



Rapport de stage de Master 1

MIASHS



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*



France
Travail

Réalisé par : TIEMTORE Wendyam Ariel

Tutrice d'entreprise : Blandine VACHER

Tuteur universitaire : Stéphane CHRETIEN

Année universitaire : 2024 / 2025

Remerciements

Je tiens à remercier particulièrement Blandine VACHER qui m'a accepté en stage chez France Travail, et m'a suivi de très près. Sans elle je n'aurais pas tant appris, ainsi qu'à Vincent GIQUET, le responsable du service d'accueil du stage, qui m'a suivi durant mon stage.

J'adresse mes remerciements également aux collaborateurs de France Travail plus spécialement aux collaborateurs du service Statistiques Etudes et Evaluations avec lesquels j'ai collaboré durant mon stage, pour leur professionnalisme, leur soutien et leur sympathie.

Je remercie également Stephane CHRETIEN, le responsable de la formation de Master 1 MIASHS qui a été disponible tout au long du stage.

Je fini par remercier chaleureusement ma famille et principalement mes parents, qui m'ont tout procuré pour que je puisse réussir.

SOMMAIRE

Remerciements	2
INTRODUCTION GENERALE.....	4
Présentation de l'organisme d'accueil.....	5
1. Introduction	6
2. Présentation de France Travail	6
3. Présentation de la région et de la Direction Régionale Auvergne-Rhône-Alpes.....	6
4. Présentation du service Statistiques, Etudes et Evaluations.....	9
Missions et déroulement du stage	11
1. Missions du stage	12
2. Déroulement du stage.....	12
2.1. Exploration des bases de données	13
2.2. Constitution d'une base unique	15
2.3. Description des établissements de la région.....	16
2.4. Autres missions	18
2.5. Rencontre et immersion	20
3. Quelques difficultés rencontrées	21
4. Les livrables.....	22
Classification des établissements de la région	23
1. Description des établissements de la région.....	24
2. Choix des variables et du modèle de classification.....	25
3. Tests et résultats.....	26
Culture d'entreprise et enseignements du stage	37
1. Culture d'entreprise.....	38
2. Encadrement et ambiance de travail.....	39
Conclusion.....	40
ANNEXES	42
Bibliographie.....	60
Sites internet.....	60

INTRODUCTION GENERALE

Le marché du travail français est en perpétuelle mutation, confronté à des défis économiques, technologiques et sociaux sans précédent. Dans ce paysage complexe, France Travail, institution publique anciennement Pôle Emploi, joue un rôle pivot en assurant la facilitation de la rencontre entre l'offre et la demande d'emploi. Pour mener à bien ces missions, France Travail s'appuie sur un maillage territorial dense, notamment via ses Directions Régionales.

C'est dans ce contexte et dans l'optique de renforcer son partenariat avec les entreprises, que j'ai réalisé mon stage au sein de la Direction Régionale Auvergne-Rhône-Alpes (DR ARA) de France Travail.

Afin de présenter de manière détaillée le déroulement et les résultats de ce stage, ce rapport s'articulera en quatre parties principales :

- La première partie présentera l'organisme d'accueil, France Travail, sa Direction Régionale ARA et le service SEE.
- La deuxième partie détaillera les missions qui m'ont été confiées et leur déroulement opérationnel, en insistant sur la méthodologie de constitution de la base de données.
- La troisième partie sera consacrée aux résultats et à l'analyse de la classification des établissements, cœur de mon étude.
- Enfin, la quatrième partie reviendra sur le bilan de cette expérience, en termes de compétences développées, de culture d'entreprise et d'enseignements personnels.

Présentation de l'organisme d'accueil

1. Introduction

Mon stage s'est déroulé à la Direction Régionale Auvergne-Rhône-Alpes (DR ARA) de France Travail. J'ai été affecté au service Statistiques, Etudes et Evaluation (SEE).

2. Présentation de France Travail

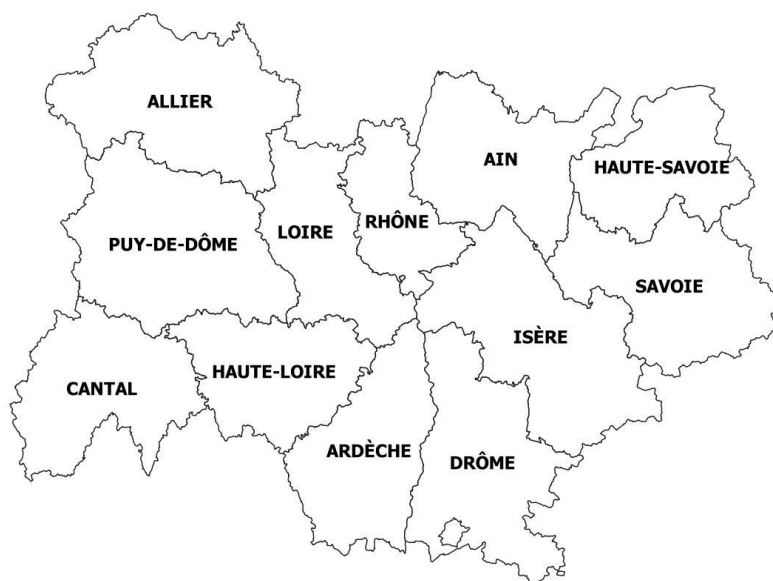
France Travail, précédemment Pôle Emploi, est une institution publique française qui a pour mission principale la facilitation de la rencontre entre les demandeurs d'emploi et les employeurs. Cette institution articule ses missions autour de trois axes (03) majeurs que sont : l'accompagnement des demandeurs d'emploi, les services aux entreprises et la gestion des aides et des allocations. Elle est dirigée par Thibaut Guilluy, le directeur général et Frank Denié, le directeur des systèmes d'information (DSI). En termes d'organisation, on retrouve, une direction générale, une direction financière, une direction de l'informatique, des directions régionales et des agences.

3. Présentation de la région et de la Direction Régionale Auvergne-Rhône-Alpes

La région Auvergne-Rhône-Alpes (ARA) est l'une des 13 régions administratives de la France métropolitaine. Elle est composée de 12 départements : Ain, Allier, Ardèche, Cantal, Drôme, Isère, Loire, Haute-Loire, Puy-de-Dôme, Rhône, Savoie et Haute-Savoie. La région est caractérisée par une grande diversité géographique, allant des montagnes des Alpes et du Massif Central aux plaines de la Saône et du Rhône. (Institut National de la Statistique et des Etudes Economiques). La région ARA est un pôle économique majeur en France, avec des secteurs d'activité variés tels que l'industrie, les services, l'agriculture et le tourisme. Les principales villes de la région sont Lyon, Grenoble, Saint-Étienne, Clermont-Ferrand, Annecy et Chambéry. La région est également connue pour ses établissements d'enseignement supérieur et de recherche, ainsi que pour ses infrastructures de transport développées. (Institut National de la Statistique et des Etudes Economiques).

Graphique 1 : Cartographie des départements de la région ARA

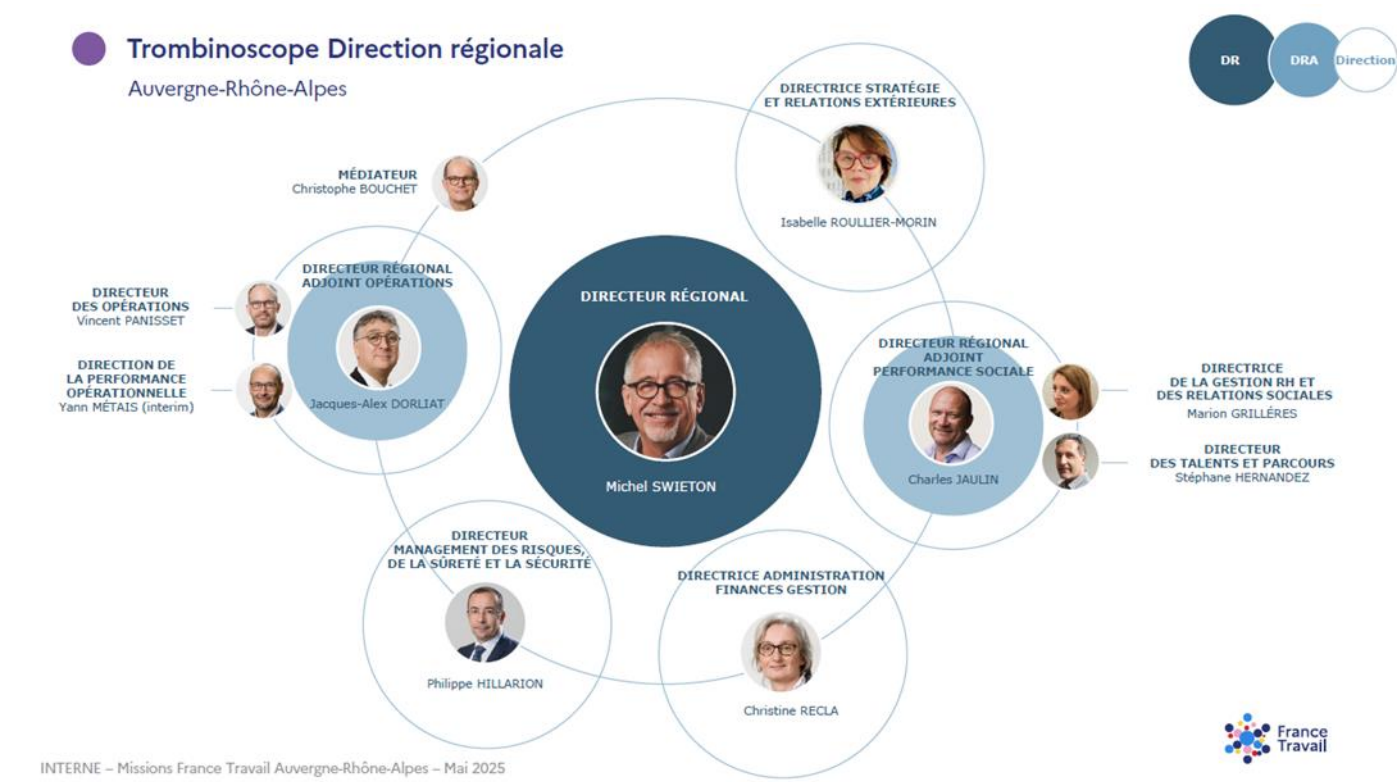
Carte de la région Auvergne-Rhône-Alpes



La Direction Régionale Auvergne-Rhône-Alpes (DR ARA) de France Travail est une structure qui pilote, coordonne et accompagne les politiques publiques de l'emploi, de la formation et de l'insertion sur l'ensemble du territoire régional. Le Directeur Régional est Michel SWIETON qui dirige environ 95 agences et plus de 6000 collaborateurs.

L'organigramme de la DR ARA repose sur une gouvernance collaborative avec des adjoints de direction, des responsables de pôles métiers et des référents fonctionnels. Chaque direction dispose de services spécialisés avec des missions précises.

Graphique 2 : Schéma des différents services de la Direction Régionale ARA et leurs responsables



La DR ARA compte cinq (05) directions qui sont :

- Direction Régionale Adjointe des Opérations ;
- Direction Administration, Finances et Gestion ;
- Direction Management des Risques, de la Sûreté et de la Sécurité ;
- Direction Stratégie et Relations Extérieures ;
- Direction Performance Sociale.

C'est au sein de la Direction Régionale Adjointe des Opérations (DRAO) que se trouve le service Statistiques, Etudes et Evaluations (SEE) (le service d'accueil pour le stage). Cette direction a la charge de l'animation du réseau des agences et directions territoriales, du pilotage de la performance opérationnelle, de la gestion de l'offre de service, du suivi des relations avec les entreprises et de l'analyse statistique et évaluative. Elle soutient les agences dans l'atteinte de leurs objectifs et le déploiement harmonisé des services de France Travail. Les principales missions de cette direction sont : accélérer le retour à l'emploi, développer l'offre de service aux demandeurs d'emploi et aux entreprises, piloter la performance opérationnelle et animer le réseau des agences locales.

En ce qui concerne les partenaires, la DR ARA peut compter sur des structures comme la région Auvergne-Rhône-Alpes, les départements, les métropoles, les chambres consulaires, les partenaires sociaux et les acteurs de la formation.

France travail met en place des projets structurants tels que :

- La digitalisation des services plus précisément le développement de l'offre multicanal ;
- L'innovation sociale grâce au laboratoire d'innovation (LAB régional) ;
- L'approche territoriale c'est-à-dire l'adaptation des services aux spécificités locales ;
- La responsabilité sociétale grâce à une politique RSE ambitieuse.

La DR ARA a pour perspective 2025-2030, le renforcement de la relation de service aux usagers, le développement des compétences numériques, l'adaptation aux transformations du marché du travail, la modernisation des outils et processus et le renforcement de l'attractivité territoriale.

4. Présentation du service Statistiques, Etudes et Evaluations

Le service Statistiques Etudes et Evaluations est constitué de Vincent GIQUET (le responsable) et ses sept collaborateurs, Saïd Hamidou ALLAOUI, Alexandre BLONDON, Delphine FAURÉ, Stéphane LASCOMBE, Isabelle OLKOWICZ, Karine SABY et Blandine VACHER qui est la tutrice du stage.

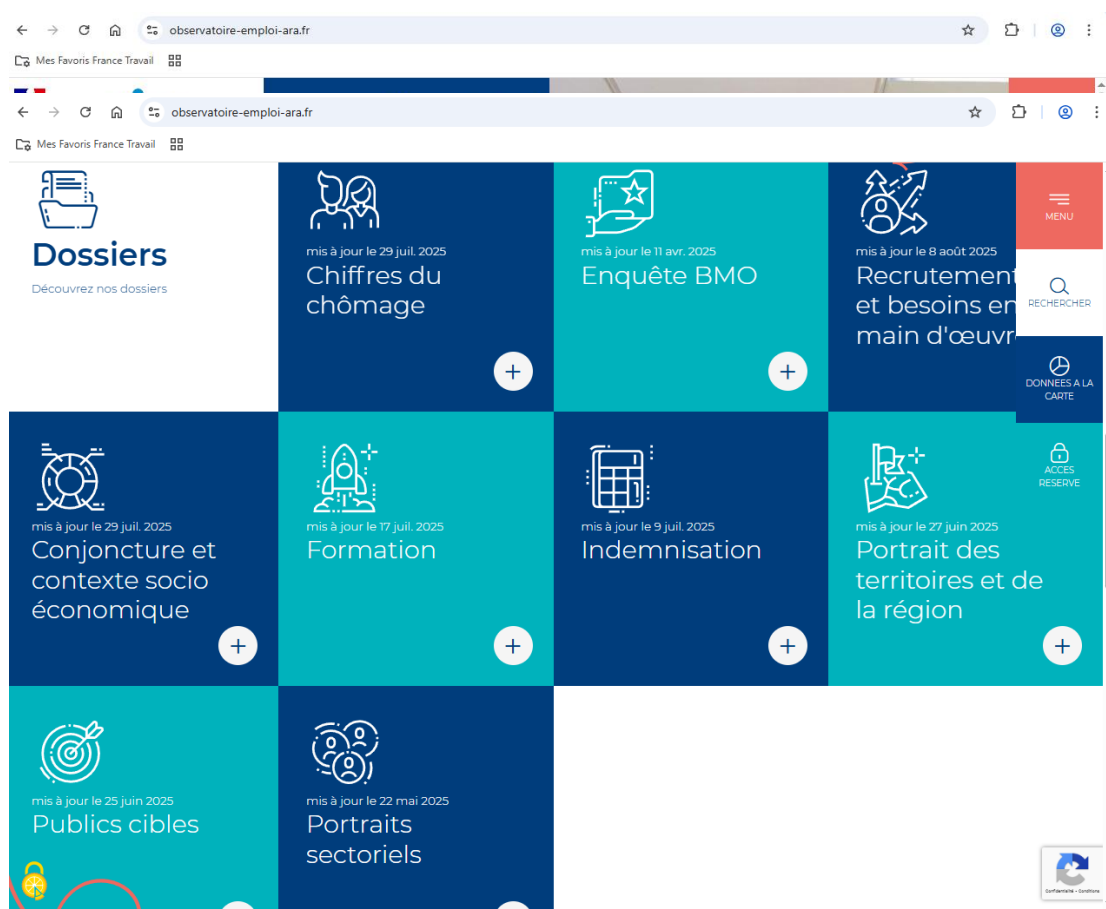
Vincent GIQUET et ses collaborateurs ont pour mission quotidienne de :

- Produire des analyses, études, prévisions et enquêtes pour comprendre l'évolution de l'emploi et du chômage local.
- Diffuser l'information sur les dynamiques du marché de l'emploi auprès des publics internes et externes.
- Contribuer au pilotage par les résultats en étudiant l'effet des actions menées par les agences.
- Évaluer les programmes et dispositifs déployés dans la région.
- Fournir des outils d'aide à la décision pour guider les stratégies régionales.
- Réaliser des diagnostics socio-économiques complets pour identifier les forces, faiblesses et opportunités des territoires.

- Fournir des données sur mesure via l'Observatoire de l'emploi Auvergne-Rhône-Alpes pour les besoins internes (diagnostics, plans d'action) ou externes (réunions avec élus, médias, partenaires).
- Assurer une communication claire grâce à des publications (rapports, portraits de territoires, notes mensuelles), des infographies et supports visuels et des vidéos explicatives pour le grand public et les partenaires. (Offre de Services, SEE ARA).

Toutes les productions et les indicateurs du service sont vulgarisés grâce au partage des publications et des données en interne et en externe au travers de l'observatoire de l'emploi (<https://www.observatoire-emploi-ara.fr/>).

Graphique 3 : Interface de l'observatoire de l'emploi



Missions et déroulement du stage

1. Missions du stage

Une des priorités de France Travail actuellement est d'aller vers les entreprises de nos territoires afin mieux faire connaître notre offre de services, révéler le potentiel de recrutement et promouvoir des profils de candidats. Afin de développer des compétences en analyse de données et de contribuer directement à une meilleure connaissance des entreprises de la région, le service Statistiques, Etudes et Evaluation de la DR ARA a recruté un stagiaire (à l'occurrence moi), pour les missions suivantes :

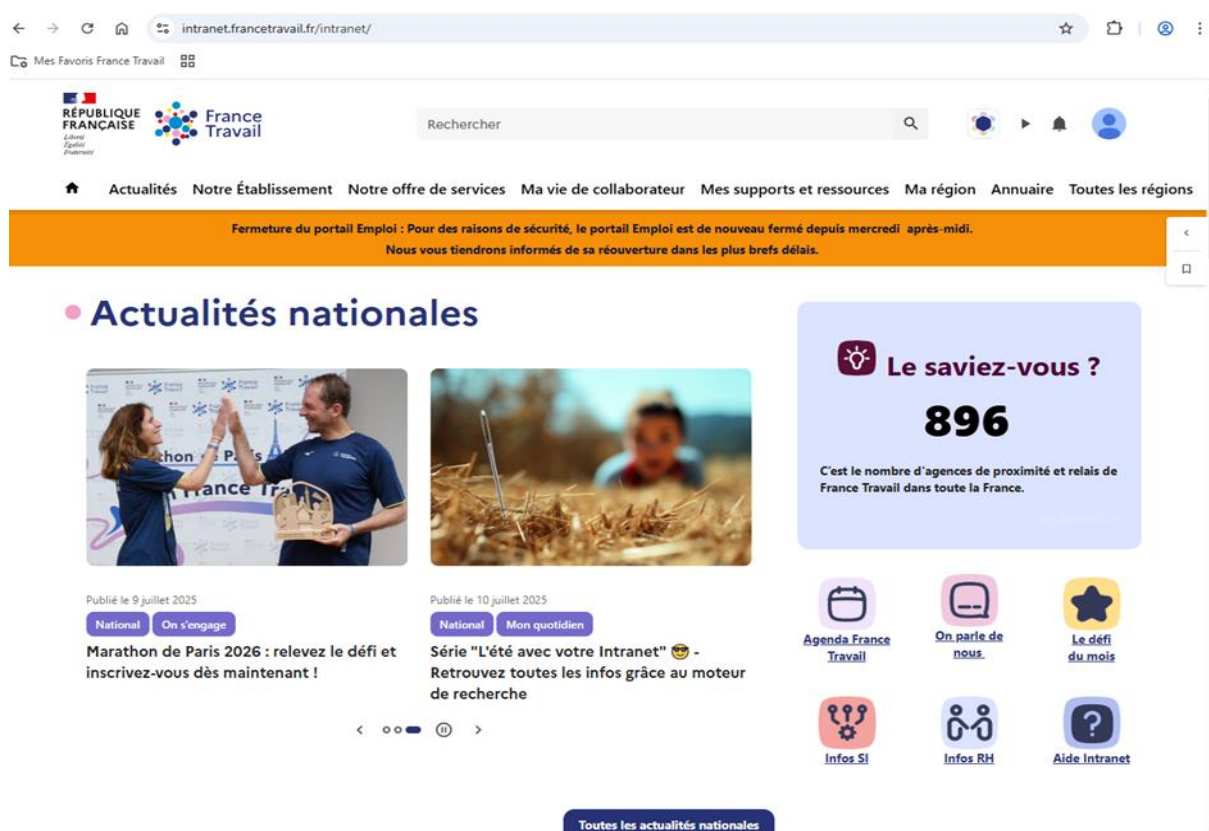
- Récupération et traitement des données sur les entreprises de la région : croiser des sources de données internes ou externes (BMO, INSEE, DPAE, offres, population mère), et constituer une base de données propre et riche.
- Étude régionale et classifications : Réalisation d'une cartographie des établissements de nos territoires pour mieux comprendre leur répartition et potentiel et améliorer le recours à France Travail. Rédaction d'une étude régionale.
- Développement d'outil pour la prospection : Conception d'un outil opérationnel (Power BI, Excel, VBA) destiné aux conseillers entreprise en agence pour faciliter, piloter et suivre leur prospection et avoir une meilleure connaissance de leur territoire.

2. Déroulement du stage

Le stage a débuté le 1er Avril 2025 pour une durée de six (06) mois au sein des locaux de la DR ARA. A mon arrivée, j'ai fait la connaissance des membres de l'équipe. Après la remise du matériel (ordinateur, etc.) et l'installation des applicatifs, j'ai pris connaissance de l'intranet de France Travail, des tâches de chaque collaborateur de l'équipe SEE et aussi du RGPD (Règlement Général sur la Protection des Données). Le RGPD est essentiel à France Travail, qui gère quotidiennement des données sensibles sur les demandeurs d'emploi, les entreprises et ses partenaires. Il garantit la confidentialité, la sécurité et la transparence des informations, tout en assurant la conformité légale et la crédibilité de l'institution. Le RGPD encadre les pratiques internes en limitant la collecte aux données nécessaires, en régulant leur conservation et en sensibilisant les équipes à leur protection. Il sécurise également le développement des services numériques et renforce la relation de confiance avec les usagers, en leur offrant des droits sur

leurs données. Cette culture de la protection des données s'intègre désormais dans toutes les activités et projets de l'établissement

Graphique 4 : Interface de l'intranet France Travail



Après les premiers jours d'intégration, j'ai commencé ma mission qui s'est structuré comme suit :

2.1. Exploration des bases de données

La première phase a été l'exploration des bases de données pour mettre en place une base de données des établissements de la région la plus exhaustive possible. Pour ce faire, nous avons eu recours à la base SIRENE de l'INSEE, la base population mère BMO (Besoin en Main d'œuvre), la base DPAE (Déclaration Préalable A l'Embauche) et la base des Offres ; les trois dernières bases étant des productions de France Travail.

La base SIRENE étant une source externe, certes en open source, a été la base la plus difficile à explorer. En effet, pour avoir accès aux données de la base, il y a deux possibilités : passer par la page internet dédiée (<https://www.sirene.fr/sirene/public/recherche>) ou passer par un téléchargement grâce à une API (Application Programming Interface).

2.1.1. Sirene.fr

La page internet dédiée nous permet de télécharger gratuitement la base SIRENE, constituer gratuitement une liste d'établissements sur mesure, rechercher par unité légale ou interroger les données SIRENE par API. Pour notre étude il était important pour nous de constituer une liste d'établissements sur mesure car nous voulons uniquement les établissements actifs et employeurs de la région ARA hors activités des ménages.

Graphique 5 : Interface de la base SIRENE de l'INSEE

The screenshot displays the Sirene.fr website interface. At the top, there's a navigation bar with links like 'Menu', 'Contenu', 'Qui sommes-nous?', and 'FR | EN'. Below this, the Sirene.fr logo is accompanied by the text 'Fournisseur officiel d'information sur chacun des 40 millions d'établissements dont 15 millions actifs. Un service de l'Insee.' The main content area is titled 'Constituer une liste' and includes a search bar 'Rechercher sur le site sirene.fr'. The interface is divided into several sections: 'Choisir les caractéristiques de votre fichier' with options for 'État administratif des établissements' (active or closed), 'Type d'établissement' (all enterprises or only headquarters), 'Appartenance au champ de l'économie sociale et solidaire' (yes, no, or not specified), and 'Appartenance au champ des sociétés à mission' (yes, no, or not specified). There's also a 'Choisir vos critères de sélection' section with filters for 'Localisation', 'Activité', 'Catégorie juridique', 'Effectif', 'Date de création', and 'Date de mise à jour'. The 'Localisation' filter is expanded, showing 'Commune', 'Département', and 'Région'. On the right side, there's a section 'Toute l'actualité de sirene.fr' with a video player and a section 'Effectuer des démarches' with buttons for 'OBTENIR', 'MODIFIER / CESSER', 'OBTENIR ou', 'CATÉGORIE D'ENTR', 'OBTENIR un', 'AVIS DE SITUATION', 'Modifier mon', 'STATUT DE DIFFUSI', and 'AUTRES DÉMARCHES' and 'FAQ'. At the bottom, there's a footer with the text 'Fraude, malveillance, factures abusives'.

Cette méthode bien qu'appropriée fut laborieuse car le téléchargement est limité à vingt-cinq mille lignes (25000), or nous avions une idée du nombre d'établissements qui est d'environ trois cent mille (300000). Nous avons alors procédé à plusieurs téléchargements en fonction des départements et ensuite fusionner les différentes bases pour obtenir une base unique qui contient les établissements des douze (12) départements.

2.1.2. API SIRENE

Une API (Application Programming Interface) ou interface de programmation d'application est une interface qui permet à un programme d'envoyer ou de recevoir des données vers/depuis un autre service ou application.

Concrètement dans notre cas, nous voulons des données de la base des établissements de l'INSEE qui est volumineuse. Nous nous sommes alors servis du programme que l'INSEE met en place, API SIRENE, pour avoir accès aux données. Pour ce faire, nous nous sommes rendus sur le site de l'INSEE dédié à ses API.

Après avoir suivi toutes les étapes, nous avons pu télécharger un fichier sous format .json et à partir de ce fichier utiliser une requête python qui nous a permis de lire le fichier téléchargé et générer une base sous format csv. Enfin, grâce aux différents filtres nous avons pu avoir une base des établissements actifs et employeurs de la région ARA hors activités des ménages. (Voir script python en annexe 1).

Finalement, nous avons continué le travail avec la base obtenue en passant par le téléchargement en plusieurs fichiers.

2.2. Constitution d'une base unique

Comme vu précédemment, nous avons eu recours à quatre (04) bases de données pour constituer notre base unique.

La base Population mère BMO est constituée via les bases ACROSS, DPAE, Population mère BMO de l'année précédente (N-1) et SIRENE. Les trois premières sont fournies par France Travail et la dernière par l'INSEE. Le périmètre de cette base est le même que celui de l'enquête BMO ARA c'est-à-dire les établissements actifs et domiciliés dans la région. La base population mère BMO est une base de données des besoins en main-d'œuvre (BMO).

La base DPAE (Déclaration Préalable A l'Embauche) est constituée de l'ensemble des déclarations d'embauche réalisées au cours de l'année 2024 dans la région. Ainsi, on note 139761 établissements qui ont fait des DPAE en 2024.

Quant à la base Offre, il s'agit de toutes les offres de la région déposées à France Travail au cours de l'année 2024 et ces offres ont été déposés par 34673 établissements.

La base SIRENE est la base de l'ensemble des établissements du pays. On y retrouve tous les établissements actifs ou non et en appliquant des filtres sur la région, l'état administratif (actif ou pas) et aussi sur les particuliers employeurs, on aboutit à une base des établissements de la région ARA.

Pour alors constituer une base unique, on prend la base « Population mère BMO » et on effectue une fusion à gauche (fusion qui garde uniquement les lignes de correspondance avec la base Population mère BMO) des autres bases (DPAE, Offres et SIRENE) grâce au SIRET. Ainsi, on obtient une base de données avec 297464 établissements. Le choix de la base « Population mère BMO » est dû au fait que c'est la base la mieux renseignée en termes d'effectif salarié.

Les variables de la base Population mère BMO sont totalement renseignées, ce qui n'est pas le cas des variables des issues des autres bases d'où le choix de cette base comme base principale.

Nombres d'observations des variables issues des différentes bases				
	BMO	DPAE	Offres	SIRENE
Nombre observations	297464	139761	34822	195429
Poids dans la base	100%	47%	12%	66%

Comme nous le montre le tableau ci-dessus, les informations issues de la base DPAE renseignent 47% des établissements de notre base ; ce qui est logique car tous les établissements de la région ne sont pas recruteurs. Aussi, on note que 12% pour la base des offres ce qui, comme la base DPAE est logique car cette base ne contient que les établissements qui ont déposés des offres au cours des 12 mois de 2024. Les données issues de la base SIRENE ne sont renseignées qu'à 66% car certains établissements de la base population mère BMO ne se retrouve pas dans SIRENE et vice versa.

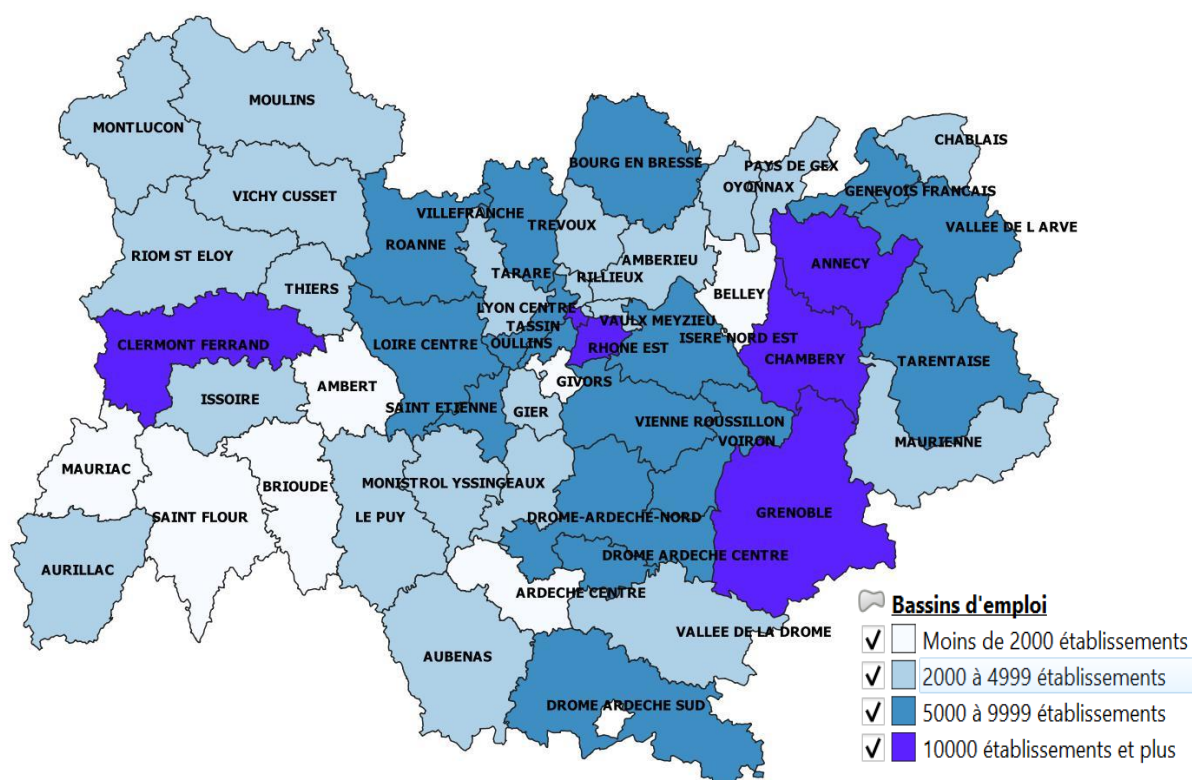
NB : Nous avons conservé toutes les variables des bases Population mère BMO, DPAE et offres. En ce qui concerne la base SIRENE, nous avons gardé que deux (02) variables, les coordonnées géographiques (longitude et latitude). On peut également consulter la liste des variables en annexe (annexe 2).

2.3. Description des établissements de la région

A partir de la nouvelle base créée, nous avons fait une étude (cette étude n'est pas encore terminée) sur les caractéristiques des établissements de la région ARA. Cela nous a permis de voir entre autres que :

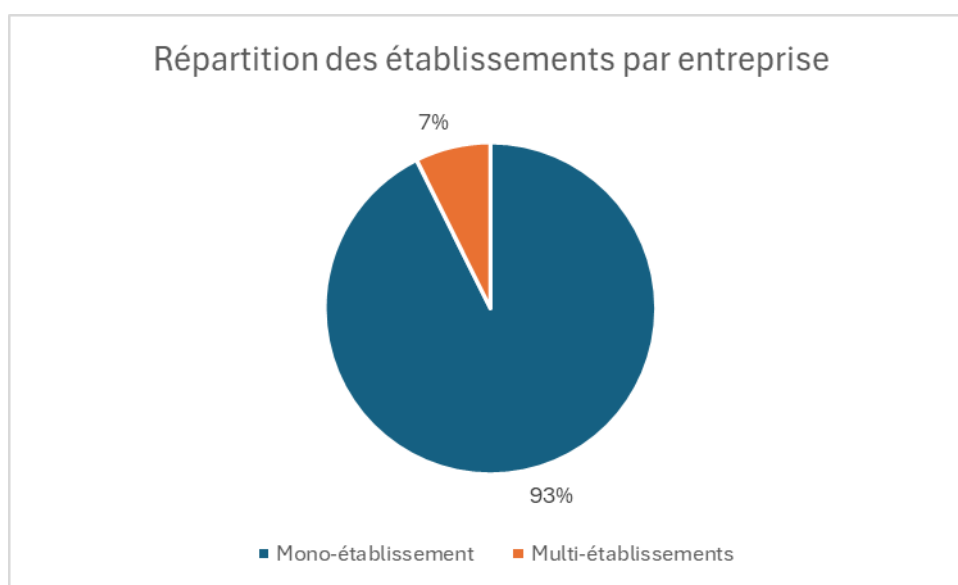
- Les dix (10) bassins d'emploi ayant le plus d'établissements regroupent à eux seuls près de la moitié (43%) des établissements de la région.

Graphique 6 : Cartographie de la région ARA avec les différents bassins d'emploi



- La région Auvergne-Rhône-Alpes présente une forte concentration d'établissements principalement dans les secteurs des services (57%) et du commerce (18%).
- Le département du Rhône est celui qui concentre le plus d'établissements.
- Les petites entreprises dominent en nombre mais sont moins actives en matière de déclarations d'embauche et de publications d'offres.

Graphique 7 : Répartition des établissements par entreprise



Après la description des établissements de la région, nous avons réalisé une classification qu'on a détaillé plus bas dans la partie Classification des établissements de la région.

2.4. Autres missions

En plus des différentes tâches énumérées plus haut, j'ai été sollicité sur des demandes concernant des cartographies et aussi concernant des données en rapport avec les entreprises. Par exemple, j'ai été sollicité pour produire une cartographie des entreprises en défaillance. Pour ce faire j'ai collaboré avec un collègue qui s'occupe de ce volet pour avoir les données nécessaires et aussi faire deux cartes (une carte pour intégrer dans une publication hebdomadaire et une autre carte interactive).

Tout d'abord il a mis à ma disposition une base de quelques entreprises en difficultés. Cependant, pour réaliser la carte, nous avons besoin des coordonnées exactes des entreprises car une même entreprise peut avoir plusieurs établissements qui sont situés dans le même département. Pour ce faire, nous avons le choix pour obtenir les coordonnées ; soit les chercher de manière manuelle en chargeant la base sur mistral intégré dans ChatFT et compléter les établissements manquants par une recherche google ; ou passer par une recherche automatique à partir d'un code python. Nous avons choisi la première méthode car nous n'avons pas trop d'entreprises aussi parce que l'objectif est de permettre au collègue d'être autonome (il ne programme pas sur python). Cela nous a permis d'avoir une base renseignée comme nous montre l'image ci-dessous :

Graphique 8 : Base de données pour la réalisation de la carte des défaillances d'entreprise

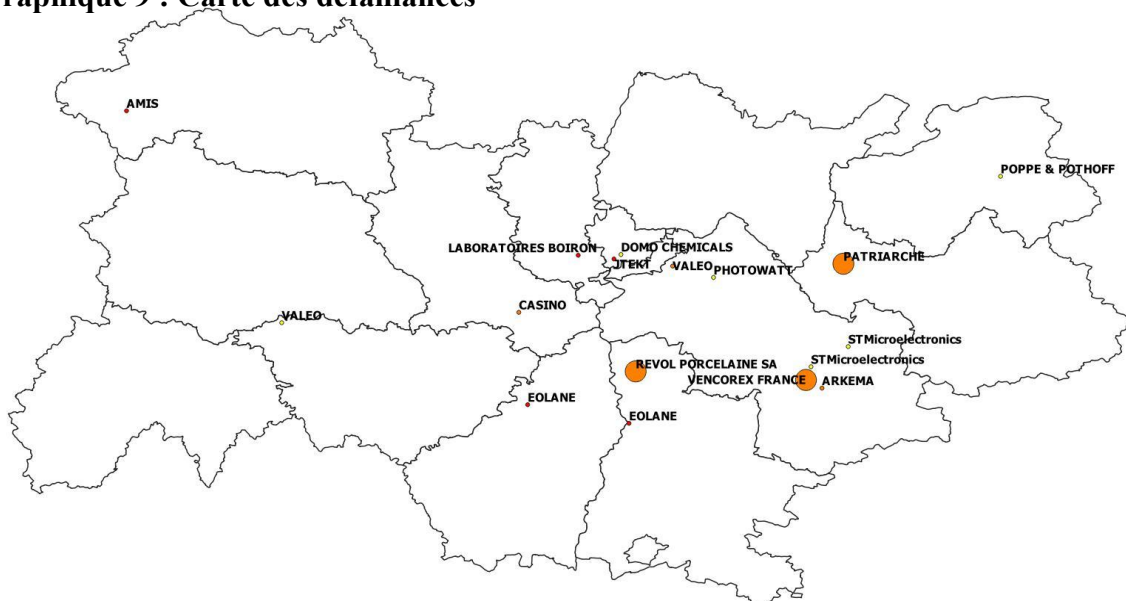
KC_ETABLIS	KC_ENTREP	SIRET	PRINC	Catégories	SIREN	PRINC	Raison social	KC_NAFREV	NAF	INSEE_COM	Code départ	Commune p	Impact DE m	Indicateur	Impact poss	longitude	latitude
11680100	11680099	4,44E+13	Top 15 des p		444187884	VENCOREX	2013B		Fabrication d	38317	38	LE PONT DE	48	2	RAS	5.6989	45.1289
25304775	25304774	8,03E+13	Top 15 des p		802586784	PATRIARCH	7111Z		Activités d'ar	73051	73	LE BOURGET	45	2	RAS	5.8694	45.6531
4743704	4743703	4,36E+13	Top 15 des p		435782172	REVOL PORC	2341Z		Fabrication d	26332	26	ST UZE (26)	40	2	RAS	4.9228	45.1677
26011379	"0001982744	3,49E+13	Top 15 des p		349361121	BEAUTY CRE	4645Z		Commerce d	74272	74	SILLINGY (74	26	1	RAS	6.0476	45.9431
1091454	1091453	3,19E+13	Top 15 des p		319086211	NOVAMOUL	2573A		Fabrication d	69091	69	GIVORS (69)	26	2	RAS	4.7677	45.5849
27053770	24959716	8,01E+13	Top 15 des p		801268848	7 FASHION	1414Z		Fabrication d	1053	1	BOURG EN B	24	2	RAS	5.2255	46.2044
42536120	42536119	8,91E+13	Top 15 des p		891373086	SABL CONST	4399C		Travaux de m	73225	73	ST BALDOPH	23	2	RAS	5.9186	45.5408
32112064	2632466	3,85E+13	Top 15 des p		385071881	OLYMPIQUE	9311Z		Gestion d'ins	69275	69	DECINES CH	22	3	RAS	4.9561	45.7656
11132846	"0011122538	5,00E+13	Top 15 des p		499872349	AMBITION T	6190Z		Autres activi	69290	69	ST PRIEST (6	21	1	RAS	4.9381	45.6966
5507592	5507591	5,62E+13	Top 15 des p		562086454	PCI-SCHEM	2841Z		Fabrication d	42218	42	ST ETIENNE	20	2	RAS	4.3872	45.4397
4319865	4319864	4,28E+13	Entreprises e		428267249	CASINO	7010Z		Activités des	42218	42	ST ETIENNE	0	2	2 200 emplois	4.3903	45.4339
3166007	3166004	3,99E+13	Entreprises e		399395581	STMicroelec	2611Z		Fabrication d	38140	38	CROLLES (38	0	3	2 000 à 3 000	5.8914	45.2794
11603066	11603065	5,05E+13	Entreprises e		504941337	STMicroelec	7219Z		Recherche-d	38185	38	GRENOBLE (3	0	3	2 000 à 3 000	5.7219	45.1885
2627530	2627529	3,85E+13	Entreprises e		384979555	EOLANE	2612Z		Fabrication d	7204	7	ST AGREVE (0	1	RJ, plus de 30	4.4306	45.0173
5188933	5188932	4,45E+13	Entreprises e		444617054	EOLANE	2612Z		Fabrication d	26362	26	VALENCE (26	0	1	RJ, plus de 30	4.8924	44.9334
86304309	7984174	4,79E+13	Entreprises e		479162695	VALEO	2931Z		Fabrication d	38449	38	ST QUENTIN	0	2	238 emplois	5.0912	45.6427
9223127	"0007984174	4,79E+13	Entreprises e		479162695	VALEO	2932Z		Fabrication d	43185	43	STE FLORINE	0	3	80 emplois st	3.3097	45.3871
1105370	1105363	3,20E+13	Entreprises e		319632790	ARKEMA	2013B		Fabrication d	38200	38	JARRIE (38)	0	2	suppression d	5.7719	45.0922
20707193	17237100	5,13E+13	Entreprises e		513281972	PHOTOWAT	2611Z		Fabrication d	38053	38	BOURGOIN J	0	3	Fermeture d	5.2776	45.5911
38213679	28812660	8,15E+13	Entreprises e		815232848	DOMO CHEN	2016Z		Fabrication d	69199	69	ST FONS (69)	0	3	Plus de 150 e	4.8558	45.6947
5652491	5652490	7,46E+13	Entreprises e		745780387	AMIS	2550A		Forge, estam	3185	3	MONTLUCO	0	1	148 emplois	2.6024	46.3428
5911420	"0005911395	9,68E+13	Entreprises e		967504697	LABORATOIR	2120Z		Fabrication d	69131	69	MESSIMY (69	0	1	145 emplois	4.6609	45.6914
5911744	5911743	9,68E+13	Entreprises e		967505967	JTEKT	2932Z		Fabrication d	69100	69	IRIGNY (69)	0	1	suppression d	4.8239	45.6744
5549852	5549849	6,06E+13	Entreprises e		605720630	POPPE & POT	2562A		Décolletage	74264	74	SCIONZIER (0	3	Fermeture d	6.5858	46.0474

Après avoir complété les coordonnées et après avoir compris ce qu'il voulait faire ressortir comme information, j'ai entrepris la réalisation des cartes sur QGIS et sur Power BI.

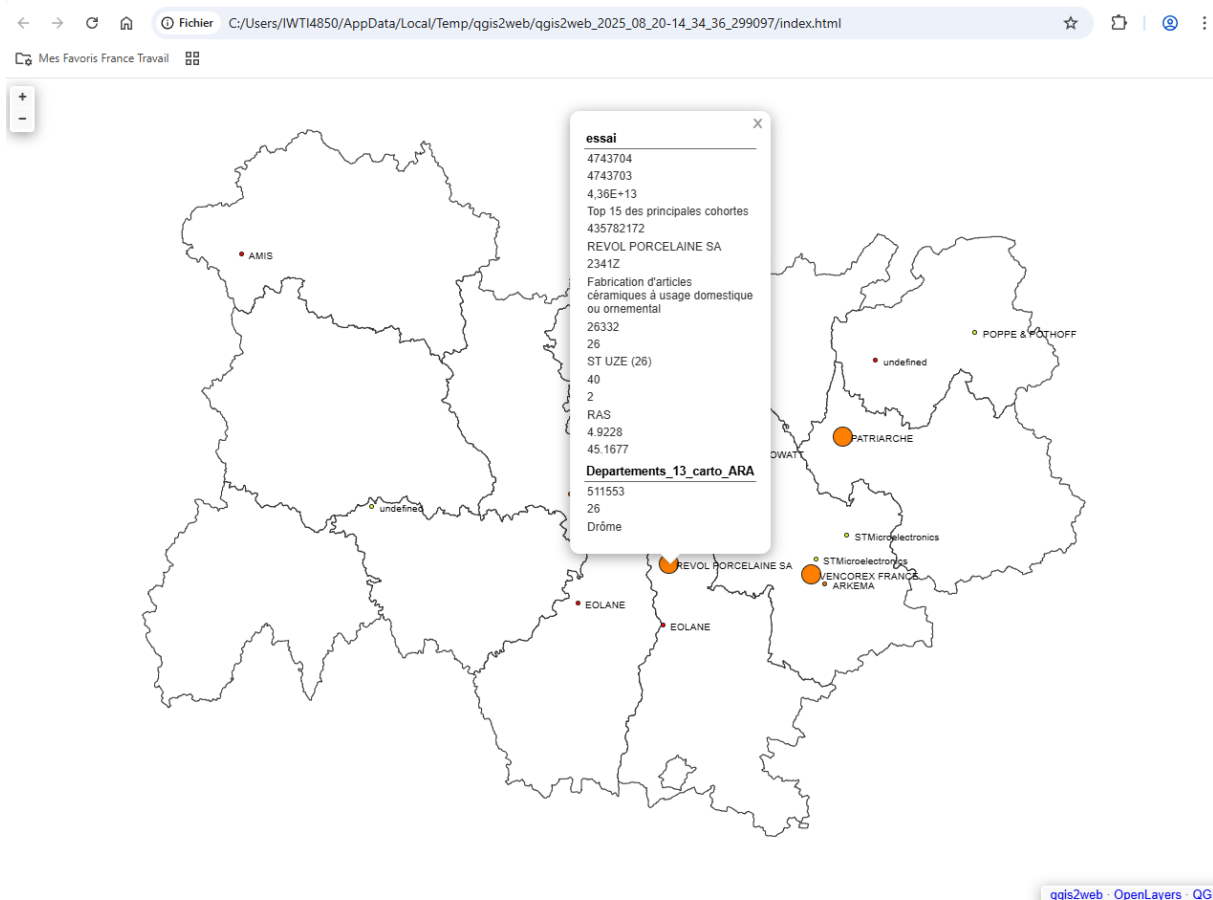
J'ai réalisé deux cartes sur QGIS, une carte à intégrer dans une publication et une autre interactive pour les personnes qui veulent plus d'informations sur l'étude. Pour réaliser ces deux cartes, j'ai installé deux extensions, quick-map-services pour accéder facilement à une large

variété de services de carte en ligne et qgis2web pour la carte interactive. Ainsi, on obtient les deux cartes suivantes :

Graphique 9 : Carte des défaillances

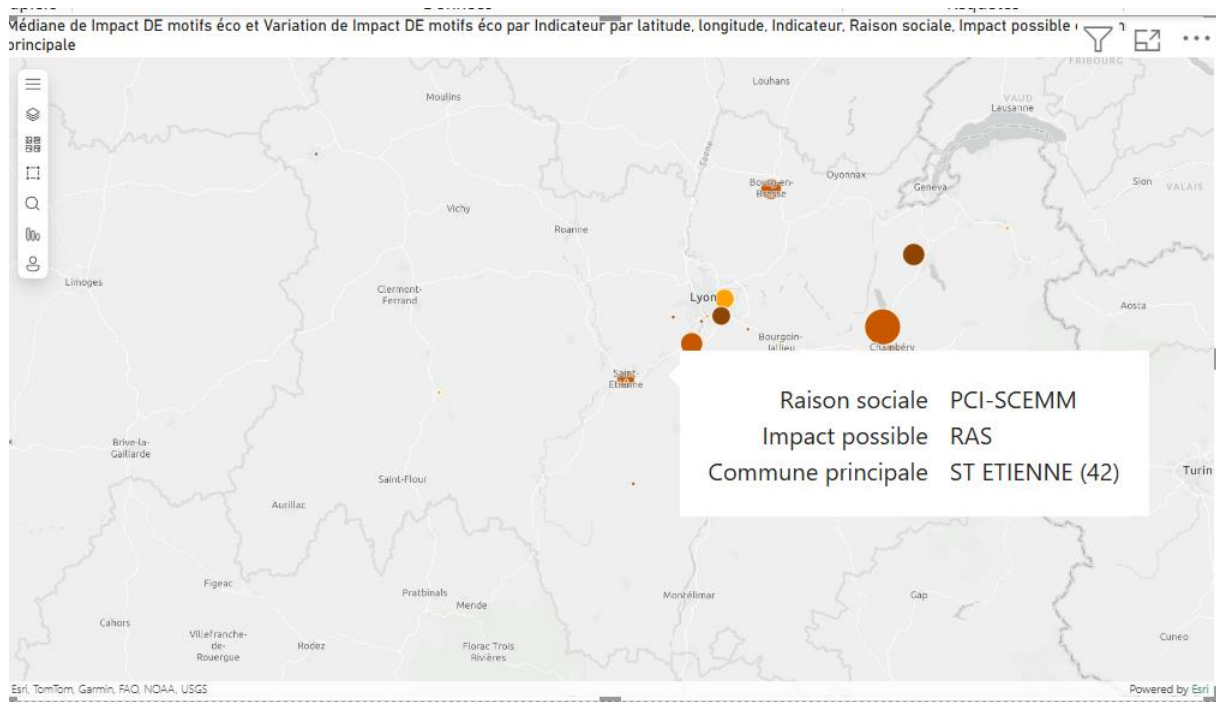


Graphique 10 : Carte interactive des défaillances



Enfin, j'ai proposé une carte interactive sur Power BI pour permettre une actualisation facile et aussi pour permettre au collègue d'avoir le choix de logiciel pour pouvoir réaliser ses cartes.

Graphique 11 : Carte interactive des défaillances réalisée avec Power BI



2.5. Rencontre et immersion

Pour mieux comprendre les données que j'analyse, notamment la base des établissements de la région, il était important de rencontrer des experts sur le volet entreprises qui ont pour mission d'accompagner les agences de proximité dans l'accélération des recrutements et la satisfaction des entreprises. C'est dans cette perspective que j'ai eu une rencontre avec un membre de l'équipe DDO (Direction des Opérations) Entreprise et une immersion en agence.

- La rencontre avec le membre de l'équipe DDO Entreprise m'a permis de voir les applicatifs qu'ils utilisent notamment Ciblage Entreprise et DUNE pour avoir un aperçu des entreprises de la région et aussi JTMO (Je Transfert Mes Offres) qui permet aux entreprises de transférer leurs offres à France Travail.

Graphique 12 : Outils Ciblage Entreprise

Ciblage entreprises

Effectuez la recherche d'établissements à partir de plusieurs critères. [Besoin d'aide ?](#)

CONSEILLER RÉFÉRENT ▼ STRUCTURE DE RATTACHEMENT ▼

MÉTIERS LES PLUS RECHERCHÉS SUR LES OFFRES ▼ SECTEUR D'ACTIVITÉ (NAF) DPAE SUR 12 MOIS ▼ AVEC PLAN D'ACTION EN COURS ▼

CLIENT FRANCE TRAVAIL ▼

Afficher plus de filtres ☐

Inclure les établissements recrutant pour un tiers ☐ [Réinitialiser tous les filtres](#)

13 903 891 établissements trouvés [EXPORTER](#)

Raison sociale	Commune (cedex)	N° SIRET	Conseiller référent	Correspondant privilégié
LA POSTE Fort potentiel d'embauche	PARIS 15 (75015)	[REDACTED]	[REDACTED]	Non VOIR LA FICHE
INFERENCE OPERATIONS	BAI MA (31130)	[REDACTED]	[REDACTED]	Oui VOIR LA FICHE

- L'immersion a eu lieu à l'agence de France Travail délocalisée au 8^{ème} arrondissement pour des raisons d'inondation. Cette immersion m'a permis de voir les applicatifs utilisés par les conseillers entreprises et aussi de voir comment ils travaillent : comment ils font les prospections des entreprises, comment ils utilisent les données produites par le service SEE au quotidien notamment un indicateur clé qui est l'ENT+ (c'est l'indicateur du recours à France Travail). À l'issue de cette immersion, j'ai beaucoup appris notamment sur la méthode de prospection, sur les services proposés aux entreprises et j'ai surtout obtenu des éclaircissements sur les questions que je me posais, notamment concernant le faible taux d'offres que j'ai observé dans l'étude.

3. Quelques difficultés rencontrées

Durant le déroulement du stage j'ai rencontré certaines difficultés :

- Au niveau logiciel : la DR ARA France Travail étant une structure gouvernementale, il y a des logiciels clés qui sont utilisés pour le traitement des données qui ne sont pas forcément les logiciels dont j'ai l'habitude d'utiliser. En effet, les principaux logiciels utilisés dans le service SEE pour le traitement et l'analyse sont Excel et SAS. Cependant, je suis plus à l'aise sur d'autres logiciels comme Python, R, Stata et SPSS. Pour pallier ce problème nous avons installé Python mais l'installation de certains packages sont difficiles parce qu'ils sont incompatibles avec les versions que nous utilisons, ce qui est un peu contraignant mais reste assez bien car certaines requêtes sont exécutables. Néanmoins, grâce aux différents collaborateurs du service, j'arrive à faire des analyses et des traitements sur les logiciels disponibles. Aussi, au niveau de la réalisation des cartes, certaines bibliothèques n'étaient pas téléchargeables. Cependant, grâce à l'IA notamment mistral de Chat FT, je suis parvenu à installer les outils équivalentes tels que qgis2web et quick map service.

- Au niveau de la constitution de la base de données, nous avons fait face à des problèmes. En effet, d'une base à l'autre, les périmètres varient. Cela pose un problème car la constitution d'une base exhaustive avec les différentes sources devient difficile. Néanmoins grâce à l'expertise de ma tutrice de stage et du responsable du service, nous avons fait un choix de base de données beaucoup plus renseignée (au niveau des effectifs salariés) que les autres que nous avons enrichi avec les autres sources.

4. Les livrables

A la fin du stage, je fournirai les éléments suivants :

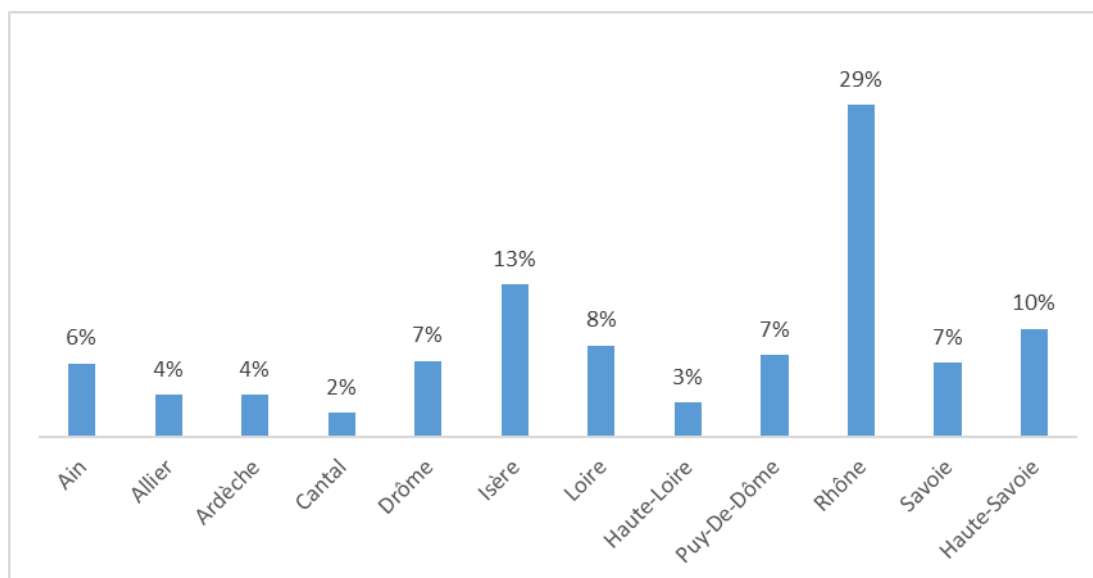
- Une étude sur les établissements de la région pour mieux comprendre leur répartition et potentiel et améliorer le recours à France Travail grâce à une classification.
- Un guide de processus détaillé pour pouvoir télécharger les données SIRENE par la méthode des API.
- Un guide de processus détaillé pour réaliser une cartographie sur QGIS telle que la cartographie défaillance évoquée plus haut.

Classification des établissements de la région

1. Description des établissements de la région

Dans la région Auvergne-Rhône-Alpes, on dénombre 243131 entreprises pour 297464 établissements. Les établissements sont repartis inégalement avec une forte concentration dans le Rhône (29,4%).

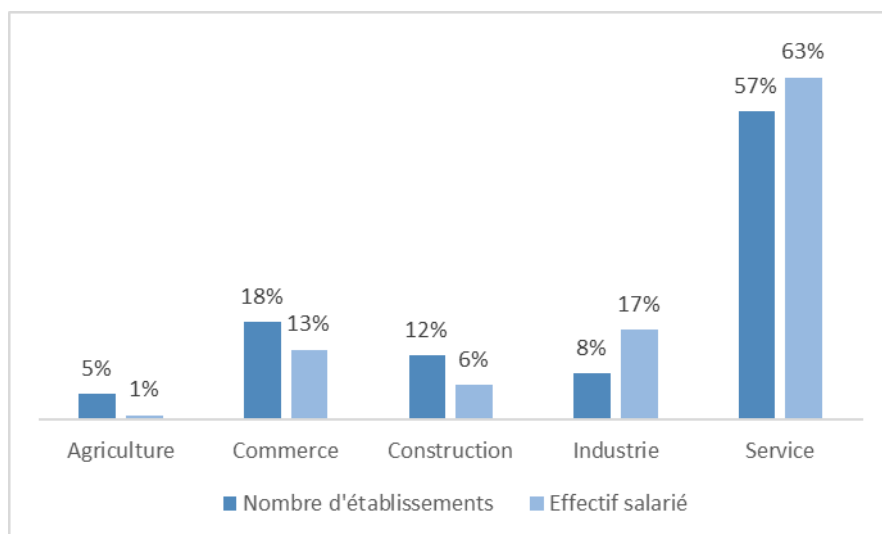
Graphique 13 : Répartition des établissements par département



La région est principalement composée de petits établissements avec 82% des établissements qui ont moins de 10 salariés. Aussi, 93% des entreprises n'ont qu'un seul établissement, même si certaines entreprises ont plus de 200 établissements, comme La Poste et EDF.

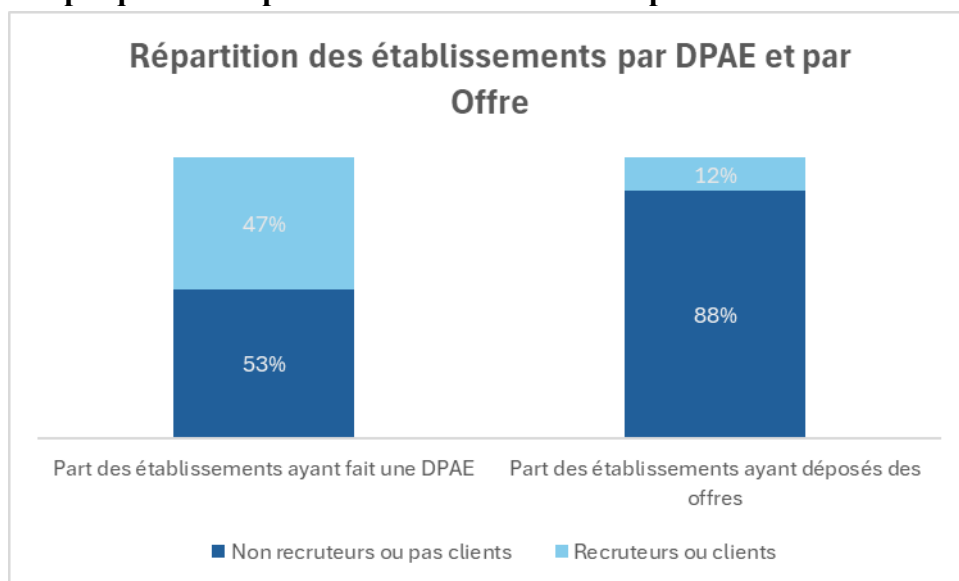
Le secteur des services domine largement avec 57% des établissements de la région et 63% des effectifs salariés. L'agriculture est le secteur le moins représenté avec 5% des établissements et 1% des effectifs salariés.

Graphique 14 : Répartition des établissements par secteur d'activité et par effectif salarié



En termes de dynamique d'emploi et d'offre, nous pouvons noter que près de la moitié des établissements de la région sont recruteurs (47%). Cependant, seulement 12% sont clients de France Travail.

Graphique 15 : Répartition des établissements par DPAE et Offre



2. Choix des variables et du modèle de classification

Nous avons choisi sept (07) variables parmi les variables de notre base par rapport à leur pertinence. Ces sept variables sont : les bassins d'emploi, les secteurs d'activité, la tranche effectif salarié, la densité, le nombre d'établissement par entreprise, les offres et les DPAE (déclaration préalable à l'embauche). On retrouve ces variables en deux types dans la base, nominal et catégoriel.

Pour faire la classification des établissements de la région nous avons utilisé une méthode de réduction de dimension, l'ACP (Analyse en Composante Principale), combinée à une méthode de clustering, K-means. On a aussi fait une AFCM (Analyse Factorielle des Composantes Multiples) avec les variables nominales pour choisir le modèle le plus adapté. L'ACP donnant un meilleur résultat notamment avec une variance expliquée à 91,21% nous l'avons donc choisi par rapport à l'AFCM (à peine 50%).

3. Tests et résultats

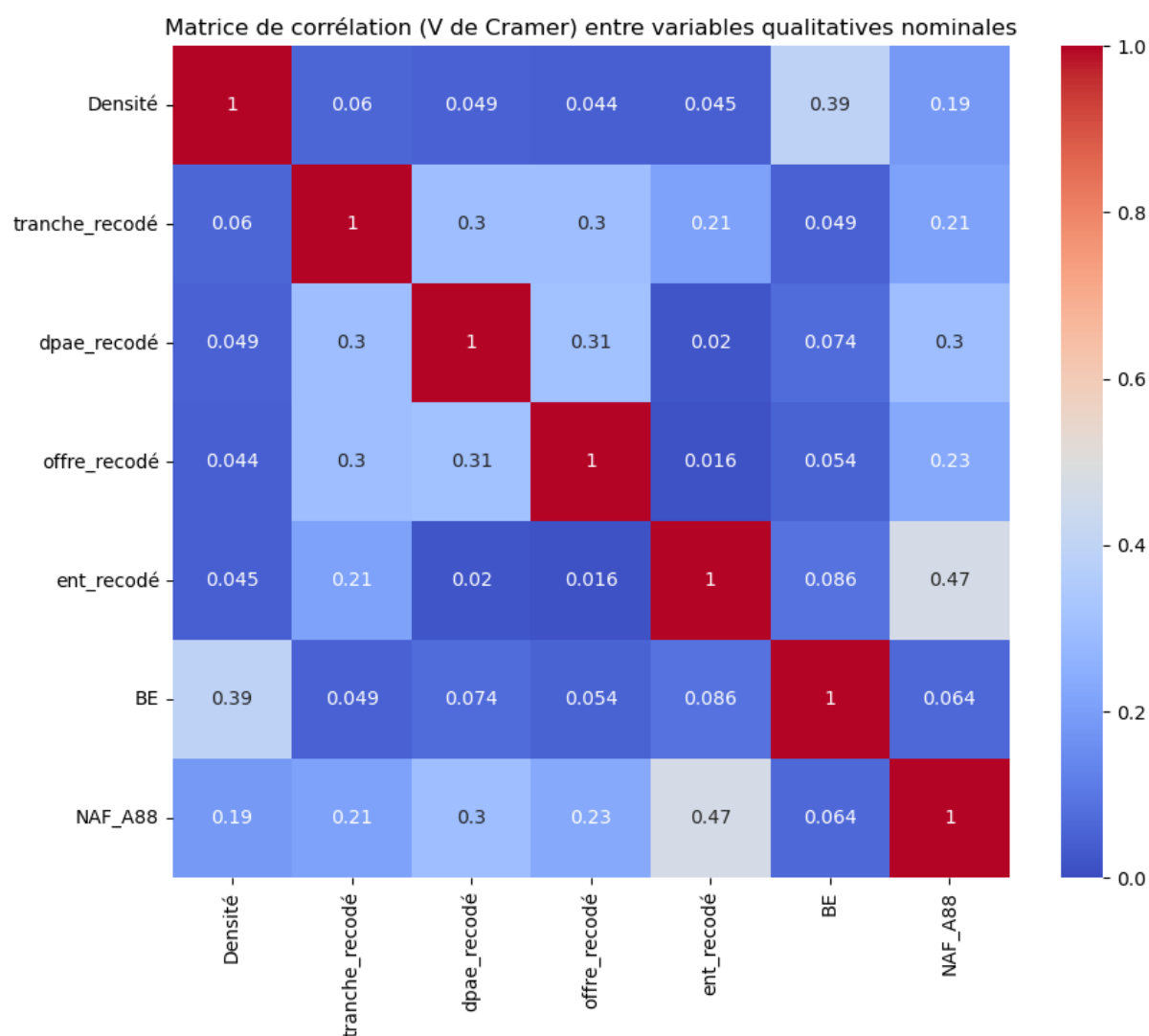
Comme décrit plus haut, nous avons fait le choix d'une ACP suivi d'un K-means pour faire notre étude. Cependant, il était important de choisir les variables de telles sortes que le modèle représente assez bien les données car au niveau géographique et sectoriel, nous avons plusieurs variables qui correspondent à différents niveaux de détail. Pour ce faire, nous avons réalisé plusieurs tests (ACP) qui nous ont permis d'avoir les résultats suivant en considérant les cinq dernières variables (05) énumérées plus haut et en faisant varier les variables géographiques et sectorielles :

- **Bassin d'emploi + NAF 38** : 66,4% de variance expliquée ;
- **CLPE + NAF 38** : 59,27% de variance expliquée ;
- **CLPE + NAF 88** : 59,19% de variance expliquée ;
- **Bassin d'emploi + NAF 88** : 91,21% de variance expliquée.

Nous avons donc choisi au niveau géographique le bassin d'emploi et au niveau sectoriel la NAF 88.

Toutes les variables étant significatives avec une $p\text{-value} < 0,5$, nous obtenons la matrice suivante :

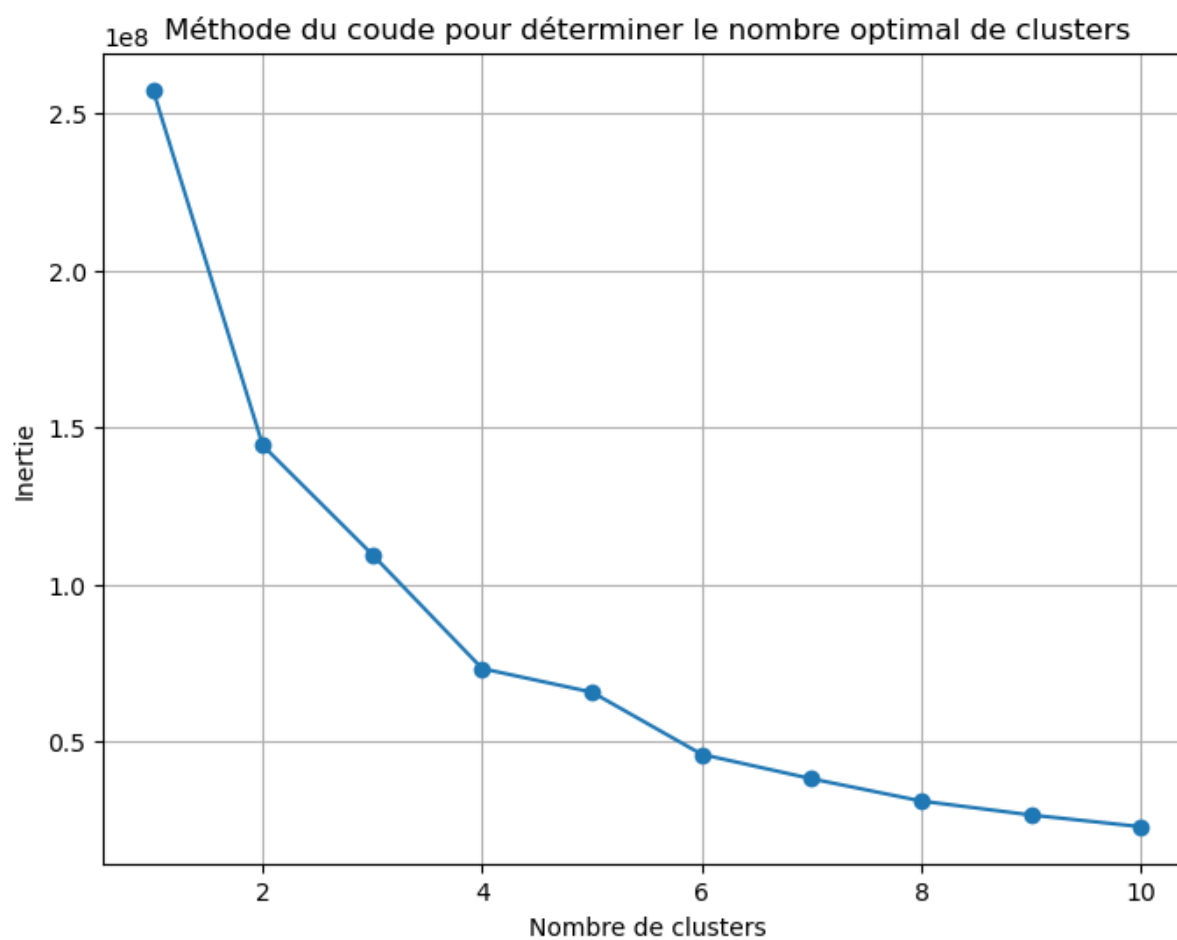
Graphique 16 : Matrice de corrélation entre les variables



La matrice nous permet de constater que la majorité des relations entre les variables est soit faible soit modérée sauf pour le bassin d'emploi et la densité (un peu plus modérée), nombre d'établissements par entreprise et la NAF 88 (un peu plus modérée), nombre d'établissements par entreprise et la DPAE (très faible) et nombre d'établissements par entreprise et offre (très faible). L'absence d'une relation forte nous permet alors de conserver les sept (07) variables dans notre modèle.

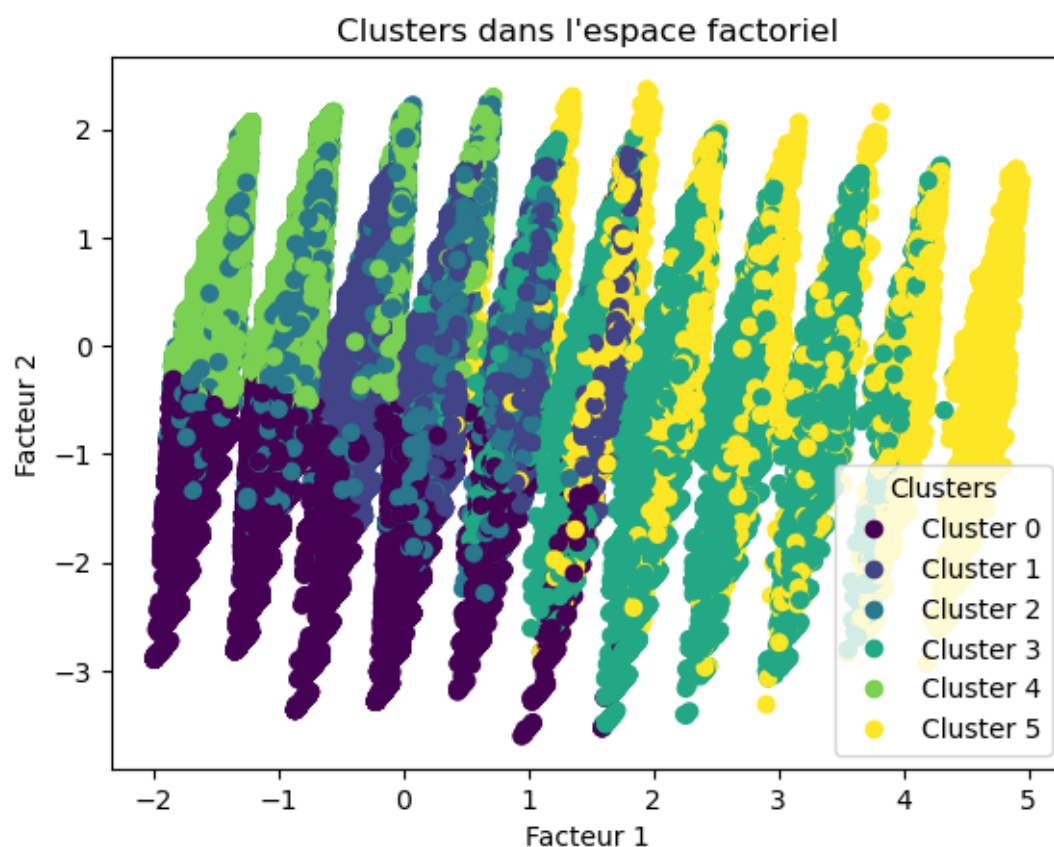
Après les tests de significativité et de corrélation, nous avons déterminé le nombre de facteur de notre modèle à travers la méthode de détermination du coude.

Graphique 17 : Méthode du coude pour déterminer le nombre optimal de cluster



Cette méthode nous a alors permis d'obtenir six (06) clusters qui sont représentés comme suit :

Graphique 18 : Distribution des établissements par cluster dans un espace factoriel

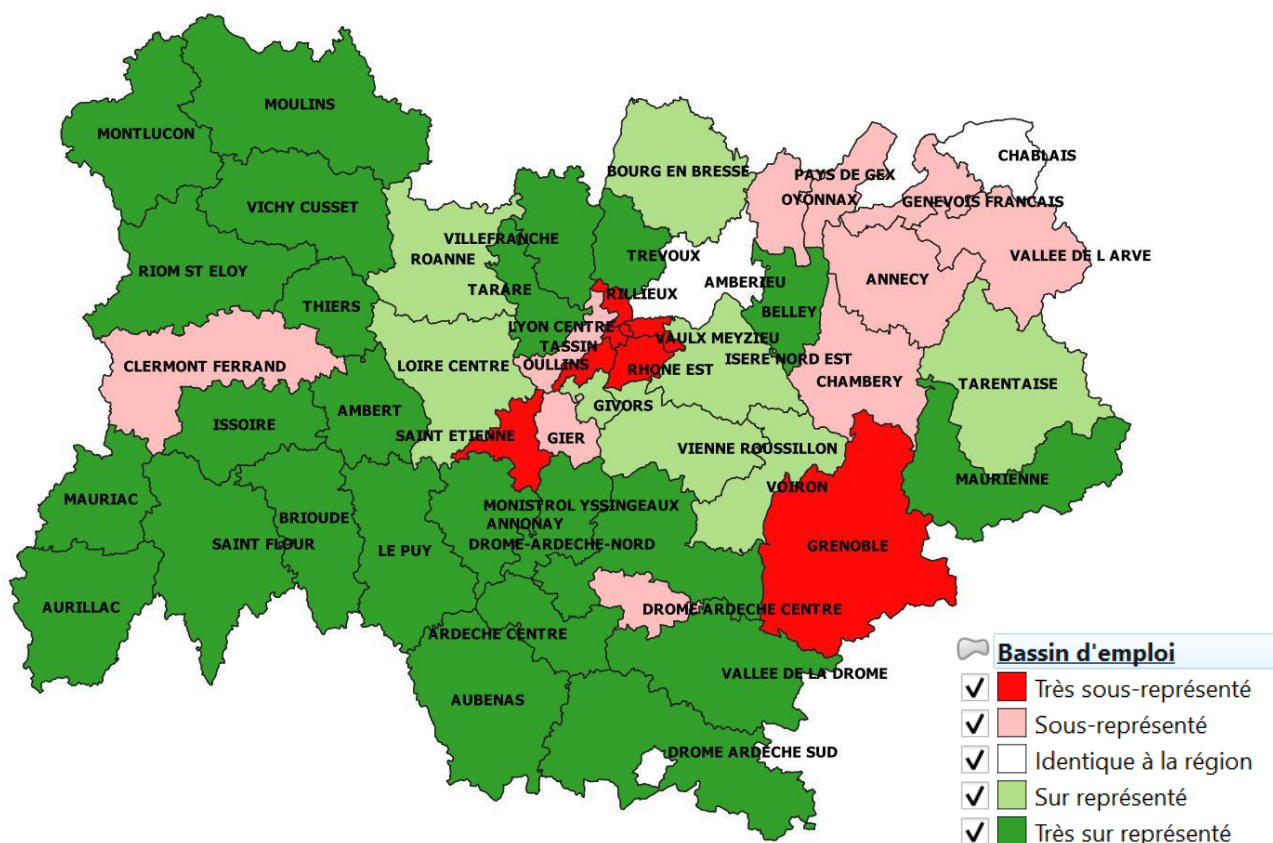


Ainsi, nous obtenons six (06) groupes distribués comme suit :

Clusters	Nombre d'établissements	Répartition des établissements par cluster
0	52605	18%
1	59444	20%
2	55936	19%
3	27348	9%
4	87202	29%
5	14840	5%

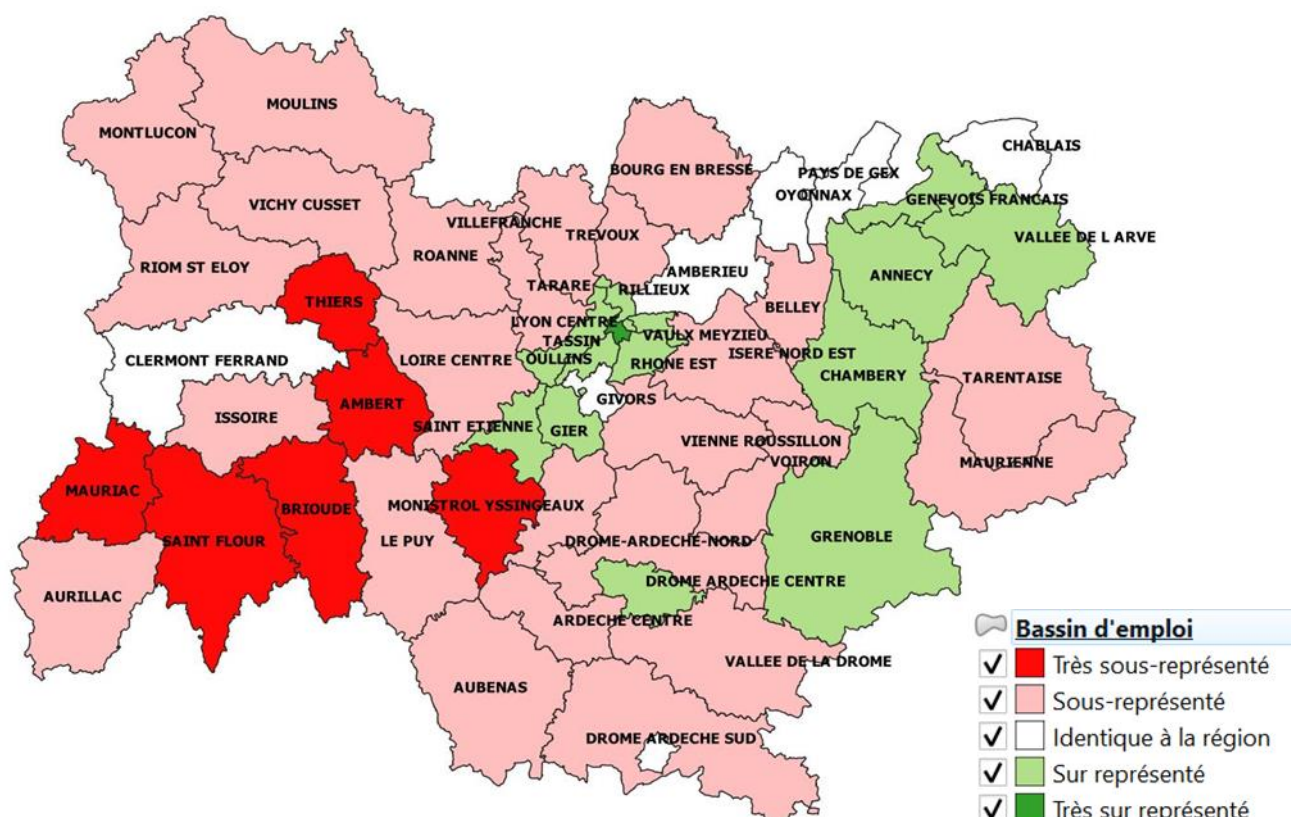
- **Cluster 0** : Ce cluster rassemble en majorité des établissements non-clients de France Travail, dont la moitié ne sont pas recruteurs, situés majoritairement dans des zones rurales (bourgs ruraux et rural dispersé). Ils sont issus d'entreprises mono-établissements, avec une forte surreprésentation des établissements de moins de 10 salariés et une dominance du secteur agricole.

Graphique 19 : Représentation des établissements par bassin d'emploi pour le cluster 0



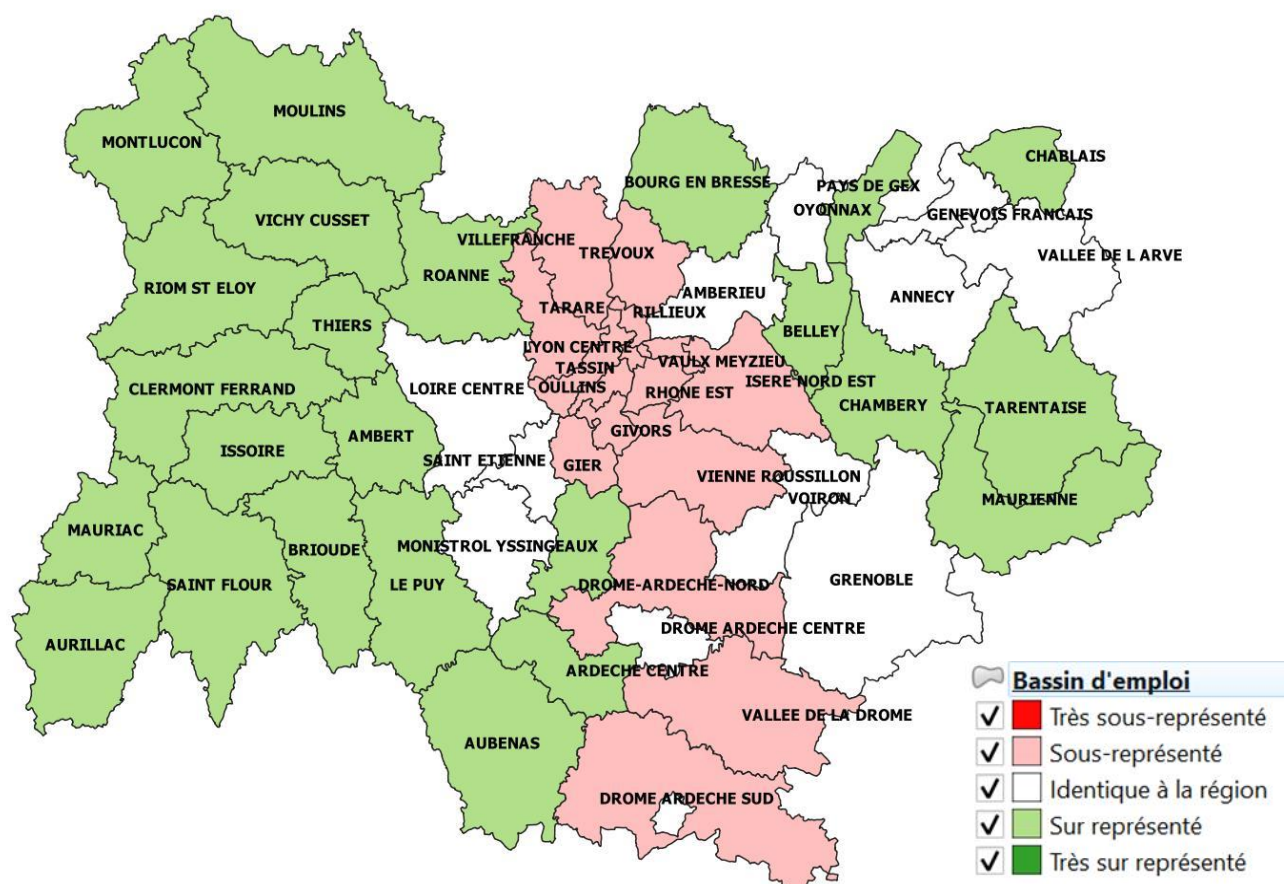
- **Cluster 1 :** Ce cluster est composé d'établissements recruteurs qui ne déposent pas d'offre à France Travail (0 % d'offres déposées, 100 % de DPAE), souvent situés dans des zones urbaines importantes (grands centres urbains et ceintures urbaines). Ils sont mono-établissements, majoritairement de taille intermédiaire (moins de 50 salariés), avec une forte présence des secteurs du commerce, des services et de la construction.

Graphique 20 : Représentation des établissements par bassin d'emploi pour le cluster 1



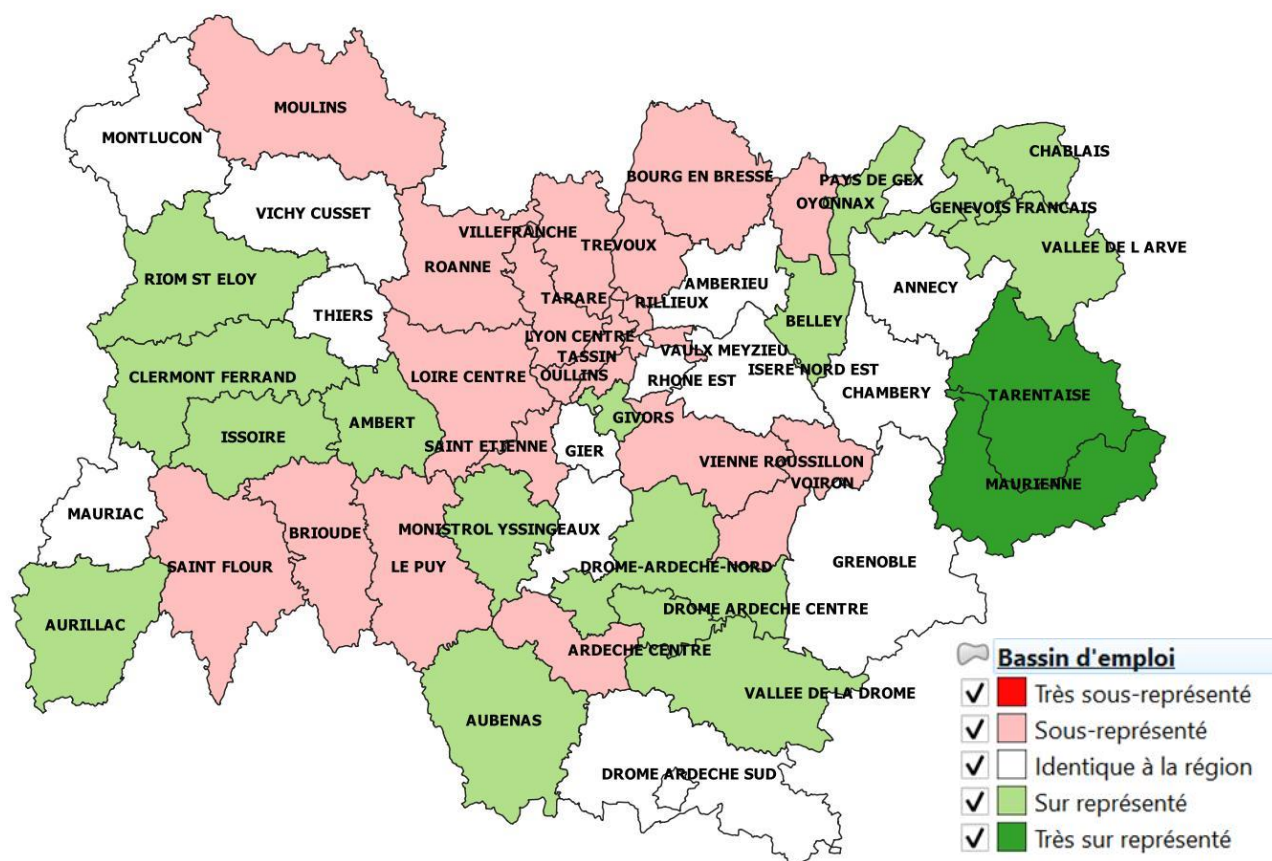
- **Cluster 2** : Les établissements de ce cluster sont non-clients de France Travail, mais un tiers (34%) sont recruteurs. On les retrouve dans des zones mixtes (urbaines intermédiaires et rurales), avec une répartition multi-établissement importante. Ils présentent une sur représentation des structures de moins de 20 salariés, dans les secteurs du commerce, de l'industrie et des services.

Graphique 21 : Représentation des établissements par bassin d'emploi pour le cluster 2



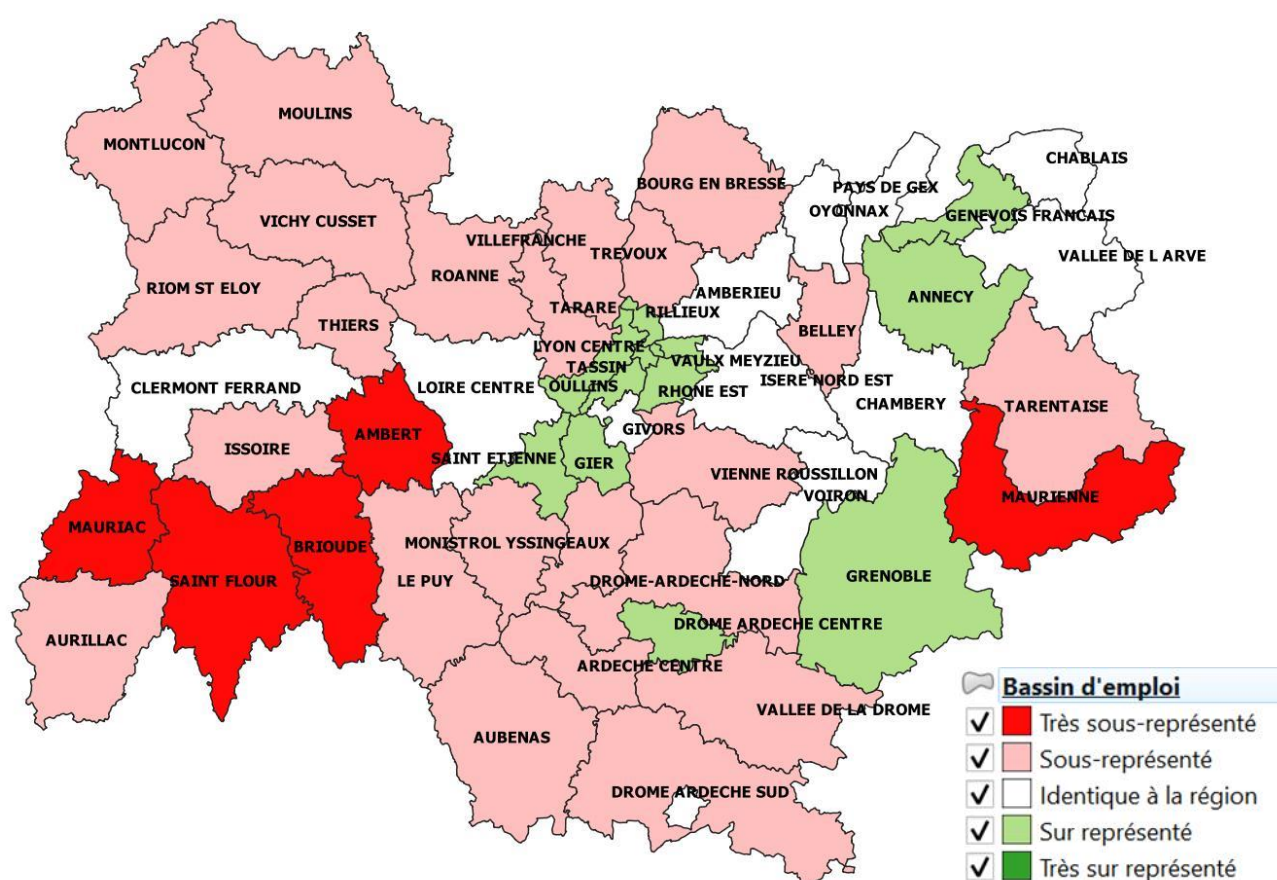
- **Cluster 3** : Ce cluster regroupe des établissements recruteurs et clients de France Travail (98 % d'offres, 91 % de DPAE), majoritairement mono-établissement et de taille intermédiaire (de 10 à 49 salariés). Ils sont localisés dans des zones urbaines et périurbaines. Les secteurs les plus représentés sont les services.

Graphique 22 : Représentation des établissements par bassin d'emploi pour le cluster 3



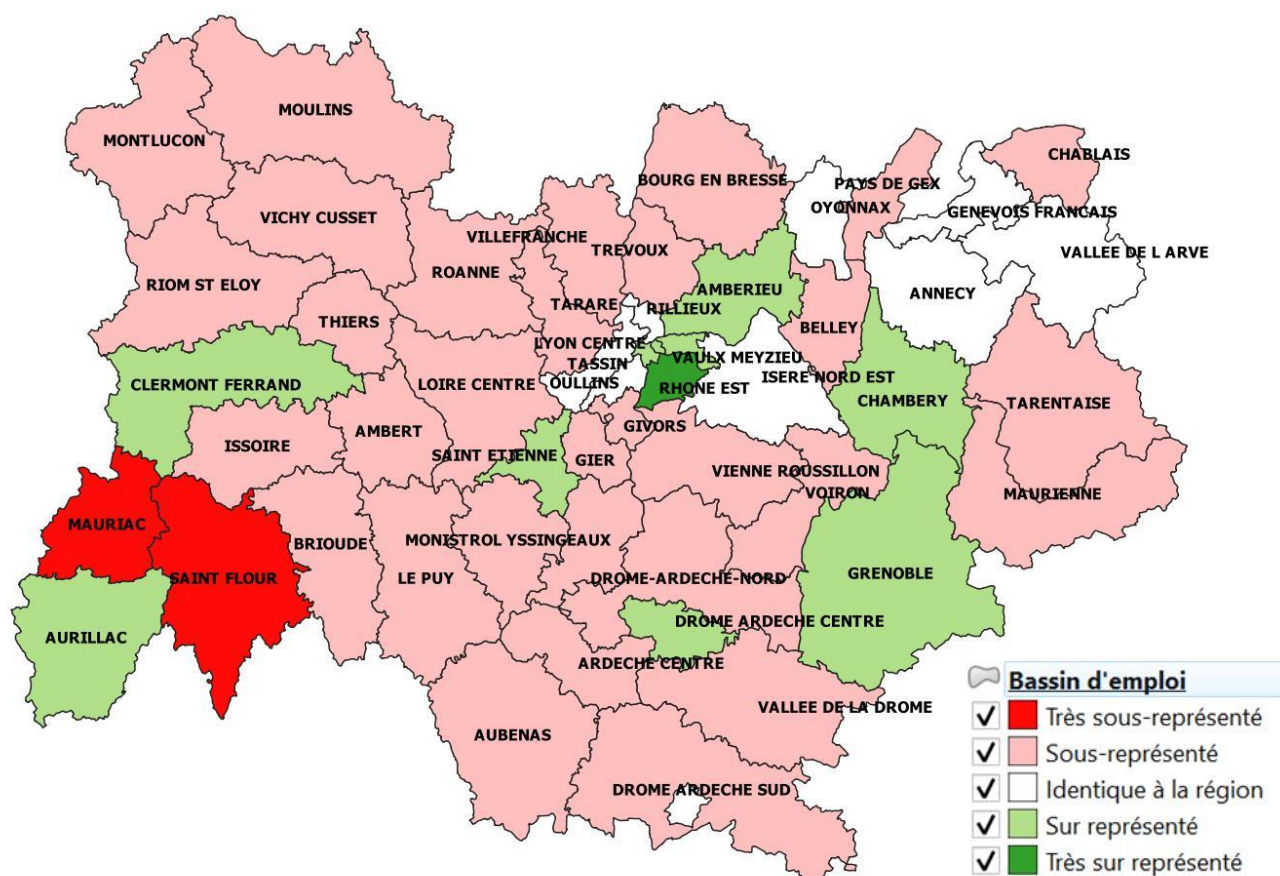
- **Cluster 4** : Les établissements de ce cluster sont majoritairement non-recruteurs et non-clients de France Travail (100 % non-offre, 100 % non-DPAE), mais localisés dans des zones urbaines ou périurbaines. Ce sont des mono-établissements, de très petite taille (moins de 10 salariés), avec une sur représentation notable des services notamment autres activités spécialisées, scientifiques et techniques, activités des sièges sociaux - conseil de gestion, etc.

Graphique 23 : Représentation des établissements par bassin d'emploi pour le cluster 4



- **Cluster 5** : Ce cluster se distingue par une forte proportion d'établissements multi-établissements (78 %) et de grande taille (33 % entre 20-49 salariés, 31 % entre 50-99, et 25 % avec plus de 100 salariés). Ces établissements sont recruteurs et la moitié d'entre eux sont clients de France Travail (49 % offres, 88 % DPAE) et localisés dans des zones urbaines denses. Les secteurs pré dominants incluent l'industrie (industrie pharmaceutique, cokéfaction et raffinage, etc.) et les services (hébergement médico-social et social).

Graphique 24 : Représentation des établissements par bassin d'emploi pour le cluster 5



En résumé nous avons six (06) groupes d'établissements qui peuvent être nommés comme suit :

- Cluster 0 : petits établissements ruraux dans le secteur agricole, non-recruteurs et non clients de France Travail ;
- Cluster 1 : établissements urbains dans les secteurs des services, commerce et construction, de taille intermédiaire, recruteurs et non clients de France Travail ;
- Cluster 2 : établissements multisites dans les secteurs des services, commerce et industrie, de petite taille non-clients de France Travail ;
- Cluster 3 : établissements de taille intermédiaire dans le secteur des services, recruteurs et clients de France Travail ;
- Cluster 4 : petits établissements urbains dans le secteur des services, non recruteurs et non clients de France Travail ;

- Cluster 5 : grandes entreprises multi-établissements dans les secteurs des services et de l'industrie, recruteurs avec une partie d'entre elles clientes à France Travail.

En outre, cette étude plus précisément le clustering, permettra de faire une segmentation fine pour un ciblage plus efficace des actions de prospection de l'entreprise. Cela permettra d'améliorer des indicateurs tels que l'ENT+ (Taux de recours à France Travail), par conséquent permettre à l'entreprise de mieux réaliser ses missions.

Culture d'entreprise et enseignements du stage

1. Culture d'entreprise

Au sein de la Direction régionale Auvergne–Rhône-Alpes de France Travail, la culture repose sur un équilibre entre performance et coopération. Chaque service, comme le Service Statistiques, Études et Évaluation (SEE) où j'ai évolué, développe une culture collaborative et orientée performance. La collaboration est favorisée par des échanges réguliers entre équipes, et par une hiérarchie qui reste accessible et à l'écoute. Cette organisation rend le travail fluide, même lorsqu'il faut traiter des demandes complexes ou urgentes. Le « Parlons Perf' » qui est une rencontre mensuelle est un exemple d'échange régulier entre équipe car c'est un échange entre les trois (03) équipes de la Direction de la Performance Opérationnelle (DPO) qui sont Statistiques Etudes et Evaluations (SEE), Qualité de Service et Pilotage Organisation Multicanal.

Quant à la prise de décision, elle suit une logique claire et structurée. Les décisions stratégiques sont discutées au niveau de la direction régionale et déclinées par les responsables de service. Pour les missions opérationnelles, une autonomie est accordée en permettant à chaque collaborateur de proposer des pistes ou solutions qui sont validées par le responsable. Pour le service Statistiques Etudes et Evaluations (SEE), une autonomie est accordée par le responsable qui, lors des points hebdomadaires, permet aux collaborateurs de choisir les demandes internes ou externes qu'ils souhaitent traiter. En ce qui me concerne, dès les premières semaines, une grande confiance m'a été accordée dans la réalisation de mes missions :

- Liberté dans l'organisation de mon temps, tant que les délais étaient respectés.
- Possibilité de proposer des méthodes d'analyse ou des visualisations adaptées aux besoins des interlocuteurs.
- Réactivité demandée pour s'adapter à des demandes urgentes de partenaires externes ou de la direction.

Cette autonomie m'a permis de développer des compétences de gestion des priorités et d'adaptation rapide, essentielles dans un environnement professionnel dynamique.

2. Encadrement et ambiance de travail

Durant le stage, les collaborateurs de France Travail, pas uniquement du service SEE, se sont montrés attentifs et réactifs. Des points réguliers ont permis de suivre l'avancement des projets et d'identifier les besoins en accompagnement. En effet, chaque semaine j'avais deux (02) points réguliers avec ma tutrice, un (01) point avec le responsable et un (01) point avec toute l'équipe du service SEE. Les retours ont été constructifs et orientés vers l'amélioration continue.

L'ambiance durant tout le stage a été caractérisée par :

- Une entraide constante pour résoudre des problématiques techniques ou organisationnelles.
- Des moments de convivialité, qui renforcent le sentiment d'appartenance à l'équipe.
- Un environnement de travail stimulant et bienveillant.
- Une ambiance professionnelle agréable, favorisant à la fois efficacité et épanouissement personnel.

Cette expérience m'a permis de :

- Développer des compétences techniques en analyse statistique et exploitation de données territoriales.
- Comprendre le fonctionnement d'une structure publique complexe et ses interactions avec des acteurs externes variés.
- Renforcer mes aptitudes en travail collaboratif.
- Gagner en autonomie et en capacité à gérer des projets dans un environnement exigeant mais formateur.

Conclusion

Ce stage de six mois au sein du service Statistiques, Études et Évaluation (SEE) de la Direction Régionale Auvergne–Rhône-Alpes de France Travail a été une expérience extrêmement formatrice, tant sur le plan professionnel que personnel. Il m’a permis de découvrir le fonctionnement d’une institution publique stratégique, au cœur des dynamiques du marché du travail, et de comprendre les enjeux liés à la production, l’analyse et la diffusion de données fiables pour éclairer les décisions des différents acteurs.

Les missions qui m’ont été confiées m’ont permis de développer des compétences techniques solides, notamment en traitement et croisement de bases de données volumineuses, en analyse statistique, en visualisation et cartographie de données, mais aussi en utilisation d’outils spécifiques comme Power BI, QGIS ou encore l’API Sirene. Par ailleurs, la dimension collaborative du service m’a appris à travailler dans un cadre structuré où la rigueur, l’autonomie et la réactivité sont essentielles, tout en favorisant le partage de connaissances et l’entraide.

J’ai également eu l’opportunité de m’immerger dans les réalités opérationnelles des agences et de dialoguer avec les équipes chargées des relations entreprises, ce qui m’a permis de donner plus de sens aux analyses réalisées et de comprendre l’impact concret des travaux du SEE sur le terrain. Cette proximité avec les utilisateurs finaux des données m’a permis de mieux cerner les besoins opérationnels et d’adapter mes productions pour qu’elles soient directement exploitables.

Au-delà des compétences techniques, cette expérience m’a appris l’importance de la flexibilité, de la gestion des priorités et de la communication professionnelle dans un environnement exigeant mais stimulant. Elle a également renforcé mon intérêt pour les métiers liés à l’analyse de données, à l’évaluation des politiques publiques et au pilotage stratégique, des domaines dans lesquels je souhaite continuer à évoluer.

Je garde de ce stage le souvenir d’une ambiance de travail bienveillante et motivante, où la hiérarchie, tout en étant attentive et disponible, encourage l’autonomie et la prise d’initiative.

Ces conditions favorables m'ont permis de progresser rapidement, de prendre confiance en mes capacités et de produire des livrables utiles pour le service.

En somme, ce stage a constitué une véritable étape dans mon parcours professionnel, en consolidant mes compétences analytiques, techniques et relationnelles, tout en m'offrant une vision concrète des enjeux liés à la gestion des données et au soutien à la décision dans une grande organisation publique.

ANNEXES

Annexe 1 : Code python pour charger la base SIRENE par API

```
import json

import pandas as pd

with open ("C:/Users/IWTI4850/OneDrive - France
Travail/Downloads/response_1745320950869.json", encoding='utf-8') as f:

    data = json.load(f)

# === Extraire les établissements ===

etablisements = data.get("etablisements", [])

# === Préparer les données pour le DataFrame ===

resultats = []

for e in etablisements:

    adresse = e.get("adresseEtablissement", {})

    unite_legale = e.get("uniteLegale", {})

    resultats.append({

        "SIREN": e.get("siren"),

        "NIC": e.get("nic"),

        "SIRET": e.get("siret"),

        "Dénomination": unite_legale.get("denominationUniteLegale"),

        "Commune": adresse.get("libelleCommuneEtablissement"),

        "Code Postal": adresse.get("codePostalEtablissement"),

        "Type Voie": adresse.get("typeVoieEtablissement"),

        "Voie": adresse.get("libelleVoieEtablissement"),

        "Numéro": adresse.get("numeroVoieEtablissement"),

        "diffusion": unite_legale.get('statutDiffusionUniteLegale'),

        "date de création": unite_legale.get('dateCreationUniteLegale'),

        "Effectif": unite_legale.get('trancheEffectifsUniteLegale'),

        "Année effectif": unite_legale.get('anneeEffectifsUniteLegale'),

        "date dernier traitement": unite_legale.get('dateDernierTraitementUniteLegale'),

        "nombre période": unite_legale.get('nombrePeriodesUniteLegale'),

        "catégorie de l'entreprise": unite_legale.get('categorieEntreprise'),
```

```

"année catégorie entreprise": unite_legale.get('anneeCategorieEntreprise'),
"périodes": unite_legale.get('periodesUniteLegale'),
"date de Debut": unite_legale.get('dateDebut'),
"etat administratif": unite_legale.get('etatAdministratifUniteLegale'),
"denomination usuelle": unite_legale.get('denominationUsuelle1UniteLegale'),
"catégorie juridique": unite_legale.get('categorieJuridiqueUniteLegale'),
"activité principale": unite_legale.get('activitePrincipaleUniteLegale'),
"NAP/NAF": unite_legale.get('nomenclatureActivitePrincipaleUniteLegale'),
"Economie Sociale Solidaire": unite_legale.get('economieSocialeSolidaireUniteLegale'),
"Coordonnée X (Lambert)": adresse.get("coordonneeLambertAbscisseEtablissement"),
"Coordonnée Y (Lambert)": adresse.get("coordonneeLambertOrdonneeEtablissement")
})

# === Créer un DataFrame et sauvegarder en Excel ===

df = pd.DataFrame(resultats)

df.to_excel("etablisements_coordonnees5.xlsx", index=False)

print("Fichier généré : etablisements_coordonnees5.xlsx")

```

Annexe 2 : Liste des variables de la base de données

Variables	Description	Modalités
SIREN	Numéro de l'entreprise	Numéro unique à 9 chiffres attribué par l'INSEE
SIRET	Numéro de l'établissement	Numéro unique à 14 chiffres (SIREN + NIC) attribué par l'INSEE
Raison sociale	Dénomination de l'établissement	Nom officiel de l'entreprise
Enseigne	Dénomination de l'enseigne	Nom commercial de l'entreprise
Adresse	Adresse de l'établissement	Adresse postale complète
Code Postal et commune	Code postal et commune de l'établissement	Code postal à 5 chiffres et nom de la commune
Code INSEE	Identifiant unique attribué à chaque commune	Code à 5 chiffres attribué par l'INSEE
Libellé Commune	Dénomination de la commune	Nom officiel de la commune
Dep	Numéro du département	Numéro à 2 chiffres des 12 départements de la région
Département	Dénomination du département	Nom officiel de chacun des 12 départements
Bassin d'emploi	Dénomination du bassin d'emploi	Nom du bassin d'emploi
Arrondissement	Dénomination de l'arrondissement	Nom de l'arrondissement
EPCI	Établissement Public de Coopération Intercommunale	Nom de l'EPCI

Variables	Description	Modalités
CLPE	Code Local du Pôle Emploi	Code attribué par Pôle Emploi
Code NAF	Code de la Nomenclature d'Activités Française	Code à 4 chiffres et 1 lettre attribuée par l'INSEE
Libellé NAF	Libellé de la Nomenclature d'Activités Française	Description de l'activité
Code NAF A129	Code détaillé de l'activité	Code à 2 chiffres et 2 lettres
Libellé NAF A129	Description détaillée de l'activité	Description détaillée de l'activité
NAF A88	Description détaillée de l'activité	Description détaillée de l'activité
Code NAF A38	Code de l'activité	Code à 2 lettres
Code NAF A5	Code de l'activité agrégée	Code 1 chiffre
NAF A5	Libellé NAF A5	Description de l'activité agrégée
Effectifs salariés	Nombre de salariés	Nombre entier
Tranche effectif créée	Effectif salarié classé en tranche	Catégorie de taille d'effectifs regroupée en 6 classes
Classe taille effectifs	Classe de taille d'effectifs	
OEE en 2024	Nombre d'offre d'emploi en 2024	Entier
DPAE en 2024	Nombre de Déclaration Préalable à l'Embauche en 2024	Entier

Variables	Description	Modalités
DC_CATEGORIESJURIDIQUES_ID_DPAE	Catégorie juridique de l'entreprise	Code de 4 chiffres
DC_NAF_ID_DPAE	Code NAF de l'entreprise	Code NAF de l'entreprise à 4 chiffres et 1 lettre
DC_CODENAF05_DPAE	Code NAF A05 positions de l'entreprise	Code NAF à 5 positions
DC_CODENAF21_DPAE	Code NAF A21 positions de l'entreprise	Code NAF à 21 positions
DC_NOMRAISONSOCIALE_DPAE	Raison sociale de l'entreprise	Nom officiel de l'entreprise
DC_ADRESSE_DPAE	Adresse de l'entreprise	Adresse postale complète
DC_LBLCOMMUNE_DPAE	Commune de l'entreprise	Nom de la commune
DC_COMMUNE_ID_DPAE	Code INSEE de la commune de l'entreprise	Code à 5 ou 6 chiffres
Code_dept_9_DPAE	Code du département de l'entreprise	Numéro à 2 chiffres ou 1 chiffre et 1 lettre
code_BE_2009_4_DPAE	Code du bassin d'emploi de l'entreprise	Code à 4 chiffres
TAILLE_DPAE	Taille de l'entreprise	Catégorie de taille d'entreprise : 8 classes
dpae_DPAE	Nombre de DPAE de l'entreprise	Nombre entier
2_CDI_DPAE	Nombre de DPAE de contrat CDI déclarés	Nombre entier
1_CDD_DPAE	Nombre de DPAE de contrat CDD déclarés	Nombre entier
3_CTT_DPAE	Nombre de DPAE de contrat CTT déclarés	Nombre entier

Variables	Description	Modalités
Latitude_SIRENE	Latitude de l'établissement	Coordonnée géographique
Longitude_SIRENE	Longitude de l'établissement	Coordonnée géographique
Nombre d'établissement par entreprise	Nombre d'établissements de l'entreprise	Nombre entier
catégorie d'établissement	Catégorie de l'établissement	Type d'établissement (ex : PME, GE, ETI)
Tranche entreprise_établissements	Tranche d'établissements de l'entreprise	Catégorie de nombre d'établissements (ex: 1, 2-5, 6-10, etc.)
Tranche effectif recodé	Tranche d'effectifs recodée	Catégorie de taille d'effectifs (ex: 1-9, 10-19, etc.)
Offre total	Nombre total d'offres d'emploi	Nombre entier
offre_CDI	Nombre d'offres de type CDI	Nombre entier
offre_CDD+12	Nombre d'offres de type CDD de plus de 12 mois	Nombre entier
offre_CDD7-12	Nombre d'offres de type CDD de 7 à 12 mois	Nombre entier
offre_CDD4-6	Nombre d'offres de type CDD de 4 à 6 mois	Nombre entier
offre_CDD1-3	Nombre d'offres de type CDD de 1 à 3 mois	Nombre entier

Variables	Description	Modalités
offre_Mission+1	Nombre d'offres de type mission de plus de 1 mois	Nombre entier
offre_CDD<1	Nombre d'offres de type CDD de moins de 1 mois	Nombre entier
offre_Mission<1	Nombre d'offres de type mission de moins de 1 mois	Nombre entier
Offre_durable	Nombre d'offres durables	Nombre entier
Offre_temporaire	Nombre d'offres temporaires	Nombre entier
Offre_occasionnelle	Nombre d'offres occasionnelles	Nombre entier
catégorie_ets	Catégorie de l'établissement	Type d'établissement (ex : PME, GE, ETI) à partir de la classification de l'INSEE
DPAE_ets	L'établissement a-t-il fait une DPAE ou pas	Variable catégorielle (OUI ou NON)
Offre_ets	L'établissement a-t-il déposé une offre ou pas	Variable catégorielle (OUI ou NON)
ent_ets	Catégorie d'établissements de l'entreprise	Variable catégorielle (mono-établissement, multi-établissement)
Densité	Regroupement par zone de densité	Variable catégorielle (Rural, urbain, etc.)

Annexe 3 : Code python pour la réalisation de l'ACP et du K-means

```
##### Importation des packages pour l'analyse

import pandas as pd

from sklearn.cluster import KMeans

from sklearn.preprocessing import StandardScaler

from sklearn.decomposition import PCA

import matplotlib.pyplot as plt

import numpy as np

from scipy.stats import chi2_contingency, fisher_exact

import seaborn as sns

##### Pour permettre d'afficher complètement les résultats dans la console

pd.set_option('display.max_columns', None) # Affiche toutes les colonnes

pd.set_option('display.max_rows', None) # Affiche toutes les lignes

##### Chargement des données

df = pd.read_excel("C:/Users/IWTI4850/France Travail/ARA - DR - Statistiques, études et
évaluations - General/STAGE 2025/TRAVAIL WENDYAM/Analyse ETS/base_py.xlsx")

##### test statistiques

def test_nominal_association(var1, var2):

    # Création du tableau de contingence

    contingency_table = pd.crosstab(var1, var2)

    # Test du Chi-carré

    chi2, p, dof, expected = chi2_contingency(contingency_table)

    # V de Cramer

    n = contingency_table.sum().sum()

    phi = np.sqrt(chi2 / n)

    r, k = contingency_table.shape

    v = np.sqrt(phi**2 / min((k-1), (r-1)))

    # Test exact de Fisher (si tableau 2x2)
```

```

fisher_p = np.nan

if contingency_table.shape == (2, 2):
    _, fisher_p = fisher_exact(contingency_table)

return {
    'Chi2': chi2,
    'p-value': p,
    'V de Cramer': v,
    'Degrés de liberté': dof
}

### Recodage des variables en variable numérique
# Suppression des doublons dans le recodage
df['dpae_recodé'] = df['DPAE_ets'].apply(lambda x: 1 if x == 'OUI' else 0)
df['offre_recodé'] = df['Offre_ets'].apply(lambda x: 1 if x == 'OUI' else 0)
df['ent_recodé'] = df['ent_ets'].apply(lambda x: 1 if x == 'mono établissement' else 2)

# Mise à jour du mapping des tranches d'effectif
tranches_effectif_mapping = {
    '0 salarié': 0,
    '1 à 9 salariés': 1,
    '10 à 19 salariés': 2,
    '20 à 49 salariés': 3,
    '50 à 99 salariés': 4,
    '100 salariés et plus': 5
}

df['tranche_recodé'] = df['Tranche effectif créée'].map(tranches_effectif_mapping)

# Conversion de la densité en numérique (en supposant que c'est une chaîne avec des chiffres)

```

```

df['Densité'] = pd.to_numeric(df['Densité'], errors='coerce')

### Teste les liaisons entre toutes les paires de variables qualitatives nominales

qual_vars = ['Densité', 'tranche_recodé', 'dpae_recodé', 'offre_recodé', 'ent_recodé', 'BE',
'NAF_A88']

results = {}

for i in range(len(qual_vars)):
    for j in range(i + 1, len(qual_vars)):
        var1 = qual_vars[i]
        var2 = qual_vars[j]
        results[f'{var1} vs {var2}'] = test_nominal_association(df[var1], df[var2])

# Affichage des résultats

print("Résultats des tests de liaison entre variables qualitatives nominales:")
results_df = pd.DataFrame(results).T
print(results_df)

### Fonction pour calculer le V de Cramer entre deux variables qualitatives

def cramers_v(var1, var2):
    contingency_table = pd.crosstab(var1, var2)
    n = contingency_table.sum().sum()
    chi2, _, _, _ = chi2_contingency(contingency_table)
    phi = np.sqrt(chi2 / n)
    r, k = contingency_table.shape
    return np.sqrt(phi**2 / min((k-1), (r-1)))

### Création de la matrice de corrélation

correlation_matrix = pd.DataFrame(index=qual_vars, columns=qual_vars)

```

```

for i in range(len(qual_vars)):
    for j in range(len(qual_vars)):
        if i == j:
            correlation_matrix.loc[qual_vars[i], qual_vars[j]] = 1.0 # Diagonale = 1
        elif j > i:
            # Utiliser le résultat déjà calculé pour éviter les doublons
            correlation_matrix.loc[qual_vars[i], qual_vars[j]] = results_df.loc[f"{qual_vars[i]} vs {qual_vars[j]}", 'V de Cramer']
        else:
            # Symétrie de la matrice
            correlation_matrix.loc[qual_vars[i], qual_vars[j]] =
correlation_matrix.loc[qual_vars[j], qual_vars[i]]

# Affichage de la matrice de corrélation
print("\nMatrice de corrélation (V de Cramer) entre variables qualitatives nominales:")
print(correlation_matrix)

# Visualisation de la matrice de corrélation
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix.astype(float), annot=True, cmap='coolwarm', vmin=0,
vmax=1)
plt.title("Matrice de corrélation (V de Cramer) entre variables qualitatives nominales")
plt.show()

### Détermination du nombre de facteurs

# Sélection des variables quantitatives et encodage des variables catégorielles
quant_vars = ['Densité', 'tranche_recodé', 'dpae_recodé', 'offre_recodé', 'ent_recodé', 'BE',
'NAF_A88']

```

```

df_quant = df[quant_vars]

# Suppression des lignes avec des valeurs manquantes
df_quant = df_quant.dropna()

inerties = []

for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=0)
    kmeans.fit(df_quant)
    inerties.append(kmeans.inertia_)

# Tracer le graphique de l'inertie
plt.figure(figsize=(8, 6))
plt.plot(range(1, 11), inerties, marker='o')
plt.xlabel('Nombre de clusters')
plt.ylabel('Inertie')
plt.title('Méthode du coude pour déterminer le nombre optimal de clusters')
plt.grid(True)
plt.show()

### Analyse factorielle (ACP)

# Standardisation
scaler = StandardScaler()
df_std = scaler.fit_transform(df_quant)

### ACP avec 6 composantes principales
pca = PCA(n_components=6)
pca.fit(df_std)

```

```

#Scores des composantes principales
factor_scores = pca.transform(df_std)
df.loc[df_quant.index, 'FAC1_1'] = factor_scores[:,0]
df.loc[df_quant.index, 'FAC2_1'] = factor_scores[:,1]
df.loc[df_quant.index, 'FAC3_1'] = factor_scores[:,2]
df.loc[df_quant.index, 'FAC4_1'] = factor_scores[:,3]
df.loc[df_quant.index, 'FAC5_1'] = factor_scores[:,4]
df.loc[df_quant.index, 'FAC6_1'] = factor_scores[:,5]

# Affichage des résultats
print("Variances expliquées:")
print(pca.explained_variance_ratio_)

#%% correlation entre les composantes principales et les variables standardisées
correlations = pd.DataFrame(df_std,
columns=df_quant.columns).corrwith(pd.DataFrame(factor_scores, columns=['FAC1_1',
'FAC2_1', 'FAC3_1', 'FAC4_1', 'FAC5_1', 'FAC6_1']))

# Création d'une matrice de corrélation pour une meilleure visualisation
correlation_matrix = pd.DataFrame(index=df_quant.columns, columns=['FAC1_1', 'FAC2_1',
'FAC3_1', 'FAC4_1', 'FAC5_1', 'FAC6_1'])

for i, component in enumerate(['FAC1_1', 'FAC2_1', 'FAC3_1', 'FAC4_1', 'FAC5_1',
'FAC6_1']):
    for var in df_quant.columns:
        correlation_matrix.loc[var, component] = np.corrcoef(df_std[:, i], factor_scores[:, i])[0,
1]

# Visualisation de la matrice de corrélation
plt.figure(figsize=(10, 8))

```

```
sns.heatmap(correlation_matrix.astype(float), annot=True, cmap='coolwarm', vmin=-1,
vmax=1)
```

```
plt.title("Matrice de corrélation entre les composantes principales et les variables
quantitatives")
```

```
plt.show()
```

```
### Clustering K-means sur les scores factoriels
```

```
kmeans = KMeans(n_clusters=6, max_iter=100, tol=1e-5, random_state=42)
```

```
df.loc[df_quant.index, 'CLUSTER_ID'] = kmeans.fit_predict(factor_scores)
```

```
### Tableaux de profils
```

```
print("\nCroisement Cluster x offre_recodé:")
```

```
print(pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'], df.loc[df_quant.index,
'offre_recodé'],
```

```
          margins=True, normalize='index'))
```

```
print("\nCroisement Cluster x dpae_recodé:")
```

```
print(pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'], df.loc[df_quant.index,
'dpae_recodé'],
```

```
          margins=True, normalize='index'))
```

```
### Statistiques descriptives par cluster
```

```
print("\nMoyennes par cluster:")
```

```
print(df.loc[df_quant.index].groupby('CLUSTER_ID')['Densité', 'tranche_recodé',
'dpae_recodé', 'offre_recodé', 'ent_recodé', 'BE', 'NAF_A88'])
```

```
### 6. Profilage des clusters
```

```
print("\nCroisement Cluster x Baasin d'emploi:")
```

```
print(pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'], df.loc[df_quant.index, 'BE'],
normalize='index'))
```

```
print("\nCroisement Cluster x Code NAF A88:")
```



```
print(pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'], df.loc[df_quant.index, 'NAF_A88'],
normalize='index'))
```

```
print("\nCroisement Cluster x densité:")
```

```
print(pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'], df.loc[df_quant.index, 'Densité'],
normalize='index'))
```

```
print("\nCroisement Cluster x tranche effectif salarié:")
```

```
print(pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'], df.loc[df_quant.index,
'tranche_recodé'], normalize='index'))
```

```
print("\nCroisement Cluster x nombre d'établissement par entreprise:')
```

```
print(pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'], df.loc[df_quant.index,
'ent_recodé'], normalize='index'))
```

```
##### Visualisation des établissements
```

```
scatter = plt.scatter(df.loc[df_quant.index, 'FAC1_1'], df.loc[df_quant.index, 'FAC2_1'],
c=df.loc[df_quant.index, 'CLUSTER_ID'], cmap='viridis')
```

```
plt.xlabel('Facteur 1')
```

```
plt.ylabel('Facteur 2')
```

```
plt.title('Clusters dans l'espace factoriel')
```

```
# Ajout de la légende
```

```
legend = scatter.legend_elements()
```

```
plt.legend(legend[0], ['Cluster 0', 'Cluster 1', 'Cluster 2', 'Cluster 3', 'Cluster 4', 'Cluster 5'],
title="Clusters")
```

```
plt.show()
```

```
##### test de Chi2
```

```
from scipy.stats import chi2_contingency
```

```

contingence = pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'],
                          df.loc[df_quant.index, 'Densité'])
chi2, p, dof, expected = chi2_contingency(contingence)
print("Chi2 =", chi2, "p-value =", p)
#%%%

contingence = pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'],
                          df.loc[df_quant.index, 'tranche_recodé'])
chi2, p, dof, expected = chi2_contingency(contingence)
print("Chi2 =", chi2, "p-value =", p)
#%%%

contingence = pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'],
                          df.loc[df_quant.index, 'offre_recodé'])
chi2, p, dof, expected = chi2_contingency(contingence)
print("Chi2 =", chi2, "p-value =", p)
#%%%

contingence = pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'],
                          df.loc[df_quant.index, 'dpae_recodé'])
chi2, p, dof, expected = chi2_contingency(contingence)
print("Chi2 =", chi2, "p-value =", p)
#%%%

contingence = pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'],
                          df.loc[df_quant.index, 'ent_recodé'])
chi2, p, dof, expected = chi2_contingency(contingence)
print("Chi2 =", chi2, "p-value =", p)
#%%%

contingence = pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'],
                          df.loc[df_quant.index, 'BE'])
chi2, p, dof, expected = chi2_contingency(contingence)
print("Chi2 =", chi2, "p-value =", p)

```

```

#%%%

contingence = pd.crosstab(df.loc[df_quant.index, 'CLUSTER_ID'],
                          df.loc[df_quant.index, 'NAF_A88'])

chi2, p, dof, expected = chi2_contingency(contingence)

print("Chi2 =", chi2, "p-value =", p)


#%%% Exportation de la base avec les clusters

df.to_csv('C:/Users/IWTI4850/Desktop/votre_fichier_traitel.csv', index=False)

df.to_excel('C:/Users/IWTI4850/Desktop/votre_fichier_traitel.xlsx', index=False)

```

Bibliographie

- [1] Statistiques - Offre de service, DR ARA, SEE
- [2] Services DR - missions Auvergne-Rhône-Alpes, DR ARA
- [3] Services DR - Trombinoscope Auvergne-Rhône-Alpes, DR ARA

Sites internet

- [4] <https://www.observatoire-emploi-ara.fr/>
- [5] <https://www.sirene.fr/sirene/public/recherche>