

12 JUILLET 2025

RAPPORT DE STAGE

STAGE DATA ANALYST

MAZARS LORIS
UNIVERSITE AIX-MARSEILLE/ENDRIX
21 rue du Chanoine Neyrand 84 700 Sorgues



Résumé

Ce rapport présente la mission réalisée au sein du cabinet Endrix, dans le cadre de mon stage de fin de Licence Professionnelle Métiers du Décisionnel et de la Statistique. La problématique posée consistait à identifier, à partir de données internes, les entreprises potentiellement éligibles à certaines aides de l'État, afin d'accompagner les experts-comptables et les conseillers dans leurs recommandations.

Pour cela, une méthode rigoureuse a été mise en place. Après avoir identifié les cinq aides les plus pertinentes en collaboration avec le service comptable et le responsable de mission, j'ai conçu un plan structuré reposant sur l'analyse de critères d'éligibilité précis. Le jeu de données a été traité dans un notebook Python : nettoyage, enrichissement avec de nouvelles variables (notamment des indicateurs d'éligibilité), et analyses statistiques ont été réalisés.

Les résultats obtenus ont ensuite été intégrés dans un tableau de bord interactif sous Power BI, pensé pour être à la fois accessible, clair et décisionnel. Ce rapport détaille chaque étape de la démarche : de la définition de la problématique à la conception du livrable final, en passant par les choix techniques, les limites rencontrées et les enseignements tirés de cette expérience professionnelle enrichissante.

Par souci de confidentialité, les données, le notebook Python et le tableau de bord présentés dans ce rapport sont entièrement fictifs. Ils ont été recréés à partir d'un jeu de données simulé, dans le respect des contraintes professionnelles et pour illustrer fidèlement la nature du travail réalisé en entreprise.

Remerciements

Je tiens à remercier chaleureusement l'ensemble de l'équipe d'Endrix pour leur accueil, leur disponibilité et leur accompagnement tout au long de mon stage.

Je remercie particulièrement Monsieur Garcia, mon tuteur de stage, pour sa confiance, ses conseils précieux et le cadre de travail stimulant qu'il m'a offert. Je remercie également le service comptable pour m'avoir fourni les données nécessaires et avoir pris le temps de m'expliquer les subtilités des dispositifs d'aide étudiés.

Je remercie enfin mes enseignants de la Licence Professionnelle MDS pour la qualité de leur enseignement, qui m'a permis d'aborder ce stage avec les compétences requises en analyse de données, visualisation et communication.

Sommaire

<i>Sommaire</i>	2
<i>Table des illustrations</i>	2
1. INTRODUCTION	3
1.1 PRESENTATION DE L'ENTREPRISE	3
1.2 ORGANISATION DE L'ENTREPRISE.....	3
3.1 CLIENTELE	4
2.1 CONTEXTE DES TRAVAUX	5
2 OBJECTIF DE LA MISSION	5
2. ANALYSE DE L'EXISTANT	6
3. PROPOSITIONS METHODOLOGIQUES	6
3.1 PREMIERE PROPOSITION	6
3.2 DEUXIEME PROPOSITION	7
4. APPLICATION	8
4.1 DEVELOPPEMENT DES CONNAISSANCES EN COMPTABILITE	8
4.2 CREATION DU NOTEBOOK	10
4.3 CREATION DU TABLEAU DE BORD POWER BI	13
4.4 RESULTATS ET INTERPRETATION	16
5. CONCLUSION	17
6. BILAN	18
7. REFERENCES BIBLIOGRAPHIQUES	19
8. ANNEXE	0
8.1 CARTE MENTALE	0
8.2 NOTEBOOK.....	0
8.3 RAPPORT POWER BI.....	0

Table des illustrations

FIGURE 1 : DIAGRAMME CIRCULAIRE DES CLIENTS	4
FIGURE 2 : CODE PYTHON POUR GENERE LE JEU DE DONNEES FICTIF.....	11
FIGURE 3 : CODE PYTHON POUR CREER LES ERREURS DU JEU DE DONNEES	12
FIGURE 4 : APERÇU DU JEU DE DONNEES FICTIF	12
FIGURE 5 : APERÇU DES DONNEES EXPORTEES DANS POWER BI	13
FIGURE 6 : PREMIERE VISUALISATION	14
FIGURE 7: EXEMPLE DE MISE EN FORME CONDITIONNELLE	14
FIGURE 8: EXEMPLE DE KPI.....	15
FIGURE 9 : CARTE MENTAL DU PROJET	0
FIGURE 10 : PAGE "ACCUEIL".....	0
FIGURE 11 : PAGE "GLOBAL"	0
FIGURE 12 : PAGE "ACRE"	1
FIGURE 13 : PAGE "CIR".....	1
FIGURE 14 : PAGE "AIDE ENERGIE"	2
FIGURE 15 : PAGE "AIDE REGIONALE"	2
FIGURE 16 : PAGE "URSSAF"	3

1. Introduction

1.1 Présentation de l'entreprise

Endrix est un cabinet d'expertise comptable, d'audit et de conseil implanté dans plusieurs régions françaises. L'entreprise accompagne des TPE, PME et ETI dans leurs démarches comptables, fiscales, juridiques et sociales. Sa mission est de fournir aux dirigeants des outils fiables d'aide à la décision et des conseils personnalisés pour optimiser la gestion et le développement de leur activité.

La structure se distingue par une forte culture de l'innovation et une volonté d'intégrer les compétences en data et en digital dans ses missions traditionnelles. Elle investit dans la formation continue de ses collaborateurs et encourage les projets internes visant à automatiser et fiabiliser les analyses. C'est dans ce contexte que ma mission a pris forme, à l'interface entre les données comptables et les outils de pilotage décisionnel.

1.2 Organisation de l'entreprise

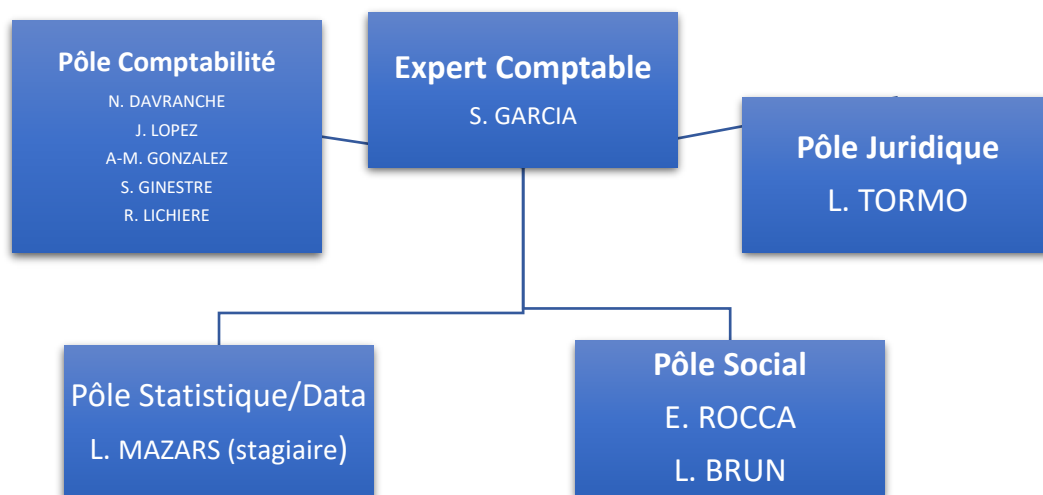
L'organigramme ci-dessous présente la structure organisationnelle du cabinet Endrix dans le cadre de mon stage. Ce cabinet fonctionne selon une répartition en pôles de compétences, chacun spécialisé dans un domaine clé de l'accompagnement des clients.

Au centre de la structure se trouve **Monsieur S. Garcia, expert-comptable**, qui coordonne l'ensemble des pôles. Il assure la supervision technique des dossiers, la relation client principale, et la validation des livrables produits par les différents services. C'est également mon tuteur principal durant le stage.

Autour de lui gravitent plusieurs **pôles spécialisés** :

- Le **pôle Comptabilité**, composé de cinq collaborateurs (N. Davranche, J. Lopez, A.-M. Gonzalez, S. Ginestre et R. Lichiere), est chargé de la gestion comptable quotidienne des clients : saisie, clôtures, déclarations fiscales, bilans.
- Le **pôle Juridique**, représenté par L. Tormo, s'occupe des questions liées au droit des sociétés, à la création ou la transformation d'entreprises, et à la conformité réglementaire.
- Le **pôle Social**, constitué de E. Rocca et L. Brun, intervient sur les problématiques liées à la paie, aux contrats de travail, aux déclarations sociales et à l'accompagnement RH.
- Le **pôle Statistique/Data**, où travaille L. Mazars, est plus récent et incarne la volonté du cabinet d'intégrer des outils décisionnels, des analyses statistiques, et des solutions de data visualisation dans ses prestations. C'est au sein de ce pôle que s'inscrit ma mission de stage.

Cette organisation permet une gestion transversale des dossiers, où les clients peuvent bénéficier d'un accompagnement global, tout en mobilisant des compétences spécialisées selon leurs besoins (comptables, juridiques, RH ou analytiques). Ma mission s'est donc déroulée dans un environnement pluridisciplinaire, avec des échanges réguliers entre les pôles afin de bien comprendre les données et de répondre aux enjeux d'identification des aides de l'État.



3.1 Clientèle

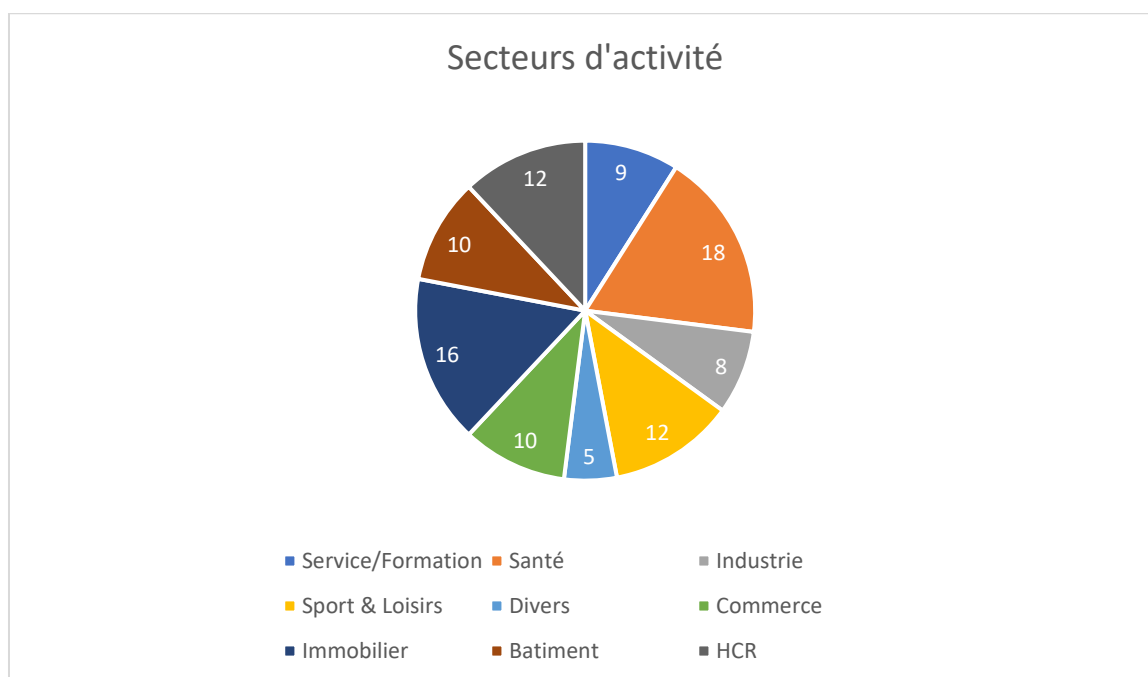


Figure 1 : Diagramme circulaire des clients

Ce diagramme circulaire présente la répartition des entreprises selon leur secteur d'activité. On observe que le secteur de la santé est le plus représenté avec 18 entreprises, suivi par le service/formation (16 entreprises), les HCR (hôtels, cafés, restaurants) et le sport & loisirs, chacun comptant 12 entreprises. Les secteurs du commerce et du bâtiment regroupent chacun 10 entreprises, tandis que l'industrie en compte 8. Les secteurs divers et immobilier sont les moins représentés, avec respectivement 9 et 5 entreprises. Cette répartition met en évidence une forte concentration dans les

domaines liés aux services et à la santé qui est dû à une forte présence de profession libérale de santé et SCI pour l'immobilier.

2.1 Contexte des travaux

Le cabinet Endrix gère un portefeuille important d'entreprises issues de secteurs variés (restauration, technologie, agriculture, commerce, etc.). En raison de la complexité croissante des aides de l'État et de la fréquence des réformes, il est devenu essentiel pour les experts-comptables de disposer d'outils leur permettant d'identifier rapidement les entreprises éligibles à certains dispositifs.

L'enjeu est double :

- Permettre un accompagnement proactif des clients dans leurs démarches administratives,
- Optimiser les conseils donnés en apportant une aide à la décision fondée sur des données fiables.

Dans ce cadre, ma mission a consisté à concevoir un outil de détection automatique de l'éligibilité des entreprises aux aides les plus pertinentes, en m'appuyant sur des analyses statistiques réalisées à partir des données du cabinet, puis à restituer ces informations dans un tableau de bord interactif. Toutes les données, images et annexes présents dans ce rapport sont toutes basées sur des données fictives. En effet l'entreprise a tenu à ce que toutes les données restent confidentielles. J'ai donc refait les démarches et raisonnement pour vous illustrer mon stage à partir d'un jeu de données fictif que j'ai généré.

2 Objectif de la mission

La question centrale était la suivante :

Comment identifier automatiquement les entreprises d'un portefeuille client éligibles à certaines aides de l'État, en s'appuyant sur leurs données comptables et sociales ?

Pour y répondre, plusieurs objectifs intermédiaires ont été définis :

- Identifier les dispositifs d'aide les plus fréquents et en déterminer les critères d'éligibilité
- Collecter, nettoyer et structurer les données comptables des entreprises
- Créer des indicateurs d'éligibilité pour chaque dispositif
- Mener une analyse statistique pour repérer des tendances ou des profils-types
- Construire un tableau de bord clair et interactif permettant une lecture directe des résultats.

L'ensemble de cette démarche devait s'inscrire dans une logique d'automatisation et de fiabilisation des processus, tout en rendant les résultats accessibles à des utilisateurs non techniques (comptables, juristes, conseillers).

2. Analyse de l'existant

Avant le début de ma mission, l'analyse de l'éligibilité aux aides était réalisée manuellement ou de manière ponctuelle, en fonction des demandes spécifiques des clients. Les experts-comptables se basaient sur leur connaissance des dispositifs et sur les échanges individuels avec les clients pour identifier les entreprises concernées. Ce processus, bien que pertinent au cas par cas, était chronophage, peu systématique et exposé à des oublis ou à une sous-exploitation des données disponibles.

3. Propositions méthodologiques

3.1 Première proposition

Dans un premier temps, j'ai envisagé d'aborder cette mission de manière séquentielle, en la divisant en trois grandes étapes : **la détermination des aides de l'État**, **la création du notebook Python** et **la conception du tableau de bord Power BI**.

- **Détermination des aides de l'État** : Il m'a semblé indispensable de commencer par identifier précisément les aides que je souhaitais intégrer dans l'analyse. En effet, il aurait été peu pertinent de réaliser une exploration des données à l'aveugle, sans avoir défini les critères d'éligibilité en amont. J'ai donc prévu de solliciter Monsieur Garcia et un comptable du cabinet afin qu'ils me transmettent les cinq aides les plus couramment mobilisées par les clients, ainsi que les critères précis associés à chacune.
- **Création du notebook** : Une fois ces aides identifiées, j'aurais importé les données issues des portefeuilles clients, fournies par le service comptable, dans un notebook Python. Ce fichier allait faire l'objet d'un nettoyage, puis d'un enrichissement avec de nouvelles variables, notamment les indicateurs d'éligibilité aux aides. J'avais également prévu de réaliser une analyse exploratoire univariée et bivariée afin de détecter des tendances intéressantes ou des profils types d'entreprises, que je pourrais ensuite mettre en évidence dans la restitution visuelle.
- **Création du tableau de bord** : Enfin, l'objectif était d'exporter le jeu de données final depuis le notebook pour l'intégrer dans Power BI. À partir de là, j'aurais conçu un tableau de bord interactif, incluant des filtres, des indicateurs clés et une mise en forme conditionnelle permettant d'identifier rapidement les entreprises éligibles, avec une colorisation spécifique. Tous les résultats de l'analyse auraient ainsi pu être restitués de façon visuelle et compréhensible.

Cette première proposition, bien que logique, s'est révélée insuffisamment précise et structurée lors de la première réunion de cadrage avec mon tuteur. Elle ne permettait pas de visualiser clairement toutes les dimensions du projet.

3.2 Deuxième proposition

À la suite de la première réunion, j'ai pris conscience de la nécessité de structurer davantage ma méthodologie.

- **Création de la carte mentale** : outil que j'ai trouvé particulièrement utile pour poser et organiser mes idées. Cette carte me permettra de visualiser l'ensemble du projet de manière synthétique, en répondant aux grandes questions fondamentales : **Qui ? Quoi ? Pourquoi ? Quand ? Comment ? Combien ?** Elle mettra en évidence les liens entre les objectifs de la mission, les acteurs impliqués, les étapes clés, les outils mobilisés et les livrables attendus. Grâce à cette carte mentale (reproduite en annexe), je pourrai redéfinir un **plan de route plus clair**, avec une meilleure anticipation des besoins intermédiaires et des validations à obtenir tout au long du projet.
- **Détermination des aides de l'État** : Comme prévu initialement, cette étape consiste à identifier, avec l'aide de Monsieur Garcia et du comptable, les cinq aides les plus pertinentes à analyser. À chaque aide sont associés des critères d'éligibilité précis, exprimés sous forme de conditions logiques simples : montant de chiffre d'affaires, ancienneté de l'entreprise, statut juridique, part des dépenses en R&D ou en énergie, etc.
- **Création du notebook** : J'importerai ensuite le jeu de données dans un notebook Python. Celui-ci contiendra toutes les étapes nécessaires : nettoyage, transformation, création des variables d'éligibilité, analyse exploratoire (statistiques descriptives, visualisations), et génération du fichier final à exploiter dans Power BI. Les fonctions seront organisées de manière modulaire afin d'assurer la reproductibilité.
- **Mise en place des critères d'éligibilités** : La mise en place des critères d'éligibilité consistera à traduire les règles de manière propres à chaque aide en conditions mesurables dans les données. Par exemple, l'aide ACRE a été définie selon l'ancienneté de l'entreprise et son statut juridique, tandis que l'aide régionale dépend de la localisation et de la taille de la structure. Chaque critère sera implémenté sous forme de variables booléennes (True/False), permettant d'identifier automatiquement les entreprises éligibles. Cette étape posera les bases de l'analyse, en rendant possible le comptage, la comparaison et la visualisation des bénéficiaires potentiels de chaque dispositif dans Power BI.
- **Conception du tableau de bord Power BI** : Après exportation, les données seront intégrées dans Power BI. Le tableau de bord inclura : des KPI dynamiques, des filtres par secteur ou par aide, des graphiques synthétiques, ainsi qu'un tableau récapitulatif interactif permettant aux comptables de détecter en un coup d'œil les entreprises éligibles à une ou plusieurs aides. Ce livrable fera l'objet de présentations intermédiaires pour recueillir les retours du tuteur et des utilisateurs finaux.

Ce second plan méthodologique s'est avéré beaucoup plus robuste, à la fois du point de vue technique et organisationnel. Il a permis de mieux communiquer sur l'avancement du projet, tout en intégrant des ajustements progressifs selon les besoins exprimés.

4. Application

4.1 Développement des connaissances en comptabilité

En concertation avec le comptable et le responsable de mission du cabinet Endrix, cinq aides de l'État ont été sélectionnées pour cette étude. Ces dispositifs ont été retenus en fonction de leur fréquence d'attribution, de leur intérêt stratégique pour les entreprises clientes du cabinet, et de la disponibilité des données nécessaires à leur évaluation. Chacune de ces aides repose sur des critères d'éligibilité mesurables à partir des données comptables et structurelles des entreprises.

Aide de l'État	Entreprises ciblées	Principaux critères d'éligibilité
ACRE (Aide à la Création ou à la Reprise d'Entreprise)	Créateurs ou repreneurs d'entreprise	Moins de 3 ans d'ancienneté, statut d'auto-entrepreneur ou assimilé, inscription à Pôle emploi
CIR (Crédit d'Impôt Recherche)	Entreprises innovantes de toutes tailles	Dépenses de R&D, présence de personnel dédié, dépôt de brevet, part R&D > 15 % du CA
Aide énergie (bouclier tarifaire)	Entreprises affectées par la hausse des coûts énergétiques	Dépenses énergie > 3 % du CA, baisse de CA ≥ 30 %, activité éligible
Aide régionale PME/TPE	Petites entreprises selon les régions	Localisation en région prioritaire, effectif < 250, seuils de CA, autonomie juridique
Report de charges sociales (URSSAF)	TPE/PME en difficulté	Baisse significative de chiffre d'affaires, dettes sociales existantes, justification financière

- La première aide retenue est l'**ACRE** (Aide à la Création ou à la Reprise d'Entreprise). Elle s'adresse aux créateurs ou repreneurs d'entreprise, notamment les auto-entrepreneurs, et repose sur des critères tels que l'ancienneté de l'entreprise (moins de trois ans), le statut juridique et, dans certains cas, la situation personnelle du dirigeant (ex. : inscription à Pôle emploi). Cette aide est fréquente dans les premières années d'activité et offre une exonération partielle des cotisations sociales.
- La deuxième aide analysée est le **Crédit d'Impôt Recherche (CIR)**, destiné aux entreprises innovantes. Elle repose sur l'évaluation des dépenses en recherche et développement (R&D), la présence de personnel dédié, ou encore le dépôt de brevets. Dans notre cas, des indicateurs tels que la part des dépenses R&D dans le chiffre d'affaires ou la présence d'employés affectés à la R&D ont été utilisés comme variables proxy pour estimer l'éligibilité.
- Le troisième dispositif est l'**aide énergie**, liée au bouclier tarifaire et aux dispositifs d'aide mis en place face à la flambée des prix de l'énergie. Les entreprises concernées doivent justifier de dépenses énergétiques importantes (supérieures à 3 % du chiffre d'affaires) et, dans certains cas, d'une baisse significative de chiffre d'affaires sur l'année. Ce critère a pu être simulé par le calcul du ratio dépenses énergie / CA et par une variable représentant l'évolution du CA entre 2022 et 2023.

- La quatrième aide retenue concerne les **aides régionales destinées aux TPE/PME**. Ces dispositifs varient selon les régions mais ont en commun des critères comme l'effectif de l'entreprise (souvent inférieur à 250 salariés), la localisation géographique (présence en zone prioritaire ou en région ciblée), ainsi que des conditions d'autonomie juridique ou d'indépendance. Les variables "effectif", "région", "zone" et "statut" ont permis d'approcher ces critères.
- Enfin, la cinquième mesure étudiée est le **report ou l'étalement de charges sociales** accordé par l'URSSAF aux entreprises en difficulté. Elle concerne principalement les structures ayant subi une baisse marquée de chiffre d'affaires ou présentant des dettes sociales. À cette fin, j'ai intégré des variables telles que la variation de chiffre d'affaires et le montant des dettes sociales déclarées.

L'un des apports majeurs de ce stage a été le développement de mes connaissances en comptabilité, domaine que je maîtrisais peu avant cette expérience. Même si ma mission s'est concentrée sur le traitement et l'analyse de données, elle s'est inscrite pleinement dans un contexte comptable réel, ce qui m'a amené à manipuler, interpréter et comprendre de nombreux indicateurs financiers.

Dès les premières semaines, j'ai été amené à me familiariser avec des notions fondamentales comme le chiffre d'affaires, les charges, les dettes sociales, les dépenses de R&D, les statuts juridiques d'entreprise ou encore les régimes fiscaux spécifiques. Comprendre ces éléments était indispensable pour construire des variables d'analyse cohérentes et surtout pour évaluer les critères d'éligibilité aux différentes aides de l'État. Par exemple, il m'a fallu distinguer les caractéristiques propres aux auto-entrepreneurs, identifier les seuils de chiffre d'affaires liés à certains dispositifs, ou encore repérer la logique des ratios utilisés dans le conseil fiscal.

Les échanges avec le comptable référent ont été particulièrement instructifs : ils m'ont permis de mieux comprendre l'utilité de chaque variable, les interprétations possibles et les limites à ne pas dépasser dans une analyse économique. J'ai appris à distinguer les indicateurs comptables des indicateurs de gestion, à poser les bonnes questions sur la structure juridique d'une entreprise, et à m'interroger sur les causes possibles d'une variation de chiffre d'affaires ou d'une charge inhabituelle.

Par ailleurs, l'intégration de règles d'éligibilité aux aides m'a conduit à me plonger dans des textes officiels, des guides administratifs et des ressources issues de sites institutionnels (URSSAF, BPI, services de l'État), ce qui a complété ma compréhension du cadre réglementaire autour des aides publiques. Cette recherche m'a aidé à mieux relier les données comptables aux décisions concrètes d'accompagnement que peut proposer un cabinet.

En résumé, ce stage a enrichi ma culture comptable de manière très concrète. Il m'a permis de mieux comprendre le rôle de la comptabilité dans la gestion d'entreprise, les liens entre indicateurs financiers et stratégie d'accompagnement, et les attentes spécifiques d'un cabinet d'expertise dans le traitement des données de ses clients. Ces acquis me seront utiles non seulement pour des analyses futures, mais aussi pour mieux dialoguer avec des interlocuteurs métiers dans mes prochains projets professionnels.

4.2 Création du notebook

Le traitement des données a été effectué dans un environnement Jupyter Notebook. J'ai utilisé plusieurs bibliothèques Python : pandas pour la manipulation de données, numpy pour les calculs numériques, matplotlib pour les visualisations, ainsi que faker pour simuler le jeu de données pour le rapport de stage.

Les étapes principales ont été les suivantes :

- **Importation du jeu de données** : L'importation du jeu de données a constitué la première étape concrète du projet. Le fichier source, au format CSV, a été chargé dans un environnement Jupyter Notebook à l'aide de la bibliothèque pandas. Une fois les données importées, j'ai procédé à une vérification rigoureuse des colonnes : noms, types de données, doublons éventuels, cohérence des intitulés, etc. Ce travail de contrôle initial m'a permis de détecter certaines valeurs manquantes ou incohérentes. J'ai alors engagé un nettoyage approfondi du jeu de données, en traitant les champs vides ou aberrants (comme des chiffres d'affaires négatifs ou des effectifs irréalistes), en corrigeant certains formats de date incorrects et en normalisant la structure du fichier afin de le rendre exploitable.
- **Création de nouvelles colonnes** : Une fois le fichier nettoyé, j'ai enrichi la base par la création de nouvelles variables, essentielles à l'analyse. Parmi elles figurait l'âge de l'entreprise, calculé à partir de la date de création. J'ai également calculé la part des dépenses de recherche et développement par rapport au chiffre d'affaires, ainsi qu'un ratio dépenses énergie sur chiffre d'affaires, utile pour évaluer l'impact énergétique de chaque structure. D'autres indicateurs simples mais pertinents ont également été ajoutés, comme des rapports entre les dettes sociales et les revenus, permettant d'alimenter les futurs critères d'éligibilité.
- **Construction de règles d'éligibilité** : Sur la base de ces variables, j'ai ensuite mis en place des règles d'éligibilité codées sous forme de conditions logiques. Pour chaque aide étudiée, j'ai créé une variable binaire (0 ou 1) reflétant la satisfaction ou non des critères définis. Par exemple, une entreprise auto-entrepreneur ou âgée de moins de trois ans était considérée comme potentiellement éligible à l'ACRE. Pour l'aide énergie, la règle nécessitait un ratio de dépenses énergétiques supérieur à un certain seuil, combiné à une baisse significative du chiffre d'affaires. Ces règles ont été implémentées en utilisant les fonctions conditionnelles de Python et validées à l'aide de cas test.
- **Analyse exploratoire** : Une fois ces enrichissements réalisés, j'ai conduit une analyse exploratoire du jeu de données. Dans un premier temps, j'ai étudié les variables de façon univariée, à travers des histogrammes, boxplots et statistiques descriptives. Cela m'a permis d'identifier des tendances générales, comme la répartition du chiffre d'affaires ou la fréquence des entreprises en zone prioritaire. J'ai ensuite approfondi l'analyse de manière croisée, en comparant les entreprises éligibles et non éligibles selon différents indicateurs. Cette phase m'a permis d'identifier des profils types d'entreprises, mais aussi de mieux comprendre les combinaisons de facteurs expliquant l'accès à certaines aides.

- **Export du fichier final** : Enfin, une fois le jeu de données complet et validé, je l'ai exporté dans un format exploitable par Power BI. L'ensemble des colonnes était renommé de manière explicite, et les indicateurs étaient prêts à être utilisés dans le cadre de visualisations interactives. Ce fichier représentait la version consolidée du travail réalisé dans le notebook Python, servant de fondation au tableau de bord final.

Pour respecter la confidentialité de l'entreprise j'ai donc généré un jeu de données fictif pour pouvoir vous présenter une simulation de mes travaux durant ce stage. J'ai utilisé la librairie faker pour générer le jeu de données en utilisant la même structure que celui qui m'a été transmis pendant le stage.

```
# Initialisation
fake = Faker()
np.random.seed(42)

# Listes utiles
regions = ['Île-de-France', 'Auvergne-Rhône-Alpes', 'Nouvelle-Aquitaine', 'Occitanie', 'PACA', 'Hauts-de-France']
secteurs = ['Industrie', 'Commerces', 'Restauration', 'Technologie', 'Agriculture', 'Construction']
statuts = ['SARL', 'SAS', 'Auto-entrepreneur', 'SASU', 'SA']
zones = ['ZFU', 'Hors ZFU']

# Génération
n = 1000
data = {
    'Entreprise_ID': [f"E{str(i).zfill(4)}" for i in range(1, n + 1)],
    'Nom_entreprise': [fake.company() for _ in range(n)],
    'Secteur': np.random.choice(secteurs, size=n),
    'Statut': np.random.choice(statuts, size=n),
    'Date_creation': pd.to_datetime(np.random.randint(2000, 2024, size=n), format='%Y') + pd.to_timedelta(np.random.randint(0, 365, size=n), unit='d'),
    'Effectif': np.random.randint(1, 300, size=n),
    'Chiffre_affaires_2023': np.random.uniform(10000, 2_000_000, size=n).round(2),
    'Depenses_energie_2023': lambda: np.random.uniform(500, 100_000, size=n).round(2),
    'Depenses_R&D': lambda: np.random.choice([0] * 700 + list(np.random.uniform(5000, 200_000, size=300)), size=n),
    'Dettes_sociales': lambda: np.where(
        np.random.rand(n) < 0.25, # 25 % avec 0
        0,
        np.random.uniform(1000, 150000, size=n).round(2)
    ),
    'Region': np.random.choice(regions, size=n),
    'Zone': np.random.choice(zones, size=n),
    'Baisse_CA_2022_vs_2023': np.random.uniform(-0.5, 0.3, size=n).round(2),
    'Employe_R&D': np.random.choice([0, 1], size=n, p=[0.75, 0.25]),
    'Possede_brevet': np.random.choice([0, 1], size=n, p=[0.9, 0.1])
}
```

Figure 2 : Code python pour générer le jeu de données fictif

Le code ci-dessus m'a permis de générer la base de mon jeu de données fictif c'est-à-dire d'avoir le même nombre de colonne, nom de colonne et nombre de lignes. Cependant pendant le stage le jeu de données que j'ai reçu comporter des erreurs et anomalies comme des doublons, valeurs aberrantes ou manquantes j'ai donc ajuste le code pour permettre de recréer le plus fidèlement possible la structure du jeu de données original.

```

# Ajout d'erreurs simulées
# 1. Valeurs manquantes
for col in ['Chiffre_affaires_2023', 'Depenses_energie_2023', 'Effectif']:
    df.loc[df.sample(frac = 0.05).index, col] = np.nan # 5% de NaN

# 2. Valeurs aberrantes
df.loc[df.sample(frac = 0.02).index, 'Chiffre_affaires_2023'] = -9999 # CA négatif
df.loc[df.sample(frac = 0.01).index, 'Effectif'] = 9999 # Effectif irréaliste

# 3. Données dupliquées
duplicates = df.sample(n = 10)
df = pd.concat([df, duplicates], ignore_index=True)

# 4. Formats erronés
df.loc[df.sample(frac = 0.01).index, 'Date_creation'] = 'inconnue' # chaîne invalide

# 5. Mauvais types
df.loc[df.sample(frac = 0.01).index, 'Depenses_R&D'] = 'non disponible'

```

Figure 3 : Code Python pour créer les erreurs du jeu de données

Une fois les erreurs les ajoutées j'ai donc obtenu un jeu de données ressemblant au jeu de données du stage et donc réalisé les étapes que je vous ai donné un peu plus haut.

	Entreprise_ID	Nom_entreprise	Secteur	Statut	Date_creation	Effectif	Chiffre_affaires_2023	Depenses_energie_2023	Depenses_R&D	Dettes_sociales
0	E0001	Stevenson, Schmidt and Clark	Technologie	SA	2015-09-17	137	1867078.12	14267.97	0.000000	0.00
1	E0002	Rodgers-Adams	Agriculture	SASU	2004-09-02	74	538337.85	98282.58	0.000000	0.00
2	E0003	Ruiz-Franklin	Restauration	Auto-entrepreneur	2008-10-01	36	1226812.59	25372.30	0.000000	73758.14
3	E0004	King LLC	Agriculture	SARL	2001-02-01	101	1360214.94	7799.72	137039.231379	51320.15
4	E0005	Cohen-Jackson	Agriculture	Auto-entrepreneur	2017-05-17	156	641923.90	66483.05	0.000000	82841.01
5	E0006	Miller-Jones	Commerces	Auto-entrepreneur	2005-01-12	255	1697737.38	59738.95	28473.739154	0.00
6	E0007	Webb, Brown and Terrell	Restauration	SA	2008-10-28	171	1894851.31	41806.99	128469.841605	117805.36
7	E0008	Glover PLC	Restauration	SASU	2000-01-16	216	1770861.68	42508.13	0.000000	49366.71
8	E0009	Sanders, Ellis and Hampton	Restauration	SAS	2008-05-08	170	1480743.28	78613.86	143532.253927	108452.73
9	E0010	Carr, Martinez and Williams	Agriculture	SASU	2015-03-04	252	562154.71	19973.41	126476.038616	99010.53

Figure 4 : Aperçu du jeu de données fictif

La partie la plus longue et la plus exigeante de cette étape a été la prise en main du jeu de données. Il m'a fallu du temps pour bien comprendre la signification de chaque variable, identifier celles qui allaient être utiles dans le cadre de l'analyse, et anticiper leur transformation. Une fois cette phase d'appropriation terminée avec l'aide du comptable et de mon tuteur de stage, la construction du notebook s'est révélée beaucoup plus fluide, me permettant d'appliquer les compétences acquises en manipulation et traitement de données.

J'ai particulièrement apprécié la phase d'analyse statistique. C'est à ce moment-là que tout le travail de nettoyage, souvent invisible, prend tout son sens : la qualité des visualisations et des relations observées dépend directement de la rigueur appliquée en amont. Voir les données "parler" à travers des graphiques bien construits a été une vraie source de satisfaction.

Ce travail m'a permis de consolider mes connaissances en Python, notamment sur les bibliothèques Pandas, NumPy et Matplotlib. J'ai également renforcé mes réflexes en logique conditionnelle et en structuration de projet analytique. Travailler avec Python dans un contexte professionnel a été une première pour moi, et j'ai réellement pris plaisir à explorer les possibilités offertes par ce langage puissant et polyvalent.

4.3 Création du tableau de bord Power BI

La dernière étape du projet consistait à créer un **tableau de bord interactif sous Power BI** à partir du fichier enrichi et exporté depuis le notebook. Ce tableau de bord devait répondre à plusieurs objectifs :

Offrir une lecture intuitive, permettre un filtrage dynamique et mettre en valeur les entreprises éligibles aux aides identifiées.

Les principales étapes de construction ont été les suivantes :

- **Importation et modélisation des données** : La phase de création du tableau de bord a débuté par l'importation du fichier CSV finalisé dans Power BI. Une fois les données chargées, j'ai pris soin de vérifier les types de données associés à chaque colonne : numériques, booléennes, textuelles ou temporelles. Cette étape de validation était essentielle pour garantir la cohérence des futures visualisations. J'en ai profité pour renommer certaines colonnes, afin de les rendre plus explicites et compréhensibles pour les utilisateurs finaux, notamment les comptables et les conseillers non techniques.

Entreprise_ID	Nom_entreprise	Secteur	Statut	Date_creation	Effectif	Chiffre_affaires_2023	Depenses_energie_2023	Depenses_R&D	Dettes_sociales
E0001	Hughes, Wilson and Gonzalez	Divers	Auto-entrepreneur	37086	180	1073311,67	4818,38	0	127121,39
E0002	Smith-Reyes	Sport & Loisirs	SARL	43958	170	1314888,04	92502,01	0	0
E0003	Lawrence, Boone and Washington	Bâtiment	Auto-entrepreneur	40576	271	776939,28	91952,01	0	60195,81
E0004	Davis-Espinoza	HCR	Auto-entrepreneur	38528	54	1366411,2	25672,52	0	119797,01
E0005	Horton, Norman and Jennings	HCR	SARL	44583	101	687842,89	69693,41	140817,78591428	23337,7
E0006	Petersen Inc	HCR	Auto-entrepreneur	37859	153	528782,11	8005,74	0	0
E0007	Mcmahon-Compton	Service/Formation	SA	44719	53	997114,53	48280,695	0	108615,63
E0008	Richards-Drake	Divers	SAS	40216	60	1388851,81	22072,51	0	108285,44
E0009	Hunter-Hill	Santé	SAS	45156	108	703189,84	29802,15	13942,9469489034	96531
E0010	Martin Group	Santé	SARL	42734	5	1873929,82	99585,22	140817,78591428	104398,32
E0011	Hall-Gallagher	Service/Formation	SASU	38403	103	87980,79	69844,04	0	81865,94
E0012	Aguirre Group	Service/Formation	SARL	45190	196	841712,6	38728,08	0	0
E0013	Duran, Jennings and Gonzalez	Bâtiment	SASU	38269	262	1935485,3	73841,52	160311,996256406	52508,7
E0014	Hicks, Meyer and Roman	Service/Formation	SAS	43492	47	1100464,05	91567,81	0	28058,06
E0015	Cooper-Kramer	Industrie	SARL	37117	144	852707,18	95890,89	0	136359,13
E0016	Cook-Collins	Santé	SA	38629	269	1141355,36	6257,46	119029,390660085	87925,38
E0017	Moore LLC	Bâtiment	Auto-entrepreneur	44482	124	1156089,9	39754,92	0	60726,86
E0018	Mcbride, Burke and Harris	Bâtiment	SASU	40275	106	1465979,03	11122,23	194756,51810767	69838,86
E0019	Morris-Harrell	Santé	Auto-entrepreneur	42288	158	264102,56	33898,18	0	0
E0020	Parker, Taylor and Chambers	Service/Formation	Auto-entrepreneur	42321	153	507532,73	17383,15	0	23849,36
E0021	Petersen Ltd	Sport & Loisirs	SARL	36829	145	1165282,58	64861,41	0	88348,24
E0022	Fields Inc	Industrie	Auto-entrepreneur	39594	153	1735562,05	39132,7	45780,5042603103	76377,41
E0023	Mueller-Anderson	Service/Formation	SA	38356	275	1128114,72	23324,78	0	0
E0024	Martinez LLC	HCR	Auto-entrepreneur	42039	92	484807,75	48280,695	0	0
E0026	Fields-Rodriguez	Santé	SA	43514	229	1482418,43	26365,01	0	139885,62
E0027	Dickson and Sons	HCR	SAS	37638	117	484089,94	45597,46	72785,7468241233	85204,84

Figure 5 : Aperçu des données exportées dans Power BI

- **Création des visualisations** : Après la modélisation des données, j'ai conçu plusieurs visualisations dynamiques adaptées à la nature de chaque variable. J'ai d'abord utilisé des diagrammes circulaires pour représenter la répartition des entreprises éligibles selon les différentes aides étudiées. Ces graphiques offrent une vue d'ensemble immédiate de la proportion d'entreprises concernées. Ensuite, j'ai mis en place des histogrammes pour explorer les distributions de certaines variables continues telles que le chiffre d'affaires, l'effectif ou encore la répartition par secteur d'activité. Afin d'introduire une dimension géographique à l'analyse, j'ai ajouté des cartes de chaleur permettant de localiser les entreprises selon leur région ou leur présence en zone prioritaire. Enfin, j'ai intégré des tableaux croisés dynamiques, filtrables par critère (aide, secteur, taille, région), afin de faciliter la navigation dans les données et permettre une exploration détaillée selon les besoins spécifiques de l'utilisateur.

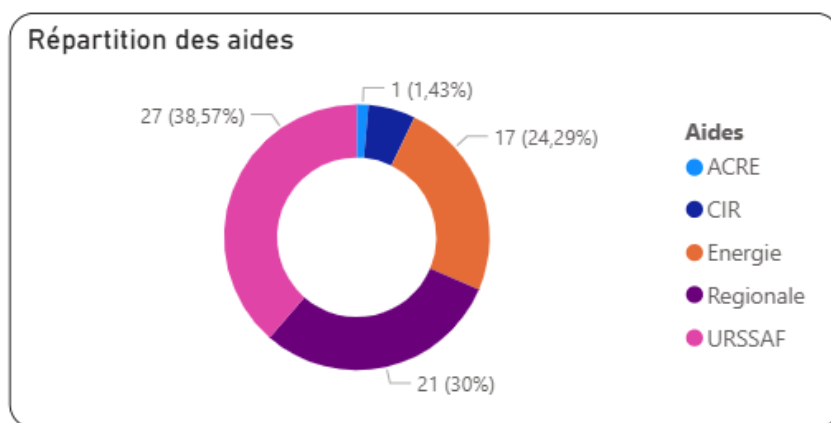


Figure 6 : Première visualisation

- **Mise en forme conditionnelle** : L'un des éléments clés du tableau de bord a été la mise en forme conditionnelle appliquée au tableau récapitulatif. J'ai défini un code couleur simple mais efficace : vert pour les entreprises considérées comme éligibles à au moins une aide, rouge pour celles qui ne remplissent aucun critère. Ce système visuel permet une lecture rapide et intuitive des résultats, même sans connaissances techniques approfondies.

Entreprise_ID	Nom_entreprise	Somme de Chiffre_affaires_2023	Somme de Eff
E0001	Hughes, Wilson and Gonzalez	1 073 311,67	
E0002	Smith-Reyes	1 314 888,04	
E0003	Lawrence, Boone and Washington	776 939,28	
E0004	Davis-Espinoza	1 366 411,20	
E0005	Horton, Norman and Jennings	687 842,89	
E0006	Petersen Inc	528 782,11	
E0007	Mcmahon-Compton	997 114,53	
E0008	Richards-Drake	1 388 851,81	
E0009	Hunter-Hill	703 189,84	
E0010	Martin Group	1 873 929,82	
E0011	Hall-Gallagher	87 980,79	
Total		103 357 980,56	24

Figure 7: Exemple de mise en forme conditionnelle

- **Ajout de KPI** : En complément, j'ai intégré plusieurs indicateurs de performance clés (KPI), tels que le nombre total d'entreprises du portefeuille, la part d'entreprises éligibles, ou encore le taux d'éligibilité par dispositif. Ces KPI sont positionnés en haut du tableau de bord pour fournir une synthèse immédiate dès l'ouverture. Grâce à cette approche, le tableau de bord Power BI constitue non seulement un outil de visualisation, mais aussi un véritable support d'aide à la décision, pensé pour s'intégrer facilement au quotidien des utilisateurs du cabinet.



Figure 8: Exemple de KPI

Enfin, j'ai présenté une version intermédiaire du tableau de bord à mon tuteur, qui m'a permis de valider la pertinence des filtres, la clarté des titres, et la lisibilité globale. Des ajustements ont été réalisés (notamment sur le regroupement des zones géographiques et l'ordre des aides) avant de finaliser la version livrable.

4.4 Résultats et interprétation

L'analyse réalisée a mis en lumière plusieurs **résultats intéressants** (à partir du jeu de données fictif et non des vraies données, je n'ai pas obtenu les mêmes résultats durant le stage) et a permis de dégager des **enseignements utiles** pour le cabinet. Sur l'échantillon traité (environ 100 entreprises), **près de 40 %** se sont révélées éligibles à au moins une aide de l'État.

Voici les principaux constats :

- **L'ACRE** est l'aide la moins utilisée (seulement une entreprise pouvait en bénéficier), en effet le cabinet de sorgues n'a que très peu de jeunes entreprises et donc dès le premier critère la plupart des entreprises n'étaient pas éligibles à cette aide.
- **L'aide énergie** concerne majoritairement les PME du secteur industriel ou agricole, ayant des charges élevées en gaz ou électricité.
- Les entreprises implantées dans les **zones franches urbaines (ZFU)** sont nettement plus représentées parmi les cas éligibles à deux aides ou plus.
- Les entreprises de plus grande taille (effectif > 100) sont moins souvent éligibles, sauf dans le cas du **CIR** (crédit d'impôt recherche), souvent lié à leur capacité à investir en innovation.

L'analyse croisée a également permis d'identifier des **profils types d'entreprises** ayant un fort potentiel d'éligibilité, mais n'ayant pas encore sollicité les aides. Par exemple, certaines structures cumulent un fort taux de R&D, un effectif inférieur à 50 salariés, et une localisation en zone prioritaire, mais n'ont aucune aide déclarée à ce jour.

Enfin, le tableau de bord permet à l'équipe comptable d'**explorer rapidement les profils clients**, de prioriser leur accompagnement, et de **générer des recommandations personnalisées** pour les dirigeants concernés.

Ces résultats ont été présentés à mon tuteur et à plusieurs membres de l'équipe, qui ont salué la clarté du tableau de bord et sa capacité à faire émerger de nouveaux axes de conseil pour les clients du cabinet.

5. Conclusion

Le projet mené au sein du cabinet Endrix m'a offert une opportunité concrète et stimulante d'appliquer les compétences acquises tout au long de ma formation en Licence Professionnelle Métiers du Décisionnel et de la Statistique. Il s'inscrit pleinement dans les enjeux actuels de l'analyse de données au service de la prise de décision, en particulier dans un environnement professionnel réel, avec des contraintes métiers précises.

L'objectif principal était d'identifier automatiquement, à partir de données comptables internes, les entreprises susceptibles d'être éligibles à certaines aides de l'État. Pour cela, un cadre méthodologique rigoureux a été mis en place. Il a débuté par l'identification, en collaboration avec le comptable et le responsable de mission, des cinq dispositifs d'aide les plus pertinents dans le contexte du cabinet. Cette étape a permis de poser les bases de l'analyse, en définissant pour chaque aide une série de critères mesurables et opérationnels.

Le cœur technique du projet s'est appuyé sur un traitement des données dans un notebook Python. Cette phase a nécessité une bonne compréhension des données disponibles, leur nettoyage approfondi, ainsi que l'enrichissement du jeu avec de nouvelles variables calculées, comme l'âge de l'entreprise, la part des dépenses en R&D, ou encore des indicateurs de performance énergétique. Des règles d'éligibilité ont ensuite été codées de manière systématique, permettant de classer les entreprises selon leur conformité aux critères définis pour chaque aide.

Une fois les données prêtes, elles ont été exploitées dans Power BI à travers un tableau de bord interactif. Ce tableau a été conçu pour répondre aux attentes des utilisateurs non techniques : comptables, conseillers, experts-comptables. Il présente une vision synthétique du portefeuille client, permet de filtrer dynamiquement les résultats, de repérer visuellement les entreprises éligibles grâce à une mise en forme conditionnelle, et d'extraire des profils types d'entreprises selon leur taille, secteur ou localisation. Plusieurs indicateurs clés (KPI) y ont également été intégrés pour faciliter la lecture globale.

Au-delà des résultats techniques, ce projet m'a apporté une vision claire de ce que représente une mission de data analyse en entreprise : savoir transformer un besoin métier en solution analytique, dialoguer avec des interlocuteurs non spécialisés, faire preuve de rigueur dans le traitement des données, et surtout rendre les résultats compréhensibles et exploitables pour ceux qui en ont besoin. Ce stage a également renforcé mes compétences techniques, notamment en Python et Power BI, tout en me confrontant à des situations concrètes de gestion de projet.

En somme, ce projet m'a permis de participer activement à une démarche à fort impact pour le cabinet : mieux identifier les entreprises en difficulté ou à fort potentiel, et leur proposer un accompagnement ciblé et proactif. Il constitue une première expérience professionnelle complète et valorisante, et confirme mon intérêt pour les métiers mêlant analyse de données, accompagnement décisionnel et développement d'outils opérationnels. Ce stage marque ainsi une étape importante dans mon parcours de futur data analyst.

6. Bilan

Ce stage a été pour moi une expérience extrêmement enrichissante, tant sur le plan technique que personnel. Il m'a permis de découvrir concrètement le fonctionnement d'un cabinet d'expertise comptable, mais surtout de mettre en application des compétences que j'avais principalement étudiées en contexte académique. Le fait de travailler sur une problématique réelle, avec de vrais enjeux métiers, m'a donné une meilleure compréhension de ce que signifie être data analyst dans un environnement professionnel.

Sur le plan technique, j'ai pu consolider mes acquis en manipulation de données avec Python, notamment dans l'utilisation des bibliothèques pandas, numpy et matplotlib. J'ai appris à structurer mon code, à documenter mes étapes et à organiser mes fonctions de manière logique et réutilisable. La phase de nettoyage et d'enrichissement du jeu de données m'a confronté à des difficultés concrètes : formats incohérents, valeurs manquantes, doublons ou aberrations statistiques. Ce sont précisément ces difficultés qui m'ont permis de progresser le plus, en développant une approche rigoureuse et méthodique du traitement des données.

La découverte et la prise en main de Power BI ont également constitué un apport important. J'ai appris à concevoir un tableau de bord lisible, interactif et utile pour les utilisateurs finaux. Cela m'a permis de mieux comprendre les attentes d'un public non technique, et de traduire des résultats statistiques en informations visuelles et exploitables pour la prise de décision. Cette compétence de restitution me semble essentielle dans le métier de data analyst.

Sur le plan humain, ce stage m'a aussi permis de gagner en autonomie et en confiance. J'ai dû prendre des initiatives, proposer une méthodologie, recueillir des retours et adapter mon travail en conséquence. Les échanges réguliers avec mon tuteur et les membres de l'équipe m'ont appris à mieux communiquer sur l'état d'avancement d'un projet, à écouter les besoins, et à justifier mes choix techniques de manière claire.

Les choses que je souhaiterais améliorer sont principalement sur le rapport Power BI, en effet j'aurais voulu l'automatiser et créer différents rôles selon pour différencier les données selon le besoin métier c'est-à-dire faire apparaître certaines données pour les comptables et d'autres pour les experts comptables. Je n'ai pas non plus eu le temps de voir les retours d'utilisation et d'améliorer mon travail car tout a été au pôle data basé à Lyon pour juger si ce procédé peut être mis en place à l'avenir ou non.

Enfin, ce stage a confirmé mon attrait pour les projets à la croisée de la technique et de la décision. J'y ai trouvé un équilibre motivant entre rigueur analytique, logique métier et communication de résultats. Cette expérience renforce ma volonté de poursuivre dans cette voie, et m'a donné de nouvelles clés pour évoluer dans le domaine de la data, que ce soit dans un environnement comptable, financier ou dans d'autres secteurs où les données sont au cœur des enjeux.

7. Références bibliographiques

- **URSSAF.** Aides et exonérations pour les entreprises. Consulté sur : <https://www.urssaf.fr>
- **Service Public.** Aides à la création, reprise et développement d'entreprise. Consulté sur : <https://www.service-public.fr/professionnels-entreprises>
- **Bpifrance.** Financements, aides publiques et accompagnement des entreprises. Consulté sur : <https://www.bpifrance.fr>
- **Légifrance.** Textes officiels sur les dispositifs fiscaux et sociaux. Consulté sur : <https://www.legifrance.gouv.fr>
- **INSEE.** Définitions statistiques : chiffre d'affaires, effectif, secteur d'activité, etc. Consulté sur : <https://www.insee.fr>
- **OpenClassrooms.** Cours en ligne : « Analyse de données avec Python », « Créez des tableaux de bord avec Power BI ». Consulté sur : <https://openclassrooms.com/fr/>
- **Documentation Microsoft Power BI** (en français). Consulté sur : <https://learn.microsoft.com/fr-fr/power-bi/>
- **Cours Python – France Université Numérique (FUN).** Introduction à la science des données. Consulté sur : <https://www.fun-mooc.fr>

8. Annexe

8.1 Carte mentale

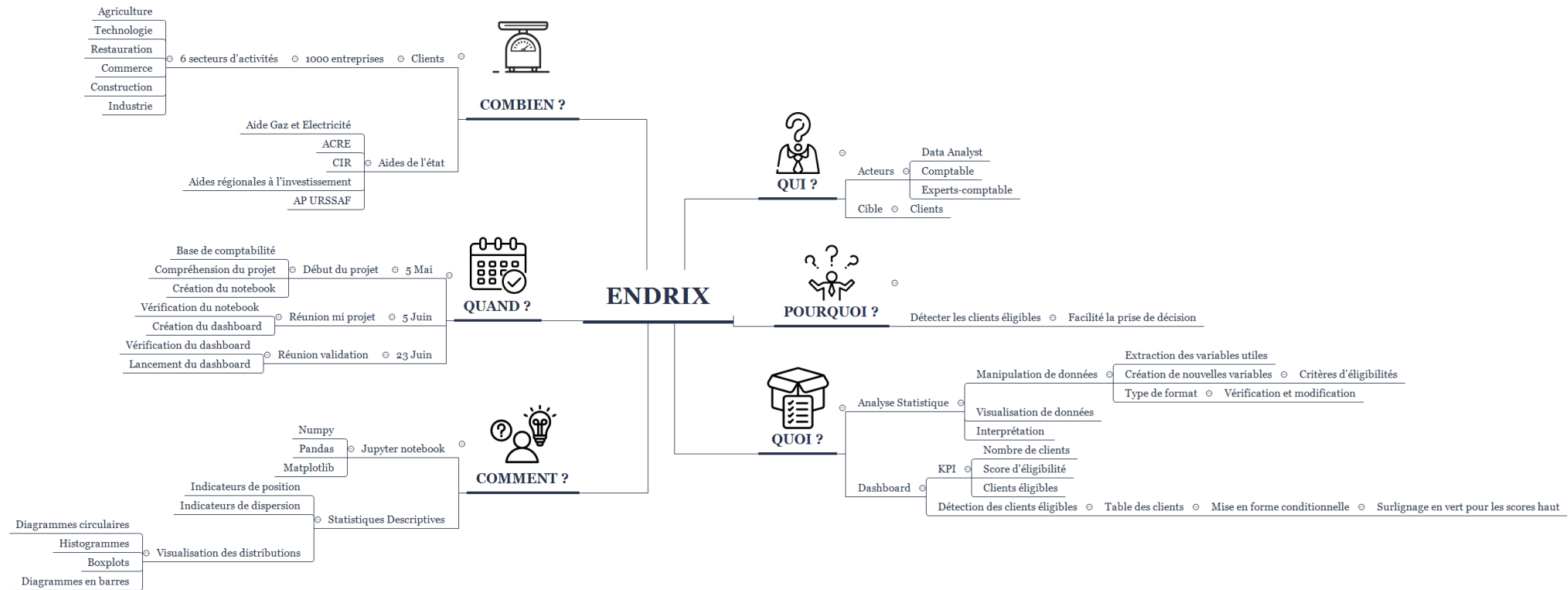


Figure 9 : Carte mentale du projet

8.2 Notebook

[Notebook Endrix](#)

8.3 Rapport Power BI

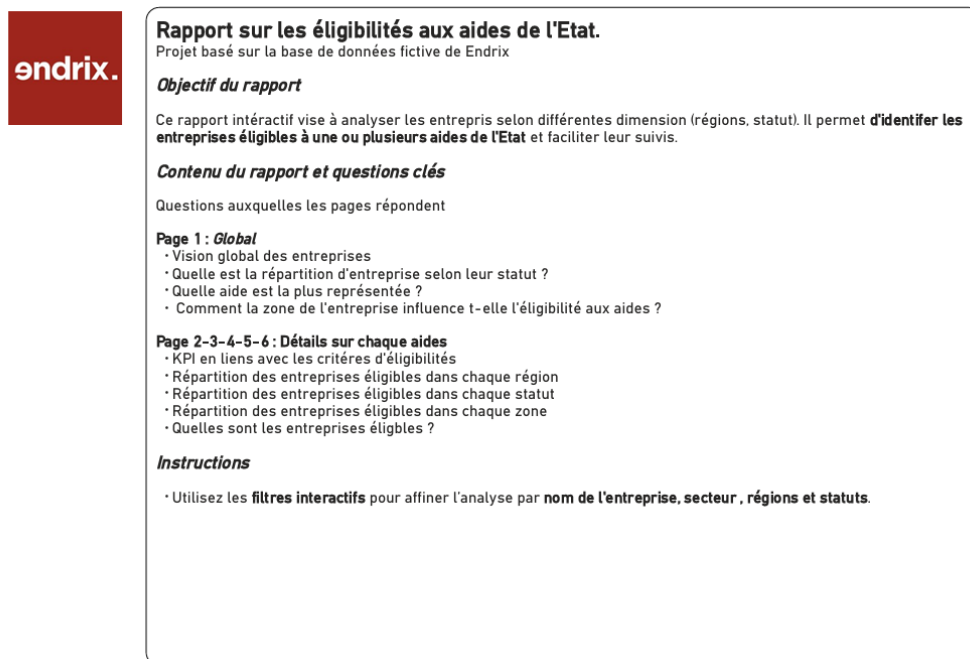


Figure 10 : Page "Accueil"

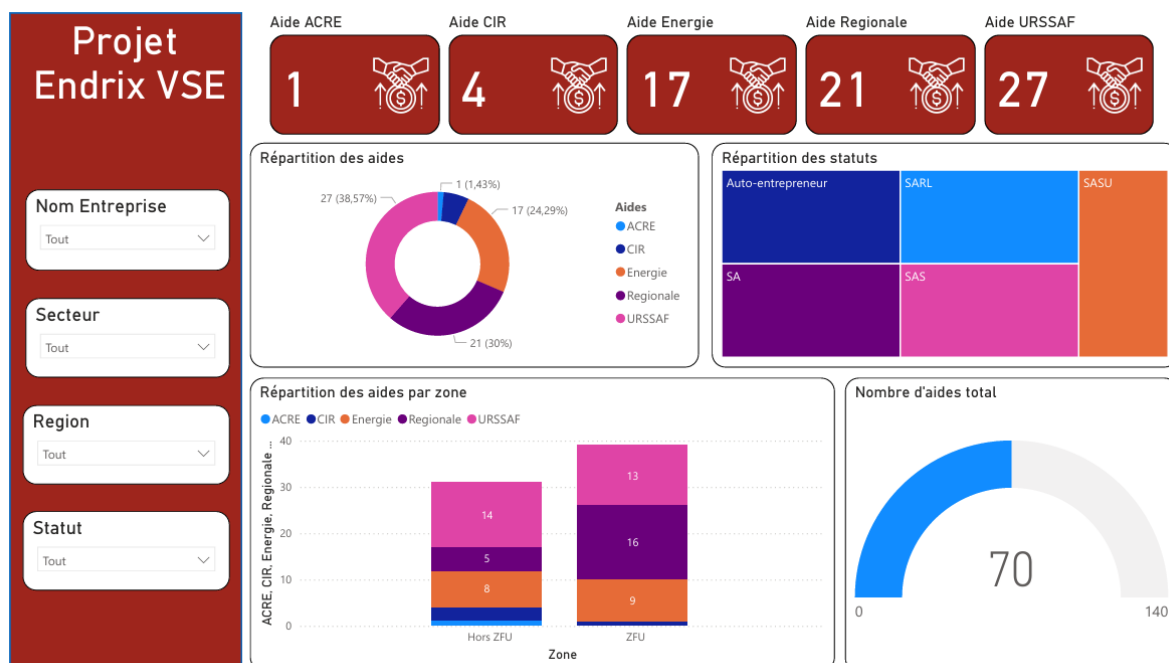
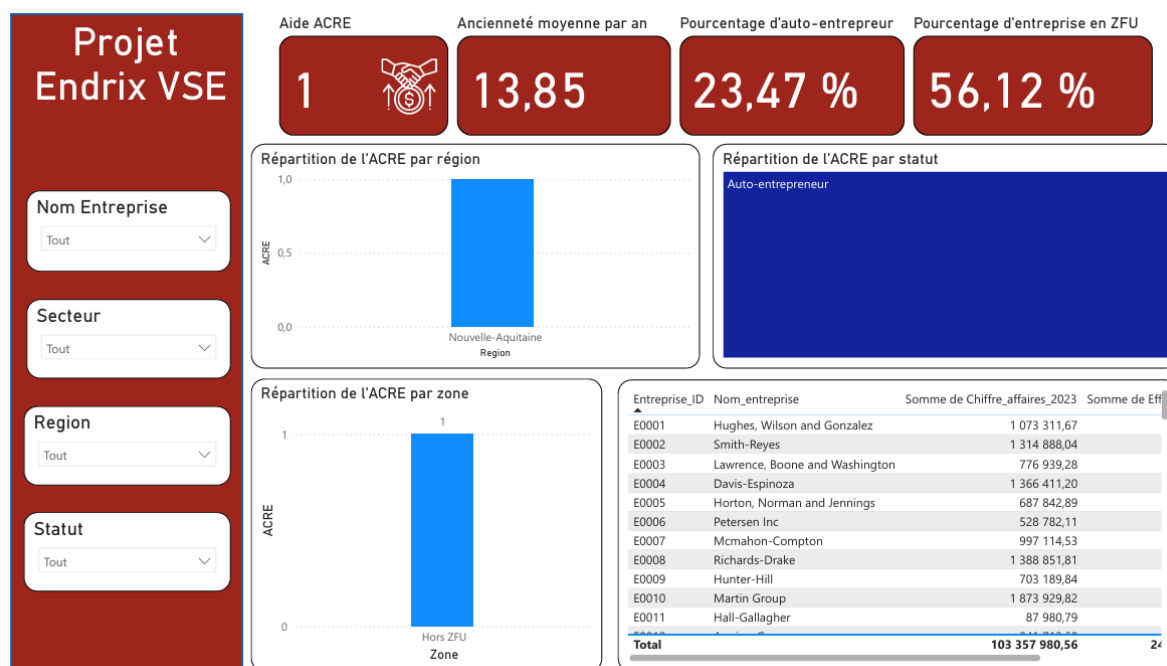


Figure 11 : Page "Global"



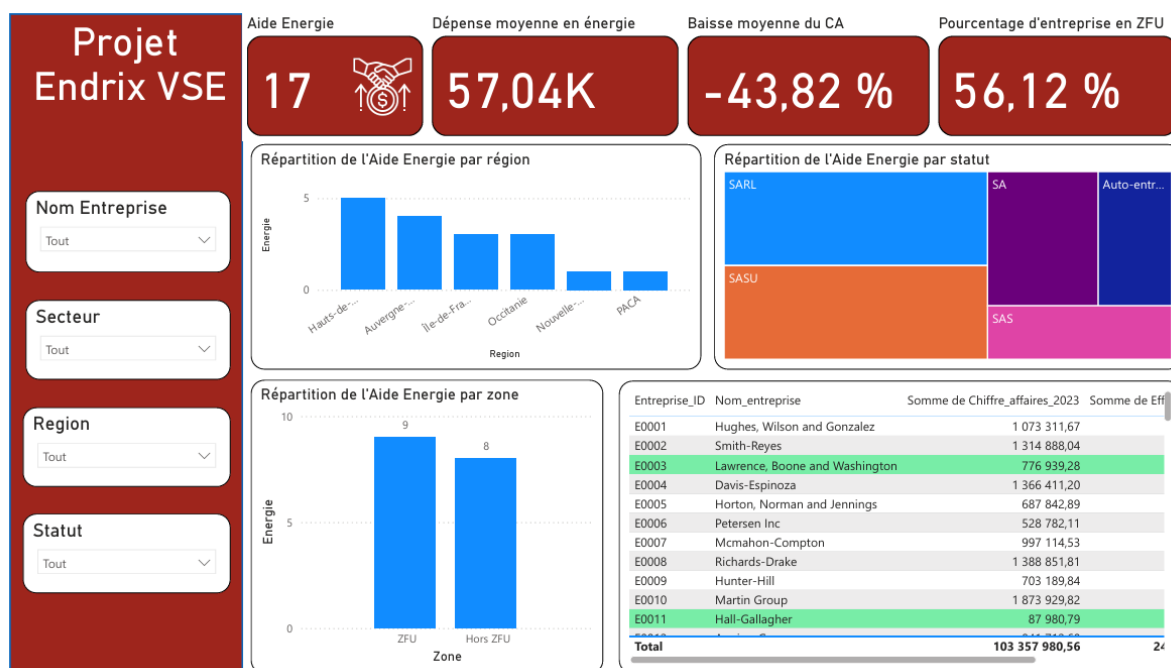


Figure 14 : Page "Aide Energie"

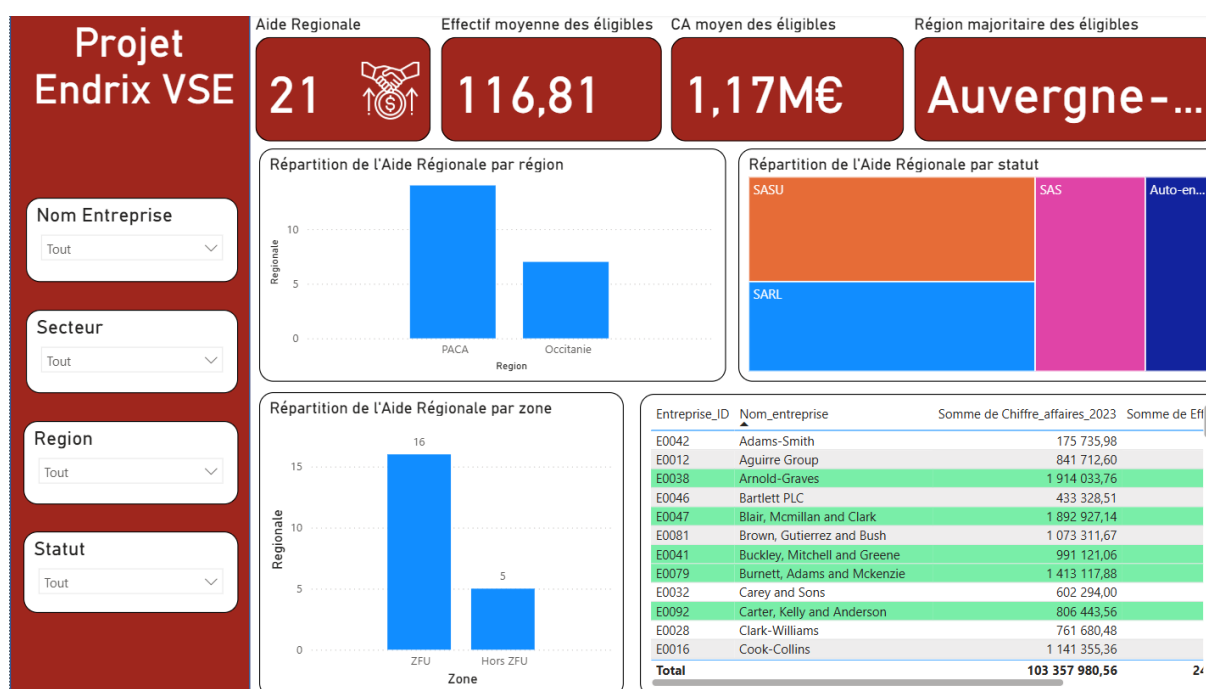


Figure 15 : Page "Aide Régionale"

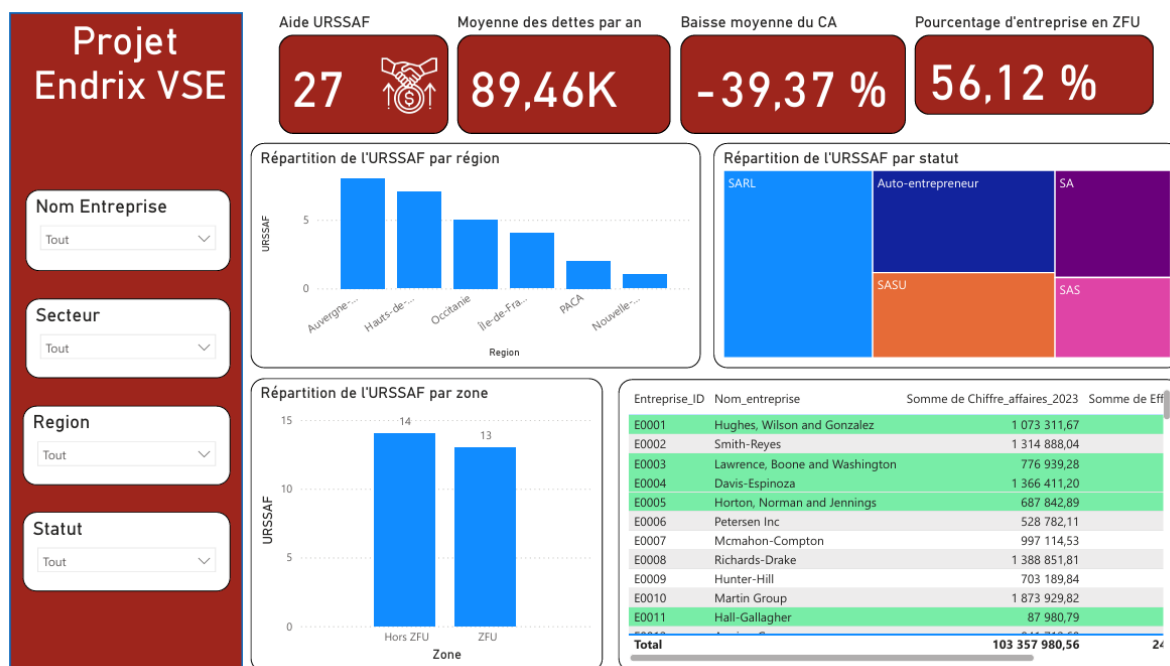


Figure 16 : Page "URSSAF"