

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Établissement Privé de formation Supérieure Arrêté N° 342 du 11 avril 2018

Institut de Management INSIM SUP



# Gestion et manipulation de fichiers Excel

Mémoire de fin d'étude en licence

Présenté par : Mohamed Ryad Khennouchi

Filière : **Informatique**

Spécialité : **Informatique de gestion**

Encadré par : Dr. Yasser Moussa Berghout

M. Yacine Tehmi

Année universitaire 2022-2023

## Engagement de non plagiat

---

Je soussigné, Ryad Khennouchi

N° carte d'étudiant : L-IG-2924

Déclare avoir pris connaissance de la charte des examens et notamment du paragraphe spécifique au plagiat. Je suis pleinement conscient que le plagiat de documents ou d'une partie de document publiés sous quelque forme que ce soit (ouvrages, publications, rapports d'étudiant, internet ; etc.) constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée.

En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour produire et écrire ce document.

Fait le 26 juin 2023.

Signature

## Dédiasses

*À e dédie ce modeste travail à :*

*A celle qui a bercé mes rêves : ma mère ;*

*A celui qui a nourri mes ambitions : mon père ;*

*A mes anges gardiens : Grand-père et Grand-mère ;*

*A celle qui a soulevé bien des fardeaux avec moi : ma  
sœur ;*

*A toute ma famille ;*

*A toute mes amis sans exception.*



## Remerciements

---

Chers Parents, famille, amis, professeurs et encadrants,

Je tenais à prendre un moment pour remercier tout ce beau monde. Votre présence a joué un rôle crucial dans le succès de mon projet, et je suis infiniment reconnaissant pour cela.

À ma mère, je voudrais dire combien ta présence, ton amour et ton soutien ont été essentiels dans chaque étape de ma vie. Ta force et ta bienveillance m'ont toujours inspiré à donner le meilleur de moi-même, et je ne pourrais jamais assez te remercier pour tout ce que tu as fait.

À toute ma famille et mes amis, vous avez été mes piliers, mes sources de réconfort et d'encouragement. Votre confiance en moi et votre soutien indéfectible ont été d'une importance capitale. Je vous suis extrêmement reconnaissant.

À mon professeur encadrant et à mon maître de stage, je suis reconnaissant pour votre expertise, votre guidance et vos précieux conseils. Votre confiance en mes compétences et votre suivi constant ont été essentiels pour mener à bien ce projet. Je suis honoré d'avoir pu bénéficier de votre encadrement.

À toute l'équipe du département de formation et du département informatique, je tiens à vous remercier pour votre professionnalisme. Votre expertise et votre collaboration ont grandement contribué au succès de mon projet.

Enfin, à tous mes professeurs, je suis reconnaissant pour votre enseignement de qualité et votre dévouement à former les esprits. Grâce à vous, j'ai acquis les connaissances et les compétences nécessaires pour mener à bien ce projet et atteindre mes objectifs.

Mes sincères remerciements à chacun de vous pour votre contribution précieuse à mon parcours.

Bonne lecture.



## Résumé

---

Ce mémoire porte sur la gestion et la manipulation de fichiers Excel à l'aide de Python. Il aborde la problématique rencontrée par le département de formation de l'entreprise Sonatrach, qui éprouve des difficultés dans le traitement de multiples données au format Excel, entraînant des erreurs, des doublons et une perte de temps. L'objectif de cette étude est de créer une application de bureau permettant de manipuler et de gérer efficacement ces fichiers Excel. L'application offre également la possibilité aux utilisateurs d'insérer des données directement depuis l'application et de détecter les doublons, les cellules vides et autres incohérences. Pour mener cette étude, une approche mixte a été utilisée, combinant des entretiens avec les employés pour comprendre la problématique et des observations personnelles. Le langage de programmation Python a été choisi pour résoudre ce problème. Les implications de ce travail sont une optimisation et une facilitation de l'utilisation des fichiers Excel, ainsi qu'une réduction des erreurs engendrées. Dans le domaine de la data science, cette étude apporte une contribution significative en termes d'efficacité et de fiabilité dans la gestion des données Excel.









# Sommaire

<b>Dédiasses</b> .....	3
<b>Remerciements</b> .....	5
<b>Résumé</b> .....	7
<b>Sommaire</b> .....	11
<b>Introduction</b> .....	12
<b>Chapitre 1 Présentation de l'organisme d'accueil</b> .....	15
<b>Partie 1 – Sonatrach</b> .....	16
Introduction .....	16
Présentation de l'entreprise.....	16
Définition de Sonatrach .....	18
Historique de Sonatrach .....	18
Les principales Activités de la Société.....	19
Les usages des produits pétroliers raffinés .....	20
Logistique de Sonatrach.....	20
Perspectives d'avenir.....	23
Conclusion.....	23
<b>Partie 2 – Problématique</b> .....	24
Introduction générale .....	24
Problématique de recherche.....	25
Contextualisation du problème .....	26
Méthodologie de recherche.....	26
Objectif de la recherche.....	29
Conclusion.....	30
<b>Chapitre 2 Domaine de recherche</b> .....	31
Introduction .....	31
Définition de la Data science.....	32
L'histoire de la Data science .....	32
Définition d'un Data scientist.....	32
Domaine de la Data Science .....	33
Les métiers et leurs rôles de la Data science.....	34
Le rôle de la data science : .....	35
Les avantages qu'offre la data science .....	36
Types de données utilisées en Data science .....	37
Les sources de données et leur collecte .....	37
Technique et outils de la data science .....	39
Les enjeux et défis de la data science .....	41
Le future de la Data science : .....	42
<b>Chapitre 3 - Implémentation et conception</b> .....	43
Introduction .....	43
Présentatin du système .....	44
<b>Modélisation</b> .....	50
Implémentation.....	51
Conclusion.....	67
<b>Chapitre 4 – Conclusion générale</b> .....	68
<b>Bibliographie</b> .....	74
<b>Tables</b> .....	80
Tables des illustrations .....	81
Table des matières .....	83

## Introduction

---

Dans le domaine de l'informatique de gestion, la gestion et la manipulation de fichiers Excel joue un rôle crucial dans de nombreuses entreprises. Notre mémoire de fin d'études se concentre sur ce sujet, en particulier dans le contexte du département de formation de Sonatrach.

Sonatrach, une entreprise de premier plan dans son secteur d'activité, est confrontée à des défis majeurs lors du traitement de multiples données au format Excel. Le département de formation de Sonatrach rencontre des difficultés en raison de nombreuses erreurs, doublons et pertes de temps lors du traitement de ces fichiers.

L'objectif de notre étude est de concevoir et développer une application desktop qui permettra la gestion et la manipulation efficaces des fichiers Excel. Cette application offrira également la possibilité à l'utilisateur d'insérer des données directement à partir de l'application, de détecter les doublons, les cellules vides et les incohérences.

Pour atteindre cet objectif, nous adopterons une approche mixte de recherche, combinant des entretiens avec les employés du département de formation de Sonatrach pour comprendre en profondeur leurs besoins et problématiques, ainsi que nos propres observations personnelles. Nous avons choisi le langage de programmation Python pour résoudre ce problème, en raison de sa popularité et de ses nombreuses bibliothèques spécialisées dans le traitement des données.

Les implications et les contributions de notre travail dans le domaine de l'informatique de gestion sont significatives. Notre application contribuera à l'optimisation et à la facilité d'utilisation des fichiers Excel, réduisant ainsi les erreurs générées lors du traitement des données. Cela permettra au département de formation de Sonatrach de gagner en efficacité et en productivité dans leurs tâches quotidiennes.

La structure et l'organisation de notre mémoire se composent de plusieurs chapitres. Dans le premier chapitre, nous présenterons l'entreprise Sonatrach, définirons son domaine d'activité, sa structure et discuterons de ses objectifs futurs. Nous aborderons également la problématique liée à la gestion des fichiers Excel au sein du département de formation.

Le deuxième chapitre sera consacré à la présentation du domaine de la Data Science, mettant en avant ses avantages, sa structure et ses différents domaines d'utilisation.

Nous examinerons la data science sous différents angles pour comprendre comment elle peut contribuer à la résolution des problèmes rencontrés.

Dans le troisième chapitre, nous présenterons les outils de programmation et les méthodes de conception que nous avons utilisés dans le développement de notre application. Nous mettrons en évidence les fonctions clés qui ont été implémentées pour la manipulation des fichiers Excel.

Enfin, dans le dernier chapitre, nous conclurons notre mémoire en discutant des résultats obtenus et en proposant des perspectives d'avenir pour l'amélioration continue de l'application, ainsi que pour l'exploration de nouveaux défis dans le domaine de la gestion et de la manipulation des fichiers Excel.

Notre mémoire vise donc à apporter une contribution significative à l'amélioration des processus de gestion des fichiers Excel dans le département de formation de Sonatrach, tout en offrant des perspectives prometteuses pour l'avenir de la Data Science dans le domaine de l'informatique de gestion.



## Chapitre 1

### Présentation de l'organisme d'accueil

---

## Partie 1 – Sonatrach

### Introduction

Algérie est l'un des pays les plus riches en énergie fossile. En 2015, elle occupait la 18ème place en tant que producteur de pétrole, la 101ème place en tant que producteur de gaz naturel et était un important exportateur mondial de gaz naturel. Cette notoriété sur la scène internationale en fait un membre essentiel de l'Organisation des pays exportateurs de pétrole et de gaz, ainsi que du Forum.

Dans ce chapitre, on parlera d'une société Majeure dans le domaine des énergies qui depuis plus de 50 ans joue un rôle central en tant que moteur de l'économie nationale en Algérie. Sa mission principale est de valoriser les vastes réserves d'hydrocarbures du pays. En tant qu'acteur clé de l'industrie pétrolière, elle est surnommée la "major africaine" et tire sa force de sa capacité à être un groupe intégré couvrant l'ensemble de la chaîne de valeur des hydrocarbures.

### Présentation de l'entreprise

SONATRACH est une entreprise publique algérienne fondée en 1963 dans le but d'exploiter les ressources en hydrocarbures du pays. Ses activités couvrent l'ensemble de la chaîne de production, de l'exploration à l'exploitation, en passant par le transport et le raffinage. Au cours des 50 dernières années, elle a joué un rôle majeur en tant que moteur de l'économie nationale, avec pour mission de valoriser les importantes réserves d'hydrocarbures de l'Algérie.

SONATRACH occupe une place prépondérante dans le secteur du pétrole et du gaz. Elle est actuellement la principale compagnie d'hydrocarbures en Afrique et en Méditerranée. Ses activités se concentrent sur quatre domaines principaux : l'amont (exploration et production), l'aval (raffinage et distribution), le transport par canalisation et la commercialisation. Elle s'est également diversifiée dans des secteurs tels que la pétrochimie et le dessalement de l'eau de mer. En termes de taille et de création de ressources, elle est la première entreprise en Algérie.

Elle a réussi à intégrer les nouvelles technologies, à établir une présence solide et fiable sur les marchés internationaux des hydrocarbures liquides et gazeux, et à capitaliser sur son expérience de partenariat avec des compagnies internationales de différentes tailles et origines géographiques. La société a participé à plusieurs projets en collaboration avec différents partenaires en Amérique latine et en Europe.



En outre, elle s'est investie dans d'autres secteurs économiques tels que les mines, la production d'électricité, l'eau, le transport aérien et maritime, et bien d'autres.



## ORGANIGRAMME DE LA MACROSTRUCTURE DE SONATRACH

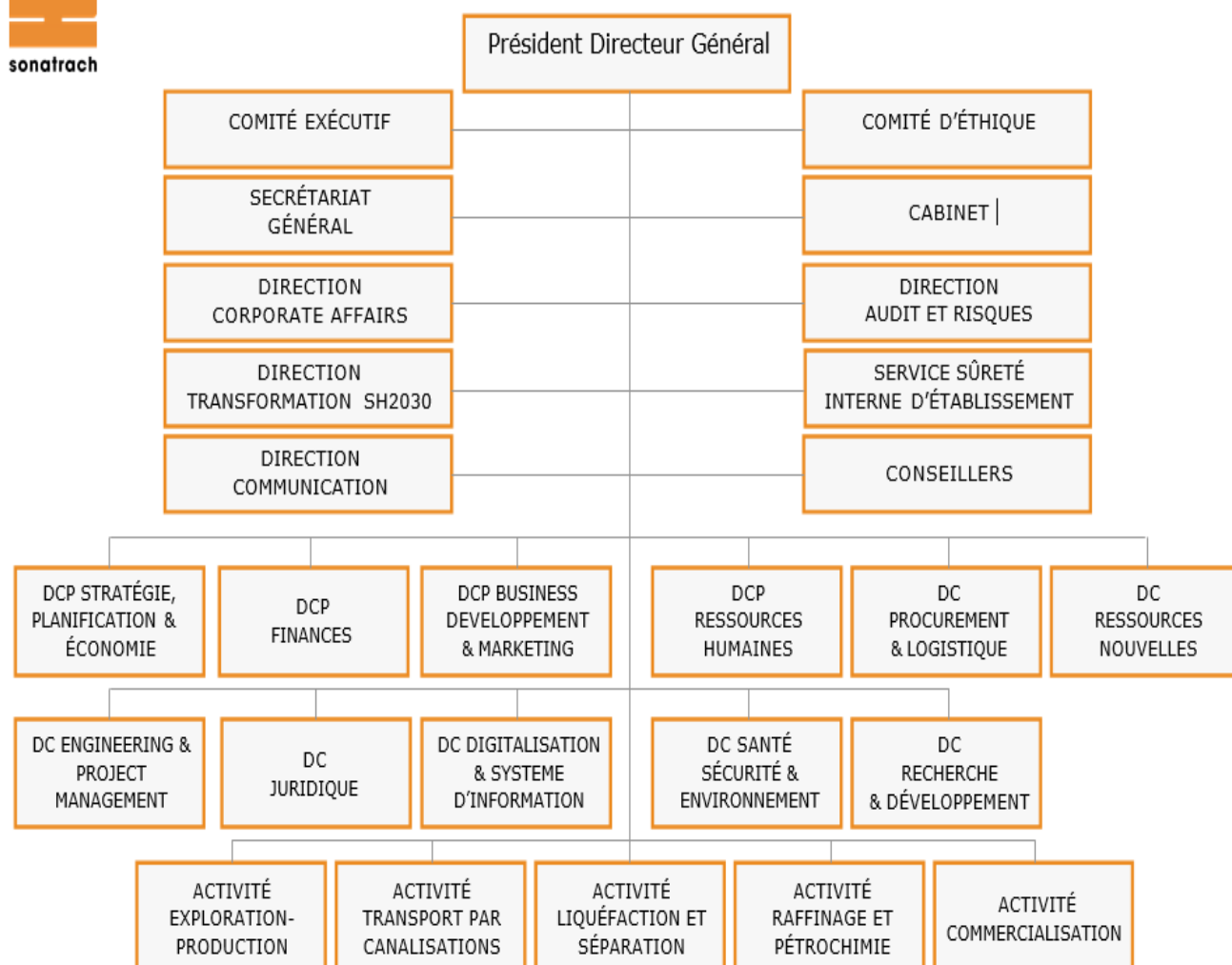


Figure 1 : Organigramme d'organisation de Sonatrach.

## Définition de Sonatrach

SONATRACH, dont l'acronyme signifie "Société nationale pour la recherche, la production, le transport, la transformation et la commercialisation des hydrocarbures", est une entreprise publique algérienne fondée le 31 décembre 1963. C'est un acteur clé de l'industrie pétrolière, largement reconnu en tant que "major africaine". En tant que première entreprise d'Afrique, elle s'engage à valoriser de manière optimale les ressources nationales en hydrocarbures, créant ainsi des richesses au service du développement économique et social du pays.

## Historique de Sonatrach

Sonatrach a été créée le 31 décembre 1963 dans le but principal d'assurer le transport et la commercialisation des hydrocarbures. Au fil des années, elle s'est étendue à d'autres domaines de l'industrie pétrolière. En 1964, elle a construit son premier oléoduc, l'OZ1, reliant Haoud El Hamra à Arzew sur une distance de 805 km, ce qui marqua le début de son implication dans le secteur du gaz. En 1979, Sonatrach a mis en service le complexe de liquéfaction de gaz naturel GL4Z (CAMEL) avec une capacité de traitement de 1,8 milliard de mètres cubes de gaz par an.

En 2014, Sonatrach est devenue la première entreprise à se lancer dans la production de gaz de schiste en Algérie, suite à l'approbation gouvernementale pour l'exploitation des hydrocarbures non conventionnels. Des changements de direction ont eu lieu, avec le limogeage du PDG Abdelhamid Zerguine en juillet 2014 et son remplacement par Saïd Sahnoun par intérim. Par la suite, Amine Mazouzi a été nommé président-directeur général de Sonatrach en mai 2015.

En mars 2016, le site gazier de Krechba a été attaqué par un groupe terroriste, entraînant une suspension temporaire de la production. Cependant, la production a repris peu de temps après. Au cours de la même année, Sonatrach a signé un accord avec la société pétrolière italienne ENI pour l'exploration offshore de nouvelles ressources pétrolières et gazières.

En mars 2017, Amine Mazouzi a été remplacé par Abdelmoumen Ould Kaddour en raison de résultats jugés insatisfaisants. Ould Kaddour a entrepris une réorganisation du groupe pour résoudre les problèmes de bureaucratie, de structure organisationnelle inadaptée, de créances, de manque de stratégie à long terme et de trop nombreuses filiales. Il a également cherché à développer le secteur du gaz de schiste.

Sonatrach a renforcé ses relations avec plusieurs compagnies étrangères, telles que Anadarko, Cepsa, Eni, Maersk, Pertamina et Talisman Energy, grâce à des accords signés en septembre 2017. L'entreprise a également signé un mémorandum avec ENI en décembre 2017, portant sur les énergies renouvelables et la recherche de sites pour la production d'énergie solaire.

En janvier 2018, Sonatrach a conclu un accord cadre avec la compagnie pétrolière libyenne NOC pour renforcer leur coopération dans la gestion des gisements transfrontaliers d'hydrocarbures. La même année, Sonatrach a acquis la raffinerie d'Augusta et trois terminaux pétroliers en Italie appartenant à Esso italiana, une filiale d'ExxonMobil. Elle a également signé un accord avec Total pour la construction d'une usine de pétrochimie à Arzew, ainsi qu'un partenariat avec Bonattia, Pertamina et Talisman pour augmenter les capacités de réinjection de gaz dans le champ pétrolier de Menzel Ledjmet nord.

Dans le cadre de sa stratégie d'expansion internationale, Sonatrach a signé un contrat en octobre 2018 avec les sociétés turques Ronasans Industrie Tesisleri Inaat Sanayi ve Ticaret et Bayegan pour la construction d'une unité de propylène en Turquie, avec une participation de 30 % de la société algérienne.

## Les principales Activités de la Société

Sonatrach est une entreprise d'envergure internationale, active dans plusieurs domaines liés à l'industrie pétrolière et gazière. La société excelle dans les métiers de la recherche et de l'exploration, où elle mène des activités visant à découvrir de nouvelles réserves d'hydrocarbures, tant sur terre qu'en mer. Grâce à son expertise en exploration, Sonatrach identifie des gisements prometteurs et met en place des opérations de production d'hydrocarbures à grande échelle.

La production d'hydrocarbures est l'un des piliers de l'activité de Sonatrach. L'entreprise utilise des technologies avancées pour extraire le pétrole et le gaz naturel de ses gisements, maximisant ainsi la récupération des ressources énergétiques. Elle se positionne également comme un acteur majeur dans le domaine du raffinage, transformant le pétrole brut en produits pétroliers raffinés tels que l'essence, le diesel, le gazole et les lubrifiants. Cette intégration verticale lui permet de contrôler toute la chaîne de valeur, du forage à la commercialisation.

Sonatrach est également impliquée dans le secteur du gaz de schiste. Pionnière en la matière, elle a été la première entreprise à se lancer dans la production de gaz de schiste en Algérie. Elle applique des techniques spécifiques pour extraire le gaz naturel des formations de schiste, contribuant ainsi à la diversification des sources d'énergie du pays.

Outre ses activités dans les hydrocarbures, Sonatrach a également développé une expertise dans le transport maritime. L'entreprise assure le transport sécurisé des hydrocarbures par voie maritime, utilisant des navires spécialisés pour acheminer les produits pétroliers vers les marchés internationaux.

Soucieuse du développement durable, Sonatrach s'intéresse aux énergies renouvelables. Elle investit dans la recherche et le développement de nouvelles sources d'énergie propres, en explorant notamment le potentiel de l'énergie solaire. Par le biais de partenariats et de collaborations, Sonatrach vise à contribuer à la transition énergétique en Algérie et à promouvoir les énergies renouvelables comme une alternative viable aux combustibles fossiles.

### Les usages des produits pétroliers raffinés

Chacun des produits raffinés issus du pétrole brut trouve un usage spécifique :

Le gaz de pétrole liquéfié (GPL) est un carburant pour les véhicules à gaz ;

Les gaz butane et propane sont utilisés pour les besoins domestiques ;

L'essence et le gazole alimentent les moteurs des véhicules automobiles ;

Le kérosène est employé comme carburant dans l'aviation ;

Le naphta est la principale matière première employée en pétrochimie ;

Le fioul domestique est un combustible Un combustible désigne tout composant ou matière solide, liquide ou gazeux susceptible de se combiner à un oxydant... de chauffage ;

Les huiles servent à fabriquer des lubrifiants ;

Le bitume est utilisé pour recouvrir les routes.

### Logistique de Sonatrach

La logistique occupe une place primordiale au sein de Sonatrach en garantissant une gestion coordonnée et efficiente de la chaîne d'approvisionnement, du transport et de la distribution des hydrocarbures. Pour assurer le transport des hydrocarbures sur de longues distances, Sonatrach dispose d'une infrastructure logistique étendue comprenant des pipelines, des terminaux de stockage, des installations portuaires et des flottes de navires pétroliers. La gestion des stocks est également une priorité pour Sonatrach, qui gère d'importantes installations de stockage comprenant des réservoirs de pétrole brut, de

GNL, de GPL et d'autres produits pétroliers. L'objectif est d'assurer un approvisionnement continu et fiable, en optimisant les niveaux de stock et en garantissant leur sécurité. En ce qui concerne la distribution, Sonatrach se charge de l'acheminement des produits pétroliers vers les clients finaux, tels que les stations-service, les industries et les fournisseurs d'énergie. Différents modes de transport sont utilisés, notamment les camions citernes, les pipelines et les navires, afin d'assurer une distribution efficace et ponctuelle des produits pétroliers. En tant qu'entreprise internationale opérant dans plusieurs pays, Sonatrach gère également la logistique de ses activités à l'étranger, en coordonnant les opérations de transport, les formalités douanières, la conformité réglementaire et les partenariats avec des entreprises locales, afin d'assurer une logistique fluide et conforme aux exigences locales. Sonatrach accorde une grande importance à l'optimisation de sa logistique, en utilisant des technologies avancées, des systèmes de suivi et de gestion des stocks, ainsi que des outils d'analyse de données pour améliorer l'efficacité de ses opérations logistiques. Cela comprend l'optimisation des itinéraires, la réduction des délais, la gestion des coûts et la minimisation des risques. En somme, la logistique occupe un rôle essentiel au sein de Sonatrach, englobant l'approvisionnement, le transport, la gestion des stocks, la distribution et la logistique internationale. L'entreprise met en œuvre des stratégies et des technologies afin d'optimiser sa logistique, garantir une chaîne d'approvisionnement fiable et efficace, et répondre aux exigences des marchés nationaux et internationaux.

Par le biais de partenariats stratégiques avec des sociétés étrangères, Sonatrach étend son rayonnement à l'échelle mondiale en participant à des projets pétroliers et gaziers dans divers pays. Grâce à ces alliances, Sonatrach explore, développe et exploite des gisements d'hydrocarbures à l'étranger, profitant ainsi de l'expertise et des ressources nécessaires pour réaliser des projets d'envergure. Ainsi, Sonatrach joue un rôle actif dans l'exploration et la production d'hydrocarbures dans différentes régions du globe, consolidant ainsi sa présence sur le marché mondial de l'énergie. La capacité de Sonatrach à s'adapter aux défis et aux opportunités internationaux atteste de sa position stratégique dans le secteur énergétique.

## Les opérations internationales de SONATRACH



Figure 2 : Opérations internationales de Sonatrach.

## Perspectives d'avenir

Sonatrach, en tant que leader national dans l'industrie des hydrocarbures, est bien positionnée pour saisir de nombreuses opportunités et faire face aux défis futurs. L'entreprise envisage de diversifier ses activités et de développer davantage les énergies renouvelables, en s'engageant dans des projets d'énergie solaire, éolienne et de stockage d'énergie. Sonatrach prévoit également de renforcer sa présence sur le marché mondial en continuant à investir dans des projets internationaux et en établissant des partenariats stratégiques. Elle vise à améliorer l'efficacité opérationnelle grâce à l'adoption de technologies de pointe, à la numérisation et à la modernisation de ses infrastructures. Parallèlement, Sonatrach accorde une attention croissante à la protection de l'environnement et à la durabilité, en mettant en œuvre des pratiques respectueuses de l'environnement et en favorisant la réduction des émissions de gaz à effet de serre. Avec une vision axée sur l'innovation et la durabilité, Sonatrach est prête à relever les défis futurs de l'industrie énergétique et à contribuer à la transition vers une économie plus propre et plus durable.

## Conclusion

Mon stage au bureau Sonatrach de Sidi Rzine m'a permis d'approfondir mes connaissances et mes compétences dans le domaine de l'informatique tout en me familiarisant avec les pratiques de gestion des Données et les techniques de codage Python. Grâce à cette expérience et au temps que j'ai passé avec eux, j'ai pu constater les avantages de ces outils en termes d'optimisation du travail, d'efficacité et d'efficience. J'ai également pu observer comment l'intégration de Python dans les flux de travail existants a permis de gagner du temps et d'améliorer la qualité de l'analyse des données. Cette expérience m'a ouvert les yeux sur la grande promesse d'utiliser Python et les opportunités qu'il présente pour améliorer les performances et la productivité dans les domaines de l'informatique et de la gestion des données.

## Partie 2 – Problématique

### Introduction générale

Au sein de Sonatrach, l'une des entreprises les plus importantes du secteur énergétique en Algérie, une préoccupation primordiale réside dans le développement professionnel et personnel de ses employés. Soucieuse d'optimiser les compétences et les performances de son personnel, Sonatrach a créé un département de formation dédié à cette mission fondamentale. Ce département se charge de former les travailleurs de manière à accroître leur efficacité, leur efficience et leur expertise, afin qu'ils puissent apporter une contribution exceptionnelle au sein de l'entreprise.

Cependant, malgré l'importance stratégique du département de formation, ses responsabilités sont considérables. En effet, il doit gérer l'ensemble des employés de l'entreprise, analyser leurs profils individuels, identifier les formations appropriées à leurs besoins spécifiques, allouer les budgets nécessaires et assurer un suivi rigoureux. Face à ce défi colossal, le département de formation de Sonatrach arbore des difficultés majeures, entraînant une perte de temps significative et un taux d'erreur préoccupant.

L'une des principales sources de cette inefficacité est le flux constant et massif de données à traiter. Le traitement manuel et individuel des informations relatives à chaque employé génère une charge de travail considérable pour le département de formation, entraînant inévitablement des retards et des erreurs. Une étude interne a révélé que le taux d'erreur atteignait jusqu'à 4-8%, mettant ainsi en évidence l'urgence d'apporter des solutions pour optimiser les processus de gestion et de traitement des données.

Dans un contexte où l'amélioration de l'efficacité et de la précision du département de formation de Sonatrach est primordiale, les avancées technologiques dans le domaine de l'informatique et des logiciels se révèlent être une solution prometteuse. L'automatisation des tâches répétitives et la mise en place d'un système de gestion des données plus performant sont des pistes à explorer pour optimiser les ressources et le temps du département. Ainsi, la problématique essentielle consiste à trouver des moyens permettant au département de formation de Sonatrach d'améliorer ses processus de gestion et de traitement des données, afin de réduire les pertes de temps et les taux d'erreur, tout en veillant à ce que chaque employé bénéficie des formations adéquates pour son développement professionnel. Pour répondre à cette problématique, l'exploration de solutions innovantes et l'adoption de technologies adaptées sont nécessaires pour garantir une excellence opérationnelle, tout en accompagnant efficacement le développement des talents au sein de l'entreprise.



## Problématique de recherche

L'objectif de la présente étude, afin de remédier aux problèmes confrontés au sein du département de formation, est la création d'une application desktop visant à gérer, manipuler et traiter les données de ce dernier.

Dans cette perspective, la problématique de recherche est formulée de la manière suivante :

### **Gestion et Manipulation des fichiers Excel.**

Il en résulte les questions suivantes :

1. Qu'elle approche prendre face à cette problématique ?
2. Quelles sont les fonctionnalités spécifiques requises dans l'application desktop pour répondre à nos besoins en terme de gestion et de manipulation des données ?
3. Quels sont les principaux problèmes et difficultés actuels liés à la gestion et à la manipulation des fichiers Excel au sein du département de formation ?
4. Quelles technologies seraient les plus appropriées pour améliorer l'efficacité du travail et résoudre ce problème ?

Hypothèses :

Afin de répondre à ces questions, nous avons les hypothèses suivantes :

1. En adoptant une approche basée sur l'automatisation et l'intégration des processus de traitement, il sera possible de résoudre efficacement la problématique de gestion et de manipulation des données au sein du département de formation.
2. L'application desktop devrait comporter des fonctionnalités telles que l'importation/exportation de fichiers Excel, la manipulation des données (l'insertion, filtrage, fusion, etc.), et la visualisation des données sous forme graphique avec la possibilité de sauvegarde.
3. L'application desktop devrait aussi comporter une option de détection de fautes tels que les doublons, les champs oubliés ou d'autres incohérences.

## Contextualisation du problème

La gestion et la manipulation des fichiers Excel jouent un rôle crucial dans le fonctionnement efficace du département formation de Sonatrach. Cependant, l'utilisation de fichiers Excel pour stocker et gérer les données a entraîné divers problèmes, tels que la perte de temps, les erreurs, les doublons et les incohérences. Ces difficultés ont un impact direct sur la productivité du département et peuvent compromettre la qualité des processus de formation.

La problématique centrale de cette recherche réside dans la nécessité de trouver des solutions pour leur faciliter le travail et optimiser l'utilisation de leur temps en ce qui concerne la gestion des fichiers Excel. Face à ces défis, une question se pose : est-il possible de développer une application desktop dédiée à la manipulation des fichiers Excel afin de simplifier les tâches de gestion, de réduire les erreurs et d'accélérer le traitement des données ?

Pour répondre à cette question, il est essentiel de recueillir des données et des informations auprès des employés du Dép. formation qui sont directement impliqués dans la manipulation des fichiers Excel au quotidien. Leurs expériences, leurs observations et leurs suggestions fourniront des indications précieuses sur les problèmes rencontrés, les aspects du processus de travail qui peuvent être améliorés et les fonctionnalités spécifiques nécessaires dans une application desktop dédiée.

Le contexte général de cette recherche repose sur la volonté de faciliter le travail du Dép. formation, d'optimiser leur temps et de réduire les taux d'erreur associés à la gestion des fichiers Excel. En développant une application adaptée à leurs besoins et à leurs défis spécifiques, il est envisageable d'améliorer l'efficacité opérationnelle du département et de garantir la précision et la fiabilité des données manipulées.

En résumé, cette étude vise à explorer la possibilité de créer une application desktop dédiée à la gestion et à la manipulation des fichiers Excel au sein du département formation de Sonatrach. Elle s'appuie sur la problématique liée à la perte de temps, aux erreurs et aux inefficacités actuelles, avec pour objectif final de faciliter le travail, d'optimiser le temps et de réduire les taux d'erreur du département formation.

## Méthodologie de recherche

La méthodologie de recherche adoptée dans cette étude vise à explorer la gestion et la manipulation des fichiers Excel dans le domaine de la Data Science, en mettant l'accent sur l'optimisation des processus et l'analyse des problèmes courants rencontrés par l'équipe du Département formation.

Pour atteindre ces objectifs, différentes méthodes de collecte et d'analyse de données ont été utilisées propre au domaine de la Data Science. Tout d'abord, des entretiens ont été menés avec le personnel du département de Sonatrach, y compris la chef de département, les employés et les sous-traitants. Ces entretiens ont permis de recueillir des informations précieuses sur leurs méthodes de travail actuelles, les défis auxquels ils sont confrontés et leurs besoins en termes de gestion des fichiers Excel. Les entretiens ont été structurés autour des questions clés liées à l'optimisation des processus, à l'intégration des données et à la détection des erreurs.

En plus des entretiens, une observation directe du travail effectué par le personnel du département de formation a été réalisée. Cette observation a permis d'observer les étapes de manipulation des fichiers Excel, les procédures d'analyse et de traitement des données, ainsi que les lacunes et les inefficacités dans ces processus. Des exemples concrets des fichiers Excel utilisés dans le département, tels que les fichiers TBF, PMT par action et PMT par CSP, ont été sélectionnés pour une étude détaillée de leurs composants et de leurs caractéristiques avec l'aide du Glossaire fourni.

La collecte de données a également impliqué l'analyse approfondie des fichiers Excel fournis. Ces fichiers ont été examinés en détail pour comprendre leur structure, leurs caractéristiques et les types de données manipulées. Cela a permis d'identifier les principaux problèmes et les erreurs courantes dans la gestion de ces fichiers.

La méthodologie de recherche adoptée dans cette étude se fonde sur une approche mixte, combinant des entretiens, des observations et des analyses des fichiers Excel. Ces différentes méthodes de collecte de données permettent d'obtenir une vision complète des défis auxquels le département de formation est confronté dans la gestion des fichiers Excel.

En conclusion, cette méthodologie de recherche nous a permis d'approfondir notre connaissance et compréhension du problème ce qui nous mèneras à trouver et offrir une solution plus adaptée au problème posé.

GLOSSAIRE_SONATRACH - Excel (Échec de l'activation du produit)						
Fichier	Accueil	Insertion	Mise en page	Formules	Données	Révision
Police	Alignement	Nombre	Style	Cellules	Formules	Données
124	A	B	C	D	E	F
		CODE DOMAINE	DOMAINE	Catégorie de formation	Type Formation	Code Formation
4		1	1.Exploration	1. Actions d'adaptation au poste de travail	1. Formation & recrutement	CDA
5		1.1	1.1.Shales gaz	2. Actions liées à l'évolution des métiers & technologies	2. Perfectionnement	CDI
6		1.2	1.2.Off-shore	3. Actions liées au développement des compétences	3. Formation de reconversion	CDE
7		2	2.Forage		4. Stages Fournisseurs	LDA
8		3	3.Réservoir Engineering		5. Formation Induction	LDI
9		4	4.Exploitation/Développement Gis		6. Formation Corporate	LDE
10		5	5.Transport HC			
11		5.1	5.1.Corrosion			
12		6	6.Transformation HC			
13		6.1	6.1.Raffinage/pétrochimie			
14		7	7.Commercialisation HC			
15		7.1	7.1.Négociation			
16		8	8.Maintenance Industrielle			
17		9	9.Sécurité Industrielle			
18		10	10.Environnement			
19		11	11.Etudes/Economie			
20		12	12.Management de Projets			
21		13	13.Business Développement			
22		14	14.Organisation/Planification			
23		15	15.Contrôle de qualité / Coûts			
24		16	16.N.T.I.C			
25		17	17.Management			
26		18	18.Ressources Humaines			
27		19	19.Finances/Comptabilité			
28		20	20.Fiscalité/Assurances			
29		21	21.Juridique			
30		22	22.Audit			
31						

Figure 3 : Glossaire.

TBF DU MOIS D'AVRIL ACTIVITE RPC - Excel (Échec de l'activation du produit)										
Fichier	Accueil	Insertion	Mise en page	Formules	Données	Révision	Affichage	Données	Révision	Affichage
Police	Alignement	Nombre	Style	Cellules	Formules	Données	Révision	Affichage	Données	Révision
AD7	A	B	C	D	E	F	G	H	I	J
6	IDENTIFICATION AGENT									
	STRUCTURE	UNITE	MATRICULE (1)	NOM PRENOM	Sexe	CSP (2)	CODE DOMAINE	METIER	DOMAIN	INTITULE DE LA FORMATION
7	LF	Siege RPC	70833H	HAMMOUM TOUFIK	M	M	21	FST	Juridique	ANALYSE ET EVALUATION D
8	LF	Siege RPC	12908Q	TIENTI TEWFIK	M	C	26	FST	Communication	COMMUNICATION INTERPE
9	LF	Siege RPC	55790L	MARICHE ALI	M	C	11	FST	ETUDES/ECONOMIE	PROCUREMENT
10	LF	Siege RPC	73146K	REZZAZ SAMIA	F	C	26	FST	Communication	REDIGER DES COURRIERS E
11	LF	Siege RPC	73871X	BENGHERIFA SAIDA	F	C	26	FST	Communication	REDIGER DES COURRIERS E
12	LF	Siege RPC	76592K	SADOUNI KHALED	M	C	26	FST	Communication	REDIGER DES COURRIERS E
13	LF	Siege RPC	55743V	LABED MOHAMED LAMINE	M	C	19	FST	Finances/Comptabilité	EVALUATION ET COMPTABI
14	LF	Siege RPC	63398T	BEN ABDESSELAM AMOKRANE	M	C	19	FST	Finances/Comptabilité	EVALUATION ET COMPTABI
15	LF	Siege RPC	53335B	CHERCHAR SAFIA	F	C	21	FST	Juridique	PASSATION DES CONTRATS
16	LF	Siege RPC	53095B	FAID WAHIBA	F	C	18	FST	Ressources Humaines	PARCOURS RESSOURCES HL
17	LF	Siege RPC	41410V	BACHA ABOU SOFIANE	M	C	18	FST	Ressources Humaines	PARCOURS RESSOURCES HL
18	DIV RAFF	RA1Z	77162E	BENABDELKRIM MOHAMMED	M	C	9	FCM	Sécurité Industrielle	SYSTEME DE PERMIS DE TR
19	DIV RAFF	RA1Z	56611M	BENOUDANE HABIB	M	C	9	FCM	Sécurité Industrielle	SYSTEME DE PERMIS DE TR
20	DIV RAFF	RA1Z	67745U	SEGHIR MUSTAPHA	M	C	9	FCM	Sécurité Industrielle	SYSTEME DE PERMIS DE TR
21	DIV RAFF	RA1Z	54697A	CHERIEF BENAISSA	M	C	9	FCM	Sécurité Industrielle	SYSTEME DE PERMIS DE TR
22	DIV RAFF	RA1Z	63464F	KHATIR ABDERRAHMANE	M	M	9	FCM	Sécurité Industrielle	SYSTEME DE PERMIS DE TR
23	DIV RAFF	RA1Z	56136N	MIMOUN SOUFIANE	M	E	9	FCM	Sécurité Industrielle	SYSTEME DE PERMIS DE TR
24	DIV RAFF	RA1Z	56118J	CHAFI MOHAMMED	M	M	9	FCM	Sécurité Industrielle	SYSTEME DE PERMIS DE TR
25	DIV RAFF	RA1Z	77537M	KHELFALLAH MOHAMED	M	C	9	FCM	Sécurité Industrielle	SYSTEME DE PERMIS DE TR
26	DIV RAFF	RA1Z	63447D	HORRI NOUREDDINE	M	M	9	FCM	Sécurité Industrielle	SYSTEME DE PERMIS DE TR
27	DIV RAFF	RA1Z	63447D	HORRI NOUREDDINE	M	M	9	FCM	Sécurité Industrielle	SYSTEME DE PERMIS DE TR

Figure 4 : Exemple du fichier TBF.

PMT\_PAR\_TYPE\_DE\_FORMATION - Excel (Échec de l'activation du produit)

	A	B	C	D	E	F	G	H	I	J	K
	Structure	Type de formation	Désignation	Rappel Réalisations 2021	Prévisions notifiées 2022	Prévision de Clôture 2022	Prévisions 2023	Prévisions 2024	Prévisions 2025	Prévisions 2026	Prévisions 2027
9											
10		Formation	Effectif								
11		Recrutement (1)	Participation								
12			H/J								
13			Coût								
14		Perfectionnement (2)	Effectif				1				
15			Participation				2				
16			H/J				9				
17			Coût								
18		Reconversion (3)	Effectif								
19			Participation								
20			H/J								
21			Coût								
22	Lignes fonctionnelles (LF)	Fournisseurs (4)	Effectif								
23	Activités/Centrales		Participation								
24			H/J								
25			Coût								
26		Induction (5)	Effectif								
27			Participation								
28			H/J								
29			Coût								
30		Corporate (6)	Effectif								
31			Participation								
32			H/J								
33			Coût								

PMT PAR TYPE DE FORMATION

Figure 5 : Exemple du fichier PMT.

## Objectif de la recherche

L'objectif principale de cette recherche consiste utiliser de manière efficace la technologie dont on dispose pour l'élaboration d'une application desktop, qui soit optimale et efficiente en terme de manipulation et gestion des données présentes dans les fichiers Excel.

L'application des outils précédemment évoqués, conjuguée aux techniques de Data Science et aux connaissances acquises par le biais de la recherche et de la méthodologie, en plus de notre expertise en programmation, sera cruciale et essentiel dans la réalisation réussie de cette solution.

## Conclusion

Après avoir introduit l'entreprise et exposé en détail la problématique sous différents angles, en précisant les méthodes de recherche utilisées, nous constatons que le projet que nous étudions présente une problématique réglable. Il est envisageable de suggérer une solution adaptée qui répondra aux besoins et aux exigences formulées par le Département formation de Sonatrach.

## Chapitre 2

### Domaine de recherche

---

#### Introduction

Après avoir établi les fondements de notre projet en introduisant l'entreprise, en exposant la problématique de manière approfondie et en la traitant en tenant compte du contexte de travail, nous abordons à présent un nouveau chapitre qui se consacre à la présentation de notre domaine d'étude. Ce domaine de travail sera essentiel pour le développement de notre application de bureau.

Dans ce chapitre, nous explorerons en détail le domaine de la Data Science, une discipline qui joue un rôle central dans notre projet. La Data Science englobe un ensemble de techniques, de méthodes et d'approches permettant d'explorer et d'exploiter les données afin d'en extraire des connaissances exploitables. Grâce à cette discipline, nous serons en mesure d'analyser les données brutes disponibles dans notre entreprise et de les transformer en informations précieuses.

Nous mettrons l'accent sur les concepts fondamentaux de la Data Science, en soulignant son importance croissante dans les entreprises modernes. Nous l'explorerons sous différents angles, on parlera de son histoire, qu'est-ce qu'un Data Scientist, on discutera de la donnée et de sa récolte.

On abordera aussi ses divers domaines d'utilisation ainsi que les techniques associées, ce qui nous permettra de mener à bien notre projet de développement.

Ce chapitre servira de base solide pour notre travail à venir, en fournissant une compréhension approfondie du domaine de la Data Science et en établissant les fondements nécessaires pour développer notre application de bureau de manière efficace et pertinente.

## Définition de la Data science

La Data Science, également connue sous le nom de science des données, représente un ensemble multidisciplinaire qui allie l'inférence de données, le développement d'algorithmes et l'utilisation de technologies avancées. Ces différentes disciplines convergent dans le but de résoudre des problèmes analytiques complexes. Au cœur de cette convergence se trouvent les données, ces abondantes informations brutes stockées dans les entrepôts de données des entreprises. En pratique, la Data Science consiste à exploiter de manière innovante ces données afin de générer de la valeur commerciale.

## L'histoire de la Data science

La data science a connu une expansion fulgurante au cours des dernières décennies. Son origine remonte aux années 1960, lorsque les statisticiens ont commencé à exploiter les capacités des ordinateurs pour analyser et modéliser des données. Au fil du temps, l'évolution technologique, la prolifération des données et les avancées en matière d'apprentissage automatique ont profondément influencé le paysage de la data science.

L'avènement d'Internet et l'explosion des données numériques ont joué un rôle clé dans le développement de la data science. Dans les années 1990 et 2000, des entreprises technologiques telles que Google et Amazon ont joué un rôle de pionniers en exploitant intensivement les données pour améliorer leurs produits et services.

Les progrès rapides du matériel informatique et l'émergence de nouvelles technologies telles que le Big Data et le cloud computing ont révolutionné la manière dont nous stockons, gérons et traitons d'énormes quantités de données. Cette évolution a ouvert de nouvelles perspectives passionnantes pour la data science, offrant des opportunités sans précédent d'explorer, d'analyser et d'exploiter intelligemment les données dans différents domaines d'application.

## Définition d'un Data scientist

Un data scientist est un expert spécialisé dans l'analyse de données, possédant des compétences approfondies en mathématiques, statistiques et informatique. Son rôle consiste à recueillir, nettoyer et préparer les données en vue de leur analyse. Grâce à l'utilisation de techniques avancées telles que l'apprentissage automatique et l'intelligence artificielle, il identifie des schémas et des tendances au sein des données. Les data scientists développent des modèles prédictifs et des algorithmes



pour résoudre des problèmes complexes et prendre des décisions éclairées. De plus, ils doivent être capables de communiquer efficacement les résultats de leurs analyses aux parties prenantes concernées.

## Domaine de la Data Science

Il existe pratiquement une infinité d'applications potentielles pour la science des données. Son utilisation trouve sa pertinence partout où d'abondantes données sont générées et où des décisions doivent être prises en se basant sur ces données.

La science des données revêt une importance capitale dans divers secteurs et domaines d'activité, notamment :

- Commerce électronique et vente au détail : Elle est utilisée pour analyser les données des clients, segmenter la clientèle, recommander des produits, pratiquer une tarification dynamique, gérer les stocks et prévoir la demande.

- Finance et services bancaires : Dans ces domaines, elle est employée pour détecter les fraudes, analyser les risques, modéliser les prix des actifs financiers, personnaliser les offres et améliorer la gestion des portefeuilles.

- Santé et médecine : Elle est appliquée à la gestion des dossiers médicaux électroniques, à l'analyse des données cliniques, à la recherche biomédicale, à la prédiction des maladies, à la personnalisation des traitements et à la santé numérique.

- Marketing et publicité : Elle est utilisée pour analyser les données des clients, segmenter la clientèle, prédire les tendances du marché, personnaliser les campagnes publicitaires et optimiser les dépenses marketing.

- Sciences sociales et études démographiques : Elle est employée pour analyser les données socio-économiques, modéliser les comportements sociaux, analyser les sentiments sur les médias sociaux, étudier les migrations et planifier l'urbanisme.

- Industrie et fabrication : Dans ce domaine, elle est utilisée pour optimiser les processus de production, effectuer une maintenance prédictive, gérer la chaîne d'approvisionnement, analyser les défaillances et optimiser la qualité.

- Transport et logistique : Elle est utilisée pour optimiser les itinéraires, planifier les transports, analyser les flux de trafic, gérer les flottes et prévoir la demande.

## Les métiers et leurs rôles de la Data science

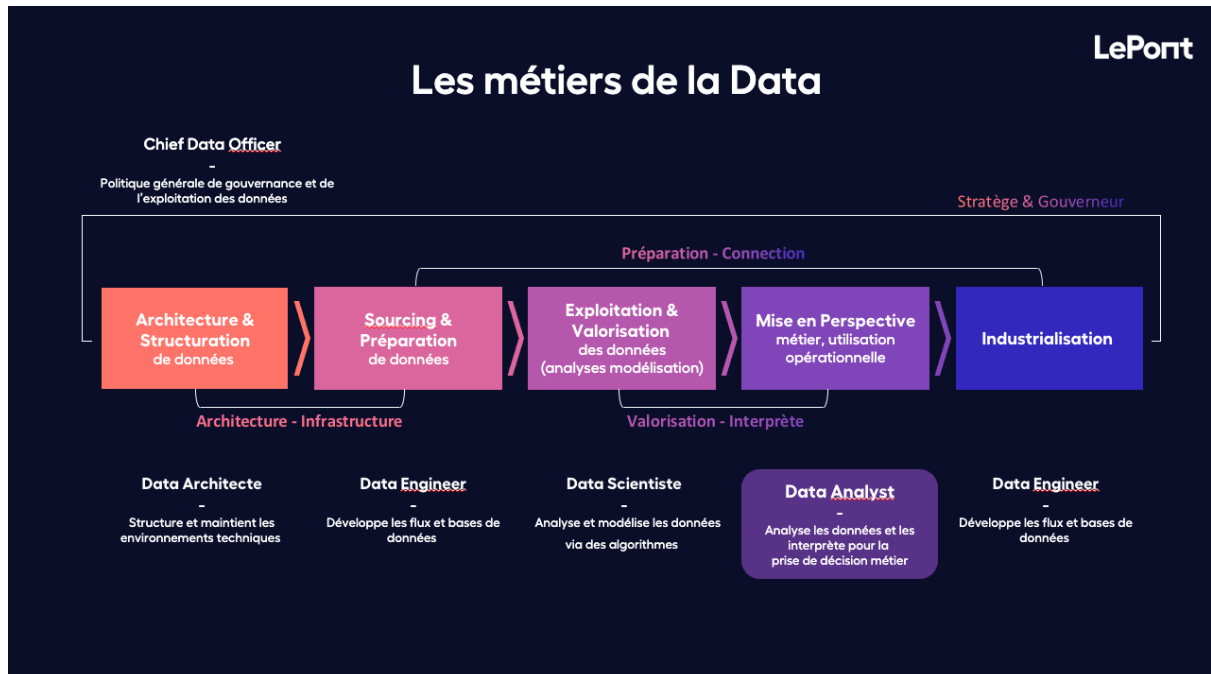


Figure 6 : Les métiers relatifs à la Data Science.

Leurs rôles :

La phase d'architecture et d'infrastructure des données englobe tout d'abord l'extraction et la structuration des données. Cette étape cruciale, réalisée par le Data Architecte, vise à assurer la mise en place et la maintenance des environnements techniques nécessaires. En effet, il est essentiel de prendre en compte les différentes méthodes d'extraction et de manipulation en fonction du volume des données à traiter.

La phase de préparation des données comprend tout d'abord le sourcing et la préparation des données. Cette étape, connue sous le nom de Data Cleaning (ou cleansing), vise à améliorer la qualité des données avant de les enrichir. C'est à ce stade que le Data Engineer intervient en développant des flux de bases de données. Il analyse et modélise les données à l'aide d'algorithmes.

La phase de valorisation et d'interprétation des données est axée sur l'exploitation et la valorisation des données, également appelée Data enrichment. Cette étape revêt une importance capitale, voire la plus importante, car c'est à ce moment que le Data Scientist utilise l'ensemble des outils et techniques à sa disposition pour donner de la valeur à la donnée (comme Croisement, combinaison, transformation, clustering des données).

En fonction du type de données et des problématiques métier auxquelles le Data Scientist est confronté, différentes méthodes peuvent être utilisées, telles que le clustering, les statistiques, le data mining ou le text mining. Selon les circonstances, la phase de Data Enrichment peut se chevaucher avec celle de l'analyse.

Pendant la phase d'analyse des données, le Data Analyst va au-delà de l'exploitation basique des données. Il peut utiliser des algorithmes d'optimisation ou d'apprentissage automatique pour effectuer des prédictions ou des recommandations.

Enfin, les méthodes et outils de visualisation des données sont essentiels pour le Data Analyst, car ils lui permettent de communiquer facilement les résultats de son analyse.

### Le rôle de la data science :

La data science joue un rôle central dans l'extraction d'informations significatives et exploitables à partir des données, en utilisant des méthodes, des outils et des techniques spécifiques. Voici un résumé des principaux rôles de la data science :

1. Analyse des données : La data science permet d'analyser les données en utilisant des techniques statistiques, des algorithmes d'apprentissage automatique et d'autres méthodes avancées. Cela révèle des modèles, des tendances et des relations cachées, fournissant ainsi des informations précieuses pour la prise de décisions.

2. Prévision et modélisation : Grâce à des techniques de modélisation statistique et d'apprentissage automatique, la data science permet de créer des modèles prédictifs. Ces modèles peuvent être utilisés pour anticiper des événements futurs, estimer des valeurs inconnues ou optimiser des processus.

3. Optimisation des processus : La data science identifie les inefficacités, les goulots d'étranglement et les opportunités d'amélioration dans les processus opérationnels. En analysant les données et en identifiant les facteurs clés, les data scientists recommandent des mesures correctives pour optimiser les performances et réduire les coûts.

4. Prise de décisions basée sur des données : La data science fournit des informations factuelles et quantitatives pour éclairer la prise de décisions. Les data scientists analysent les données pertinentes et fournissent des recommandations fondées sur des preuves solides, réduisant ainsi les conjectures et les décisions basées uniquement sur l'intuition.

5. Développement de produits et de services : La data science joue un rôle essentiel dans le développement de produits et de services innovants. En analysant les données des utilisateurs, les data scientists comprennent les besoins, les préférences et les comportements des clients, ce qui permet de créer des produits et des services adaptés à leurs attentes.

6. Analyse des tendances et du marché : La data science analyse les tendances du marché, prévoit les comportements des consommateurs et anticipe les demandes futures. Ces informations sont précieuses pour élaborer des stratégies commerciales, planifier les ressources et prendre des décisions marketing.

### Les avantages qu'offre la data science

L'utilisation et l'implémentation de la Data Science offre une multitude d'avantages :

1. Optimisation de la chaîne d'approvisionnement pour améliorer l'efficacité et la fiabilité des processus logistiques.

2. Amélioration de la qualité des produits grâce à l'identification des problèmes potentiels, l'analyse approfondie des données et l'application de techniques de contrôle de la qualité basées sur les données.

3. Satisfaction des besoins des clients en comprenant mieux leurs préférences, leurs comportements d'achat et leurs attentes, permettant ainsi la personnalisation des produits et services.

4. Détermination des paramètres commerciaux clés tels que les prix optimaux, les stratégies de tarification, les promotions efficaces et les prévisions de la demande en utilisant l'analyse des données et les modèles prédictifs.

5. Amélioration de la rétention des employés en identifiant les facteurs influençant leur satisfaction et engagement, en mettant en place des mesures appropriées pour favoriser leur fidélisation et en anticipant les risques de départ.

En exploitant la puissance de la science des données, les entreprises peuvent prendre des décisions plus éclairées, optimiser leurs opérations et stratégies, et ainsi acquérir un avantage concurrentiel sur le marché.

## Types de données utilisées en Data science

La Data Science utilise plusieurs types de données au cours de son traitement, chaque donnée est différente et est structurée différemment, voici le type de données utilisées :

- Le Big data représente des énormes volumes d'informations incompatibles avec les bases de données relationnelles standards.
- Les données peuvent être structurées, non structurées ou semi-structurées, se différenciant par leur modèle et leur organisation prédéfinis.
- Les données horodatées sont ordonnancées selon la séquence de capture ou de collecte.
- Les données machine sont les émissions numériques générées par les systèmes, les technologies et les infrastructures des entreprises.
- Les données spatio-temporelles décrivent à la fois le lieu et l'heure d'un événement, permettant d'observer les changements dans un lieu physique au fil du temps.
- Les données spatiales décrivent des emplacements ou des trajectoires, formant des objets géographiques comme des routes ou des empreintes de bâtiments.
- Les données temporelles contiennent des informations de date et d'heure dans un horodatage, distinguant le temps réel écoulé du moment où un fait est connu dans la base de données.
- Les données en Open data sont librement accessibles et peuvent être utilisées avec possibilité d'analyse, mais elles doivent être partagées dans un format standardisé et traçable, sans restrictions de droits d'auteur ou de brevets.

## Les sources de données et leur collecte

La collecte de données en data science varie en fonction des différentes sources de données. Ce paragraphe présente les principales sources de données et de leurs méthodes de collecte :

1. Données structurées : Ces données sont organisées de manière prévisible dans des tableaux, des bases de données relationnelles ou des fichiers CSV. Elles peuvent être collectées à partir de sources internes telles que les systèmes d'information de l'entreprise ou les bases de données, ou à partir de sources externes telles que des sites web publics ou des fournisseurs de données tiers. La collecte de ces

données se fait généralement via des requêtes SQL, des scripts d'extraction de données ou des outils d'intégration de données.

2. Données non structurées : Il s'agit de données dans des formats tels que le texte, les images, les vidéos, les fichiers audio, les e-mails, les médias sociaux, les documents PDF, etc. La collecte de ces données peut être réalisée à l'aide de techniques de web scraping pour extraire des informations à partir de sites web, d'outils d'analyse de texte pour traiter des documents textuels, de la reconnaissance d'images ou de la vision par ordinateur pour analyser des images, ou de techniques de traitement du signal pour les données audio et vidéo.

3. Données en temps réel : Ces données sont générées en continu et nécessitent une collecte et une analyse en temps réel. Elles peuvent provenir de capteurs, de dispositifs IoT (Internet des objets), de flux de médias sociaux, de transactions financières en direct, etc. La collecte de ces données se fait généralement en utilisant des technologies de streaming de données telles qu'Apache Kafka, Apache Flink ou des API spécifiques.

4. Données externes : Ces données proviennent de sources tierces telles que des fournisseurs de données, des gouvernements, des organisations de recherche, des API publiques, etc. Elles peuvent être collectées via des requêtes d'API, des téléchargements de jeux de données, des services de diffusion de données ou des accords de partage de données avec des partenaires externes.

5. Données des réseaux sociaux : Les données provenant des plateformes de médias sociaux telles que Twitter, Facebook, Instagram, etc. peuvent être collectées en utilisant des API spécifiques aux médias sociaux qui permettent d'extraire des données telles que des publications, des commentaires, des likes, des profils d'utilisateurs et autres.

## Technique et outils de la data science

La data science tire parti d'une gamme variée de techniques et d'outils pour la collecte, le nettoyage, l'analyse et l'interprétation des données. Voici un aperçu des techniques et outils couramment utilisés dans ce domaine :

**Langages de programmation :** Python et R sont largement adoptés en data science, offrant des bibliothèques spécialisées telles que NumPy, Pandas, scikit-learn, TensorFlow et Keras. Ces outils facilitent la manipulation des données, l'analyse statistique, la visualisation et la création de modèles d'apprentissage automatique.

**Manipulation de données :** Des outils tels que SQL, Pandas et dplyr permettent de filtrer, trier, agréger et transformer les données, les préparant ainsi pour l'analyse.

**Analyse statistique :** Les techniques statistiques, comme les tests d'hypothèses, les analyses de variance, les régressions linéaires et logistiques, ainsi que les analyses de séries temporelles, sont utilisées pour explorer les relations, identifier les modèles et tester les hypothèses sur les données.

**Apprentissage automatique :** L'apprentissage automatique utilise des algorithmes pour entraîner des modèles à partir des données et les utiliser pour effectuer des prédictions ou prendre des décisions. Les techniques d'apprentissage automatique incluent la régression, la classification, les arbres de décision, les réseaux de neurones, le clustering et les machines à vecteurs de support (SVM), entre autres.

**Deep Learning :** Cette branche de l'apprentissage automatique utilise des réseaux de neurones profonds pour résoudre des problèmes complexes tels que la vision par ordinateur, la reconnaissance vocale, la traduction automatique et la génération de texte.

**Visualisation des données :** Des outils tels que Matplotlib, Seaborn, ggplot et Tableau permettent de créer des graphiques interactifs, des diagrammes et des tableaux de bord pour communiquer les résultats et les insights obtenus à partir des données.

**Big Data :** Les outils de traitement du Big Data, comme Apache Hadoop, Apache Spark et les bases de données NoSQL, sont utilisés pour gérer et analyser de grandes quantités de données massivement parallèles, y compris des données en temps réel, non structurées et distribuées sur des clusters informatiques.

**Text Mining et NLP :** Les techniques de text mining et de traitement du langage naturel (NLP) permettent d'extraire des informations significatives à partir de données textuelles, réalisant des tâches telles que la classification de texte, la détection de sentiments, la reconnaissance d'entités nommées et la génération automatique de texte.

Automatisation et workflows : Des outils tels qu'Apache Airflow, KNIME et Alteryx permettent de créer des workflows automatisés pour gérer et exécuter efficacement des tâches de data science de manière reproductible.



Figure 7 : Cycle de vie de la Data Science



## Les enjeux et défis de la data science

La data science présente divers enjeux et défis significatifs, parmi lesquels :

**Qualité des données :** L'un des défis majeurs de la data science réside dans la qualité des données. Cela implique des données incomplètes, inexactes, erronées ou comportant des valeurs aberrantes, ce qui peut affecter la fiabilité des résultats. Assurer la collecte de données de haute qualité et mettre en place des processus de nettoyage rigoureux sont essentiels pour garantir la fiabilité des analyses et des modèles.

**Confidentialité et protection des données :** Avec la préoccupation croissante concernant la confidentialité des données personnelles, la protection des données constitue un enjeu majeur. Les data scientists doivent se conformer aux réglementations en matière de confidentialité des données, telles que le Règlement général sur la protection des données (RGPD), et prendre des mesures pour garantir la sécurité des données lors de leur collecte, stockage et traitement.

**Éthique de la data science :** L'utilisation des données peut soulever des questions éthiques complexes, notamment en ce qui concerne le biais, la discrimination et la prise de décisions automatisées. Les data scientists doivent être conscients de ces questions et travailler à minimiser les biais et à assurer une utilisation éthique des données dans leurs analyses et leurs modèles.

**Complexité des données :** Les données peuvent être volumineuses, non structurées, hétérogènes et complexes. Les data scientists doivent être capables de gérer et de traiter efficacement ces données en utilisant des techniques et des outils appropriés, tels que le Big Data et les outils d'analyse distribuée, afin d'extraire des informations pertinentes.

**Interprétation des résultats :** L'interprétation des résultats de l'analyse des données peut s'avérer complexe. Les data scientists doivent être en mesure de comprendre les modèles, les relations et les prédictions générés par les modèles d'apprentissage automatique, et de les expliquer de manière claire et compréhensible aux parties prenantes, notamment lorsqu'il s'agit de prises de décisions critiques.

**Évolution technologique rapide :** La data science est un domaine en constante évolution, avec l'émergence de nouvelles techniques, méthodes et outils. Les data scientists doivent rester à jour avec les avancées technologiques, acquérir de nouvelles compétences et s'adapter rapidement aux changements pour rester compétitifs.

Collaboration interdisciplinaire : La data science nécessite souvent une collaboration étroite entre les data scientists, les experts métiers et les professionnels des technologies de l'information. Une communication efficace et une compréhension mutuelle des enjeux et des besoins de chaque partie sont essentielles pour réussir les projets de data science.

## Le future de la Data science

Le champ de la data science présage un avenir extrêmement prometteur, offrant des opportunités exceptionnelles. Grâce à l'évolution rapide de l'intelligence artificielle, de l'automatisation et de l'analyse en temps réel, la data science continuera de révolutionner de nombreux secteurs et de façonner notre société. Les avancées technologiques permettront des analyses de données plus sophistiquées et des prédictions plus précises, débloquent de nouvelles découvertes et des solutions innovantes. Cependant, il est primordial de maintenir un équilibre en intégrant des considérations éthiques et une gouvernance rigoureuse des données pour garantir la transparence, la responsabilité et la protection de la vie privée. La collaboration interdisciplinaire et l'exploitation de nouvelles sources de données joueront également un rôle crucial dans l'évolution de la data science. En fin de compte, la data science continuera de transformer notre manière de travailler, de prendre des décisions et d'interagir avec le monde qui nous entoure, offrant ainsi d'innombrables opportunités d'innovation et de progrès.

## Chapitre 3 - Implémentation et conception

---

### Introduction

L'utilisation de la Data Science est devenue chose courante de nos jours pour la majorité des pays avancés, certains même n'arrivent plus à s'en passer puis ce quelle réponds à diverses demandes dans divers domaines plus innovateurs les uns que les autres. De plus, puisque ce domaine est nouveau, il ne cesse d'innover et de se développer.

A ce jour, il propose des solutions dans maintes professions, il réduit le temps de travail, améliore la précision, facilite la manipulation de données et optimise l'analyse. Il est donc primordial que les grandes sociétés algériennes comme Sonatrach commencent à l'implémenter au sein de leurs entreprises.

Notre travail se porte sur la manipulation et la gestion des fichiers et des données au sein de Sonatrach. L'objectif de ce travail est de profiter des avantages que nous octroie la Data Science et ses diverses utilisations ; dans notre cas : le traitement de données, minimalisation des fautes, repérage des incohérences et la facilité et rapidité d'insertion de données.

Pour atteindre cet objectif, nous avons fait appel au langage Python, réputé pour son efficacité et sa richesse en bibliothèques utiles dans le domaine de la Data Science. Ainsi, nous avons pu tirer parti d'un éventail de nouvelles méthodes et techniques que nous avons implémentées et utilisées dans notre programme pour répondre à nos besoins spécifiques.

Le présent chapitre est organisé de la manière suivante : Dans la première section, nous procéderons à la présentation des outils de développement, en discutant notamment de l'environnement de développement intégré (IDE) choisi. Ensuite, nous aborderons le langage de programmation utilisé, ainsi que ses bibliothèques associées, qui jouent un rôle clé dans notre approche. Enfin, nous explorerons le Framework qui encadre notre démarche.

Dans la deuxième section, nous examinerons en détail l'application que nous avons développée, en mettant l'accent sur ses programmes et les fonctions spécifiques que nous avons mises en œuvre. Nous fournirons des explications approfondies sur la logique et le fonctionnement de notre application, en mettant en évidence ses contributions et son utilité pratique.

En conclusion, nous dresserons un bilan de notre recherche, en soulignant les principales réalisations et découvertes. Nous discuterons également des perspectives envisagées pour l'avenir, en identifiant les domaines d'amélioration potentiels et les opportunités de développement ultérieur.

## Présentatin du système

### *1. Éléments caractéristiques de l'environnement de développement :*

L'environnement de développement adopté se fonde sur Windows 10 professionnel, accompagné de l'IDE (Integrated Development Environment) VS code, établissant un appui solide pour le langage de programmation Python et son Framework associé, à savoir Tkinter.

### *2. Les outils de développement :*

L'instauration d'un tel système requiert l'usage d'un langage de programmation sophistiqué ainsi que d'un Framework approprié. Par conséquent, après de multiples investigations, notre choix s'est porté sur Python en tant que langage de programmation, accompagné de Tkinter en tant que Framework.

Cette section du chapitre se déploiera de la manière suivante : tout d'abord, nous exposerons l'environnement de développement intégré (IDE), puis nous présenterons en détail notre langage de programmation, en mettant en évidence les diverses bibliothèques utilisées, pour enfin conclure par la présentation exhaustive du Framework.

### *L'IDE VS code*

Visual Studio Code, communément appelé VS Code, est un environnement de développement intégré (IDE) contemporain et polyvalent présenté pour la toute première fois lors de la conférence des développeurs Build d'avril 2015 comme un éditeur de code multiplateforme, open source et gratuit, supportant une dizaine de langages. Grâce à son interface conviviale et instinctive, il procure aux développeurs une plateforme éminente pour rédiger, éditer et déboguer du code. Avec une pléthore d'extensions et de fonctionnalités sur mesure, VS Code accroît l'efficacité et la productivité tout au long du processus de développement. Qu'il s'agisse de programmation web, d'analyse de données, d'intelligence artificielle ou de tout autre projet, VS Code est un compagnon fidèle et performant qui facilite la création logicielle et offre une expérience de programmation d'une remarquable fluidité.

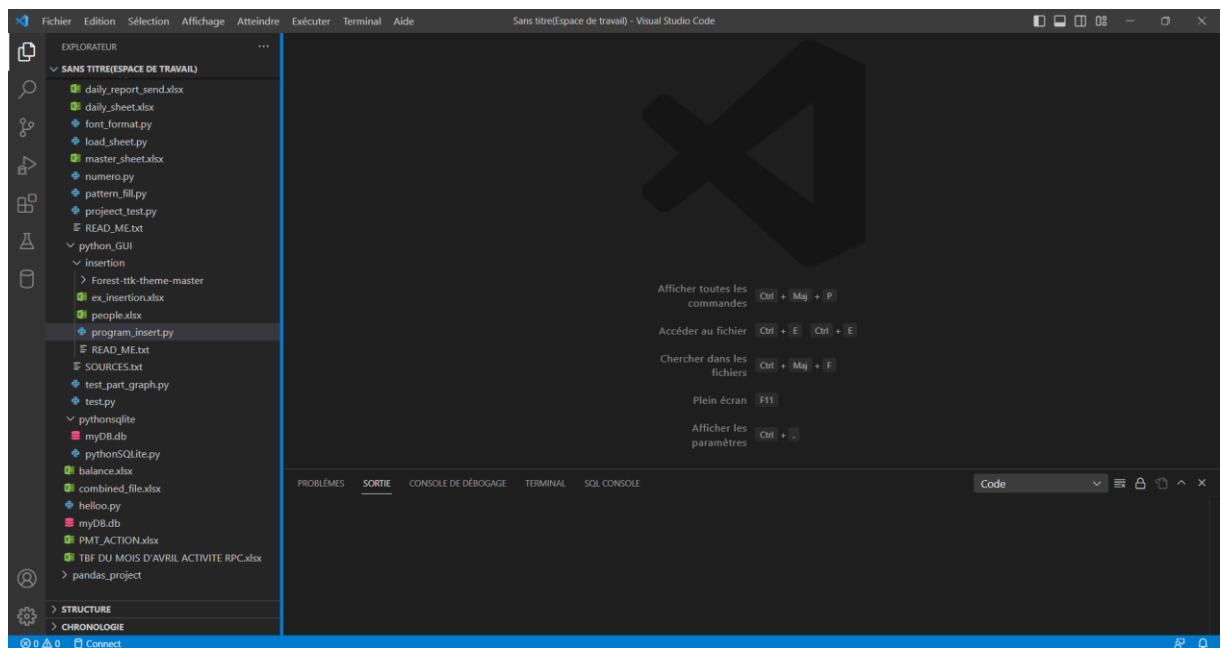


Figure 8 : La fenêtre principale de Visual Studio Code.

### Extensions :

Les extensions de Visual Studio Code offrent une multitude de fonctionnalités supplémentaires pour améliorer l'expérience de programmation. Pour les développeurs Python, plusieurs extensions sont particulièrement utiles, voici celles qu'on a utilisé :

L'extension "autoDocstring" simplifie la génération automatique de docstrings pour les fonctions et les classes. En écrivant simplement une ligne de commentaire, cette extension peut générer une documentation complète, ce qui facilite la compréhension du code par les autres développeurs.

"Code Runner" permet d'exécuter rapidement du code Python directement depuis l'éditeur. Il suffit de sélectionner le code à exécuter, puis d'utiliser un raccourci clavier pour le lancer. Cela permet de tester rapidement des portions de code sans quitter l'interface de développement.

L'extension "Jupyter" permet d'utiliser des notebooks Jupyter directement dans Visual Studio Code. Les notebooks Jupyter offrent une manière interactive et visuelle d'exécuter du code, d'afficher des résultats et de documenter le processus de développement. Cette extension facilite l'utilisation des notebooks Jupyter pour les projets Python.

"Pylance" est une extension qui améliore les fonctionnalités d'auto complétion, de suggestion de code et de débogage pour Python. Elle offre également des informations contextuelles sur les types de données, ce qui facilite la compréhension du code et aide à éviter les erreurs.

L'extension "Python" fournit un support complet pour le développement Python dans Visual Studio Code. Elle offre des fonctionnalités telles que la coloration syntaxique, la vérification des

erreurs, la navigation dans le code et l'intégration avec des outils externes tels que les testeurs unitaires.

Enfin, l'extension "python indent" améliore l'indentation automatique du code Python dans l'éditeur. Elle garantit une indentation cohérente et correcte, ce qui facilite la lecture et la compréhension du code.

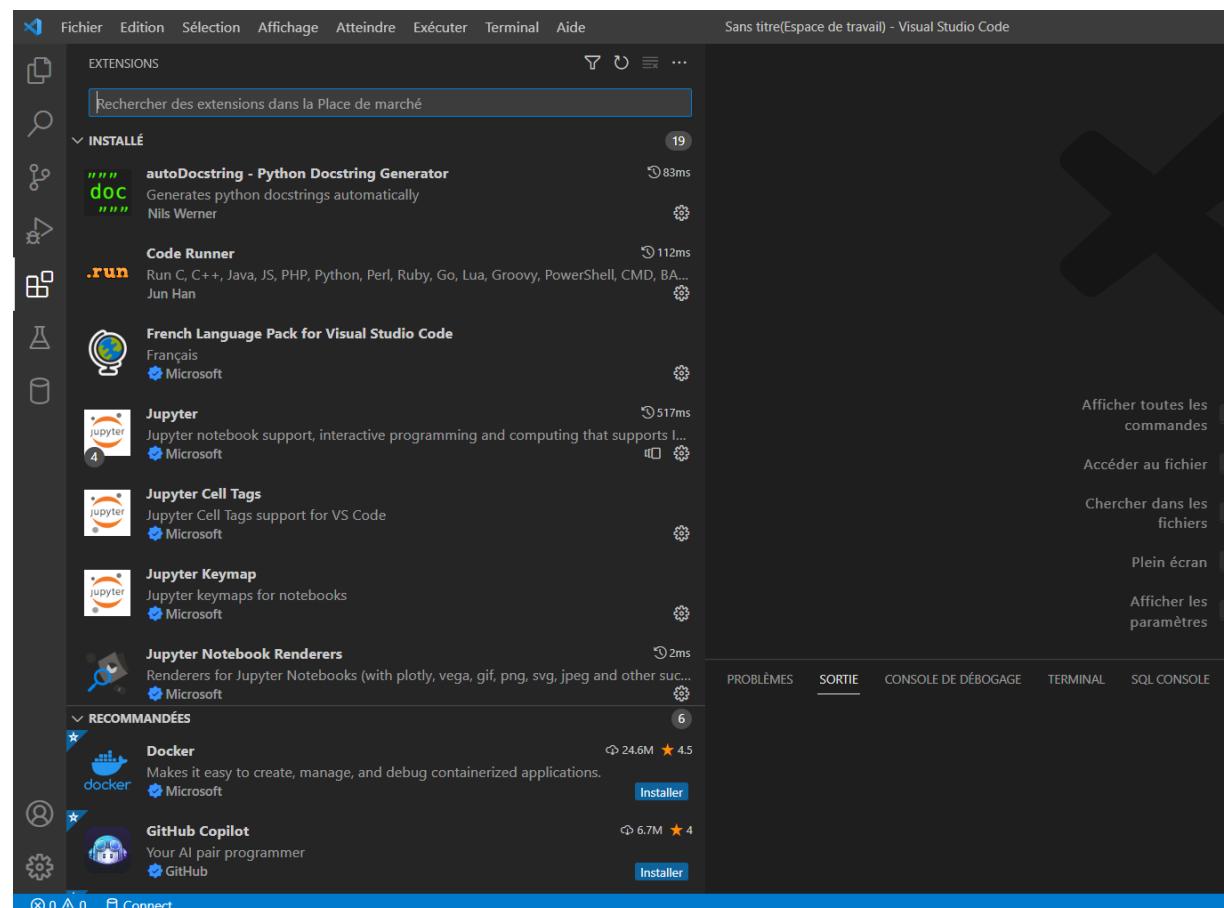


Figure 9 : Menu d'extensions Visual Studio

### Avantages de son utilisation :

Visual Studio Code octroie à ses utilisateurs maintes avantages quand ils attaquent le développement de leurs applications, tout comme :

*Collaborate and code remotely* : cette fonction permet de collaborer à distance avec vos enseignants ou camarades de classe grâce à l'extension gratuite Live Share. Elle permet d'effectuer des modifications et déboguer votre code en temps réel, tout en permettant l'utilisation des fonctionnalités de discussion et d'appel pour poser des questions ou échanger des idées collectivement. Cette fonction vous octroie la possibilité d'inviter plusieurs personnes à rejoindre votre session et à coder ensemble.

*Easy code fix* : Cette fonctionnalité rends le l'IDE user friendly, que vous soyez ancien ou débutant en programmation Visual Studio Code met en évidence les mots-clés de votre code ainsi que les fonctions avec différentes couleurs afin de vous aider à mieux identifier les motifs de programmation et à apprendre plus rapidement. De plus, vous pouvez bénéficier des fonctionnalités avancées telles que IntelliSense et Peek Definition qui vous permettent de mieux appréhender l'utilisation des fonctions et leur interrelation.

*Corrigez les erreurs pendant que vous codez* : Lorsque vous rédigez du code, VS code vous offre des suggestions et des correctifs pour compléter et rectifier votre code. De plus, l'utilisation du débogueur permet d'examiner chaque ligne de code et d'appréhender son fonctionnement.

*Comparaison de modification* : On peut profiter de la fonction de contrôle de source intégrée pour sauvegarder notre travail au fur et à mesure afin d'éviter de perdre nos avancées. Il permet de visualiser les différentes versions de votre code côte à côte pour faciliter la comparaison de modification.

*Code direct à partir du Notebook* : Si on souhaite tester un projet en Data Science ou visualiser les données on peut utiliser les fonctions du Notebook de Jupyter directement à partir de VS code, on peut exécuter notre code, visualiser et interagir avec nos données graphes et autres.

## *Le langage de programmation Python*

### Présentation :

Python est un langage de programmation qui a été créé par Guido van Rossum en l'an 1991 et a été rendu public sur le forum Usenet alt.sources. Avec sa syntaxe à la fois simple et élégante, il offre à ses utilisateurs la possibilité de créer des programmes dans divers domaines tels que le développement d'applications web, l'analyse de données, ainsi que la programmation et le contrôle de robots. Quel que soit le domaine d'application, Python propose une solution adaptée.

La force de Python réside également dans sa vaste communauté de développeurs, qui met à disposition de nombreuses ressources et bibliothèques prêtes à l'emploi. Cette richesse permet de gagner du temps et d'accélérer le processus de développement de projets.

Python est également reconnu pour sa facilité d'apprentissage, ce qui en fait un choix idéal pour les débutants en programmation. Son approche privilégiant la simplicité encourage les bonnes pratiques de programmation, facilitant ainsi l'écriture de code clair et organisé.

En outre, Python offre un environnement convivial et flexible pour la programmation, ce qui s'avèrera très utile pour mener à bien nos futurs projets.

### Bibliothèques utiles :

Python nous confère une pléthore d'options en matière de programmation grâce à sa vaste communauté de développeurs, qui met à notre disposition un éventail considérable de bibliothèques facilitant la programmation.

Dans notre cas, notre attention se porte davantage sur les bibliothèques en corrélation avec notre projet de développement axé sur la Data Science. Voici une présentation plus détaillée des principales bibliothèques que nous allons exploiter :

**Openpyxl** : Cette bibliothèque offre une solution pratique pour lire et écrire des fichiers Excel. Elle nous permet d'extraire des données à partir de feuilles de calcul existantes, de les modifier et de créer de nouveaux classeurs Excel. Avec Openpyxl, nous pouvons effectuer des opérations complexes sur des données tabulaires, telles que le tri, le filtrage et la mise en forme conditionnelle.

**Pandas** : Considérée comme une référence dans le domaine de l'analyse de données, Pandas fournit des structures de données puissantes, notamment les DataFrames, qui permettent de manipuler facilement des données tabulaires. Cette bibliothèque offre des fonctionnalités avancées pour le nettoyage des données, la fusion de datasets, le calcul de statistiques descriptives et bien plus encore. Grâce à Pandas, nous pouvons effectuer des transformations complexes sur nos données avec une syntaxe concise et intuitive.

**Numpy** : En tant que fondement de nombreuses bibliothèques de calcul scientifique en Python, Numpy offre des fonctionnalités pour la manipulation de tableaux multidimensionnels et des opérations mathématiques avancées. Grâce à Numpy, nous pouvons effectuer des calculs vectorisés rapides et efficaces, ainsi que des opérations linéaires et algébriques sur les tableaux. Cette bibliothèque est essentielle pour les tâches nécessitant des calculs numériques complexes, tels que le traitement d'images, l'apprentissage automatique et la modélisation mathématique.

**Matplotlib** : En matière de visualisation de données, Matplotlib est une bibliothèque incontournable. Elle offre une large gamme de fonctionnalités pour la création de graphiques et de visualisations, allant des simples graphiques en barres aux graphiques en nuage de points complexes. Avec Matplotlib, nous pouvons personnaliser tous les aspects de nos graphiques, tels que les axes, les légendes, les couleurs et les styles. Cette bibliothèque nous permet de représenter visuellement nos données de manière claire et percutante, facilitant ainsi l'interprétation et la communication des résultats.

**OS** : La bibliothèque OS de Python est un outil puissant pour interagir avec le système d'exploitation. Elle nous permet de manipuler des fichiers, gérer les répertoires et travailler avec les variables d'environnement. La bibliothèque OS est portable et compatible avec différentes plateformes, offrant ainsi une solution fiable et flexible pour la gestion des fichiers et les interactions avec le système d'exploitation.

Ces bibliothèques Python, parmi d'autres, sont des outils précieux qui nous permettent d'explorer, analyser et visualiser les données de manière approfondie. Leur combinaison nous offre une puissance de calcul, une flexibilité et une expressivité qui nous permettent de mener à bien nos projets avec efficacité et précision.



## *Le Framework Tkinter*

### *Présentation :*

Tkinter est un Framework Python largement utilisé pour créer des interfaces graphiques conviviales. Il est inclus dans la distribution standard de Python, ce qui signifie qu'il est facilement accessible sans nécessiter d'installation supplémentaire. Tkinter s'appuie sur le toolkit Tk pour développer des interfaces graphiques multiplateformes. Il offre une large gamme de widgets prêts à l'emploi et permet une organisation flexible de ces widgets pour créer des interfaces utilisateur attrayantes. Tkinter est apprécié pour sa simplicité d'utilisation, sa portabilité sur différentes plates-formes, ainsi que ses fonctionnalités avancées telles que la gestion des événements et la personnalisation visuelle. En résumé, Tkinter est un choix populaire pour créer des interfaces graphiques interactives et esthétiques en Python.

### *Modules et classes :*

Tkinter offre une multitude de classes qui facilitent la création d'interfaces graphiques riches en fonctionnalités. Parmi ces modules, voici ceux qu'on a utilisé :

Le module `messagebox` : il fournit des boîtes de dialogue standard pour afficher des messages à l'utilisateur. Il offre des options telles que les boîtes de dialogue d'information, de confirmation, d'avertissement et d'erreur, qui permettent de communiquer efficacement avec l'utilisateur.

Le module `simpledialog` permet de créer des boîtes de dialogue pour obtenir des entrées de l'utilisateur. Il propose des boîtes de dialogue simples pour demander des valeurs numériques, des chaînes de caractères ou d'autres types de données spécifiques, facilitant ainsi l'interaction avec l'utilisateur.

`Customtkinter` est une extension de Tkinter qui offre des fonctionnalités supplémentaires pour personnaliser et styliser les widgets. Il permet de créer des widgets personnalisés, d'appliquer des styles personnalisés et de modifier le comportement par défaut des widgets.

Le module `ttk` (Themed Tkinter) fournit des widgets avec une apparence plus moderne et cohérente sur différentes plates-formes. Il utilise les styles du système d'exploitation pour rendre les interfaces graphiques plus attrayantes et professionnelles.

Le module `filedialog` permet de créer des boîtes de dialogue pour la sélection et l'enregistrement de fichiers. Il facilite l'interaction avec le système de fichiers, permettant à l'utilisateur de naviguer dans les dossiers, de sélectionner des fichiers et de spécifier des emplacements de sauvegarde.

## Modélisation

La modélisation est l'étape la plus importante de toutes études de performances. Elle permet de représenter les différents aspects à étudier en faisant abstraction du système réel. Dans notre cas, il s'agit de créer une application desktop qui traite des fichiers Excel.

Notre système aura besoin d'une interface graphique appelé main, dans laquelle on trouveras plusieurs options qui nous seraient utiles pour aboutir à une application concrète, on devra donc créer les programmes et fonctions suivantes :

a.           Fonction sélection et affichage : On doit permettre à l'utilisateur de visualiser son travail, et donc le fichier Excel traité. Cette fonction devra donc permettre à l'utilisateur de choisir un fichier Excel et de l'afficher dans l'application.

b.           Fonction repérage des doublons et duplicates : dans cette fonction, l'application devra permettre à l'utilisateur de choisir un fichier Excel, choisir une colonne à traiter pour repérer et mettre en évidence les doublons existants et pouvoir ensuite les corriger.

c. Fonction de fusion : Cette fonction devra donner la possibilité à l'utilisateur de fusionner deux fichiers Excel pour avoir un fichier combiné des deux.

d.           Fonction de somme : Cette fonction devra permettre à l'utilisateur d'additionner et ajouter des valeurs d'une même colonne pour avoir une somme globale.

e.           Fonction de tri : Dans cette fonction, on doit permettre à l'utilisateur de trier des lignes dans Excel dans un ordre précis (croissant ou décroissant) pour lui faciliter l'organisation de ses données dans Excel.

f.           Fonction repérage des vides : Pour minimiser les erreurs, l'utilisateur aura besoin d'une fonction pour repérer les vides et les oublis d'ajout de valeur dans son fichier Excel pour minimiser les erreurs et les incohérences durant son traitement des multiples fichiers qu'il a à gérer.

g.           Fonction d'insertion : Lors de l'exécution de cette fonction, elle doit permettre à l'utilisateur d'insérer des valeurs dans son tableau Excel avec une rapidité et une facilité d'insertion tout en évitant les fautes qu'il peut commettre avec une insertion manuelle.

## Implémentation

L'implémentation est l'étape où l'on met en pratique nos notions de développement tout en les expliquant, dans cette phase on va expliquer comment on a implémenté et introduit chaque fonction dans notre programme en expliquant les méthodes utilisées :

### *Programme « main » :*

Ce programme utilise le module Tkinter afin de créer une interface graphique globale pour nos programmes Python. Voici une description des fonctionnalités et des boutons présents dans cette interface :

#### I. Bibliothèques de ce code :

**Tkinter** : on a importé d'abord la bibliothèque Tkinter avec deux modules qui sont :  
**filedialog** : Ce module nous permet de communiquer avec l'utilisateur en lui fournissant des boîtes de dialogues afin de sélectionner des fichiers ou des répertoires à l'aide de l'interface graphique.

**Tk** : Ce module nous fournit les éléments de base pour la création d'une interface graphique, comme des fenêtres, des boutons, des entrées, des listes déroulantes, des cases à cocher et autres.

**Ttk** : Ce module nous permet d'utiliser des thèmes customisés sur tkinter en fournissant des widgets avec un aspect plus moderne et cohérent, il permet de personnaliser les styles des widgets ttk en utilisant les objets de style (.style) pour modifier la taille, la couleur et les aspects de nos widgets en fournissant des versions modernes (Pour les boutons, check box, les entrées, treeviews ect...).

**Pandas** : on a importé cette bibliothèque afin de manipuler les fichiers Excel, on l'a importé sous l'alias « pd » pour utiliser ses fonctionnalités.

**Subprocess** : L'utilisation de cette bibliothèque nous permet d'exécuter des commandes système et des programmes externes en interagissant avec eux depuis notre programme.

#### II. Elaboration du code :

Dans cette phase on va discuter et expliquer notre code :

- a. On crée une fenêtre principale à l'aide de la fonction « root = tk.TK », ce code fait appel au module (tk) pour créer une instance de la classe Tk qui représente la fenêtre principale.

Ensuite on utilise « root.title("Gestion et manip générale") » pour donner un nom à la fenêtre qui sera « Gestion et manip générale ». On utilise aussi « root.geometry("700x400") » pour définir la taille de cette fenêtre.

- b. On utilise la « def » pour définir les fonctions qu'on va utiliser dans ce programme sous la forme : **def nom ():** *Instructions*.

On crée la première fonction avec def qui va s'appeler « select\_file » qui permet de sélectionner un fichier en utilisant la fonction Tkinter « .askopenfilename() » qui possède l'option « filetypes » pour préciser le format du fichier à sélectionner donc Excel (.xlsx) dans notre cas qu'on va attribuer à **file\_path**.

On utilise ensuite une boucle pour que si file\_path n'est pas une chaîne vide elle exécute ce code :

Utiliser la bibliothèque **pandas** pour lire le fichier Excel sélectionné avec l'aide de la fonction « read\_excel() ».

Les données du fichier Excel sont stockées dans un objet DataFrame appelé **df**. Ensuite, les headers des colonnes du DataFrame sont extraites à l'aide de la méthode « columns.tolist() » pour obtenir une liste de noms de colonnes qui sera assignée à la variable **header**. Les données du DataFrame seront converties à une liste assignée à la variable data à l'aide de « values.tolist() ». Enfin, on fait appel à la fonction « create\_table(header, data) » pour les headers et les données des colonnes dans un tableau dans le GUI.

La seconde fonction def fait appel à « run\_file(file\_path) » qui prends file\_path comme le chemin d'accès à un fichier. Ensuite on utilise la fonction **subprocess** pour exécuter file\_path avec la commande python.

La troisième fonction def définit une fonction appelée « create\_table » qui va être utilisée pour créer un tableau dans notre interface graphique. Elle crée d'abord une séparation verticale à l'aide du module « ttk.Separator » pour la placer dans la fenêtre principale. A l'aide de (tk.Frame) on paramètre son positionnement.

Un module « treeview » est ensuite créé pour afficher les headers des colonnes.

On configure ensuite la taille des colonnes à l'aide de « treeview.column » et « treeview.headers ».

On ajoute aussi un scrollbar qui nous permet de défiler les données dans notre tableau à l'aide de « ttk.scrollbar » en définissant son axe et sa commande.

On crée ensuite une boucle **for** pour insérer les données dans notre tableau à partir des données du fichier de file\_path à l'aide de la fonction « treeview.insert ».

```
def select_file():
    file_path = filedialog.askopenfilename(filetypes=[("Excel Files", "*.xlsx")])
    if file_path:
        #lire le fichier excel et l'afficher
        df = pd.read_excel(file_path)
        header = df.columns.tolist()
        data = df.values.tolist()
        create_table(header, data)
```

Figure 10 : Exemple création de la fonction « Def ».

- c. Pour la création des boutons et leurs programmations, on utilise la classe « tk.Button » pour créer les boutons et les placer dans la fenêtre (root) tout en utilisant la fonction (.grid) pour préciser leurs positionnement.
- Chaque bouton a un texte affiché et une commande associée qui sera exécutée lorsque le bouton sera cliqué. La commande est définie en utilisant la syntaxe (lambda) pour créer une fonction anonyme qui appelle la fonction « run\_file » avec un chemin de fichier spécifique en argument.
- Sur notre interface on aura 7 boutons (Sélectionner le fichier, Trouver doublons, Fusionner deux fichiers, Total Global, Trier les valeurs, Repérage des vides, Insertion). Lorsque l'utilisateur clique sur un bouton, la fonction « run\_file » qui exécute le chemin du fichier correspondant, ce qui déclenchera l'exécution du fichier Python.

```
#bouton: Select File
select_file_btn = tk.Button(root, text="Sélectionner le fichier", command=select_file)
select_file_btn.grid(row=1, column=0, padx=10, pady=5, sticky="w")

#bouton: Doublons
doublons_btn = tk.Button(root, text="Trouver doublons",
                        command=lambda: run_file(r"C:\Users\user\Documents\PYTHON\python_
doublons_btn.grid(row=2, column=0, padx=10, pady=5, sticky="w")
```

Figure 11 : Exemple création de bouton.

- d. La fonction « root.mainloop() » lance la boucle principale d'événements de l'application Tkinter. Elle permet d'exécuter en continu les événements et de gérer les interactions de l'utilisateur dans l'interface graphique (GUI).

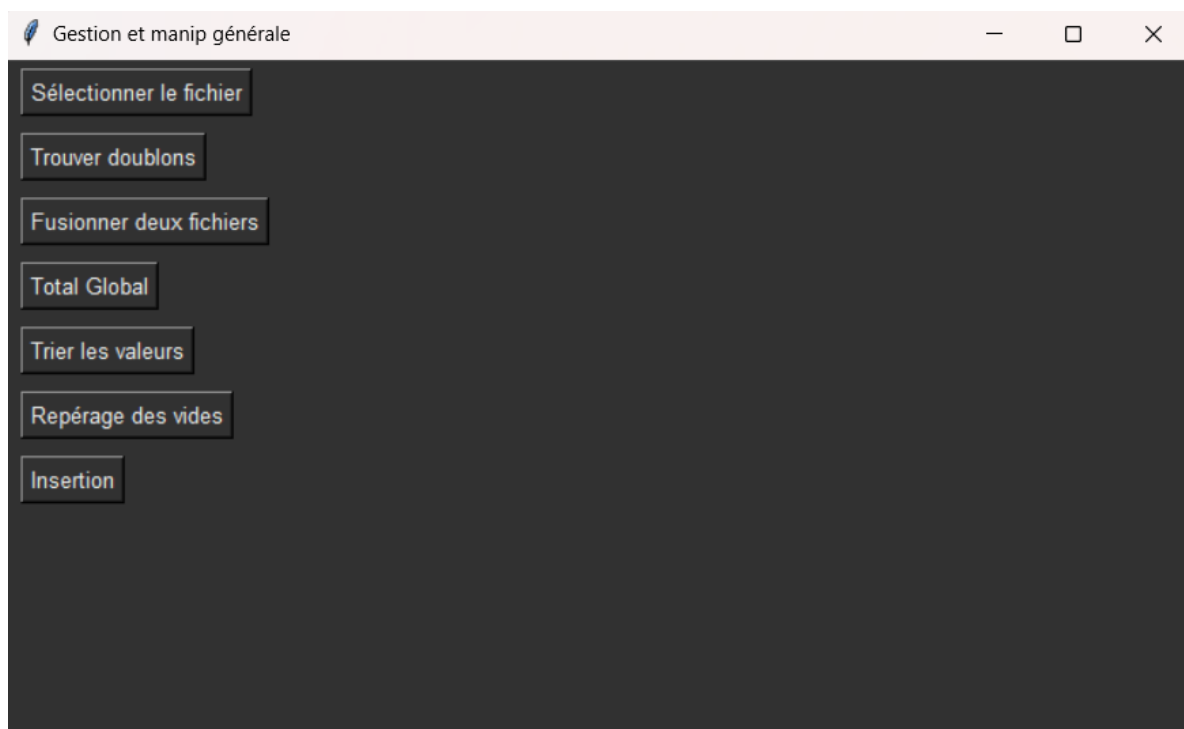


Figure 12 : Exécution du programme « main.py ».

MATRICULE	NOM PRENOM	GENRE	RANG	CODE	DOMAINE	MODE FORMATI	COUT GLOBALE
76988T	Wendy McQuak	Female	E	19	Etudes	E-Learning	19
91713I	Elfreda McBeith	Female	E	21	Juridique	Présentiel	28
90454H	Clair Lenoir	Female	C	20	Etudes	Présentiel	29
59029I	Rouvin Beau	Male	E	13	Etudes	Présentiel	6
06532L	Siobhan Gianotti	Female	C	15	Juridique	E-Learning	26
48065W	Wheeler Tousy	Male	M	20	Juridique	Présentiel	2
84177M	Tanner Selburn	Male	C	19	Communication	Présentiel	4
15733V	Benito Caldicot	Male	E	4	Etudes	E-Learning	21
54031I	Leoine Nesbitt	Female	M	10	Communication	E-Learning	1
38964C	Ky Switzer	Male	E	12	Communication	E-Learning	14
71180Y	Evanne Berninst	Female	M	10	Juridique	E-Learning	18

Figure 13 : Exécution du programme « main.py » avec l’affichage du tableau.

### Programme « reperage\_doublons » :

Ce programme utilise le module Tkinter avec ses modules « tk » et « filedialog » présentés précédemment. Il utilise aussi deux autres bibliothèques qui sont openpyxl et OS qui ont comme modules :

#### I. Bibliothèques de ce code

**Openpyxl** : C’est une bibliothèque Python utilisée pour lire, écrire et manipuler des fichiers au format (.xlsx), elle nous permettra de créer, modifier et extraire des données à partir de fichiers Excel.

**PatternFill** : C’est une classe de la bibliothèque openpyxl qui permet de définir les motifs de remplissage des cellules dans un classeur Excel tout en les personnalisant de cette façon : « pattern\_fill = Pattern Fill(fill\_type="gray125") ».

**Counter** : La classe Counter est utilisée pour compter le nombre d'occurrences des éléments dans une séquence ou un conteneur. Elle crée un dictionnaire spécialisé où chaque élément unique de la séquence est associé à son compteur, représentant le nombre de fois où il apparaît.

**OS** : La bibliothèque os fournit des fonctionnalités permettant d'interagir avec le système d'exploitation.

#### II. Elaboration du code :

Dans cette phase on va discuter et expliquer notre code :

a. On commence par la définition des fonctions **def** :

Def select\_file : Cette fonction permet à l'utilisateur de sélectionner un fichier Excel en utilisant une boîte de dialogue en utilisant le module « filedialog » et « .askopenfilename ».

On utilise aussi une boucle **if** pour vérifier qu'un fichier Excel a été choisi.

Def\_trouver\_doublons : On fait appel à la fonction select\_file pour permettre à l'utilisateur de sélectionner un fichier Excel à traiter. On va stocker le fichier sélectionné dans file\_path, si file\_path = True on continue l'exécution.

On charge le fichier à l'aide d'openpyxl en utilisant la fonction « .load\_workbook ». Le classeur est assigné à la fonction workbook et la feuille à la fonction créée 'Sheet'. Une fenêtre tkinter est ensuite créée pour permettre à l'utilisateur de sélectionner une colonne à analyser sous le format 'Alphabet' (A, B, C, ect...).

Dans cette même def, on crée une nouvelle fonction def appelée « handle\_trouver\_doublons », une fois exécutée cette fonction lis et récupère la colonne sélectionnée en tant que column\_var.

La fonction « selected\_column.isalpha() » vérifie si la colonne est une lettre alphabétique, si c'est le cas elle exécute la fonction « highlight\_doublons() » qui fait que chaque doublon (donc deux mots de la même valeur) seront highlightés et colorés de la même couleur pour faciliter leur repérage, si l'opération est réussie, une fenêtre Tkinter sera affichée en disant « Opération réussie », Sinon un message d'erreur est affiché.

```
def trouver_doublons():
    file_path = select_file()
    if file_path:
        workbook = openpyxl.load_workbook(file_path)
        sheet = workbook.active

        column_window = tk.Toplevel(root)
        column_window.title("Selection colonne")

        def handle_trouver_doublons():
            selected_column = column_var.get()
            if selected_column.isalpha():
                if highlight_doublons(workbook, sheet, file_path, selected_column):
                    show_message("Opération réussie")
                    column_window.destroy()
                    open_excel_file(file_path)
            else:
                show_message("Valeur entrée incorrecte, réessayez")
        else:
            show_message("Valeur entrée incorrecte, réessayez")

        column_var = tk.StringVar()
        column_label = tk.Label(column_window, text="Selection colonne (A, B, C...):")
        column_label.pack()
        column_entry = tk.Entry(column_window, textvariable=column_var)
        column_entry.pack()
        trouver_doublons_btn = tk.Button(column_window, text="OK", command=handle_trouver_doublons)
        trouver_doublons_btn.pack()
```

Figure 14 : Fonction « trouver\_doublons ».

Def\_highlight\_doublons : Cette fonction prends en paramètres le classeur « workbook », la feuille « sheet », le chemin « file\_path » et la colonne « selected\_column ». Cette fonction traite les cellules de la colonne qui sont remplies donc ignore les cellules vides.

On crée une liste appelée « duplicate » en utilisant la classe « Counter » pour faire un comptage du nombre d'occurrences de chaque élément dans la liste « column\_values ». Si un élément a un compteur supérieur à 1, il sera ajouté à la liste « doublons ».

On crée aussi une liste appelée « colors » qui contient des codes de couleurs prédéfinie pour attribuer une couleur à chaque doublon.

Après sa on crée une boucle pour parcourir chaque cellule de la colonne. Si la cellule = non nulle et que la valeur de cette cellule est présente dans la liste « doublons » on applique dessus ce processus : Si la valeur de la cellule n'est pas déjà présente dans le dictionnaire « color\_mapping », une couleur est attribuée et celle-ci est ensuite incrémentée. On applique cette couleur en utilisant la méthode « cell.fill = PatternFill() ».

En utilisant « workbook.save(file\_path) », on permet la sauvegarde de nos modifications. La fonction finit par retourner 'True' pour marquer le succès de l'opération.

```
def highlight_doublons(workbook, sheet, file_path, selected_column):
    column_values = [cell.value for cell in sheet[selected_column] if cell.value is not None]

    doublons = [item for item, count in Counter(column_values).items() if count > 1]

    colors = ["FF0000", "00FF00", "0000FF", "FFFF00", "FF00FF", "00FFFF"]
    color_mapping = {}
    color_index = 0

    for cell in sheet[selected_column]:
        if cell.value is not None and cell.value in doublons:
            if cell.value not in color_mapping:
                color_mapping[cell.value] = colors[color_index % len(colors)]
                color_index += 1
            cell.fill = PatternFill(start_color=color_mapping[cell.value], end_color=color_mapping[cell.value], fill_type="solid")

    workbook.save(file_path)
    print("Doublons trouvés et fichier sauvegardé")
    return True
```

```
(function) def len(
    _obj: Sized,
    /
) -> int
Return the number of items in a container.
```

Figure 15 : Fonction « highlight doublons ».

Def\_show\_message : Cette fonction définit une fonction appelée show message qui affiche une fenêtre de message (hors la fenêtre root). Lorsque la fonction est appelée, une nouvelle fenêtre est créée avec un titre fixe "Message". Le texte spécifié est affiché dans la fenêtre à l'aide d'un widget Label. Un bouton "OK" est ajouté à la fenêtre, et lorsqu'il est cliqué, la fenêtre de message se ferme.

Def\_open\_excel\_file : Lors de l'appel de cette fonction, elle ouvre le fichier Excel de l'attribut « file\_path ». Dans cette fonction, on utilise aussi « os.startfile(file\_path) » comme fonction pour ouvrir le fichier spécifié de « file\_path » dans Excel.

- b. On crée une fenêtre Tkinter avec la fonction « root= tk.Tk() ». On lui administre un nom en utilisant la fonction « root.title(" ") » et on lui donne une taille de '300x150' en utilisant la fonction root.geometry(" ")

On sélectionne le style des icônes et widgets avec la fonction « root.tk.call("source", " ") »

- c. Afin de créer le bouton "Trouver doublons" on utilise la fonction « tk.button » et pour préciser le padding on utilise la fonction (.pack).



- d. La fonction « `root.mainloop()` » lance la boucle principale d'événements de l'application Tkinter. Elle permet d'exécuter en continu les événements et de gérer les interactions de l'utilisateur dans l'interface graphique (GUI).

### *Programme « prog\_combine » :*

Ce programme utilise le module Tkinter avec ses modules « tk » et « filedialog » présentés précédemment. Il utilise les bibliothèques Pandas et OS.

#### I. Elaboration du code :

Dans cette phase on va discuter et expliquer notre code :

##### a. On commence par définir 3 fonctions **def** :

**Def\_choose\_file** : On définit cette fonction qui va permettre à l'utilisateur de choisir une fonction à l'aide du module « filedialog » et de sa méthode (`.askopenfilename`) .

**Def\_show\_message** : Pour définir cette fonction on aura besoin de la fonction (`.Toplevel`) dans la méthode « tk », la fonction « `tk.Toplevel()` » ouvrira une fenêtre au dessus de la fenêtre principale.

On utilise une fonction `tk.label` pour donner une étiquette a cette dernière fenêtre. Ensuite, on adopte la fonction `tk.button` pour donner une fonction au bouton qu'on va créer par la suite, le texte du bouton sera 'OK' et il permettra de fermer cette fenêtre.

**Def\_fusion\_files** : Tout d'abord, cette fonction nous permet de choisir un premier fichier Excel en utilisant la fonction « `choose_file()` » puis de choisir un deuxième fichier Excel en utilisant cette même fonction.

Ensuite, cette fonction 'def' lit les deux fichiers Excel choisis en utilisant une méthode de la bibliothèque pandas pour lire qui s'appelle (`.read_excel()`) .

Le contenu des fichiers est stocké dans les objets `df1` et `df2` qui sont des `DataFrame`. De plus, les noms des colonnes de chaque `DataFrame` seront extraits et stockés dans les listes `header1` et `header2` pour ensuite effectuer une vérification et comparaison des colonnes (si elles sont égales) ensuite des noms de chaque colonne (les headers). Si les fichiers, colonnes et headers matchent, alors on procède à la combinaison des deux fichiers utilisant la méthode (`.concat`) de la bibliothèque Pandas.

Après avoir exécuté la fusion des deux fichiers, on crée un nouveau fichier Excel qu'on va nommer 'combinaison.xlsx' en utilisant la méthode `to_excel()` de pandas sur le `DataFrame df_combined`.

Après avoir créé le nouveau fichier Excel, la fonction utilise « `os.startfile()` » pour ouvrir automatiquement le fichier dans l'application par défaut associée aux fichiers Excel.

Enfin, la fonction affiche un message de succès à l'utilisateur en appelant `show_message()` avec le message "Opération réussie".

PS : Si l'étape de vérifications échoue, la fonction affiche un message d'erreur à l'utilisateur en appelant « `show_message()` » avec le message "Valeurs différentes, réessayez".

- b. On crée une fenêtre Tkinter avec la fonction « `root= tk.Tk()` ». On lui dispose un nom en utilisant la fonction « `root.title(" ")` » et on lui donne une taille de '200x100' en utilisant la fonction `root.geometry(" ")`

On sélectionne le style des icônes et widgets avec la fonction « `root.tk.call("source", " ")` ».

Afin de créer le bouton "Choisir fichiers" on utilise la fonction « `tk.button` » et pour préciser le padding on utilise la fonction `(.pack)`.

- c. La fonction « `root.mainloop()` » lance la boucle principale d'événements de l'application Tkinter. Elle permet d'exécuter en continu les événements et de gérer les interactions de l'utilisateur dans l'interface graphique (GUI).

```
def fusion_files():
    #choisir le premier fichier
    file_path1 = choose_file()
    if not file_path1:
        return

    #choisir le deuxieme fichier
    file_path2 = choose_file()
    if not file_path2:
        return

    # Lire le premier fichier
    df1 = pd.read_excel(file_path1)
    header1 = df1.columns.tolist()

    #Lire le de le deuxieme fichier
    df2 = pd.read_excel(file_path2)
    header2 = df2.columns.tolist()

    # vérification du nombre de colonnes et de leurs nomination
    if header1 == header2 and df1.shape[1] == df2.shape[1]:
        df_combined = pd.concat([df1, df2], ignore_index=True) # Combiner les tableau avec les headers
        # creer un nouvzl fichier
        output_file = "combinaison.xlsx"
        df_combined.to_excel(output_file, index=False)

        os.startfile(output_file) # Ouvrir le nouvel fichier

        show_message("Opération réussie") #réussie
    else:
        show_message("Valeurs différentes, réessayez") #erreur
```

Figure 16 : Fonction "Def\_fusion\_files".

### Programme « prog\_total\_glob » :

Ce programme utilise le module Tkinter avec ses sous-modules tk et filedialog pour créer une interface graphique. Il fait également appel aux bibliothèques Pandas et OS pour effectuer des opérations de manipulation de données et d'interaction avec le système d'exploitation.

#### I. Elaboration du code :

Dans cette phase on va discuter et expliquer notre code :

- a. Def\_select\_file : La fonction 'choose\_file' est définie pour permettre à l'utilisateur de sélectionner un fichier à l'aide du module 'filedialog' et de sa méthode 'askopenfilename'.

Def\_calculate\_sum : Cette fonction fait appel à la fonction select\_file pour permettre à l'utilisateur de choisir un fichier Excel. Quand le fichier est sélectionné, il est ouvert à l'aide de la bibliothèque openpyxl en utilisant la fonction « load\_workbook » et le chemin du fichier sélectionné. Et est défini tant que « file\_path ».

Une nouvelle fenêtre « column\_window » est créée à l'aide de « tk.Toplevel » pour permettre à l'utilisateur de sélectionner la colonne sur laquelle il souhaite effectuer le calcul de la somme. Une étiquette 'column\_label' est affichée dans la fenêtre pour demander à l'utilisateur de sélectionner une colonne sous forme d'alphabet (A, B, C...).

Un champ de saisie « column\_entry » est affiché pour permettre à l'utilisateur de saisir le nom de la colonne.

Un bouton « calculate\_sum\_btn » est affiché dans la fenêtre pour lancer le calcul de la somme. Lorsque l'utilisateur clique sur le bouton "Calculate Sum", la fonction « handle\_calculate\_sum » est utilisée avec les paramètres nécessaires : le classeur (workbook), la feuille (sheet), le chemin du fichier (file\_path) et le nom de la colonne sélectionnée (column\_var.get()).

Def\_handle\_calculate\_sum : Cette fonction est appelée lorsque l'utilisateur clique sur le bouton "Calculer la somme". Pour effectuer le calcul, cette fonction utilise la fonction « add\_values » qui est appelée avec ses paramètres (workbook, sheet, file\_path et selected\_column), elle est responsable de calculer la somme des valeurs de la colonne sélectionnée de la feuille Excel. Si le calcul de la somme des valeurs est effectué avec succès (donc « add\_values = True ») les étapes suivantes sont exécutées :

À l'aide de la fonction « show\_message » Un message sera émis qui affichera "Opération réussie". Ensuite le fichier Excel est ouvert à l'aide de la fonction « open\_excel\_file ».

Sinon la fonction affichera le message "Valeur invalide, réessayez".

Def\_add\_values : Les paramètres de cette fonction sont 'workbook', 'sheet', 'file\_path' et 'selected\_column'.

On utilise une compréhension de liste pour extraire les valeurs non nulles et numériques de la colonne sélectionnée, puis on stocke ces valeurs dans la fonction « column\_values ».

La somme des valeurs de la colonne est calculée en utilisant la fonction « sum » appliquée à la liste « column\_values ». Le résultat est stocké dans la variable « total\_sum ».

L'index de la dernière ligne est obtenu en utilisant la longueur de la liste sheet « selected\_column ».

Deux nouvelles cellules sont créées :

Une pour la somme des valeurs et une pour écrire "Total Globale". Ces cellules sont positionnées après la dernière ligne de la colonne sélectionnée en utilisant « last\_row\_index+2 » et « last\_row\_index+3 » respectivement.

La modification est sauvegardée avec la méthode (.save) en utilisant le chemin du fichier « file\_path ».

La fonction renvoie True pour indiquer que l'opération d'ajout du Totale global a été réussie.

Def\_open\_file : Le fichier Excel est chargé en utilisant la fonction « openpyxl.load\_workbook(file\_path) » qui ouvre le fichier excel de « file\_path », cela crée l'objet « workbook » qui représente le fichier Excel.

Ensuite le code utilise la méthode « workbook.active » pour obtenir la feuille active du fichier, ce qui retourne un objet « sheet » qui est la feuille de calcul.

La fonction utilise une boucle « for » pour lire chaque ligne de la feuille 'sheet'. L'instruction « sheet.iter\_rows(values\_only=True) » retourne un itérable fournissant chaque ligne sous forme d'un tuple de valeurs.

À chaque itération de la boucle, les valeurs de la ligne sont affichées à l'aide de print(\*row, sep="\t").

Def\_show\_message : Pour définir cette fonction on aura besoin de la fonction (.Toplevel) dans la méthode « tk », la fonction « tk.Toplevel("") » ouvrira une fenêtre au-dessus de la fenêtre principale.

On utilise une fonction tk.label pour donner une étiquette à cette dernière fenêtre. Ensuite, on adopte la fonction tk.button pour donner une fonction au bouton qu'on va créer par la suite, le texte du bouton sera 'OK' et il permettra de fermer cette fenêtre.

Def\_open\_excel\_file : La fonction utilise « os.system » pour ouvrir un fichier Excel spécifié par 'file\_path' dans l'application Excel pour afficher le résultat à l'utilisateur.

- b. On crée une fenêtre Tkinter avec la fonction « root= tk.Tk() ». On lui donne un nom en utilisant la fonction « root.title(" ") » et on lui donne une taille de '300x150' en utilisant la fonction root.geometry(" ")

On sélectionne le style des icônes et widgets avec la fonction « root.tk.call(" source", " ") ».

Afin de créer le bouton "Calculer la somme" on utilise la fonction « tk.button » et pour préciser le padding on utilise la fonction (.pack).

- c. La fonction « root.mainloop() » lance la boucle principale d'événements de l'application Tkinter. Elle permet d'exécuter en continu les événements et de gérer les interactions de l'utilisateur dans l'interface graphique (GUI).

### Programme « prog\_ranking » :

Ce programme utilise la bibliothèque Tkinter pour créer une interface graphique en utilisant les modules « tk » et « filedialog ». Il utilise également les bibliothèques Pandas et OS pour effectuer des opérations de manipulation de données et interagir avec le système d'exploitation.

#### I. Elaboration du code :

Dans cette phase on va discuter et expliquer notre code :

- a. Def\_choose\_file : La fonction choose\_file est définie pour offrir à l'utilisateur la possibilité de sélectionner un fichier en utilisant le module filedialog et sa méthode askopenfilename.

Def\_sort\_data : Tout d'abord, l'utilisateur est invité à choisir le fichier Excel à trier à l'aide de la fonction « choose\_file ». Si aucun fichier n'est sélectionné, la fonction se termine. Ensuite, le fichier Excel est lu en utilisant la bibliothèque Pandas, et les headers des colonnes sont récupérés.

Une nouvelle fenêtre Tkinter est créée pour afficher les options de tri. Cette fenêtre affiche deux cases à cocher ; une pour l'ordre décroissant et une pour l'ordre croissant. Lorsque l'utilisateur coche l'une des cases, la fonction « sort\_and\_save » est appelée avec le type de tri correspondant en tant que paramètre.

Dans la fonction « sort\_and\_save », les données sont triées en fonction du précédent. Si l'option est "décroissant" les données sont triées par ordre décroissant, si l'option est "croissant" les données sont triées par ordre croissant. Si une option de tri invalide est sélectionnée, un message est affiché et la fenêtre de tri est fermée.

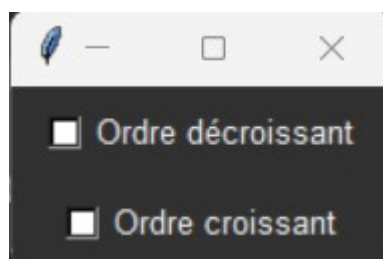


Figure 17 : Fenêtre de tri.

Les données triées sont ensuite sauvegardées dans le même fichier Excel, en remplaçant les données existantes. puis la fenêtre de tri est fermée et le fichier Excel modifié est ouvert à l'aide de la bibliothèque OS.

- b. On crée une fenêtre Tkinter avec la fonction « root= tk.Tk() ». On lui dispose un nom en utilisant la fonction « root.title(" ")».

On sélectionne le style des icônes et widgets avec la fonction « root.tk.call(" source", " ")».

Afin de créer le bouton "Choisir fichier" on utilise la fonction « tk.button » et pour préciser le padding on utilise la fonction (.pack).

- c. La fonction « root.mainloop() » lance la boucle principale d'événements de l'application Tkinter. Elle permet d'exécuter en continu les événements et de gérer les interactions de l'utilisateur dans l'interface graphique (GUI).

### *Programme « prog\_cellules\_vides » :*

Ce programme utilise le module Tkinter avec ses modules 'tk', 'ttk' et 'filedialog' pour créer une interface graphique. Il fait également appel aux bibliothèques Openpyxl et OS pour effectuer des opérations de manipulation de données et d'interaction avec le système d'exploitation.

- I. Elaboration du code :

Dans cette phase on va discuter et expliquer notre code :

- a. Def\_select\_file : Cette fonction est définie pour octroier à l'utilisateur la possibilité de sélectionner un fichier en utilisant le module 'filedialog' et sa méthode (.askopenfilename). L'objet workbook est utilisé pour accéder à la feuille de calcul active en utilisant workbook.active qui nous donne un objet sheet qui représente la feuille Excel. Une boucle imbriquée est utilisée pour itérer sur toutes les cellules de la feuille en utilisant « sheet.iter\_rows() ». A chaque fois cette fonction vérifie si la valeur de la cellule est nulle en utilisant « cell.value ». Si la cellule est vide, elle est mise en évidence en la coloriant avec une couleur (jaune) avec la méthode « cell.fill = PatternFill() », la méthode 'PatternFill' spécifie la couleur de remplissage. Ensuite, en utilisant « workbook.save(file\_path) » on sauvegarde les modifications apportées dans le fichier Excel. Un message est affiché pour indiquer que l'opération a été un succès et que le fichier Excel a été sauvegardé. Enfin, la fonction « open\_file(file\_path) » est appelée pour ouvrir le fichier Excel modifié.  
  
Def\_open\_file : Cette fonction prends le chemin « file\_path » puis utilise la méthode (.startfile) de la bibliothèque OS pour ouvrir le fichier et l'afficher à l'utilisateur.
- b. On crée une fenêtre Tkinter avec la fonction « root= tk.Tk() ». On lui dispose un nom en utilisant la fonction « root.title(" ") » et on lui donne une taille de '500x250' en utilisant la fonction root.geometry(" ")

On sélectionne le style des icônes et widgets avec la fonction « root.tk.call("source", " ") ».

Afin de créer le bouton " Selection du fichier Excel" on utilise la fonction « tk.button » et pour préciser le padding on utilise la fonction (.pack).

- c. La fonction « `root.mainloop()` » lance la boucle principale d'événements de l'application Tkinter. Elle permet d'exécuter en continu les événements et de gérer les interactions de l'utilisateur dans l'interface graphique (GUI).

#### *Programme « `program_insertion` » :*

Ce programme utilise le module Tkinter avec ses modules 'tk' et 'ttk' pour créer une interface graphique. Il fait également appel à la bibliothèque Openpyxl pour effectuer des opérations de manipulation de données.

##### I. Elaboration du code :

Dans cette phase on va discuter et expliquer notre code :

- a. **Def\_load\_data :** La variable 'path' contient le chemin d'accès au fichier Excel utilisé. On utilise la bibliothèque Openpyxl pour stocker le fichier Excel dans « `workbook` ». La feuille 'data' est récupérée et stockée grâce à la méthode « `workbook.active` ». Les valeurs de la feuille 'data' sont extraites sous forme de liste en utilisant la méthode (`.values`), la liste des valeurs est stockée dans la variable (`list_values`). On extrait les headers des colonnes grâce à la première ligne de (`list_values`), on utilise ces headers pour le module (`treeview`). On récupère et insère les données de (`list_values`) dans « `treeview` » sauf les headers, on utilise une boucle pour parcourir les fichiers et procéder à leurs récupérations

**Def\_show\_message :** La fonction « `show_message` » permet d'afficher une fenêtre de message secondaire de la fenêtre principale. Lorsqu'elle est appelée, une nouvelle fenêtre est créée. Le texte spécifié est affiché dans la fenêtre à l'aide d'un widget Label. Un bouton "OK" est ajouté à la fenêtre qui permet de fermer la fenêtre de message.

**Def\_insert\_row :** Tout d'abord, cette fonction vérifie si les champs nécessaires (MATRICULE, nom, CodeDomaine et Cout) sont tous remplis. Si l'un des champs est vide, un message d'erreur est affiché à l'aide de la fonction « `show_message` » et la fonction se termine.

Ensuite, les valeurs des champs sont récupérées à partir des widgets correspondants (MATRICULE\_entry, nom\_entry, CodeDom\_entry, Cout\_entry, status\_combobox, status\_combobox1, status\_combobox2). Ces valeurs sont assignées aux variables correspondantes (MATRICULE, NOM, genre, rang, CodeDomaine, choix\_domaine, mode\_status, Cout).

Ensuite, le fichier Excel est chargé à l'aide de la bibliothèque openpyxl et sa méthode 'workbook' et 'workbook.active'. La feuille 'data' est récupérée et la dernière ligne est déterminée en utilisant la méthode « `last_row = data.max_row + 1` ».

Les valeurs sont ensuite insérées dans la méthode « `treeview` » avec les headers correspondants tout en ajoutant une ligne à chaque ajout de données à chaque utilisation de la méthode « `insert` ».

- b. **Style :** Cette fonction nous permet d'utiliser un thème personnalisé en faisant appel à la commande « `root.call("source", ...)` ». Ensuite on associe la méthode 'ttk.style' pour définir et gérer les widgets.

La méthode (.theme\_use) est ensuite utilisé pour appeler la fonction « style » et appliquer le theme choisi.

- c. Widgets et GUI : On commence par créer 3 Combobox (combo\_list, combo\_list1, combo\_list2) sous la forme « combo\_list = ["choix0", "choix1", "choix2", ect...].

Puis on crée une Frame pour regrouper nos widgets dedans et la méthode 'frame.pack' pour l'afficher dans notre fenêtre principale. On nommeras cette frame 'Infos Personnelles' grâce à la méthode « ttk.LabelFrame » tout en utilisant la méthode (.grid) pour positionner notre Frame.

On va aussi créer 4 entrées de texte (MATRICULE\_entry, nom\_entry, CodeDom\_entry, Cout\_entry) en utilisant la méthode « ttk.entry ».

Ensuite on utilisera la méthode (.insert) pour afficher un texte par défaut dans le champs de texte (ex : nom\_entry.insert(0, 'Nom et Prénom') ). On ajoutera a ca la méthode (.bind) qui nous permettra d'effacer la valeur par défaut quand l'utilisateur commencera a insérer une valeur dans le champs de texte, elle sera sous forme « .bind("<FocusIn>", lambda e: nom\_entry.delete('0','end')) », exemple : ( nom\_entry.bind("<FocusIn>", lambda e: nom\_entry.delete('0','end')) ).

Pour positionner le champs d'insertion on utilisera la méthode (.grid), exemple : (nom\_entry.grid(row=1, column=0, padx=5, pady=(0, 5), sticky="ew") )

Pour créer un DropDownMenu qui sera associé aux combobox déjà créés, on utilise la méthode (ttk.combobox) qui sera sous la forme de « ttk.Combobox(widgets\_frame, values=combo\_list) » ou l'on remplacera la champs 'values' par les fonctions associées pour chaque DropDownMenu (en l'occurrence combo\_box, combo\_box1, combo\_box2). On utilisera aussi la méthode (.current) pour définir la valeur par défaut sélectionnée par la liste déroulante et on la mettra zéro (.current(0)) pour prendre le premier choix des combo\_list.

On devra créer 3 Widget Frames, une pour la Combobox 'Genre', une pour 'Rang' et une dernière pour 'Domaine'. On appliquera la méthode « ttk.LabelFrame » pour donner un nom à notre Frame ( "Genre", "Rang", "Domaine" respectivement) et on réutilisera la fonction (.grid) présenté précédemment pour positionner nos Frames.

On appliquera la methode (ttk.Chekbutton) pour créer deux boutons (pour faire un choix entre l'option "E-Learning" et l'option "Présentiel") avec la méthode (.grid) pour le positionnement des boutons.

Pour faire fonctionner nos deux CheckBox, on va ajouter la méthode (tk.BooleanVar()) et les assigner a deux variables qu'on va nommer « a » et « b » respectivement pour « E-Learning » et « Présentiel ».

On auras besoin de faire l'usage d'une boucle pour éviter les erreurs qu'on peut rencontrer en créant ces deux Checkbox (pour éviter que l'utilisateur ne les coches les deux), la boucle prendra cette forme : si la variable 'a' = 1, alors 'b' = 0, sinon si 'b' = 1, alors 'a'=0.

On intégrera les deux Chekbutton, la liste défilante (combo\_list2) et l'entrée de texte (CodeDom\_entry) dans le même frame qui est celle de « Domaine ».

On intégrera aussi à ce Frame un bouton « insertion » qui sera créé à l'aide de la méthode (ttk.Button) tout en utilisant la méthode (.grid) pour le positionner.



Quand le bouton « insérer » sera actionné, la méthode utilisera la fonction « insert\_row » vue précédemment qui permettra à l'utilisateur d'insérer et d'ajouter des valeurs.

- d. Pour la manipulation du tableau, on a fait usage de la méthode « `ttk.Frame(frame)` » pour la création d'un cadre à l'intérieur du cadre 'frame' déjà existant tout en le positionnant grâce à la méthode `(.grid)`.

Pour créer une barre de défilement verticale dans le cadre (treeFrame) qui nous permettra de défiler nos données, on a fait appel à la méthode `(ttk.Scrollbar(treeFrame))` qu'on a nommé « treeScrolly » et une deuxième barre de défilement horizontale créée de la même façon qui a été nommée « treeScrollX ». On utilisera ensuite la méthode `(.pack)` pour positionner les Scrollbar à droite et en bas du cadre respectivement pour 'treeScrolly' et 'treeScrollX'.

On va créer une variable qu'on va appeler « cols » qui va regrouper tous les noms des colonnes du tableau (donc : "MATRICULE", "NOM PRENOM", "GENRE", "RANG", "CODE DOMAINE", "DOMAINE", "MODE FORMATION", "COUT GLOBALE").

« `treeview = ttk.Treeview(treeFrame, show="headings", yscrollcommand=treeScrolly.set, xscrollcommand=treeScrollX.set, columns=cols, height=10)` »: Cette fonction crée le tableau (treeview) à l'intérieur du cadre 'treeFrame'. Les options spécifiées incluent le paramètre `show="headings"` pour afficher les headers de colonne, les paramètres `yscrollcommand` et `xscrollcommand` pour lier les barres de défilement verticale et horizontale au tableau, les colonnes 'columns' et 'height' pour définir les colonnes du tableau et sa hauteur.

On utilise la méthode « `treeview.column("col", width=100)` » pour Définir la largeur de la colonne, on exécute cette procédure pour chacune des colonnes qu'on doit traiter.

Pour positionner le tableau 'treeview' dans le cadre 'treeFrame', on fait usage de la méthode `(treeview.pack())`.

Pour configurer la méthode de défilement des barres 'a' et 'b', on utilise la méthode « `.config(command=treeview.view)` ».

Enfin on fait appel à la fonction « `load_data()` » pour charger les données à afficher dans le tableau.

- e. La fonction « `root.mainloop()` » lance la boucle principale d'événements de l'application Tkinter. Elle permet d'exécuter en continu les événements et de gérer les interactions de l'utilisateur dans l'interface graphique (GUI).

The screenshot shows a Tkinter application window titled 'tk'. On the left, there is a form titled 'Infos Personnelles' with fields for 'Matricule', 'Nom et Prénom', 'Genre' (a dropdown menu showing 'Non Précisé'), 'Rang' (a dropdown menu showing 'C'), and 'Domaine'. The 'Domaine' section includes a 'Code Domaine' field, a 'Non Précisé' dropdown, a 'Couts' field, and two checkboxes for 'E-Learning' and 'Présentiel'. An 'Insérer' button is at the bottom of the form. On the right, there is a table with the following data:

MATRICULE	NOM PRENOM	GENRE	RANG	CODE	DOMAINE	DOMAINE	MODE FORMATI	COUT GLOBALE
76988T	Wendy McQuak	Female	E	19		Etudes	E-Learning	19
91713I	Elfreda McBeith	Female	E	21		Juridique	Présentiel	28
90454H	Clair Lenoir	Female	C	20		Etudes	Présentiel	29
59029I	Rouvin Beau	Male	E	13		Etudes	Présentiel	6
06532L	Siobhan Gianotti	Female	C	15		Juridique	E-Learning	26
48065VW	Wheeler Tousy	Male	M	20		Juridique	Présentiel	2
84177M	Tanner Selburn	Male	C	19		Communication	Présentiel	4
15733V	Benito Caldicot	Male	E	4		Etudes	E-Learning	21
54031I	Leoine Nesbitt	Female	M	10		Communication	E-Learning	1
38964C	Ky Switzer	Male	E	12		Communication	E-Learning	14
71180Y	Evanne Beminst	Female	M	10		Juridique	E-Learning	18

Figure 18 : Fenêtre représentant l'exécution du programme d'insertion.

### Style utilisé (forest-dark) :

Pour appliquer le thème de notre application, dans chaque programme on a utilisé la fonction « root.tk.call » pour faire appel au thème qu'on veut utiliser (dans notre cas c'est le thème forest-dark du créateur 'rdbende').

Ensuite nous avons utilisé le module « style.theme\_use » pour appliquer ce dernier suivi de la fonction style.configure pour configurer les propriétés de nos widgets.

```
# Appliquer le style forest dark
root.tk.call("source", r"C:\Users\user\Documents\PYTHON\

style = ttk.Style(root) # Configuration du style
style.theme_use("forest-dark")
style.configure(".", font=("Arial", 10))
```

Figure 19 : Exemple d'application du thème.

## Conclusion

Dans ce chapitre, nous avons présenté l'environnement d'implémentation et de développement de notre application, en se focalisant sur les techniques de programmation utilisées pour implémenter notre programme et résoudre le problème. La description de notre application s'est faite en présentant les interfaces essentielles tout en parlant du code, mentionnant et expliquant toutes les bibliothèques et méthodes utilisées.

### **Conclusion générale :**

En conclusion, ce mémoire de fin d'étude s'est concentré sur la création et le développement d'une application desktop pour la gestion et la manipulation des fichiers Excel à l'aide du langage Python. Tout au long de cette recherche, j'ai acquis une solide expérience en programmation et une meilleure compréhension de Python. Malgré les défis liés au manque de documentation sur la problématique abordée, j'ai réussi à résoudre cette problématique avec succès.

Les contributions de cette étude sont significatives pour le domaine de la gestion des données. L'application que j'ai développée répond aux besoins spécifiques du département de formation de Sonatrach en optimisant et en réduisant le temps d'insertion des données complexes. Elle facilite également le tri, l'identification des doublons, des vides et des incohérences, ce qui résout les difficultés rencontrées par le département de formation.

Sur le plan théorique et pratique, cette application présente des implications considérables. Elle permet au département de formation de Sonatrach d'améliorer son efficacité opérationnelle et d'optimiser son flux de travail. Cela entraîne une réduction des erreurs et une augmentation de la productivité.

Cependant, il est important de noter certaines limites et lacunes de cette étude. L'interface graphique aurait pu être mieux développée si j'avais disposé de plus de temps. De plus, l'application aurait pu être optimisée pour gérer plus de flux de données simultanément ou pour ajouter des fonctionnalités supplémentaires, telles que l'implémentation de données SQL.

Pour les futures recherches dans ce domaine, il est recommandé de fournir une meilleure documentation et de choisir soigneusement le sujet de recherche. Travailler en équipe et s'appuyer sur les études et mémoires existants peuvent également aider à approfondir les connaissances et à développer des applications plus avancées.

Ce mémoire de fin d'étude revêt une importance particulière pour moi, car il m'a permis d'acquérir des compétences précieuses en optimisation et en programmation. Il a également renforcé mon intérêt pour le domaine de la science des données, ce qui oriente mes perspectives d'avenir vers une carrière dans ce domaine en pleine expansion.

Les implications plus larges de cette étude se situent dans le contexte professionnel de Sonatrach. En offrant une solution efficace et fiable pour la gestion des données, cette application contribue à améliorer les processus de travail, à réduire les erreurs et à augmenter la productivité du département de formation. Cela a un impact positif sur l'ensemble de l'organisation.

En conclusion, la réalisation de ce mémoire de fin d'étude témoigne de la volonté de se surpasser et d'oser viser plus haut. Dans un monde en constante évolution technologique, il est essentiel d'être courageux et d'embrasser de nouvelles opportunités. Je souhaite transmettre ce message clé : "**Vivif post funera virtus**" - le courage survit à la mort. Soyez courageux, osez vivre et visez toujours plus haut.











## Bibliographie

---

titouan : <https://github.com/Titouan-Legend/Discord-Bot-Python-Tutoriel> 27/05/2023

github : <https://github.com/Titouan-Legend/Discord-Bot-Python-Tutoriel>

Formasys: <https://www.youtube.com/@formasys8974> 27/05/2023

ALGOMIUS: [https://www.youtube.com/watch?v=H\\_pVrg1\\_49w&ab\\_channel=Algomius](https://www.youtube.com/watch?v=H_pVrg1_49w&ab_channel=Algomius) 20/05/2023

site: <https://www.algomius.fr> 20/05/2023

github : <https://github.com/Algomius/Tri> 21/05/2023

CODE FIRST WITH HALA : <https://www.youtube.com/@codefirstwithhala> 15/06/2023

linkdin : <https://www.linkedin.com/in/hala-saadeh/?originalSubdomain=lb> 17/06/2023

NEURALNine : <https://www.youtube.com/@NeuralNine> 30/05/2023

GitHub: <https://github.com/NeuralNine> 30/05/2023

Rdbende : <https://github.com/rdbende> 15/05/2023

Python Bits:<https://www.youtube.com/@python-bits> 3/06/2023

Twitter: <https://twitter.com/pythonbits1> 3/06/2023

tres facile : <https://www.youtube.com/@debutants> 3/06/2023

FreeCodeCamp: site: FreeCodecamp.org 5/06/2023

youtube: <https://www.youtube.com/@freecodecamp> 5/06/2023

Python: <https://www.python.org/> 30/05/2023

VScode: <https://code.visualstudio.com/> 30/05/2023

memoire université guelma : <https://dspace.univ-guelma.dz/jspui/bitstream/123456789/1253/1/memoire.pdf>  
19/06/2023

memoire université ummto-tiziouzou :<https://www.ummto.dz/dspace/bitstream/handle/ummto/12526/lamracheFatima.pdf?sequence=1> 19/06/2023

documentation sonatrach : sonatrach.com 10/06/2023

wiki : <https://fr.wikipedia.org/wiki/Sonatrach> 10/06/2023

vscode:<https://code.visualstudio.com/learn> 8/06/2023

<https://code.visualstudio.com/docs/editor/extension-marketplace> 8/06/2023

[https://fr.wikipedia.org/wiki/Python\\_\(langage\)](https://fr.wikipedia.org/wiki/Python_(langage)) 8/06/2023

[https://fr.wikipedia.org/wiki/Python\\_\(langage\)#Historique](https://fr.wikipedia.org/wiki/Python_(langage)#Historique) 8/06/2023

openpyxl : <https://openpyxl.readthedocs.io/en/stable/> 8/06/2023

<https://pypi.org/project/openpyxl/> 8/06/2023

pandas : <https://pandas.pydata.org/docs/> 21/04/2023

<https://pandasguide.readthedocs.io/en/latest/> 21/04/2023

numpy: <https://numpy.org/doc/> 21/04/2023

<https://docs.scipy.org/doc/> 21/04/2023

matplotlib: <https://matplotlib.org/stable/index.html> 29/04/2023

<https://devdocs.io/matplotlib~3.1/> 29/04/2023

<https://pythonforge.com/module-os-systeme-dexploitation/#~:text=Le%20module%20os%20de%20Python,im- porte%20quelle%20syst%C3%A8me%20d'exploitation.> 29/04/2023

<https://docs.python.org/fr/3/library/os.html> 30/04/2023

[https://paul-peton.medium.com/de-quoi-le-futur-de-la-data-sera-t-il-fait-sous-un-intervalle-de-pr%C3%A9vision- assez-large-f1d4d0efa109](https://paul-peton.medium.com/de-quoi-le-futur-de-la-data-sera-t-il-fait-sous-un-intervalle-de-pr%C3%A9vision-assez-large-f1d4d0efa109) 10/06/2023

<https://www.quai-des-entrepreneurs.com/lintegration-de-la-data-science-au-sein-dune-entre- prise/#~:text=La%20data%20science%20permet%20aux,la%20demande%20des%20clients%20potentiels.>

10/06/2023

<https://ledigitalpourtous.fr/2020/02/24/la-data-science-un-metier-ses-enjeux-et-cas-dusages/> 12/06/2023

<https://www.lepont-learning.com/fr/data-science-definition-enjeux/#:~:text=La%20normalisation%20des%20donn%C3%A9es%20sera,%C3%A0%20se%20poser%20en%20amont.> 12/06/2023

<https://openclassrooms.com/fr/courses/4011851-initiez-vous-au-machine-learning/5868381-selectionnez-les-outils-de-data-science-appropries#:~:text=Les%20plus%20reconnues%20sont%20Google,et%20la%20mesure%20de%20performance.> 12/06/2023

<https://www.techtitude.com/ml/informatique/diplome-universite/diplome-univeriste-exploitation-information-data-science> 12/06/2023

<https://datascientest.com/les-outils-dans-le-quotidien-dun-data-scientist#:~:text=Matplotlib%2CSeaborn%20sont%20des%20outils,et%20un%20partage%20d'informations.> 12/06/2023

<https://ledigitalpourtous.fr/2020/02/24/la-data-science-un-metier-ses-enjeux-et-cas-dusages/> 12/06/2023

<https://www.lebigdata.fr/data-science-definition> 13/06/2023

<https://seleritysas.com/blog/2019/12/06/preparing-for-the-future-of-data-analytics/> 13/06/2023

<https://blogrecrutement.bpce.fr/differents-types-de-data-1/> 13/06/2023

<https://www.usine-digitale.fr/article/data-scientist-un-metier-d-elite-en-voie-de-democratisation.N444472>  
13/06/2023

<https://liveroomlk.medium.com/analyzing-big-data-techniques-related-to-data-science-9b528f77e230>  
13/06/2023

<https://www.kdnuggets.com/2018/06/data-science-predicting-future.html> 13/06/2023

<https://seleritysas.com/blog/2019/12/06/preparing-for-the-future-of-data-analytics/> 13/06/2023

## Table des matières

---

<b>Engagement de non plagiat .....</b>	<b>2</b>
<b>Dédiasses.....</b>	<b>3</b>
<b>Remerciements.....</b>	<b>5</b>
<b>Résumé .....</b>	<b>7</b>
<b>Sommaire.....</b>	<b>11</b>
<b>Introduction.....</b>	<b>12</b>
<b>Chapitre 1 Présentation de l’organisme d’accueil.....</b>	<b>15</b>
<b>Partie 1 – Sonatrach .....</b>	<b>16</b>
<b>Introduction.....</b>	<b>16</b>
<b>Présentation de l’entreprise.....</b>	<b>16</b>
<b>Définition de Sonatrach .....</b>	<b>18</b>
<b>Historique de Sonatrach .....</b>	<b>18</b>
<b>Les principales Activités de la Société .....</b>	<b>19</b>
<b>Les usages des produits pétroliers raffinés.....</b>	<b>20</b>
<b>Logistique de Sonatrach .....</b>	<b>20</b>
<b>Perspectives d’avenir.....</b>	<b>23</b>
<b>Conclusion .....</b>	<b>23</b>
<b>Partie 2 – Problématique.....</b>	<b>24</b>
<b>Introduction générale .....</b>	<b>24</b>
<b>Problématique de recherche.....</b>	<b>25</b>
<b>Contextualisation du problème .....</b>	<b>26</b>
<b>Méthodologie de recherche.....</b>	<b>26</b>
<b>Objectif de la recherche.....</b>	<b>29</b>
<b>Conclusion .....</b>	<b>30</b>
<b>Chapitre 2 Domaine de recherche .....</b>	<b>31</b>
<b>Introduction .....</b>	<b>31</b>
<b>Définition de la Data science.....</b>	<b>32</b>
<b>L’histoire de la Data science.....</b>	<b>32</b>
<b>Définition d’un Data scientist .....</b>	<b>32</b>
<b>Domaine de la Data Science.....</b>	<b>33</b>
<b>Les métiers et leurs rôles de la Data science.....</b>	<b>34</b>
<b>Le rôle de la data science :.....</b>	<b>35</b>
<b>Les avantages qu’offre la data science .....</b>	<b>36</b>
<b>Types de données utilisées en Data science .....</b>	<b>37</b>
<b>Les sources de données et leur collecte.....</b>	<b>37</b>
<b>Technique et outils de la data science.....</b>	<b>39</b>

Les enjeux et défis de la data science.....	41
Le future de la Data science.....	42
<b>Chapitre 3 - Implémentation et conception.....</b>	<b>43</b>
Introduction .....	43
Présentatin du système .....	44
Modélisation.....	50
Implémentation.....	51
Conclusion .....	67
<b>Chapitre 4 – Conclusion générale .....</b>	<b>68</b>
<b>Bibliographie.....</b>	<b>74</b>
<b>Tables .....</b>	<b>80</b>
Tables des illustrations .....	81
Table des matières.....	83

---

## Tables

---



## Tables des illustrations

Figure 1 : Organigramme d'organisation de Sonatrach.....	17
Figure 2 : Opérations internationales de Sonatrach.....	22
Figure 3 : Glossaire.....	28
Figure 4 : Exemple du fichier TBF.....	28
Figure 5 : Exemple du fichier PMT.....	29
Figure 6 : Les métiers relatifs à la Data Science.....	34
Figure 7 : Cycle de vie de la Data Science.....	40
Figure 8 : La fenêtre principale de Visual Studio Code.....	45
Figure 9 : Menu d'extensions Visual Studio.....	46
Figure 10 : Exemple création de la fonction « Def ».....	52
Figure 11 : Exemple création de bouton.....	53
Figure 12 : Exécution du programme « main.py ».....	53
Figure 13 : Exécution du programme « main.py » avec l’affichage du tableau.....	54
Figure 14 : Fonction « trouver_doublons ».....	55
Figure 15 : Fonction « highlight doublons ».....	56
Figure 16 : Fonction "Def_fusion_files".....	58
Figure 17 : Fenêtre de tri.....	61
Figure 18 : Fenêtre représentant l’exécution du programme d’insertion.....	66
Figure 19 : Exemple d'application du thème.....	66



## Table des matières

<b>Dédiasses</b> .....	<b>3</b>
<b>Remerciements</b> .....	<b>5</b>
<b>Résumé</b> .....	<b>7</b>
<b>Sommaire</b> .....	<b>11</b>
<b>Introduction</b> .....	<b>12</b>
<b>Chapitre 1 Présentation de l'organisme d'accueil</b> .....	<b>15</b>
<b>Partie 1 – Sonatrach</b> .....	<b>16</b>
Introduction .....	<b>16</b>
Présentation de l'entreprise.....	<b>16</b>
Définition de Sonatrach .....	<b>18</b>
Historique de Sonatrach .....	<b>18</b>
Les principales Activités de la Société.....	<b>19</b>
Les usages des produits pétroliers raffinés .....	<b>20</b>
Logistique de Sonatrach.....	<b>20</b>
Perspectives d'avenir.....	<b>23</b>
Conclusion.....	<b>23</b>
<b>Partie 2 – Problématique</b> .....	<b>24</b>
Introduction générale .....	<b>24</b>
Problématique de recherche.....	<b>25</b>
<i>Hypothèses :</i> .....	<b>25</b>
Contextualisation du problème .....	<b>26</b>
Méthodologie de recherche.....	<b>26</b>
Objectif de la recherche.....	<b>29</b>
Conclusion.....	<b>30</b>
<b>Chapitre 2 Domaine de recherche</b> .....	<b>31</b>
Introduction .....	<b>31</b>
Définition de la Data science.....	<b>32</b>
L'histoire de la Data science.....	<b>32</b>
Définition d'un Data scientist.....	<b>32</b>
Domaine de la Data Science.....	<b>33</b>
Les métiers et leurs rôles de la Data science.....	<b>34</b>
<i>Leurs rôles :</i> .....	<b>34</b>
Le rôle de la data science : .....	<b>35</b>
Les avantages qu'offre la data science .....	<b>36</b>
Types de données utilisées en Data science .....	<b>37</b>
Les sources de données et leur collecte .....	<b>37</b>
Technique et outils de la data science .....	<b>39</b>
Les enjeux et défis de la data science .....	<b>41</b>
Le future de la Data science .....	<b>42</b>
<b>Chapitre 3 - Implémentation et conception</b> .....	<b>43</b>
Introduction .....	<b>43</b>
Présentatin du système .....	<b>44</b>
<b>1. Éléments caractéristiques de l'environnement de développement :</b> .....	<b>44</b>
<b>2. Les outils de développement :</b> .....	<b>44</b>
<b>L'IDE VS code</b> .....	<b>44</b>
<b>Le langage de programmation Python</b> .....	<b>47</b>
<b>Présentation :</b> .....	<b>47</b>
<b>Le Framework Tkinter</b> .....	<b>49</b>
<b>Présentation :</b> .....	<b>49</b>
<b>Modules et classes :</b> .....	<b>49</b>
<b>Modélisation</b> .....	<b>50</b>

Implémentation.....	51
<i>Programme « main » :</i> .....	51
<i>Programme « reperage_doublons » :</i> .....	54
<i>Programme « prog_combine » :</i> .....	57
<i>Programme « prog_total_glob » :</i> .....	59
<i>Programme « prog_ranking » :</i> .....	61
<i>Programme « prog_cellules_vides » :</i> .....	62
<i>Programme « program_insertion » :</i> .....	63
<i>Style utilisé (forest-dark) :</i> .....	66
Conclusion.....	67
Chapitre 4 – Conclusion générale .....	68
Bibliographie.....	74
Tables.....	80
Tables des illustrations .....	81
Table des matières .....	83

---

