel considérable d'innovations : digital, sécurité sanitaire, utilisation de l'eau pure, récupération et renforcement physique des agents

Ses réponses :

- Changer en profondeur, bouleverser l'expérience client et collaborateur

- Accompagner cette transformation des environnements de travail grâce à une plus grande flexibilité, agilité et réactivité.

- Apporter au marché une connaissance de la transformation de l'externalisation des SI au Cloud avec son expérience de la relation client et collaborateur.

- Les agents incarnent la volonté de rendre la vie plus facile aux clients donneur d'ordre et plus agréable à leurs collaborateurs.

- Qualité de service inspirée de l'accueil et de l'hôtellerie haut de gamme donne aux collaborateurs clients le sentiment d'être accueillis et bien traités.

  - Intégrer le numérique dans la chaine de valeur (du salarié au client)

C'est sur ce dernier point que je suis intervenu.

Les chapitres qui vont suivre relatent les travaux du projet général d'acquisition de nouveaux clients.

Inbound Marketing

L'inbound marketing est une méthodologie consistant à attirer des prospects en créant du contenu utile et des expériences personnalisées afin de les convertir en leads et in fine en clients.

Dans cette optique, RnPur disposait d'un site dont la stratégie de contenu reposait principalement sur la thématique du bien-être et d'une page LinkedIn vierge en plus de la page de Giles Costa qui avait-elle une stratégie de contenu plus axée sur les enjeux de Responsabilité Sociétale des Entreprises (RSE).

La priorité qui m’a été déléguée sur le sujet a été la gestion du site qui requérait une remise à niveau ainsi que quelques modifications.

Site web

Mises en place diverses :

Publier un site de test :

Avant d’effectuer toute modification sur le site principale, il m’a été de remettre en service un site de test pour plus de précautions pour prévenir le site original de toute modification pouvant lui nuire.

J’ai ensuite pu :

. Installer les mises à jour afin de s’assurer que l’optimisation de son référencement via moteur de recherche (SEO) ne soit pas impactée et tout simplement que le site ne crash pas

. Créer une page pour le recrutement. Afin de rendre le site plus vivant en publications, le système de référencement des moteurs de recherche de même que les leads et clients étant sensibles à l’ancienneté des publications.

.Nourrir le blog de publications issues de LinkedIn pour les mêmes raisons

En plus de ces derniers points, j’ai dû m’atteler à la gestion des cookies du site

Gestion des cookies :

Mesure d’audiences

Les Cookies peuvent être vus comme des boîtes noirs déposées par un site sur l’ordinateur d’un utilisateur afin par exemple, que lors d’une prochaine visite sont identifiant utilisateur et ses préférences puissent récupérées pour personnaliser le comportement du site.

Ils sont principalement controversés, car un même cookie de traçage déposé sur plusieurs sites peut permettre à une GAFAM qui en détient la propriété (principalement Google) de recomposer l'identité et la vie numérique d'un internaute intentant alors au principe de la protection des données personnelles de la CNIL.

Il existe cependant plusieurs types de cookies, dont certains ne demandent aucune mesure légale et d'autres requérant la mise en place une bannière de consentement permettant de les désactiver au gré de l'utilisateur.

Tandis que certaines sont tout bonnement interdits avec ou sans consentement.

Le premier type de cookies est celui que la CNIL (Commission Nationale de l'Informatique et des Libertés) considère comme étant strictement nécessaires à la bonne administration du site et ne requière donc normalement aucun consentement de la part d'un utilisateur du moment qu'il est laissé à la disposition de ce dernier la possibilité de les retirer.

En voici la liste générale :

- la mesure de l'audience, page par page ;
- la liste des pages à partir desquelles un lien a été suivi pour demander la page courante (parfois nommé « referrer ») que ce soit interne ou externe au site, par page et agrégée de manière journalière ;
- les type de terminal, navigateur et taille d'écran des visiteurs, par page et agrégé de manière journalière ;
- des statistiques de temps de chargement des pages, par page et agrégée de manière horaire ;
- des statistiques de temps passé sur chaque page, de taux de rebond, de profondeur de défilement, par page et agrégée de manière journalière ;
- des statistiques sur les actions utilisateurs (clic, sélection), par page et agrégée de manière journalière ;
- des statistiques sur la zone géographique d'origine des requêtes, par page et agrégée de manière journalière.

Le site d'RnPur ne revendant pas les informations de ces utilisateurs à des fin publicitaires et n'ayant pas de cookies tiers en général à part son cookie de mesure d'audience, il n'aurait pas dû y avoir de problème.

Cependant il s'est avéré que Google Analytics notre outil de mesure d'audience était précisément dans la ligne de mire de la CNIL.

Du fait de la possibilité pour Google de non seulement porter atteinte à la vie privée des internautes mais aussi d'exporter leurs données en dehors de la juridiction européenne.

L’utilisation de Google Analytics sans serveur mandataire a été proscrite par la CNIL sous peine de mise en demeure du responsable du site jusqu’à la contrainte d’une amende de 4% du chiffre d’affaires de la société responsable.

Un serveur mandataire, ou proxy est un serveur servant d’intermédiaire entre un utilisateur et un prestataire.

Du fait d’une mise en place chronophage et coûteuse, j’ai proposé à Giles la désinstallation de Google Analytics après sécurisation des mesures d’audiences déjà relevées par l’outil.

J’ai ensuite implémenté Wysistat un prestataire français, gratuit, nativement conforme RGPD(Règlement Générale sur la protection de données) pour reprendre le flambeau laissé par Google Analytics.

Une fois ces œuvres réalisées, il nous a fallu choisir un prestataire délivrant un plugin de gestion du consentement utilisateur.

Bannière de consentement

La bannière de consentement est un point crucial lorsqu’un internaute se rend sur un site.

C’est le premier contact effectué entre un prestataire et un lead.

Offrir une bonne expérience de consentement RGPD est donc primordiale pour éviter les taux de rebond sur son site (sortir d’un site sans avoir consulter son contenu).

Axceptio grâce à son design plus sympathique et une meilleure tarification que ses concurrents dans notre cas est le service tiers qui a été choisi.

Une fois les travaux urgents de ce chapitre effectués,
J’ai pu m’atteler aux enjeux de communication sur LinkedIn.

LinkedIn

Tout d’abord nous avons dû établir une stratégie de contenu avant de nous lancer dans Une campagne de publication.

J’ai donc dû appréhender le fonctionnement du référencement des publications avant de partager mon opinion à mon manager sur le sujet.

En voici les notes :

SEO

SEO comme concisément précisé au sein du chapitre précédent tient pour Search Engine Optimisation.

C’est le système de référencement que va employer une plateforme hébergeant du contenu Qui va privilégier un contenu plutôt qu’un autre.

Par exemple la SEO de Google qui est l’une des principales raisons de son essor repose essentiellement au jour où j’écris ces mots sur le backlinking.

Plus il y aura de contenus (en l’occurrence des sites renvoyant un même autre, plus ce dernier sera bien référencé car jugé intègre)

Lorsqu’on se penche sur le SEO d’un réseau social dont une part du business model repose sur la publicité qu’il parvient à afficher à ses utilisateurs.
Ce dernier aura tendance à privilégie lui les publications susceptibles de susciter le plus d’intérêt afin de nous retenir le plus longtemps possible et donc de nous passer le plus de contenu sponsorisé possible.

Les critères qu’emploie la filiale de Microsoft pour juger l’intérêt éprouvé par un utilisateur pour un contenu (publication) sont entre autres les suivants :

. Le Dwell Time : Temps passé sur une publication

. Le nombre de réactions (commentaires, mentions) au cours de la première heure de publication et celles qui s’ensuivent

La SEO va lors de la première heure de publication tester la réactivité des connexions des premiers degrés de l’auteur puis en fonction de ces dernières va présenter la publication à des connexion de degrés plus éloignés (contacts de contacts).

Il est à noter que pour la raison évoquée ci-dessus, étendre son réseau trop rapidement sur Linkedin peut être à double tranchant.
En effet, une connexion de premier niveau ignorant votre publication diminuera la visibilité de cette dernière.

Voici une brève synthèse de la SEO.

C’est donc une fois après avoir effectué ces recherches que nous avons entamé l’élaboration de notre stratégie de contenu.

Campagne SMART

Nous avons décidé de garder une ligne éditoriale orientée RSE et actualités de l’entreprise
Et avons élu le compte du CEO pour les premières publications, ce compte comptant plus de connexions et abonnés actifs que le compte de l’entreprise.

De plus, après quelques recherches, il s’est avéré que les gens portaient en général plus d’intérêt à un profil utilisateur qu’à un compte d’entreprise

Le compte RnPur a tout même été employé afin de reposter les postes du compte Giles Costa avec beaucoup moins d’impact cependant. Ce pour les raisons déjà énoncées en plus du fait que la republication d’un poste est bien moins référencée par la SEO de LinkedIn qu’une publication originelle.

Ensuite il a été admis par le feed (file d’actualité) de Giles Costa et des persona ciblés (CEO de PME, happiness managers, responsables des environnements de travail…),
Que les types de contenu que nous devions réaliser était de format texte + image et vidéo.

Le premier type est en effet le meilleur format dans notre cadre pour sa simple réalisation
En plus d’attirer l’œil tout en optimisant le dwell time, la deuxième demande de débloquer un budget pour réaliser des vidéos de qualité mais est un type de contenu à haut taux d’attractivité en plus d’être plébiscité par la SEO.

Afin de d’établir des objectifs Specific Mesurable Achievable Relevant Time(SMART), il m’a ensuite été demander de définir les KPI de nos postes sur court et moyen terme afin d’en établir le suivi :

Pour appréhender l’ampleur multicanale de notre impact,

J’ai couplé les KPI de LinkedIn avec ceux du service de mesure d’audience de notre site implémenté au préalable (Wysistat).

Le suivi a été matérialisé en un tableur Excel de 18 colonnes permettant soit d’identifier chaque poste par un identifiant unique, soit de relater leur thème, intention, style d’écriture et format ou encore leur date, heure de publication puis enfin les interactions qu’ils ont suscité

une heure puis une semaine après leur parution sur les comptes de l'entreprise et sur le site d'après l'outil d'analyse de trafic Wysistat cf partie **site web**.

Chaque publication LinkedIn a vu son doublon posté sur le blog du site action qui nous a permis de gagner des visiteurs venus par voie organique(référencement naturel).

Une nette amélioration a pu être constatée au fil du temps et nous avons remarqué que les postes parlant de la vie de l'entreprise étaient ceux qui donnaient les meilleurs retours.

Nous n'avons cependant expérimenté que le format image + texte, et l'étude n'as été effectuée que sur une courte période pour affirmer la bonne stratégie à adopter.

I : Présentation du Répertoire Stage-2023

1 : Documentation

Contient toute la documentation nécessaire à la prise en main de la base de données Sirene

. INSEE Documentation API Sirene Variables :

Décrit les variables d'une base de données téléchargée via Sirene.fr

. Mode opératoire constitution de listes Sirene :

Décrit le processus d'acquisition d'un jeu de données sur Sirene.fr

. Nomenclatures_NAF_Reedition_2020 :

Décrit les catégories de l'arborescence NAF

. RNERegistre nationnal des entreprises :

Décrit la composition générale des codes d'identification des entreprises

2 : Entreprises :

Contient les jeux de données propres à l'établissement de l'annuaire des sièges d'entreprises correspondants à nos critères

Dossiers :

a : AnnuaireSegmente

Est la version segmentée par tranche de salariés d'AnnuaireFiltre.

Les jeux de ce dossier ont été réalisés via le script annuaireSegmenteScript au format IPYNB (JupyterNotebook, Google Collab…) présent dans le dossier Scripts.

b **:** Dossier Annuaire_INSEE

Contient les premières modifications réalisées sur le jeu de données via Excel

. Annuaire_INSEE :

Classeur contenant la base de données brute (la feuille etablissements) et ses premières modifications réalisées sur Excel ainsi que les feuilles collaborateurs qui sont les versions segmentées par tranches de collaborateurs de la feuille EtablissementsModifies.

Annuaire_INSEE comporte aussi la nomenclature des 5 niveaux NAF qui a permis de référencer les codes associés.

. AnnuairesSegmenteExcel :

Première version d'AnnuaireSegmente.

Réalisée via Excel la segmentation a été opérée par tranches d'effectifs de salariés

. EtablissementsModifies :

Version extraite de la feuille correspondante d'Annuaire_INSEE

. Feuilles collaborateurs :

Contient le descriptif des feuilles collaborateurs d'Annuaire_INSEE

c : Maps Concatenation

Contient les bases de données segmentées de finalMaps issues du redémarage multitiple du programme ainsi que le fichier aide ayant servi à leur concaténation.

Documents

. AnnuaireFiltre :

Première version filtrée de la base originelle réalisée via Python.

. AnnuaireFinal :

Version contenant uniquement les sièges d'AnnuaireFiltre ainsi
Que les colonnes nécessaires à la création des recherches à destination du scraping du site.

. finalMaps :

Base de données récupérée à la suite du scraping du site aux clefs de recherche issues d'AnnuaireFinal

. finalMapsAndAnnuaireFinal :

Concaténation de finalMaps et AnnuaireFinal pour croisement des données

3 : Dossier GroupesScolairesPrives

a : adressesEcoles

Noms des établissement récoltés dans EtablissementScolairesPrive et fr-annuaire-education
Groupés par adresse commune.

b : EtablissementScolairesPrives

Données issues du scraping d' https://www.enseignement-prive.info

c : etablissementsSirene

Données ciblant tous les établissements référencé Education par le NAF de leur section via Sirene.fr en Île-de-France

d : fr-annuaire-education

Données ciblant tous les établissements scolaires franciliens privés téléchargée via l'API De https://www.education.gouv.fr/annuaire

4 : Scripts

Contient les squelettes des scripts du projet au format IPYNB (JupyterNotebook, Google Collab..).

Dossier

. ScriptEtablissementScolaires :

Contient les scripts relatifs au croisement des bases de données des établissements recensant les établissements scolaires privés et leurs coordonnées.

Document

. annuairesSegmenteScript.ipynb :

Contient le script ayant permis de créer les jeux de données du dossier AnnuaireSegmente

II Entreprises

Afin de créer un vivier de sièges d'entreprises pertinents le processus s'est déroulé en 3 étapes.

1 Extraction des données

L’extraction de cette liste s’est déroulée via Siren.fr.

Pour ce faire, j'ai dans un premier temps récolté le registre de tous les établissements comportant un minimum de 50 salariés afin d’estimer une surface à nettoyer suffisamment rentable pour nos interventions.

Cette opération a été réalisée l’aide du site Siren.fr
Elle correspond à la feuille “établissements” du classeur “Annuaire_INSEE”

La fiche technique de l’extraction est disponible via le document ‘mode opératoire constitution de listes Sirene du dossier Documentation.

2 Ecrémage des données

Chaque unité légale et établissement dépendant de cette même unité légale, possède un code APE (Activité Principale des Entreprises) nommé aussi code NAF (Nomenclature des Activités Françaises).

Ces codes étaient présents dans la base de données originelle mais non référencés à leur activité lisible.

J’ai donc téléchargé la base de données NAF via https://www.insee.fr/fr/information/2120875, pour ensuite affilier chaque partie de l’identifiant à son sous-groupe associé.
cf feuille ‘établissements modifiés’ du même classeur.

Pour plus de précisions sur la nomenclature arborescente NAF,
cf Nomenclatures_NAF_Reedition_2020 du dossier Documentation.

Cela étant fait, j’ai commencé à filtrer les données sur Excel ; premièrement dans l’optique de filtrer les établissements en fonction de la pertinence de leur activité et deuxièmement dans l’optique de réaliser trois fichiers segmentés en fonction du nombre de salariés présents au sein de chaque établissement.

Cependant Excel s’est trouvé être trop rigide et chronophage pour réaliser efficacement ces tâches.

J’ai donc entrepris de reprendre le traitement des données avec Python et la bibliothèque Pandas.

Voici les filtres que j’ai appliqué à la base de données nommée ‘AnnuaireFinal‘

```
DenominationUniteLegale = ['PUBLIC', 'PUBLIQUE', 'NATIONAL',
'NATIONALE']
ActivitePrincipaleUniteLegaleSousClasse = ['Nettoyage', 'consulaires']
ActivitePrincipaleUniteLegaleSection = ['Agriculture',
'Administration', 'Enseignement']
```

Pour chaque colonne en bleu les valeurs contenant une des chaînes de caractère orange, Sont supprimées.
Pour etablissementSiege la valeur True est celle pour laquelle les lignes correspondantes ont été conservées.

A noter qu'un filtre sur les Activités est disponible lors du téléchargement de la base, bien que les données téléchargées à la suite du filtre ne contiennent pas le référencement associé à son code NAF.

Le cas des établissements scolaires a été traité séparément (cf : chapitre IV) du fait de l'impossibilité de déterminer le type de contrat de chaque établissement, notre cible ne concernant que les établissements privés.

Enfin de cette liste je n'ai récupéré que les établissements étant des sièges.
Ayant établi que le responsable de l'entretien de l'espace de travail d'un siège était le responsable de tous les établissements de la branche.

De là, j'ai entamé la deuxième étape du projet.

3 : Vérification et enrichissement

a : Présentation de la base finalMaps

Afin de déterminer si les adresses sont bien les bonnes afin d'entamer la campagne d'evoie de cadeaux d'entreprise, le scraping d'un site a été entrepris afin de croiser les données avec les données de l'INSEE.

Le squelette du programme est présent dans le script scrapMaps du dossier Scripts.

De la base de données précédemment énoncée, j'ai extrait les éléments les plus pertinents pour une recherche sur le site ; à savoir la dénomination de l'unité légale ou de l'établissement le plus reculé dans la base de données si n'est pas un sigle, associé de l'adresse complète de l'établissement.

(cf fonction Search() du script scrapMaps du dossier Scripts)

Le programme va alors lancer la recherche s'apparentant à une clef composée chacun de ses éléments appartenant à une colonne distincte.

Les données finales extraites par scrapMaps composent la base finalMaps directement accessible via le dossier Entreprises et sont ordonnées en 5 colonnes :

mapsName: nom du lieu retourné par la recherche.

Cette variable a deux fonctions ;

Premièrement si aucun nom de lieu n'est retourné par la recherche ;
Toutes variables lui correspondant se verront attribuer la mention 'NA' ce qui se traduit par une ligne vide dans Excel.

Dans un second temps, elle permettra lors du croisement des données de vérifier si la dénomination correspond bien à la recherche effectuée

mapsAddress : Cette variable a elle aussi le sacerdoce de vérifier que l'adresse récupérée correspond bien à la recherche effectuée

mapsSite : contient le DNS du site de 5300 entreprises.

A défaut de ne pas nous laisser à disposition leur mail sur Goolgle Maps, nos prospects référencent leur site, ce qui permet d'envisager la récupération des mail manquant suite au scraping d'une autre plateforme une fois ce dernier réalisé.

mapsPhone : contient 5465 numéros de téléphone du standard des prospects.

Cette variable pourrait être employée dans le cadre d'une campagne de Calling.

firstMatch : contient les valeur booléennes True, False
et l'opérateur NA (not available):

Lorsqu'une recherche est effectuée en fonction de sa correspondance avec la base de données du site il nous renverra sur :

. Une correspondance directe => firstMatch = True

Les adresses les plus sécurisées dans le cadre d'une campagne de marketing direct

. Une liste d'établissements ; là le programme ira récupérer le premier établissement de la liste => firstMatch = False

Ces données sont toujours utiles au projet d'enrichissement via une autre plateforme mais nécessitent un travail de comparaison plus poussé que pour les adresses de première catégorie pour le directe marketing.

. Aucune correspondance => firstMatch = NA (ou vide via Excel)

2 : Synthèse des données récoltées

Adresses sièges entreprises :

Sur les 6428 recherches effectuées :
Groupe A

. 4918 sont issues d'une recherche directe

Groupe B

. 1085 sont issues d'une recherche en deux temps

Groupe C

. 413 n'ont données aucun résultat

. 12 n'ont pas d'adresses (5 des recherches directes 7 pour les recherches en deux temps)

Note : Ces données sont complémentaires (4918 + 1085 + 413 + 12 = 6428)

- ⇨ Dans le cadre d'une campagne d'envoi de cadeaux d'entreprises cela représente jusqu'à 1510 retours à l'envoyeur évités

Coordonnées :

	Groupe A	Groupe B	Total
Téléphone :	4507	958	5465
Site :	4322	978	5330

IV : Etablissements Scolaires Privés

1 : Présentation de la base EtablissementScolairesPrives

Comme énoncé dans la partie II 2 Ecrémage des données;
Pour dénicher les contrats privé des établissements scolaires, la première base de données n'était pas suffisante.

Afin de déterminer les écoles privées basées en Île-de-France, j'ai donc commencé par scraper le site https://www.enseignement-prive.info qui m'est apparu comme la source la plus probante étant biaisé par le SEO Google et la structure lisse du site.

Le script correspondant de sa version 1 à 3 est disponible sous le dossier Script dans le sous-dossier scriptsEtablissementsScolaire sous les noms :

SrapEtablissementScolaires

ScriptEtablissementsPrivés1

ScriptEtablissementsPrivés2

La base de données qui en est issue ; EtablissementsScolairesPrives est classée dans le dossier GroupeScolairePrives.

Elle comporte deux colonnes, l'une contenant les noms des écoles récoltées, l'autre leurs adresses et fait 972 lignes.

2 : Présentation de la base fr-en-annuaire-education

J’ai découvert une fois le scraping d’enseignement-prive.info qu’il existait une API des données du ministère de l’éducation (cf : https://www.education.gouv.fr/annuaire) donnant accès aux établissements scolaires recensés et précisant leur type de contrat.

Ce jeu de données bien que semblant incomplet à la suite des explorations détaillée dans le sous chapitre suivant contient des informations pertinentes tels le mail et les numéros de téléphones référencés par les établissements scolaires.

Le jeu de données est disponible sous le nom fr-en-education dans le dossier DroupesScolairesPrives

3 : Croisement des données

1 fusion des deux bases

Après avoir filtré puis téléchargé fr-en-annuaire-education, j’en ai concaténé puis formaté les colonnes pertinentes afin de composer leurs adresses en une seule colonne puis de comparer les valeurs de cette dernière avec la base EtablissementScolairesPrivés.

J’ai ensuite réalisé trois bases de données:

. La première résultant d’un inner join d’EtablissementScolairesPrives a et de fr-en-annuaire-education ne compte que 253 lignes

. La deuxième EtablissementScolairePrives un Left Join avec fr-en-annuaire-education n’en compte donc que 516

. La Troisième un Right Join compte1393 lignes

Un formatage incomplet ou un référencement d’adresse différent peut toujours avoir survenu c’est l’une des raisons pour laquelle les opérations qui vont suivre ont été effectuées.

2 Groupe Unité Légale

Afin de déterminer les unités légales de chaque établissement et ainsi d’éviter de prospecter deux fois chez le même décideur, j’ai récupéré tous les établissements dont la section NAF est l’éducation via sirene.fr, puis une nouvelle fois formaté les adresses afin de les faire matcher avec la concaténation des trois bases précédentes.

Cette base est nommée etablissementsSirene

Je n’ai pour l’instant qu’un match, dû soit au formatage soit au non-référencement ou autre référencement des adresses recherchées (un même établissement pouvant être accessible par plusieurs voies) ou encore une mauvaise classification des établissements..

Le match entre etalissementSirene et fr-en-annuaire-education reste à être effectué grâce à leur variable Siret, cependant la base de données scrapée elle, n’en dispose pas.

Pour pallier à ce problème j’ai groupé les établissements partageant la même adresse et afin d’avoir une première idée des groupes scolaires nichés dans nos bases de données puis les aient concaténés et exportés sous le nom d’adressesEcoles au sein du même dossier.

Je compte lancer le programme de scraping de Maps pour chaque établissement recensé à chaque adresse afin de récolter plus de coordonnées et de réessayer le croisement avec etablissementsSirene.

V : Copropriétés

Mes recherches n’ont abouti pour l’instant qu’à une API gouvernementale ne permettant les extractions qu’au compte-goutte sans le compte d’une institution autorisée.

Le scraping de cette base est envisageable grâce à un croisement des données avec celui du site cependant dû au temps d’exécution, l’opération nécessitera l’usage d’un serveur virtuel privé (VPS).

Je dois cependant encore réaliser des recherches plus poussées afin de trouver une solution plus simple