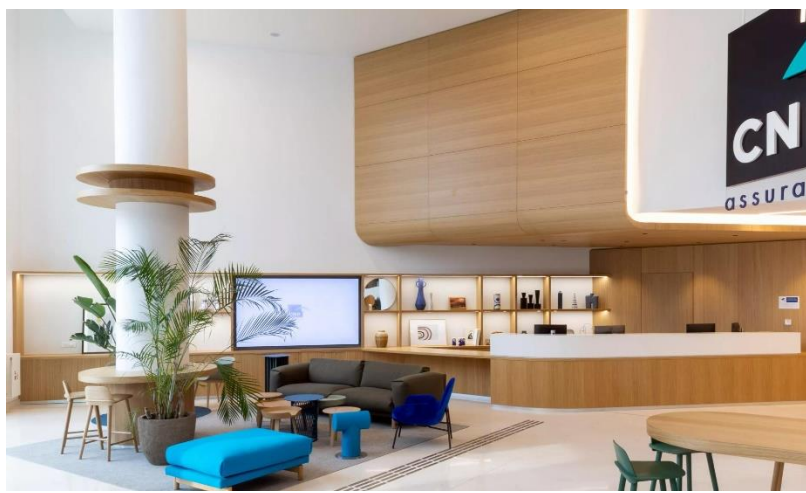


Rapport de stage

Traitement automatisé des données budgétaires avec
Python et SAS – Mission au sein de la CNP Assurances

Année 2024 – 2025
Kinaza ALI

2^{ème} année en Bachelor universitaire de technologie (BUT) Science des données à l' IUT Paris Rives de Seine –
Université Paris Cité – Parcours Exploration et Modélisation Statistique (EMS)



Tuteur : M. Guillaume THOMAIN

Maître de stage : M. Jean-Philippe KIENNER

Remerciements

Je tiens tout d'abord à remercier chaleureusement mon tuteur de stage, **Guillaume Thomain**, pour sa disponibilité, sa patience et sa pédagogie. Toujours à l'écoute, il a su m'accompagner tout au long de ma mission avec bienveillance, tout en m'apportant de solides compétences techniques, notamment en **SAS**, qu'il avait déjà su transmettre brillamment lors du **troisième semestre de formation en tant que professeur**. Son encadrement a grandement contribué à la réussite de ce stage.

Je remercie également toute **l'équipe UEX**, pour leur accueil chaleureux et leur bienveillance. Grâce à eux, je me suis sentie à l'aise dès les premiers jours, ce qui m'a permis de m'intégrer rapidement dans un environnement de travail stimulant et convivial. Un merci tout particulier à **Olivier R.** et **Nicolas G.**, les responsables de l'équipe, pour leur confiance et leur disponibilité.

Je tiens aussi à exprimer ma reconnaissance à **Monsieur Kienner**, mon maître de stage, pour avoir veillé au bon déroulement de cette expérience et s'être assuré que les conditions de mon intégration et de mon apprentissage soient optimales.

Enfin, je remercie **Hema**, stagiaire présente lors de mes débuts, pour ses conseils, son soutien et son partage d'expérience. Son accompagnement m'a permis de mieux comprendre le fonctionnement de l'entreprise et de m'y sentir pleinement intégrée.

À toutes ces personnes, je souhaite adresser un grand merci pour avoir rendu cette expérience aussi formatrice qu'enrichissante.

SOMMAIRE

Introduction	6
1 . Présentation générale de l'entreprise	7
1.1 Présentation du service UEX 1	10
2. Cartographie de l'architecture des données	12
2.2. ... Présentation des outils informatiques et statistiques (environnement technique)	14
2.2 Présentation des bases de données	16
2.2.2.2 Description de toutes les difficultés ou points de vigilances soulevées par les données bruts.....	18
3. Présentation de la mission	20
3.3 Contexte	20
3.3.3 Enjeux.....	21
3.3.3.3 Objectifs de la mission	21
Annexe	37
 Démarche portfolio Analyse du stage au regard des compétences du BUT Science des données	 40

Introduction

Du 28 avril au 4 juillet 2025, j'ai effectué mon stage de deuxième année de Bachelor Universitaire de Technologie (BUT) en Science des Données au sein de la **CNP Assurances**, à Issy-les-Moulineaux. Ce stage orienté *data* m'a offert une première immersion dans le monde professionnel et m'a permis de mieux appréhender les enjeux réels liés à la gestion, à l'exploitation et à l'automatisation des traitements des données dans une grande entreprise.

Il a constitué une opportunité concrète de mobiliser les compétences théoriques, techniques et analytiques acquises au cours de ma formation, en les appliquant à un projet réel à fort enjeu métier.

Ma mission principale porte sur la comparaison entre un script **SAS** existant, utilisé en production, et un nouveau programme à développer en **Python**. L'objectif est d'évaluer la solution la plus optimale pour automatiser le traitement et la structuration de fichiers budgétaires extraits depuis l'outil Anaplan.

Ce travail s'inscrit dans une réflexion plus large, que l'on peut résumer à travers la problématique suivante : **Comment optimiser le traitement automatisé des données budgétaires à la CNP Assurances en comparant les atouts et limites de Python et SAS dans un environnement métier contraint par la sécurité, la performance et la maintenabilité ?**

1 . Présentation générale de l'entreprise

Fondée en 1959, CNP Assurances (Caisse Nationale de Prévoyance) est issue de la fusion de plusieurs caisses d'assurance publiques. Initialement établissement public, elle devient société anonyme en 1992. En 2022, elle est entièrement intégrée au groupe La Banque Postale, devenant **ainsi un acteur central du pôle financier public français**. Aujourd'hui, la CNP Assurances est un leader de l'assurance de personnes en France, le 7^e assureur européen et le 3^e au Brésil.

Figure n°1 : La CNP en chiffre



CNP Assurances opère dans le domaine de l'assurance de personnes, offrant une gamme complète de produits.

Prévoyance

Le groupe joue un rôle majeur dans la protection sociale complémentaire des particuliers et des professionnels.

**Couverture de prêts**

Accompagnement des projets des clients :
Prêts immobiliers /Crédits à la consommation/ Prêts professionnels

**Santé et prévention**

CNP Assurances couvre les dépenses et frais de santé.

**Assurance vie**

CNP Assurances est un acteur de référence de l'assurance vie en France

**Retraite**

Le groupe accompagne plus d'un million d'assurés dans leur préparation à la retraite et dans le service de leur rente.

**Caution / Protection des biens**

CNP Caution propose des solutions en matière de cautionnement des prêts immobiliers sur le marché des particuliers.



La CNP Assurances adopte un **modèle multi-partenarial**, collaborant avec divers acteurs pour distribuer ses produits avec des banques comme la Banque Postale (actionnaire à 100 %), Groupe BPCE (Banque Populaire, Caisse d'Épargne), Crédit Agricole, BNP Paribas. Mais également des **mutuelles et institutions de prévoyance**, des **courtiers et conseillers en gestion de patrimoine**, des **Collectivités locales et association** et des **Start-up** via le programme Open CNP.

Ce modèle permet à CNP Assurances de bénéficier de canaux de distribution variés et complémentaires pour ses produits d'assurance.

Sur le marché français, les principaux concurrents de CNP Assurances incluent le **Crédit Agricole Assurances**, **AXA France**, **BNP Paribas Cardif**, **Generali France**, **Groupama** et **Allianz France**.

Ces acteurs se disputent les parts de marché dans les domaines de l'assurance vie, de la prévoyance et de l'assurance emprunteur.

Le groupe CNP Assurances mène diverses missions et projets. Parmi celles- ci se trouvent :

- **Le développement de produits d'assurance** adaptés aux besoins des clients et aux évolutions sociétales.
- **La digitalisation des services** avec des mise en place de processus de souscription en ligne, signature électronique, etc.
- **Un engagement en faveur de la finance durable**, en intégrant les critères ESG (environnementaux, sociaux et de gouvernance) dans leurs investissements .
- **Des partenariats internationaux** à titre d'exemple, avec Santander pour la distribution de produits d'assurance emprunteur en Europe.
- **La création de filiales spécialisées**, comme CNP Assurances Protection Sociale, en partenariat avec La Mutuelle Générale, pour renforcer l'offre en santé et prévoyance

1.1 Présentation du service UEX 1

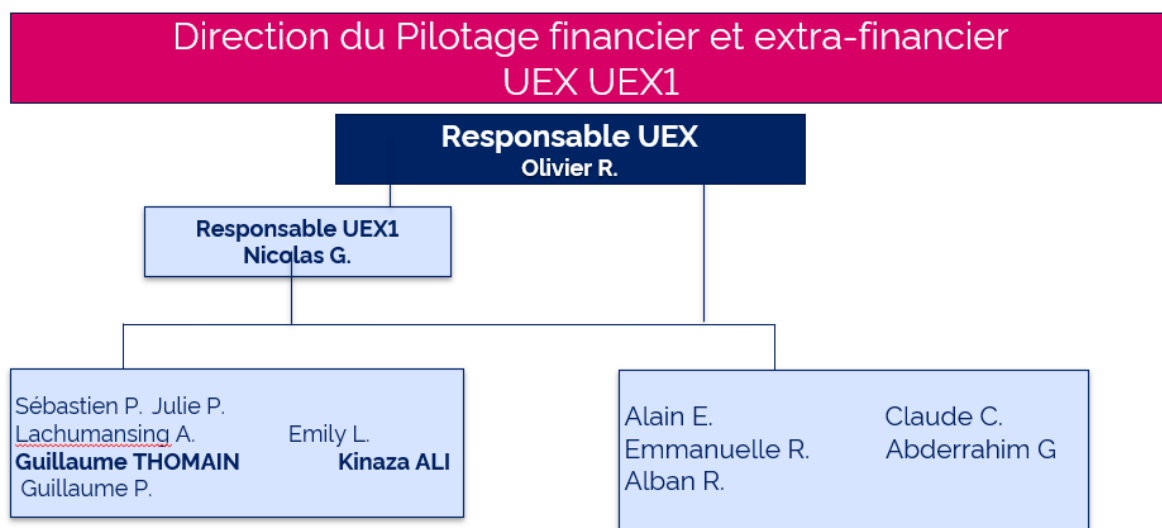
Le service dans lequel j'ai réalisé mon stage est le service **UEX1**, une entité rattachée au service **UEX**, lui-même appartenant au **Centre d'expertise Support Outils & Data**. UEX1 est la branche orientée **décisionnel et data** : elle exploite les données nécessaires à la **prise de décision**.

De son côté, le service UEX collabore étroitement avec les équipes du **contrôle de gestion** et de l'**actuariat**, afin d'assurer un pilotage des données à la fois **en temps réel** et **prévisionnel**. Il a pour mission de répartir les frais et de modéliser les coûts pour les différentes **BU** (Business Units ou unités commerciales) de la **CNP Assurances**.

Le service UEX (cf. fig. 2) fait partie de la **Direction de la Transformation Opérationnelle (DTO)**, dont les principales missions sont de renforcer l'**efficacité opérationnelle** et de superviser le **déploiement technique** au sein du groupe. La DTO regroupe plusieurs services spécialisés.

Plus globalement, **UEX** et la **DTO** sont rattachés à la **Direction du Pilotage Financier et Extra-Financier** (cf. fig.4).

Figure n°2 : Organigramme du service UEX

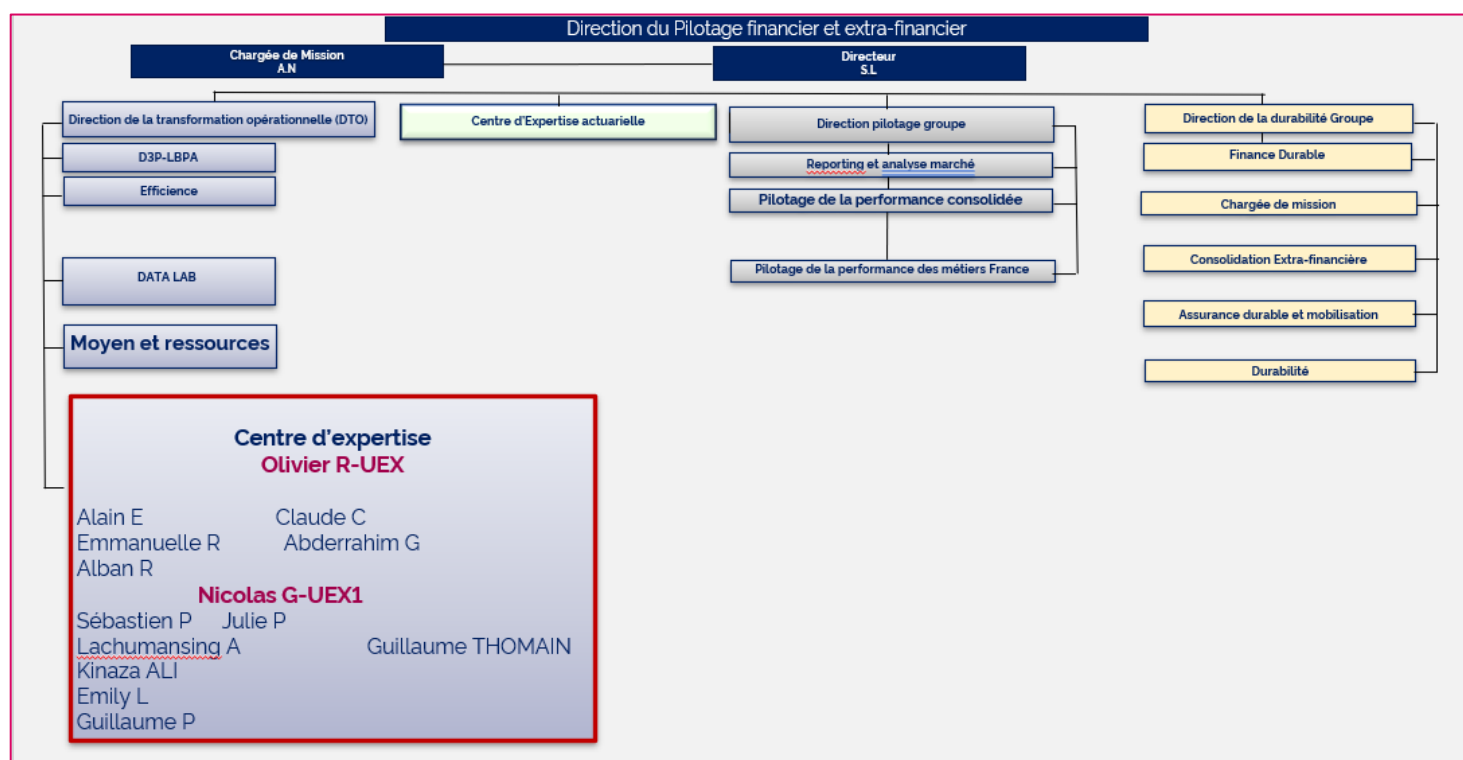


UEX1 reste étroitement lié au reste du service et constitue une composante intégrante de **UEX**. Il s'agit bien d'un seul et même service, dont la mission est de mettre en place des données **fiables** et **exploitées** par les services **comptables** et de **contrôle de gestion** de **CNP Assurances**. UEX1 représente ainsi la **brique technique et préliminaire**, fournissant des données **conformes** et **structurées** aux autres équipes.

Les outils utilisés par le service UEX incluent : **Anaplan**, **Tableau**, **Excel**, **SAS**, **PyCharm**, **SQL Server**, **Knime**, et **BusinessObjects (BO)**.

Par ailleurs, l'équipe UEX organise des **réunions quotidiennes** réunissant l'ensemble du service. Ces réunions permettent de partager les **directives du moment**, de suivre les **évolutions en cours**, et de maintenir une **communication fluide**. Cela illustre bien l'**aspect transverse** du service, où chaque membre, qu'il appartienne à UEX ou UEX1, collabore étroitement au quotidien.

Figure n°3 : Organigramme de la direction du pilotage financier et extra-financier



2. Cartographie de l'architecture des données

Durant le stage, ma mission principale consiste à **réécrire en Python un programme qui existe initialement en langage SAS**. Ce programme a pour objectif de **trier et structurer des données**.

Les données à ma disposition sont réparties en **trois grandes catégories**, correspondant à **trois bases de données distinctes** :

- les **engagements**,
- les **factures**,
- les **factures provisionnées**.

Ces données sont saisies par les **services de contrôle de gestion** directement dans **Anaplan** (progiciel de gestion des données du contrôle de gestion). Chaque **unité commerciale** y entre manuellement différentes **dépenses**, en les rattachant aux **postes budgétaires** qui les concernent. Un poste budgétaire regroupe les dépenses relevant d'une même **famille de biens ou de services**.

En plus des saisies manuelles, **Anaplan** intègre également des données issues de **fichiers plats** ou de **fichiers Excel**, ce qui permet de centraliser les informations provenant de différentes sources.

Le **service de contrôle de gestion** saisit au fur et à mesure les informations dans le système. Ensuite, une **consolidation des données** est effectuée par **Anaplan**, qui se charge de générer des **indicateurs** à partir des données enregistrées. Il y a une agrégation et structuration des données saisies.

Cette consolidation permet un **pilotage de l'activité** en fonction des résultats obtenus pour chaque **poste budgétaire** et pour chaque **unité commerciale**.

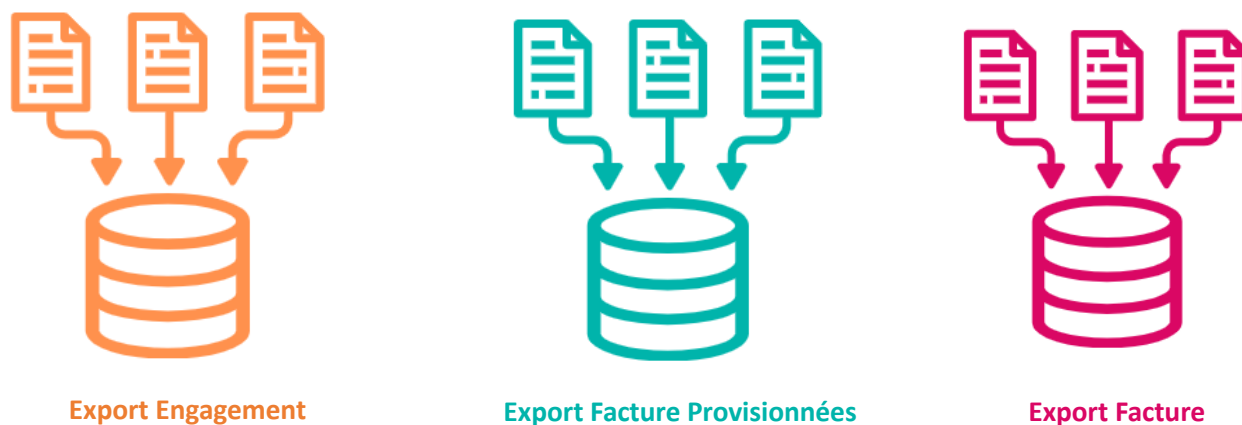
Après cette étape, les données sont **réutilisées** par une partie du service de **contrôle de gestion** ainsi que par les services de **comptabilité**, afin d'en extraire les **informations nécessaires à la gestion efficace des budgets**. Ces données permettent notamment de suivre, **en temps réel**, les montants dépensés par poste budgétaire au cours de l'année, de vérifier si les dépenses sont **maîtrisées** ou **excessives** en fonction de la période, et ainsi de mesurer la **performance budgétaire à un instant T**.

Les données saisies initialement sont qualifiées de **données brutes**, c'est-à-dire qu'elles n'ont subi **aucune transformation** depuis leur enregistrement. Une fois intégrées dans **Anaplan**, une **mécanique de traitement** permet de générer des **indicateurs métiers** adaptés aux problématiques du contrôle de gestion, indispensables pour le **pilotage opérationnel** de l'activité.

La problématique étant : **Comment concilier exigences techniques (fiabilité du traitement des données, compatibilité avec Anaplan) et contraintes métiers (besoin de lisibilité, fiabilité des indicateurs) dans un projet de refonte du programme SAS existant ?**

Figure n°4 : Schéma de l'origine des données des 3 fichiers manipulés durant la mission.

Etape 1 : Saisie et centralisation de diverse données vers Anaplan par les contrôleurs de gestion



Les données issues d'Anaplan sont exportées sous format Excel en 3 fichiers distincts

Ces bases de données constituent la matière première utilisée pour le stage. Le processus décrit précédemment correspond à la manière dont ces données sont mises à disposition, mais **les étapes de saisie et de consolidation ne font pas partie de ma mission.**

Je me retrouve donc avec **trois fichiers de données** regroupant des informations issues d'Anaplan, mais **non réparties** selon les différents **organismes** et/ou **centres budgétaires** de la CNP Assurances.

Autrement dit, avant que les équipes de contrôle de gestion puissent analyser ces données, il est **nécessaire de les trier et de les classer** en fonction des centres concernés.

2.2. Présentation des outils informatiques et statistiques (environnement technique)

Pour réaliser ma mission, dont les détails seront présentés plus loin, il est nécessaire de disposer d'un certain nombre d'outils informatiques adaptés et optimaux pour assurer le bon déroulement du travail.

Dans un premier temps, j'ai à ma disposition un ordinateur professionnel ainsi qu'un **poste personnel sécurisé** rattaché à l'environnement informatique de l'entreprise. La CNP Assurances étant un groupe particulièrement sensibilisé à la **sécurité des données** et à la **cybersécurité**, de nombreuses restrictions et contrôles sont appliqués aux postes utilisateurs, en raison de la manipulation de **données à caractère sensible**.

Ces données sont qualifiées de sensibles au sens du **Règlement Général sur la Protection des Données (RGPD)**. Il s'agit de données personnelles appartenant à une catégorie particulière, dont la divulgation pourrait porter atteinte de manière irréversible à une personne. Elles sont également **confidentielles**, et doivent le rester pour garantir le bon fonctionnement de l'activité. Leur diffusion publique pourrait engendrer des risques importants pour l'entreprise.

L'ordinateur mis à ma disposition est ainsi équipé d'un **VPN (Virtual Private Network)**. Le VPN chiffre les données et masque l'adresse IP, garantissant une connexion sécurisée, même hors du réseau Wi-Fi de l'entreprise. Celui utilisé à la CNP Assurances s'appelle **Systancia**, et il permet de maintenir un réseau privé sécurisé. **Sans ce VPN, il est impossible d'accéder à mon poste à distance.**

SAS Entreprise Guide :



Le développement de mon programme **s'appuie** sur un programme initial conçu sous SAS. J'ai donc dû demander l'accès à SAS Enterprise Guide sur mon poste dans le cadre de mon stage.

SAS n'est pas un logiciel libre (open source), son utilisation nécessite **l'achat d'une licence**. Une licence SAS permet d'utiliser les différentes solutions proposées par l'éditeur, comme SAS Enterprise Guide, selon des conditions spécifiques (durée d'abonnement, nombre d'utilisateurs autorisés, droits d'accès et fonctionnalités disponibles).

Ainsi, pour pouvoir accéder à SAS sur mon poste, j'ai dû solliciter un droit d'accès auprès des personnes habilitées au sein de la CNP Assurances. Cette démarche m'a amenée à échanger avec plusieurs services, ce qui m'a permis de découvrir **l'importance de la collaboration interservices** et de la **gestion des droits d'accès dans un environnement sécurisé**.

Pycharm :



PyCharm est un environnement de développement intégré (IDE) dédié à la programmation en **Python**. La majeure partie de ma mission a été réalisée en langage Python, à l'aide de cet environnement.

C'est l'occasion pour moi de **découvrir le fonctionnement de l'installation des packages Python**, étape essentielle dans la mise en place d'un environnement de travail. Cependant, en raison des **mesures de sécurité strictes mises en place par la CNP Assurances**, j'ai rapidement rencontré un obstacle. En effet, l'entreprise utilise un **proxy**, qui permet de filtrer les connexions Internet afin de protéger la vie privée, sécuriser les données et appliquer les politiques de sécurité informatique. Ce système de proxy bloque l'accès à certains sites, dont **PyPI** (Python Package Index), qui est la bibliothèque officielle permettant d'installer des modules avec la commande **pip install**.

Cette restriction a provoqué des erreurs lors de l'installation des packages nécessaires à ma mission. Après avoir sollicité l'**équipe informatique**, j'ai appris qu'il était possible de **créer un environnement virtuel préconfiguré**, contenant déjà les packages requis. Il suffisait ensuite, dans PyCharm, de **remplacer l'interpréteur Python par défaut** par celui de cet environnement spécifique.

Cette solution, qui contourne les blocages liés au proxy en **évitant les téléchargements directs sur Internet**, m'a permis de débloquer la situation.

Ce problème, survenu au début de mon stage, a été particulièrement formateur. Il m'a sensibilisée aux **enjeux de sécurité informatique**, à l'importance de **la collaboration avec les services techniques**, et m'a permis de **mieux comprendre le fonctionnement des environnements virtuels et de la gestion des packages en Python**.

Excel :



Le dernier logiciel que j'utilise est **Excel**. En effet, les **instructions** concernant mon programme ainsi que le format des trois fichiers exportés et des fichiers générés en sortie de mon programme sont au format Excel.

Je travaille donc régulièrement sur Excel afin de **vérifier le bon fonctionnement de mon programme** et de contrôler la qualité des données produites.

En résumé concernant la partie technique :



La base sur laquelle je dois m'appuyer et avec laquelle je compare mes résultats.



Le langage de développement de mon programme, et mon principal outil de travail.



Contient mes fichiers sources et sert à vérifier la bonne réalisation de mon programme.

2.2.2. Présentation des bases de données

Les bases de données à ma disposition (fichiers export) sont au format **xlsx** et **évoluent quotidiennement** en nombre d'individus, c'est-à-dire en nombre de lignes. Cette évolution est due aux nouvelles saisies effectuées chaque jour par le service contrôle de gestion.

- **Fichiers des données brutes export**

Les trois fichiers « Export », issus d'Anaplan, constituent les données initiales à la réalisation de ma mission.

Les bases de données **Drill « Facture »** et **Drill « Facture provisionnée »** sont composées chacune de 30 variables. Ces colonnes sont les suivantes :

N° Facture e-Achat	Centre OCEAN	Libellé Fournisseur
N° Facture	Type de Frais	Code Projet OCEAN
N° Commande	Rubrique	Libellé Projet OCEAN
N° Pièce	Compte BI	Année Comptable
Entité AP	Compte OCEAN	Date Facture
Direction	Montant HT	Date Saisie
Département	Code TVA	Date Paiement
Centre BI	Montant TTC	Période
Libellé Centre BI	Description Facture	Nature de Compte
Année	Code Fournisseur	Utilisateur

Les variables surlignées correspondent aux attributs numériques. Toutes ces variables permettent de situer la nature des dépenses, le moment où elles ont été engagées, leur montant, ainsi que leur justification.

Les deux fichiers « Facture » partagent donc une structure de données identique, même si l'axe d'analyse peut différer selon les fichiers, en fonction des missions des contrôleurs de gestion ou de la comptabilité.

Quant à la dernière base de données, **Drill « Engagement »**, elle est composée de 27 variables, qui sont les suivants :

Société Juridique	Enveloppe	Code Projet OCEAN
Direction	Rubrique	Libellé Projet OCEAN
Département	Libellé Compte BI	Description Commande
Centre BI	Compte OCEAN	Exercice
Libellé Centre BI	Libellé Compte OCEAN	Période
Centre OCEAN	Montant engagé TTC (euros)	Date Echéance Commande
N° Commande	Montant Facturé	Statut Commande
Code de la DA	Solde Engagement	Date de Création de la DA
Type de Frais	Libellé Fournisseur	

- Fichiers créés à la sortie du programme**

Sans entrer trop dans les détails, étant donné que ces fichiers seront présentés dans la partie consacrée à **l'explication de la mission**, les fichiers issus du programme sont **également mis à jour quotidiennement**. Cela signifie que d'un jour à l'autre, le nombre de lignes peut varier. En revanche, un changement est visible concernant le nombre d'attributs en sortie ainsi que leurs intitulés.

Les bases de données Excel relatives aux fichiers « Facture » et « Facture provisionnée » sont désormais composées de 26 attributs, portant les noms suivants :

Cen Bud BI	Libellé Compte BI	Libellé Axe
Libellé Centre BI	Compte OCN	Année comptable
Centre OCN	Montant HT	Date Facture
N° pièce	Code TVA	Date Saisie
N° fact e-Achat	Montant TTC	Date Paiement
Nature charges	Description fact	Période
N° Commande	Code Fsieur	Nature de compte
N° fact	Libellé Fournisseur	Utilisateur
Rubrique / Nature	Axe OCN	

Les fichier « Engagement » sont maintenant composés de 18 attributs avec les nominations suivantes :

Centre BI	Cpt Com	Axe Pro
Libellé centre BI	Lib Cpt Com	Lib Axe Pro
Centre OCEAN	<u>Mont TTC Réparti</u>	Description Cde
Num Cde	<u>Mont Facturé</u>	Exercice
Rubrique	<u>Sold Eng</u>	Période
Libellé compte BI	Libellé Fournisseur	Date Ech Cmde

Ces fichiers sont classés selon des répertoires relatifs aux différents centres budgétaires, ce qui permet de mettre en place le pilotage des activités de chaque centre.

2.2.2.2 Description de toutes les difficultés ou points de vigilances soulevées par les données bruts

Au cours de la gestion des données, quelques anomalies sont identifiées en lien avec le programme utilisé. En effet, ce dernier requiert, en plus du tri des données par centre, un reformatage spécifique de certaines colonnes. De plus, des suppressions de colonnes et des sélections d'éléments selon les besoins imposent un formatage particulier.

J'ai pu remarquer que, pour deux colonnes, **les espaces insérés lors du reformatage n'avaient pas été pris en compte dans le programme mis en production**, entraînant la présence d'espaces superflus en début de chaque cellule dans les fichiers triés. Bien que ces détails puissent paraître **mineurs**, leur identification est une **plus-value** afin de garantir la qualité des données en sortie.

Par ailleurs, plusieurs **difficultés techniques** sont rencontrées en relation avec la manipulation des données issues d'Anaplan :

- Les données exportées depuis Anaplan vers Excel peuvent contenir des caractères particuliers, qui peuvent générer une modification de certains libellés.
- Il est indispensable de vérifier que les données respectent les formats requis, notamment via des outils de gestion des données (Data Management), afin d'assurer la cohérence des traitements ultérieurs.
- Un problème récurrent concerne le format des dates, qui n'étaient pas toujours correctement formatées dans les fichiers Excel en sortie.

Ces observations soulignent l'importance d'une vigilance accrue dans la manipulation et le contrôle des données tout au long du processus afin d'assurer la robustesse et la fiabilité des résultats.

Une autre difficulté majeure rencontrée concerne le déploiement et l'exécution automatique quotidienne du programme sur **le serveur de production**. En effet, afin d'assurer la mise à jour régulière des données, le service informatique place le programme sur un serveur de production, qui présente des contraintes techniques plus strictes, notamment **des contraintes spécifiques sur le nom des variables** en SAS.

De ce fait, une fois le programme finalisé, **une phase de collaboration étroite** avec le service informatique est nécessaire pour assurer son bon fonctionnement sur ce serveur.

Cette contrainte impose **une vigilance particulière** lors du développement et des tests, ainsi qu'une communication efficace avec le service informatique, afin de garantir la fiabilité et la pérennité du processus automatisé.

3. Présentation de la mission

3.3 Contexte

Dans la perspective d'une évolution future des outils internes, mon stage s'inscrit plus précisément dans un projet de **traduction d'un programme existant initialement développé en SAS vers le langage Python**. L'objectif est de **comparer les deux langages** en termes de simplicité de développement, de lisibilité, de performance (temps d'exécution), mais aussi en termes de **coûts**, Python étant un langage **open source** ne nécessitant pas de licence payante contrairement à SAS. Mon travail porte ainsi à la fois sur :

- **La reproduction fidèle des traitements existants,**
- **L'optimisation du nouveau script Python,**
- **Et l'analyse comparative** des deux approches.

L'objectif principal de mon programme est de **générer automatiquement, quotidiennement, des fichiers Excel à partir des fichiers « Export »**, en les **répartissant selon les règles des centres budgétaires**.

Ce traitement permet de **créer une arborescence de dossiers**, chacun étant **associé à une entité spécifique du groupe CNP Assurances**.

En effet, CNP Assurances étant un groupe structuré autour de plusieurs entités, l'accès aux données doit être **restreint et sécurisé** en fonction des droits d'accès de chaque structure.

Pour illustration, les données relatives au centre **BPCE** sont automatiquement stockées dans un **dossier nommé "BPCE"**, contenant uniquement les trois fichiers correspondant à cette entité : **"Engagement"**, **"Facture"** et **"Facture approvisionnée"**, chacun filtré pour ne contenir **que les informations propres à BPCE**.

Ce fonctionnement répond à un **impératif de confidentialité et de protection des données**, en garantissant que chaque entité ne puisse accéder qu'aux informations qui les concernent.

Ce programme, une fois finalisé, est **déployé sur un serveur de production** par le service informatique, afin qu'il **s'exécute automatiquement chaque jour**. Cette automatisation permet de **fiabiliser le processus** tout en **gagnant du temps** pour les équipes de contrôle de gestion, qui disposent ainsi **quotidiennement de fichiers prêts à l'emploi**.

3.3.3 Enjeux

Le stage s'inscrit dans une logique d'optimisation des processus de traitement des données budgétaires au sein de CNP Assurances. Les principaux enjeux identifiés sont les suivants :

- **La fiabilisation des données** : garantir l'exactitude, la complétude et la cohérence des informations issues des exports quotidiens de la plateforme Anaplan. Cela passe par un nettoyage rigoureux, un formatage homogène et une structuration adaptée des fichiers.
- **Le gain de temps** : réduire significativement le temps passé par les équipes de contrôle de gestion à retraiter manuellement les fichiers. Le programme développé permet d'automatiser des tâches répétitives, jusque-là chronophages.
- **Une meilleure traçabilité** : produire des fichiers lisibles, homogènes et bien organisés, facilitant leur exploitation par les différents acteurs métiers (contrôle de gestion, comptabilité). Chaque fichier est structuré selon des règles claires et récurrentes.
- **De l'automatisation** : industrialiser le traitement des données en déployant le programme sur un serveur de production. Ce dernier permet une exécution automatique et quotidienne, sans intervention manuelle, tout en respectant les règles d'accès et de confidentialité propres aux différentes entités du groupe.
- **La comparaison des langages** : dans le cadre d'une potentielle évolution technologique, ma mission de stage vise également à comparer un programme existant développé en SAS avec sa réécriture en Python. L'objectif est d'évaluer la simplicité de développement, la lisibilité du code, les performances d'exécution, ainsi que les avantages économiques (Python étant un langage open source) pour orienter les choix futurs du service.

3.3.3.3 Objectifs de la mission

Le programme développé en Python devait permettre de lire, nettoyer, structurer et trier automatiquement trois types de fichiers exportés quotidiennement :

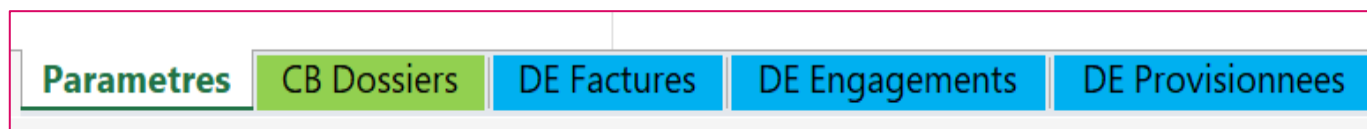
- les **engagements**,
- les **factures**,
- les **factures approvisionnées**.

L'automatisation vise à **réduire le temps d'intervention manuel** pour les équipes de contrôle de gestion, à **garantir la conformité des formats**, et à **préserver la confidentialité** des données en les répartissant dans des dossiers distincts selon les entités concernées.

Paramétrage à partir d'un fichier « paramètre » annexe

Pour garantir la flexibilité du programme et éviter le recours à des modifications répétées dans le code, il est décidé d'externaliser dans un fichier Excel les différentes **informations de paramétrage** nommé "**Paramètre CB des requêtes structures**". Ce fichier joue un rôle crucial dans le fonctionnement du programme. (cf figure 5 et 8)

Figure n°5 : Onglets présents dans le fichier "Paramètre CB des requêtes structures"



Dans l'onglet paramètre :

- L'**année de référence** à étudier, utile pour filtrer ou renommer les fichiers de sortie ;
- Le et **chemins de sortie** (où stocker les résultats après traitement) ;
- Le chemin des fichiers sources Export
- Le nom du fichier paramètre

Figure n°6 : Aperçue de l'onglet Parametres

Parametre	Valeur
AnneeFichier	2025
CheminFichiersSourcesANAPLAN	\\cyan\5-Partages Bureautiques\FG SIMBAD\04 Prod\02 Anaplan\01 Modèle des frais\Drills FG SIMBAD ECHANGE\Sources
CheminFichierParametres	\\cyan\5-Partages Bureautiques\FG SIMBAD\04 Prod\02 Anaplan\01 Modèle des frais\Drills FG SIMBAD ECHANGE\Parametres
NomFichierParametres	Paramètres CB des requêtes structures
CheminSorties	P:\Test_sorti_Drill

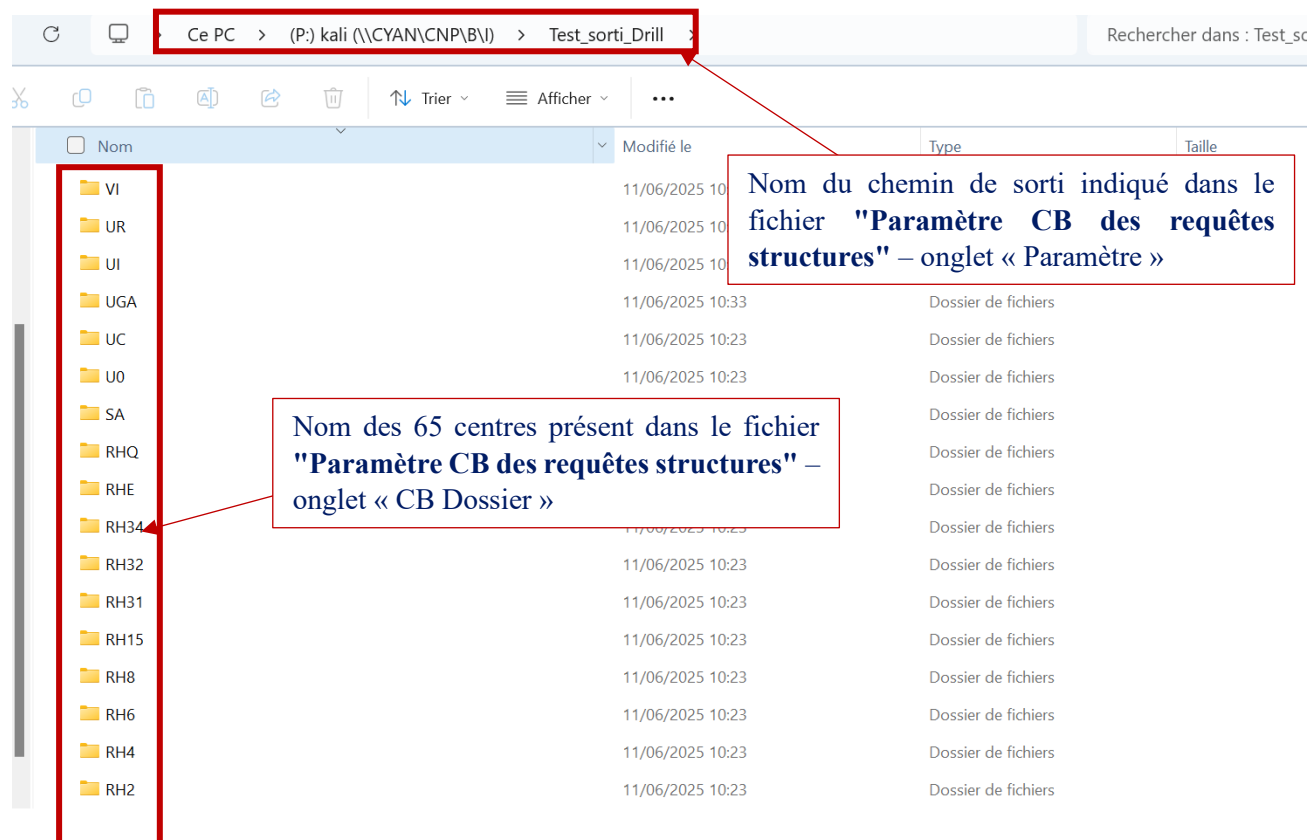
Dans l'onglet CB Dossiers

- La **liste des centres budgétaires** à prendre en compte et classer selon leur unités commerciales, avec leurs **intitulés exacts** à utiliser pour nommer les dossiers de sortie (ex. : BPCE, ACV, etc.). (cf figure 7 et 8)

Figure n°7 : Aperçue de l'onglet CB dossier

Dossier	Complement	Centres
A2180 - BC		A2180;BC
ACV		ACV
AI		AI
BD		BD0;BD1;BD2;BD3;BD4
BF		BFA1;BFA3;BFRI;BFICV1;BFICV2;BFICV3;BFICVT;BFL1;BFL2;BFL3;BFL4;BFL6;BFP10;BFP11;BFPA;BFPB;BFPC4;BFPOT P;BFPQ;BFPQ0;BFPR;BFPS;BFPT;BFPU;BFPU0;BFPU1;BFPV;BFPV1;BFPW;BFPW1;BFPZ;BFR5;BFGARG
BPCE		BPCE0;BPCE1;BPCE5;BPCE6;BPCE7;BPCE8;BPCE90;BPCE91;BPCE92;BPCE93;BPCE94;BR0;BR2;BR3;BR4;BR5;BR6;E
BPCE_PA		BPCE1;BPCE5;BPCE6
BPCE_RC		BPCE7;BPCE8;BPCE90;BPCE91;BPCE92;BPCE93;BPCE94
BPCE7		BPCE7
BPCE8		BPCE8
BR		BR0;BR2;BR3;BR4;BR5;BR6
CCOP		CCOP
CCY		CCY0;CCY1;CCY2;CCY3;ID1
DCSU		DCSU
DECK		DECK0;DECK1;DECK20;DECK21;DECK22;DECK23;DECK24;DECK25;DECK26;DECK3;DECK4;DECK5;AIT1
DECSI		DECK0;DECK1;DECK20;DECK21;DECK22;DECK23;DECK24;DECK25;DECK26;DECK3;DECK4;DECK5;DPO;PIL0;PIL1;PIL OL0;CSOL1;CSOL2;CSOL3;CSOL4;CSOL6;CSOL7;CSOL8;CSOL9;POSU0;POSU1;POSU2;POSU3;POSU4;POSU5;POSU6
DECSI0		DPO
DECSI1		PIL0;PIL1;PIL2;PIL4;PIL6;PIL8;PIL9;PIL10
DECSI2		DECSI2
DECSI3		AIT0
DECSI8		CSOL5;SIBCP0;SIBCP1;SIBCP2;SIBCP3;SIBCP4;SIBCP5;SIBCP6;SIBCP7;SIBCP8;SIBCP9;PIL9
DECSI9		SIFRI0;SIFRI1;SIFRI2;SIFRI3;SIFRI4;SIFRI5;SIFRI6;SIFRI7
DECSI10		CSOL10;CSOL11;CSOL12;CSOL13;CSOL14;CSOL15;CSOL16;CSOL17;CSOL18;CSOL19

Figure 8 : Visualisation de l'étape de la création des dossier répertoires relatives à chaque unité commerciale



Dans les onglets DE...

Trois feuilles supplémentaires décrivent les **règles de structuration des fichiers en entrée et en sortie** pour chaque fichier (engagements, factures, factures provisionnées), et l'ordre des colonnes attendues en sortie, leur libellé exact, et les règles de transformations éventuelles. (cf figure 8)

Figure n°9 : Aperçue de l'un des onglets des exigences de mise en forme

A	B	C	D
Drill Factures-ANAPLAN	Excel	Ordre	Commentaire
N° Facture e-Achat	N° fact e-Achat	5	
N° Facture	N° fact	8	
N° Commande	N° Commande	7	
N° Pièce	N° pièce	4	
Entité AP			
Direction			
Département			
Centre BI	Cen Bud BI	1	
Libellé Centre BI	Libellé Centre BI	2	
Centre OCEAN	Centre OCN	3	Premier mot de la variable "Centre OCN" par rapport au signe -
Type de Frais	Nature charges	6	Deuxième mot de la variable "Nature charges" par rapport au signe -
Rubrique	Rubrique / Nature	9	Deuxième mot de la variable "Rubrique / Nature" par rapport au signe -
Compte BI	Libellé Compte BI	10	Deuxième mot de la variable "Libellé Compte BI" par rapport au signe -
Compte OCEAN	Compte OCN	11	
Montant HT	Montant HT	12	
Code TVA	Code TVA	13	
Montant TTC	Montant TTC	14	
Description Facture	Description fact	15	
Code Fournisseur	Code Fsseur	16	
Libellé Fournisseur	Libellé Fournisseur	17	
Code Projet OCEAN	Axe OCN	18	
Libellé Projet OCEAN	Libellé Projet OCEAN	19	

Ce système permet une exploitation et **maintenabilité accrue** du script pour des personnes étrangères au codage, en assurant un fonctionnement piloté par des paramètres centralisés et faciles d'accès.

- **Livrables**

Au terme de la mission, plusieurs livrables ont été produit :

Le programme Python finalisé et documenté : script fonctionnel mais avec des limites. Ce dernier est paramétrable, conçu pour être déployé dans un environnement de production. Il est accompagné de commentaire technique détaillant son fonctionnement.

Les fichiers Excel générés automatiquement : à partir des exports bruts Anaplan, les fichiers de sortie sont nettoyés, reformatés et classés dans des répertoires spécifiques à chaque entité budgétaire. Ces fichiers sont directement exploitables par les équipes de contrôle de gestion.

Une **fiche de suivi des anomalies** a également été tenue à jour tout au long du projet afin de tracer les erreurs détectées lors des tests en environnement de production. Cette fiche, bien qu'importante pour le bon déroulement de la mission, reste un document de travail interne et ne constitue pas un livrable officiel.

- **Calendrier prévisionnel (indicatif)**

Période	Tâches principales
Semaine 1-2	Prise en main des outils et compréhension des données
Semaine 3-4	Développement initial du script Python
Semaine 5-6	Tests, améliorations et ajustements selon les retours
Semaine 7	Collaboration avec le service informatique (mise en prod)
Semaine 8	Documentation et finalisation des livrables

Figure n°10 : Calendrier prévisionnel (indicatif)

J'ai établi au début de ma mission un calendrier prévisionnel ainsi qu'une fiche de suivi du programme afin d'assurer une progression structurée du projet, tout en intégrant une certaine flexibilité face aux imprévus techniques. Le travail s'est ainsi déroulé selon plusieurs phases successives, chacune associée à des objectifs précis.

La première phase a consisté en une appropriation du contexte fonctionnel et technique. Durant cette période, j'ai étudié l'existant, pris connaissance des fichiers sources et du fonctionnement de la plateforme de ce programme ainsi que des attentes exprimées. Cette étape a également permis de comprendre la structure des fichiers d'entrée et des règles de retraitement à appliquer.

Une fois cette base acquise, la phase de développement a pu commencer. Elle s'est étalée sur plusieurs semaines et a été ponctuée par des échanges avec les référents métier et le service informatique. Le script Python a été construit de manière itérative : lecture des données, nettoyage, classification, puis génération des fichiers de sortie. Des tests intermédiaires ont été menés à chaque étape, à partir de données réelles, afin d'identifier les dysfonctionnements éventuels et affiner le comportement du programme.

Enfin, les dernières semaines de la mission ont été consacrées à la stabilisation du programme, à la rédaction de la documentation technique et à la préparation du transfert vers l'environnement de production. Cette documentation devait permettre à d'autres collaborateurs de comprendre et maintenir le script en toute autonomie à l'avenir.

- **Cahier des charges (contraintes principales)**

Le développement du programme devait respecter plusieurs contraintes strictes définies en collaboration avec les équipes métier et informatiques. Les fichiers sources, transmis quotidiennement au format .xlsx, devaient être traités automatiquement sans intervention manuelle. Cette récurrence impliquait une rigueur particulière dans la structure du script, qui devait s'adapter à l'évolution potentielle des fichiers au fil du temps.

L'environnement d'exécution prévu étant un serveur de production, une attention particulière devait être portée à la casse. Le système étant beaucoup plus sensible aux majuscules et minuscules que les environnements de développement classiques. Ce contexte technique imposait une standardisation stricte des noms utilisés dans le programme SAS.

Par ailleurs, l'intégrité des données constituait un enjeu central : aucune perte d'information n'était tolérée. Les colonnes jugées critiques devaient être rigoureusement vérifiées à la fois en entrée et en sortie afin de garantir leur complétude, leur exactitude et leur format. Certaines d'entre elles nécessitaient également des traitements spécifiques, comme le renommage, la suppression ou le reformatage, selon des règles précises établies avec les utilisateurs métier.

Le script développé devait être totalement automatisé pour s'intégrer au processus de production, ce qui supposait une exécution sans interface manuelle et une capacité à fonctionner de manière autonome, y compris en cas d'erreur, avec une gestion robuste des exceptions.

Enfin, la nature des données manipulées étant confidentielle, notamment en raison de leur caractère budgétaire et organisationnel, leur traitement devait respecter les normes de sécurité en vigueur, en particulier celles liées à la protection des données personnelles (RGPD). Cela impliquait de prendre des précautions supplémentaires dans la structuration, l'accès et le stockage des fichiers produits.

- **Collaboration avec les équipes métier et techniques**

La réussite de cette mission a reposé en partie sur les échanges constants avec les différents interlocuteurs internes à CNP Assurances. En effet, le projet impliquait à la fois une bonne compréhension des besoins fonctionnels des utilisateurs finaux (les équipes de contrôle de gestion) et une coordination technique avec le service informatique pour l'accès à certains droits et plus tard la mise en production.

D'un côté, les échanges ont permis de clarifier les attentes autour du traitement des données. Ils ont apporté des précisions essentielles concernant le format des fichiers attendus, la logique de classement par centre budgétaire, ou encore les règles de nettoyage et de renommage à appliquer.

D'un autre côté, la collaboration avec les équipes IT s'est révélée indispensable pour garantir une exécution conforme.

Enfin, ces interactions transverses ont renforcé la dimension collaborative de mon stage. Elles m'ont permis de développer mes **compétences en communication technique**, en gestion de projet, ainsi qu'en adaptation aux contraintes spécifiques de l'environnement professionnel **d'une grande entreprise**.

- **Grandes étapes du programme**

Au moment de mon arrivée, un programme développé en langage SAS était déjà en place afin de créer automatiquement les différents dossiers contenant les trois fichiers exportés depuis la plateforme Anaplan. Chaque matin, un membre de l'équipe UEX a pour mission d'extraire les nouvelles données depuis Anaplan, de les convertir au bon format (Excel), puis de les déposer dans un dossier spécifique qui est le répertoire de téléchargement (cf. annexe ii-iii-iv).

Ce programme SAS, automatisé et intégré à l'environnement de production, se chargeait ensuite de trier les fichiers selon les centres budgétaires et de les enregistrer dans l'arborescence réseau de l'entreprise. Cette automatisation permettait une exécution quotidienne sans intervention manuelle, garantissant ainsi un traitement régulier et fiable des données.

L'exécution du script est prévue chaque jour ouvré, à l'exclusion des week-ends et des jours fériés. Cette restriction, intégrée directement dans le code, répond à une volonté de limiter la consommation énergétique inutile et s'inscrit dans une démarche plus large de sobriété numérique et de responsabilité environnementale. (Cf figure 10 et 11)

Figure 11 : Schéma de l'arborescence de la demande client

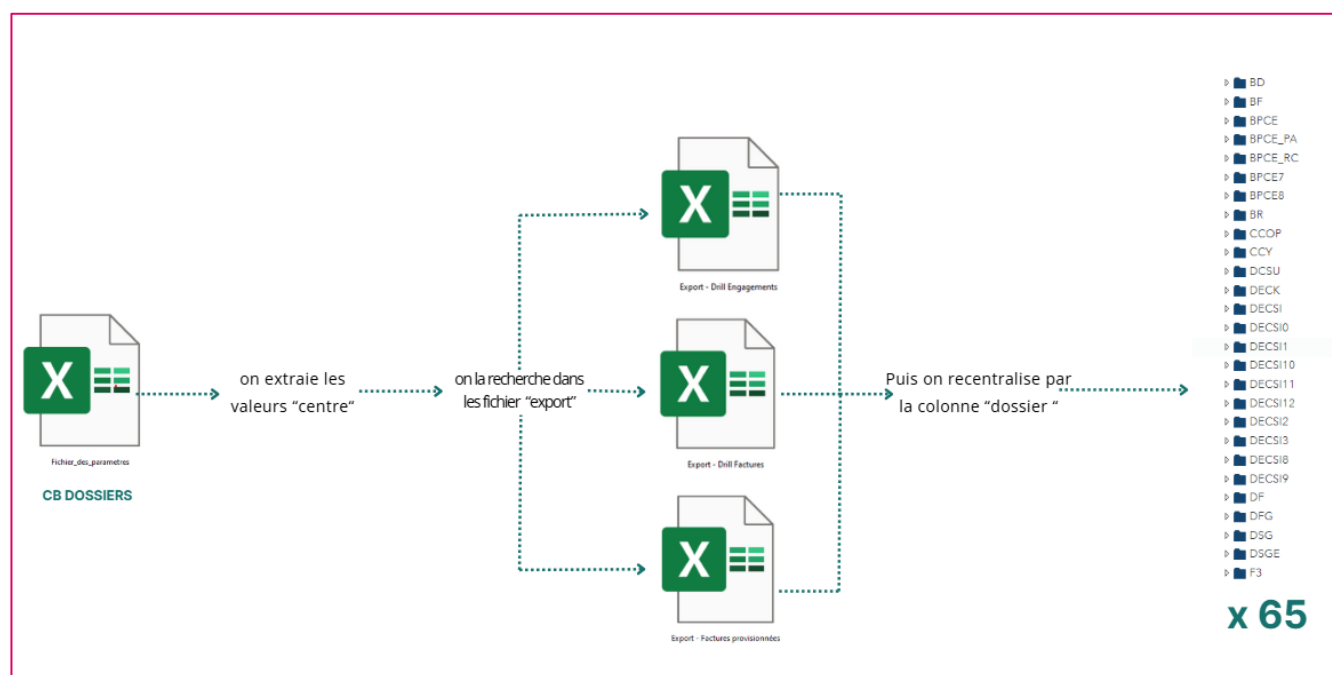
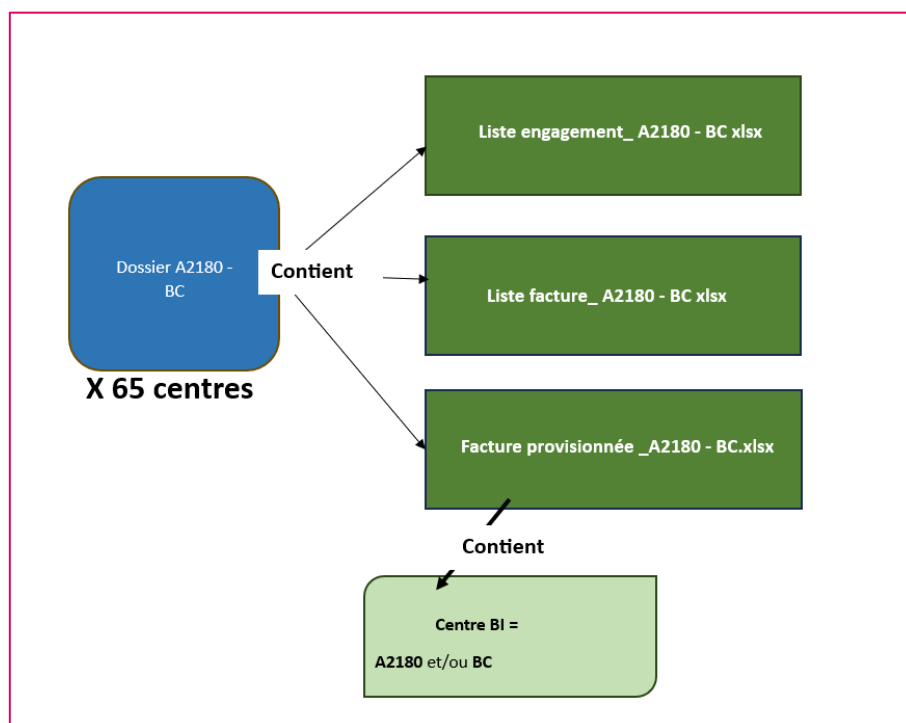


Figure 12 : Exemple de la composition d'un des fichier (centre A2180 – BC)



- **Limite de ma mission**

Certains obstacles techniques ont freiné la progression de la mission, principalement en lien avec l'environnement informatique sécurisé de l'entreprise.

Dans un premier temps, l'installation de certains packages Python s'est révélée complexe en raison des restrictions imposées par le proxy interne de la CNP. Des ajustements particuliers ont été nécessaires pour permettre au programme de fonctionner correctement dans cet environnement contraint.

Un autre point bloquant important a concerné l'envoi automatique d'un courriel à la fin de l'exécution du programme pour informer de la présence des nouveaux fichiers chaque jour aux personnes compétentes. Cette fonctionnalité, présente dans le programme SAS initial, ne pouvait pas être reproduite aussi simplement en Python. Après échanges avec le service informatique, il est apparu que l'envoi d'e-mails via un script Python nécessitait des droits spécifiques difficiles à obtenir, ou bien la mise en place d'une machine virtuelle dédiée. Ces solutions impliquent des démarches longues et fastidieuses, qui sont, au moment de la remise de ce rapport pas encore établies.

En conséquence, bien que le programme Python **remplisse pleinement les objectifs fonctionnels** attendus — tri, nettoyage, mise en forme, et génération des fichiers Excel structurés —, il n'a **pas encore pu être mis en production**. L'automatisation complète, notamment l'envoi de notifications par courriel, reste en suspens dans l'attente d'un éventuel déblocage des droits d'exécution nécessaires.

Dans ce contexte, et afin d'apporter une solution opérationnelle à court terme, le choix a été fait de revenir temporairement au programme SAS en place, tout en l'améliorant. En effet, lors de mes tests, **j'ai identifié** une anomalie liée à une mauvaise importation de certaines données, notamment des libellés récemment apparus dans les fichiers d'export. Ce dysfonctionnement, non détecté auparavant, m'a amenée à analyser en profondeur le fonctionnement du programme SAS, notamment la manière dont ce dernier interprète les caractères spéciaux à l'aide **de codes hexadécimaux**.

Cette étape a été particulièrement formatrice, car elle m'a permis de mieux comprendre les **mécanismes internes de traitement des données dans SAS**, ainsi que l'importance d'anticiper les évolutions des formats d'entrée. Le programme SAS a donc été affiné pour corriger ces erreurs, en attendant une solution plus pérenne avec le programme Python.

Ce genre de problème est dû au fait que les données utilisées lors de l'exécution sont des données mise sur le réseau CNP et non uniquement sur nos postes.

- **Tableau de comparaison des deux langages obtenus pour ma mission**

Critères	Programme SAS	Programme Python
Maintenance	Complexe, moins lisible, nécessite des compétences spécifiques en SAS	Plus lisible et modulaire ; plus facile à maintenir et à adapter
Lisibilité / clarté du code	Syntaxe spécifique, parfois difficile à lire ou adapter	Syntaxe claire, possibilité de commenter, découper et structurer le code
Automatisation	Déjà en production, planifié quotidiennement via le serveur SAS	Non encore en production, mais script prêt et automatisable
Envoi de mails	Fonctionnel et intégré dans la chaîne d'exécution	Bloqué pour des raisons de sécurité (droit restreint, proxy, etc.)
Flexibilité de traitement	Moins flexible (ex. : libellés mal gérés, encodage problématique)	Très flexible, meilleure gestion des chaînes de caractères, des formats, des règles
Temps d'exécution	Rapide, optimisé sur le serveur SAS	Très rapide également
Confidentialité / accès restreint	Respectée via le réseau interne	Respectée, à condition de bien paramétrer les chemins d'accès sécurisés

Figure 13 : Tableau de comparaison des deux langages

- **Perspective pour la mission**

À plus long terme, des évolutions importantes sont à envisager, notamment en lien avec les transformations techniques prévues au sein de l'environnement de production. En effet, d'ici la fin de l'année, le serveur actuel va migrer vers SAS Viya, une version modernisée de SAS qui intègre nativement le langage Python. Cette évolution ouvre la voie à de nouvelles possibilités hybrides, alliant la puissance des traitements SAS à la flexibilité de Python.

Dans cette perspective, il serait pertinent de réfléchir à la conception d'un programme transverse, combinant le meilleur des deux environnements. Une telle approche permettrait d'assurer le traitement principal des données en Python (notamment pour le nettoyage, le tri, et la mise en forme), un langage plus permissif et moins contraint sur les libellés, la casse et les structures de données. L'intégration avec SAS pourrait être conservée pour des tâches spécifiques, comme l'envoi des e-mails automatisés, qui reste plus facilement géré dans cet environnement sécurisé.

Un tel projet offrirait à la fois une meilleure robustesse, une maintenance facilitée, et une automatisation plus souple, tout en tirant parti des forces complémentaires de chacun des deux outils. Cette hybridation pourrait ainsi constituer une solution durable et évolutive pour répondre aux besoins du service dans un contexte de transition technologique.

- **Mise en lien de la mission avec ma formation**

Cette mission de stage m'a permis d'avoir une première immersion concrète dans les attentes et les exigences des entreprises du secteur de la data. J'ai compris que la première étape essentielle consiste à bien cerner la **demande du client** : il ne s'agit pas simplement de produire un programme, mais de s'assurer que la solution proposée est bien calibrée, ni trop complexe, ni trop simpliste. À travers les différentes demandes formulées par les clients internes du service UEX, j'ai pu observer l'importance pour le service technique — auquel j'étais rattachée — de bien **évaluer l'envergure réelle** d'une demande, même lorsqu'elle semble anodine au départ.

D'un point de vue plus technique, cette mission m'a permis de travailler pleinement sur la **programmation décisionnelle**, en mobilisant et renforçant mes compétences en **SAS** et en **Python**. J'ai appris à gérer des fichiers complexes, à les manipuler avec la bibliothèque **pandas**, et à faire des parallèles entre deux langages tout en tenant compte des **contraintes spécifiques à chaque environnement**. Par exemple, certaines règles de casse ou de format sont critiques dans un langage mais pas dans l'autre, ce qui m'a obligée à adapter mes méthodes et à faire preuve de rigueur.

Enfin, cette mission m'a appris à **assurer le suivi d'un code existant**, à **l'optimiser** et surtout à le **documenter de façon claire** afin qu'il puisse être repris facilement par un autre collaborateur ou intégré dans un processus automatisé. Ce travail de fond sur la fiabilité, la lisibilité et la maintenabilité du code correspond pleinement aux compétences que j'ai acquises dans ma formation, en particulier dans le cadre du **BUT Science des Données**.

- **Conclusion de la mission**

Cette mission m'a offert l'opportunité de mobiliser concrètement les compétences acquises au cours de ma formation, tout en découvrant les réalités du travail en environnement professionnel. Elle m'a permis de contribuer à un projet à la fois technique et structurant, en lien direct avec les besoins métiers et les enjeux de fiabilité des données (cf. annexe vi).

En développant un programme Python en parallèle du système SAS déjà en place, j'ai pu mesurer l'importance de la **polyvalence technologique**, de la **rigueur dans le traitement des données** et de la **collaboration avec les équipes métiers et informatiques**. Cette expérience m'a également sensibilisée aux contraintes propres à un environnement de production, et à l'importance de **proposer des solutions pérennes, documentées et adaptables**.

Au-delà des aspects purement techniques, cette mission m'a permis de prendre du recul sur mes pratiques, de comprendre les attentes des utilisateurs finaux, et d'envisager l'amélioration continue comme un processus indispensable dans tout projet data. Elle constitue une étape déterminante dans mon parcours, et renforce mon envie d'évoluer dans le domaine de la **science des données appliquée aux besoins concrets des organisations**.

- **Conclusion générale**

Ce stage au sein de **CNP Assurances** a été une expérience extrêmement enrichissante, à la fois sur les plans **intellectuel, technique et humain**. Il m'a permis de confronter mes compétences acquises durant mes deux années de formation en **BUT Science des Données** à la réalité du terrain, en travaillant sur un projet concret autour de l'automatisation du traitement de données budgétaires.

Au-delà des aspects purement techniques, ce stage m'a permis de découvrir **le fonctionnement d'une grande entreprise**, sa rigueur, son organisation, mais aussi sa dynamique humaine. J'ai eu la chance d'intégrer une équipe bienveillante et disponible. Travailler en équipe, échanger avec des collègues aux profils variés, participer à la vie de l'entreprise : autant d'expériences qui m'ont permis de gagner en maturité et en assurance.

J'ai également pu participer à des événements internes, comme un **webinaire** et un **séminaire en forêt**, qui ont montré à quel point l'entreprise tient à renforcer la cohésion de ses équipes tout en liant ces moments conviviaux à des **enjeux professionnels concrets**. Ces instants ont renforcé mon attachement à l'entreprise et m'ont donné l'envie de m'impliquer davantage.

Enfin, ce stage m'a permis de mieux comprendre **le secteur de l'assurance**, riche en défis analytiques et en opportunités. J'ai découvert une entreprise engagée, **accueillante**, avec un **large éventail de métiers** et un réel souci de transmission. Cette immersion m'a véritablement **épanouie**, et m'a confortée dans l'idée de poursuivre dans cette voie, en mobilisant toutes mes compétences et en continuant à apprendre avec enthousiasme.

En somme, ce stage m'a offert bien plus qu'une simple application de mes acquis : il a été une **expérience marquante**, qui a renforcé ma vocation et ma motivation à évoluer dans le

domaine de la data au sein d'environnements professionnels stimulants comme celui de CNP Assurances.

- **Ressources :**

- Documentation officielle Python
URL : <https://docs.python.org/3/>
Utilité : Compréhension des fonctionnalités DAX, création de rapports interactifs, utilisation des visuels personnalisés.
- Définition et compréhension des codes hexadécimaux
URL : https://www.conseil-creation.com/informatique/bureautique/ascii_hexadecimal.php
Utilité : Comprendre comment fonctionnent les empreintes numérique des lettres et leur utilités
- OpenClassrooms – Analyser des données avec Python
URL : <https://openclassrooms.com/fr/courses/6204541-analyser-des-donnees-avec-python>
Utilité : Approfondissement sur la manipulation des données avec pandas pour fiabiliser les bases d’indicateurs métier.
- Documentation interne CNP – Intranet
Utilité : Présentation de l’entreprise. Information pour les organigrammes.

- **Tableau des illustrations**

N°	Titre	Type	Page
Fig.1	La CNP en chiffre	Illustration	5
Fig.2	Organigramme du service UEX	Organigramme	7
Fig.3	Organigramme de la direction du pilotage financier et extra-financier	Organigramme	8
Fig.4	Schéma de l'origine des données des 3 fichiers manipulés durant la mission	Schéma	10
Fig.5	Onglets présents dans le fichier "Paramètre CB des requêtes structures"	Image	20
Fig.6	Aperçu de onglet "Parametres"	Image	20
Fig.7	Aperçue de l'onglet « CB dossier »	Image	21
Fig.8	Figure 8 : Visualisation de l'étape de la création des dossier répertoires relatives à chaque unité commerciale		22
Fig.9	Aperçue de l'un des onglets des exigences de mise en forme	Image	23
Fig.10	Calendrier prévisionnel (indicatif)	Tableau Calendrier	24
Fig.11	Schéma de l'arborescence de la demande client	Schéma	27
Fig.12	Exemple de la composition d'un des fichier (centre A2180 – BC)	Schéma	27
Fig.13	Tableau de comparaison des deux langages	Tableau	29












Annexe i : Organigramme du service UEX réalisé dans le cadre d'un séminaire du service DTO



Annexe ii : Les fichiers « Export » qui sont téléchargés chaque jour eu format .xlsx

> Téléchargements >			
Trier Trier ▼ Afficher Afficher ▼ ...			
<input type="checkbox"/>	Nom	Modifié le	Type
<input type="checkbox"/>	Export - Drill Engagements.txt	05/06/2025 16:07	Document texte
<input type="checkbox"/>	Export - Factures provisionnées.txt	05/06/2025 16:07	Document texte
<input type="checkbox"/>	Export - Drill Factures.txt	05/06/2025 16:06	Document texte
<input checked="" type="checkbox"/>	mise en garde et difference des factures .xlsx	02/06/2025 14:52	Feuille de calcul Micr...
<input checked="" type="checkbox"/>	mise en garde des factures .xlsx	02/06/2025 14:50	Feuille de calcul Micr...
<input checked="" type="checkbox"/>	Export - Drill Factures.xlsx	02/06/2025 11:41	Feuille de calcul Micr...
<input checked="" type="checkbox"/>	Export - Factures provisionnées.xlsx	02/06/2025 11:41	Feuille de calcul Micr...
<input checked="" type="checkbox"/>	Export - Drill Engagements.xlsx	02/06/2025 11:41	Feuille de calcul Micr...

Une intention particulière doit donc être faite afin que le programme prenne toujours en compte les 3 fichiers « Export » les plus récents et au bon format.

	Export - Drill Factures.txt	05/06/2025 16:06
		02/06/2025 14:52
		02/06/2025 14:50
	Export - Drill Factures.xlsx	02/06/2025 11:41
	Export - Factures provisionnées.xlsx	02/06/2025 11:41
	Export - Drill Engagements.xlsx	02/06/2025 11:41
Le mois dernier		
		13/05/2025 10:52
	Export - Drill Engagements (2) - Copie.xlsx	12/05/2025 10:40
[..	06/05/2025 15:52
		06/05/2025 10:31
Plus tôt cette année		
	src-master.zip	29/04/2025 16:13
	Export - Drill Engagements (2).xlsx	28/04/2025 16:26
	Export - Drill Factures (2).xlsx	28/04/2025 16:26
	Export - Factures provisionnées (2).xlsx	28/04/2025 16:26

Annexe iii : Visualisation du dossier téléchargement

Les fichiers les plus récents doivent être traités

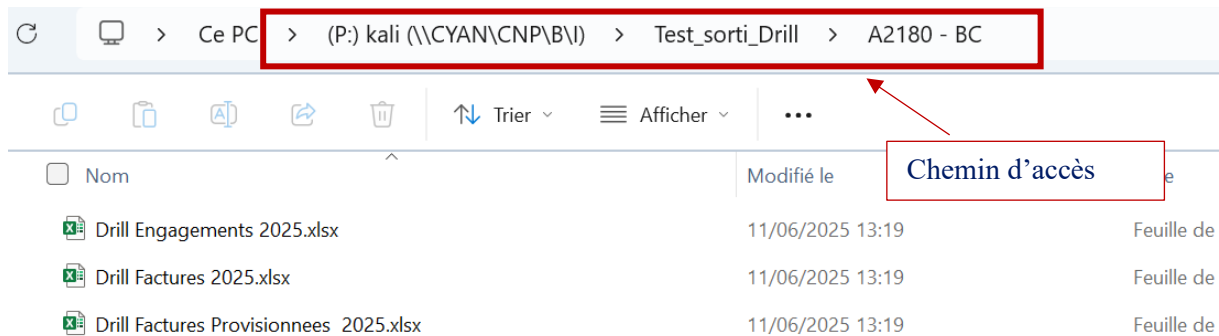
```
2 usages
def recuperer_fichiers_exports(dossier_telechargements: str, mot_cle: str = "export", extension: str = ".xlsx"):
    dossier = os.path.expanduser(dossier_telechargements)
    fichiers = [
        os.path.join(dossier, f)
        for f in os.listdir(dossier)
        if mot_cle.lower() in f.lower() and f.endswith(extension)
    ]
    fichiers_recents = sorted(fichiers, key=os.path.getmtime, reverse=True)
    #print(f"Fichiers triés par date de modification : {fichiers_recents}")

    selection_fichiers = {
        "engagement": None,
        "facture_provisionnee": None,
        "facture": None
    }
```

Instruction 'reverse = True' prend en compte les fichier importé dans le temps dans l'ordre décroissant

Annexe iv : Partie du code pyhton pour gérer l'importation des fichier les plus récents

Annexe v : Visualisation de l'étape de la création des 3 fichiers présents dans les dossiers de chaque unité commerciale



L'importance de vérifier à chaque étape d'abord de manière manuelle puis industriel pour comparer les résultats.

Annexe vi : Exemple d'un code réalisé au cour de mes traitements pour vérifier que mes résultats étaient les même que ceux déjà en production

```
#facture
import pandas as pd
chemin_fichier_Acomparer = r"P:\data_compare\BPCEBon\Liste Factures 2025.xlsx"
chemin_fichier_Averifier= r"P:\Test_sorti_Drill\BPCE\Drill Factures 2025.xlsx"
df1 = pd.read_excel(chemin_fichier_Acomparer)
df2 = pd.read_excel(chemin_fichier_Averifier)
comparison = df1.equals(df2)
if comparison:
    print("The datasets are identical!")
else:
    print("The datasets have differences.")

df1.compare(df2)
#erreur lors de l'importation des tables exports en sas car elles sont sur le réseaux
```

Démarche portfolio | Analyse du stage au regard des compétences du BUT Science des données

Compétence 1 – Traiter des données à des fins décisionnelles

Mon stage à la CNP Assurances m'a particulièrement permis de renforcer cette compétence, notamment à travers la mission principale consistant à transposer un script SAS existant en un script Python, afin d'en comparer l'efficacité. Cette tâche m'a amenée à intervenir sur toutes les étapes du cycle de vie de la donnée : extraction, transformation, structuration, documentation, et test. J'ai ainsi pris conscience de l'importance de bien identifier les langages, librairies et modèles de données adaptés aux besoins spécifiques d'un environnement professionnel. Le travail m'a aussi permis de mieux appréhender les contraintes métiers réels : rigueur dans la documentation, exigence de performance, sécurité des données. Ce contexte m'a conduite à structurer mes programmes avec attention, dans une logique de maintenabilité. J'ai mieux compris la nécessité d'une documentation claire, tant pour la relecture interne que pour la continuité des travaux après le départ d'un stagiaire. Le stage a donc concrétisé des notions apprises durant le BUT, comme le rôle central de l'entrepôt de données ou la gestion des sources hétérogènes.

Compétence 2 – Analyser statistiquement des données

Même si l'analyse statistique n'était pas le cœur de ma mission, plusieurs éléments ont enrichi ma compréhension de cette compétence. J'ai dû réaliser des analyses exploratoires pour comprendre les structures des fichiers budgétaires, mettre en évidence les incohérences ou les redondances, et choisir les méthodes les plus pertinentes pour les manipuler. J'ai ainsi mieux perçu l'intérêt des techniques d'exploration **en amont** du développement. Le stage m'a également sensibilisée à la spécificité de certaines données temporelles et budgétaires, avec leurs particularités de format, de périodicité ou de structure. J'ai pris conscience de la différence entre **modélisation** et **exploration**, en adaptant mes méthodes aux besoins immédiats des utilisateurs. Cette expérience me donne de solides bases pour aborder de futures analyses plus complexes, en comprenant les enjeux d'interprétation, de variabilité et de lisibilité des résultats pour un public non expert.

Compétence 3 – Valoriser une production dans un contexte professionnel

Le stage m'a permis de développer cette compétence à travers la restitution régulière de mon travail auprès de mon tuteur. J'ai appris à **adapter mon discours au niveau d'expertise de mes interlocuteurs**, à **justifier mes choix techniques**, et à **communiquer les résultats de façon claire et structurée**, tant à l'écrit qu'à l'oral. Ce travail de synthèse m'a demandé de faire preuve de **rigueur, d'esprit critique**. J'ai également veillé à respecter les **aspects déontologiques et réglementaires**, notamment en m'assurant que les données manipulées restaient strictement internes. Enfin, le stage m'a permis de mieux comprendre les **enjeux**

économiques de l'automatisation, en prenant en compte les gains de temps potentiels, les coûts de transition entre langages, et les impacts organisationnels. J'ai ainsi pu expérimenter l'importance de **savoir défendre ses résultats**, en les contextualisant dans les besoins opérationnels de l'entreprise.

Compétence 4 – Modéliser les données dans un cadre statistique (parcours EMS)

Bien que ma mission ne comportait pas de modélisation statistique avancée au sens académique du terme, il m'a permis de **mobiliser des réflexes de modélisation adaptés à un contexte professionnel**. Par exemple, pour réécrire le script SAS en Python, j'ai dû comprendre la **logique implicite des traitements** : filtrage conditionnel, agrégation, détection d'anomalies, autant de pratiques proches de la préparation de données en vue d'un modèle.

J'ai pu constater l'importance de **structurer les données selon une logique métier**, et d'**appréhender leurs spécificités**, notamment les données budgétaires, parfois mal formatées ou déséquilibrées. Cette étape m'a sensibilisée à **l'impact du type de données sur les traitements possibles**, et plus généralement à la nécessité d'adapter les méthodes aux enjeux du secteur de l'assurance.

Même sans réaliser de tests statistiques formels, le travail réalisé m'a permis de **m'interroger sur la qualité des données, la reproductibilité des traitements**, et la **validité des hypothèses implicites**. J'ai ainsi développé une posture critique, essentielle à toute mise en œuvre future de modèles statistiques.

Enfin, ce stage m'a permis de prendre conscience que **la modélisation ne commence pas au moment du calcul**, mais bien dès **la collecte, le nettoyage et la compréhension du terrain**. Cette compréhension globale est un prérequis indispensable pour réaliser des modélisations rigoureuses et utiles.

Apports transversaux – Savoir-être développés

Ce stage m'a enrichie sur le plan intellectuel, technique, mais aussi humain. J'ai appris à mieux gérer mon temps, à travailler de façon autonome tout en sachant solliciter de l'aide au bon moment. J'ai amélioré ma capacité d'adaptation, que ce soit face aux outils, aux méthodologies ou aux imprévus. Ce stage m'a épanouie, donné confiance en moi, et renforcé ma motivation à mobiliser toutes mes connaissances pour poursuivre sur cette voie professionnelle.