



École Nationale Supérieure d'Informatique et d'Analyse des
Systèmes - RABAT

Rapport de projet de fin d'année

FILIÈRE : BUSINESS INTELLIGENCE AND ANALYTICS

Analyse comparative des offres d'emploi LinkedIn en Business Intelligence et Analytics pour les diplômés de l'ENSIAS

Réalisé par :
BEKKAI Chamss Doha
CHOUKHANTRI Ikram

Jury :
Pr. Amrani

Encadré par :
Pr. KERAZI

SOUTENU LE 09/06/2023

Année Universitaire : 2022 - 2023

Remerciement :

Nous tenons à exprimer notre profonde gratitude envers nos parents pour leur soutien inconditionnel dans la réalisation de notre projet de fin d'année. Leur amour, leur encouragement et leur dévouement ont été des éléments essentiels de notre parcours.

Leur présence constante et leur soutien émotionnel nous ont donné confiance et motivation. Leur aide logistique, que ce soit dans la recherche de matériel ou dans l'organisation de notre travail, a été précieuse pour nous permettre de nous concentrer pleinement sur notre projet.

Nos parents ont toujours cru en nous, même lorsque nous doutions de nos propres capacités. Leurs encouragements et leurs conseils avisés nous ont poussés à dépasser nos limites et à atteindre des résultats auxquels nous n'aurions pas cru possible.

Nous sommes reconnaissants pour les sacrifices qu'ils ont consentis pour nous offrir des opportunités éducatives. Leur investissement en termes de temps, d'énergie et de ressources a été immense et nous en sommes profondément reconnaissants.

Table des matières

Introduction générale	6
1 Contexte générale du projet	8
1.1 Adéquation : Formation - Marché d'emploi	8
1.2 Plateforme LinkedIn	8
1.3 Motivation et objectifs du projet	9
1.4 Méthodologie du projet	10
1.5 Planification	11
2 Extraction des données	13
2.1 Outils d'extraction des données	13
2.2 Collecte des données	14
2.3 Traitement des données	17
2.3.1 Exploration des données	17
2.3.2 Uniformation de la date	18
2.3.3 Uniformation de la localisation	19
2.3.4 Traitement de la description	19
2.3.5 Traitement des certificats	20
2.3.6 Suppression des duplications	20
2.4 Indicateurs et axes d'analyse	20
3 Analyse des données	22
3.1 L'outil d'analyse :	22
3.2 Conception des visualisations	22
3.3 Visualisations	23
3.3.1 Graphes relatifs aux offres	23
3.3.2 Graphes relatifs aux profils	29
4 Discussion	36
4.1 Résultats importants	36
4.2 Recommandations	37
4.3 Limites	38
Conclusion générale	39

Table des figures

1.1	LinkedIn	9
1.2	Conduite du projet	10
1.3	Diagramme de Gantt	12
2.1	Outils d'extraction des données	13
2.2	Resultat des offres avant traitement	18
2.3	Resultat des profils avant traitement	18
2.4	Resultat après traitement de la date	19
2.5	Resultat des offres après traitement de la localisation	19
2.6	Resultat des profils après traitement de la description	20
3.1	Power BI	22
3.2	Nombre d'offre par chaque entreprise	23
3.3	Répartition géographique des offres d'emploi	24
3.4	nombre d'offre demandant les outils d'analyse et de visualisation des données	24
3.5	nombre d'offre demandant langages de requête et de manipulation de données	25
3.6	nombre d'offre demandant Outils de développement et de gestion de versions	26
3.7	nombre d'offre demandant langages de développement et de programmation :	27
3.8	nombre d'offre demandant outils de gestion d'entreprise et de projets	28
3.9	nombre d'offre demandant ces autres Outils	29
3.10	Carte de répartition des lauréats	29
3.11	Répartition des lauréats par entreprise	30
3.12	Répartition des lauréats par entreprise dans la France	31
3.13	Répartition des lauréats par entreprise dans Canada	31
3.14	Tableau des entreprises des lauréats sans certifications	32
3.15	Certifications par personnes	33
3.16	Domaine de certifications	33
3.17	Double diplomation	34

Résumé :

Notre projet de fin d'année vise à étudier l'adéquation entre la formation académique de l'ENSIAS et les compétences recherchées dans les domaines de la Business Intelligence et de l'Analytics. Pour cela, nous utilisons la technique de scraping des données sur LinkedIn afin d'explorer les opportunités d'emploi dans ce domaine et de comparer les compétences demandées par les employeurs avec celles développées par les étudiants.

En utilisant la bibliothèque Selenium , nous avons extrait automatiquement 150 offres d'emploi spécifiques aux les profils de "Business Intelligence" et "Analytics" sur LinkedIn dans la période du 30 mai. Parallèlement, nous avons collecté des informations sur les profils des lauréats de l'ENSIAS présents sur la plateforme à l'aide de la bibliothèque BeautifulSoup, ce qui nous permet de réaliser une comparaison concrète entre les compétences requises et celles acquises par les étudiants.

Pour analyser les résultats obtenus, nous avons utilisé Power BI, un outil Microsoft d'analyse de données. Grâce à ses fonctionnalités de visualisation avancées, nous avons identifié les compétences les plus demandées par les employeurs et ainsi mettre en évidence les écarts éventuels entre la formation académique et les compétences professionnelles requises.

L'objectif final de notre projet est de formuler des recommandations pour améliorer l'alignement entre la formation académique de l'ENSIAS et les besoins du marché du travail en constante évolution dans les domaines de la Business Intelligence et de l'Analytics.

Mots clés : Adéquation, Business Intelligence, Data Analysis, Data visualization, LinkedIn, Marché de travail , Offres d'emploi, Power BI, Web scraping.

Abstract :

Our end-of-year project aims to study the alignment between the academic curriculum at ENSIAS and the skills sought after in the fields of Business Intelligence and Analytics. To achieve this, we employ the technique of scraping data from LinkedIn to explore employment opportunities in this domain and compare the skills demanded by employers with those developed by students.

Using the Selenium library, we automatically extracted 150 job postings specific to "Business Intelligence" and "Analytics" profiles on LinkedIn during the period of May 30th. Concurrently, we gathered information about ENSIAS graduates' profiles on the platform using the BeautifulSoup library, allowing us to make a concrete comparison between the required skills and those acquired by the students.

To analyze the obtained results, we utilized Power BI, a data analysis tool by Microsoft. Thanks to its advanced visualization features, we identified the most sought-after skills by employers, thereby highlighting any discrepancies between the academic curriculum and the required professional competencies.

The ultimate objective of our project is to formulate recommendations for enhancing the alignment between ENSIAS academic curriculum and the evolving demands of the job market in the fields of Business Intelligence and Analytics.

Key words : Alignment, Business Intelligence, Data Analysis, Data Visualization, LinkedIn, Job Market, Job Postings, Power BI, Web Scraping.

Introduction générale

Dans le cadre de notre projet de fin d'année, nous avons entrepris une étude approfondie visant à explorer et comparer les opportunités d'emploi pour les profils de Business Intelligence et d'Analytics. Pour atteindre cet objectif, nous avons utilisé la technique du scraping des données à partir de LinkedIn, l'une des plus grandes plateformes professionnelles en ligne. Cette approche nous a permis d'obtenir un aperçu précis des compétences demandées par les employeurs et de les confronter à la formation dispensée par l'ENSIAS, notre institution d'appartenance.

La problématique qui sous-tend notre projet est de savoir dans quelle mesure la formation de l'ENSIAS correspond aux exigences du marché du travail dans le domaine de la Business Intelligence et de l'Analytics. Nous sommes motivés par le souhait de comprendre si les compétences enseignées aux étudiants de notre filière sont en adéquation avec les besoins et les attentes des employeurs dans le secteur. Cette démarche nous permettra d'évaluer l'efficacité de notre programme d'études et d'identifier les éventuelles lacunes ou ajustements nécessaires pour mieux préparer les étudiants à leur future carrière.

La solution que nous proposons repose sur l'exploration et l'analyse des données extraites de LinkedIn, en se concentrant principalement sur les offres d'emploi et les profils des professionnels de la Business Intelligence et de l'Analytics. Nous avons développé une méthodologie rigoureuse pour collecter, traiter et analyser ces données, en mettant l'accent sur les compétences demandées par les employeurs et celles enseignées dans notre formation à l'ENSIAS.

Notre approche se divise en plusieurs chapitres clés.

- Le chapitre 1 présente le contexte général de notre projet, en mettant en évidence les enjeux et les tendances actuelles dans le domaine de la Business Intelligence et de l'Analytics.
 - Le chapitre 2 est dédié à l'exploration des données, où nous détaillerons le processus de scraping des offres d'emploi et des profils des lauréats sur LinkedIn.
 - Le chapitre 3 se concentre sur l'analyse des données collectées, en mettant en évidence les correspondances et les écarts entre les compétences demandées et enseignées.
 - Le chapitre 4 sera consacré à la discussion des résultats obtenus, en fournissant des insights et des recommandations pour améliorer notre programme de formation.
- L'organisation de ce document suit une structure logique correspondant à nos objectifs

de recherche. En explorant les opportunités d'emploi et en comparant les compétences demandées, nous espérons mettre en lumière l'adéquation entre la formation de l'ENSIAS et les exigences du marché du travail. De plus, nous cherchons à évaluer la pertinence de notre programme d'études face aux besoins actuels de l'industrie de la Business Intelligence et de l'Analytics.

Au travers de cette étude, nous espérons fournir des informations précieuses tant aux étudiants et enseignants de l'ENSIAS qu'aux professionnels du domaine. Nous croyons fermement que cette analyse comparative aidera à améliorer l'alignement entre la formation académique et les compétences

Contexte générale du projet

Ce chapitre, présente le cadre du projet, spécifie les objectifs visés et met en évidence la démarche du travail.

1.1 Adéquation : Formation - Marché d'emploi

De nos jours, nous sommes confrontés à un problème majeur : l'évaluation de l'adéquation entre la formation des étudiants et les offres d'emploi disponibles sur le marché du travail. Cette situation soulève la nécessité d'analyser attentivement cette problématique afin de prendre des mesures visant à améliorer la qualité de la formation dispensée, dans le but de mieux préparer les étudiants à leur future carrière.

L'évolution rapide du monde du travail, les avancées technologiques et les changements économiques constants créent un écart entre les compétences acquises par les étudiants et celles requises par les employeurs. Cette inadéquation entre la formation et les besoins du marché peut entraîner une situation préoccupante où les étudiants diplômés se retrouvent confrontés à des difficultés pour trouver un emploi correspondant à leurs compétences et aspirations professionnelles.

Pour résoudre cette problématique, il est primordial de mener une analyse approfondie de l'offre d'emploi actuelle et des compétences requises par les employeurs suivant plusieurs indicateurs .

1.2 Plateforme LinkedIn

Le choix de la plateforme LinkedIn pour le scraping des offres d'emploi repose sur plusieurs facteurs clés. LinkedIn est un réseau social professionnel majeur, utilisé par des millions de professionnels et d'entreprises à travers le monde. Cette plateforme offre une vaste base de données d'offres d'emploi provenant de divers secteurs d'activité et de niveaux d'expérience, ce qui en fait une source de données riche et pertinente pour notre étude.

En utilisant LinkedIn pour le scraping, nous avons accès à des informations détaillées sur les offres d'emploi, telles que les titres de poste, les descriptions, les compétences requises, les niveaux d'expérience et les emplacements géographiques. Ainsi, fournit des informations sur les lauréats de la filière BI&A grâce à leurs comptes sur la plateforme. Cela nous permet d'identifier des éléments clés tels que leurs noms, leurs entreprises, leurs



FIGURE 1.1 – LinkedIn

localisations et même les certifications qu'ils ont obtenues. Cette richesse d'informations nous permet d'analyser et de comprendre les tendances du marché du travail, les exigences des employeurs et les compétences recherchées et enseignées dans différents domaines.

De plus, LinkedIn propose des fonctionnalités de recherche avancées qui nous permettent de spécifier des critères de recherche précis, tels que des mots-clés, des emplacements et des filtres de compétences. Ces fonctionnalités nous aident à affiner nos résultats et à cibler spécifiquement les offres d'emploi qui correspondent à nos intérêts et à nos objectifs de recherche.

LinkedIn étant une plateforme en ligne, le scraping des offres d'emploi et des profils des lauréats est réalisé de manière automatisée en utilisant des techniques de programmation. Cela nous permet de collecter rapidement et efficacement un grand volume de données, ce qui facilite notre analyse et nos études comparatives.

1

1.3 Motivation et objectifs du projet

Notre projet de fin d'année trouve sa motivation dans la volonté de comprendre les exigences du marché du travail dans le domaine de la Business Intelligence, en relation avec notre formation à l'ENSIAS. En tant qu'étudiants de première année dans la filière Business Intelligence, nous sommes conscients de l'importance de développer les compétences nécessaires pour réussir dans ce secteur en constante évolution. C'est pourquoi nous avons entrepris de scraper les offres d'emploi actuelles afin de déterminer les compétences recherchées par les employeurs.

L'objectif principal de notre projet consiste à analyser les compétences demandées dans les offres d'emploi en Business Intelligence. Nous avons recueilli un échantillon d'offres d'emploi provenant de LinkedIn, et nous nous apprêtons à examiner attentivement les compétences spécifiques recherchées par les employeurs. Cette analyse approfondie nous permettra d'obtenir un aperçu précis des compétences clés qui sont actuellement en demande sur le marché du travail.

Notre principal objectif est de mettre en évidence les correspondances et les écarts entre la formation dispensée par l'ENSIAS et les compétences recherchées par les employeurs sur LinkedIn. Cette démarche vise à améliorer la formation des étudiants de la filière BI&A et à faciliter leur intégration sur le marché du travail.

Nous souhaitons aussi évaluer si la formation répond aux exigences actuelles de l'industrie de la Business Intelligence et de l'Analytics. Cette évaluation sera essentielle pour

1. Le lien de linkedIn :<https://www.linkedin.com/feed/>

actualiser la formation académique et l'adapter aux nouvelles technologies.

1.4 Méthodologie du projet

Pour mener à bien notre projet, nous avons adopté deux approches distinctes. La première consiste à analyser les offres d'emploi en Business Intelligence et Data Analyst publiées sur la plateforme LinkedIn. La deuxième approche se base sur l'analyse des profils des diplômés en BI&A, afin de procéder à une analyse approfondie des données. Pour visualiser ces données, nous utiliserons des graphiques, des illustrations et des visuels interactifs. Pour ce faire, nous utiliserons Python pour collecter les données et Power BI comme outils d'analyse et de visualisation.

Nous avons divisé le projet en plusieurs étapes.

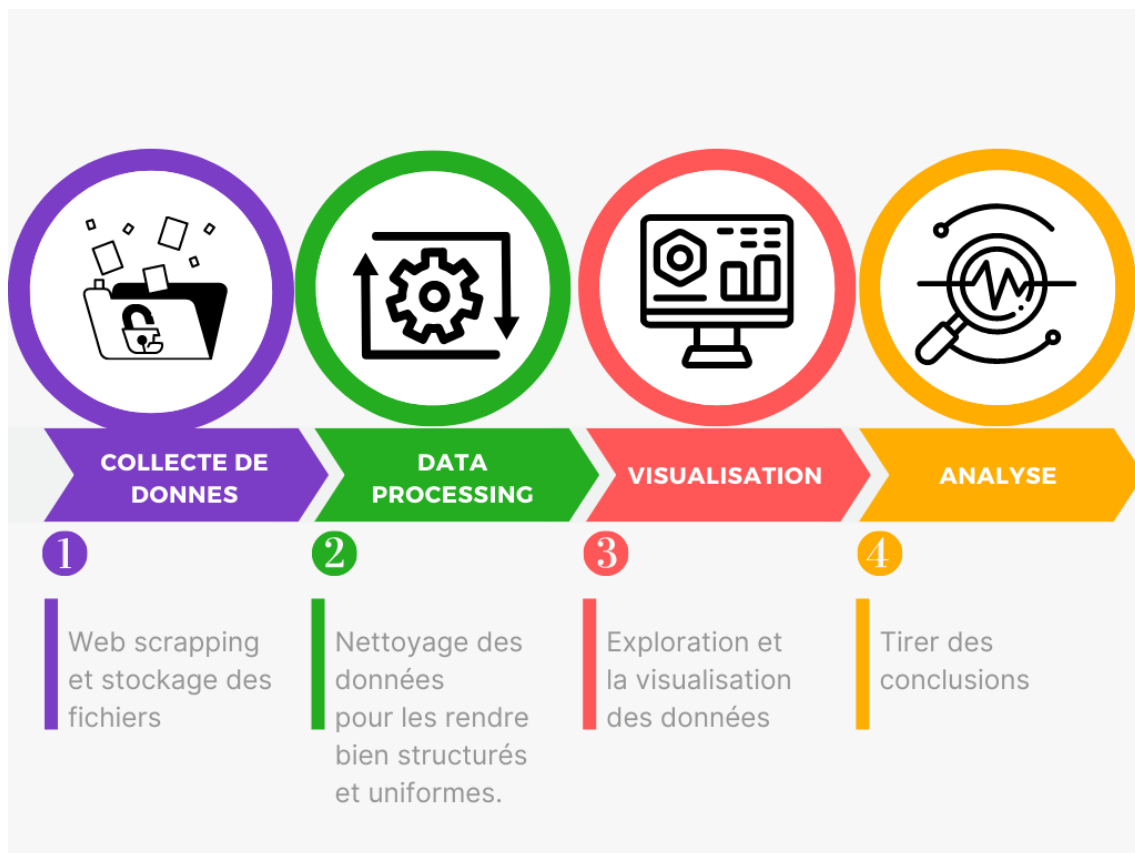


FIGURE 1.2 – Conduite du projet

L'architecture fonctionnelle de notre projet, représentée par la Figure 1.2 intitulée "Conduite de projet", comprend quatre étapes essentielles : la collecte de données, le traitement des données, la visualisation et l'analyse.

La première étape de notre projet est la collecte de données. Cette étape consiste à rassembler les offres d'emploi en Business Intelligence et des profils des lauréat à partir de LinkedIn et les stockées dans des fichiers Excel.

La deuxième étape de notre projet consiste au prétraitement des données, une étape essentielle pour nettoyer, organiser et transformer les données brutes extraites. Les données collectées à partir de différentes sources sont souvent dans un état non structuré

et inadapté à l'analyse ou à la modélisation des données. Le prétraitement des données implique des tâches telles que la suppression des doublons, la correction des erreurs, la gestion des valeurs manquantes, la normalisation des formats, la conversion des types de données, ainsi que la détection d'anomalies et la réduction de dimension. Ces étapes garantissent des données de meilleure qualité, cohérentes et prêtes à être utilisées dans les étapes suivantes du projet.

La troisième étape de notre architecture fonctionnelle est la visualisation. Après avoir traité les données, nous utiliserons l'outil de visualisation des données Power BI pour représenter graphiquement les informations extraites. Nous créerons des tableaux de bord interactifs et des graphiques visuels qui permettront de mieux comprendre les tendances, les relations et les schémas présents dans les offres d'emploi et les profils analysés.

Enfin, la dernière étape de notre architecture fonctionnelle est l'analyse. Une fois que les données ont été collectées, traitées et visualisées, nous procéderons à une analyse approfondie des compétences demandées dans les offres d'emploi. Nous utiliserons des techniques d'analyse des données pour identifier les compétences techniques spécifiques ainsi que les compétences transversales recherchées par les employeurs dans le domaine de la Business Intelligence. Tout d'abord, la collecte des données a été réalisée en utilisant LinkedIn comme source pour extraire les données pertinentes. Ensuite, nous avons sélectionné les outils de travail et les bibliothèques associées pour passer à l'implémentation du code.

Cette approche en plusieurs parties nous a permis de travailler de manière méthodique.

1.5 Planification

- Définition des objectifs : 11 avril 2023
- Collecte des ressources : 26 avril 2023
- Établissement d'un calendrier : 30 avril 2023
- Collecte des offres d'emploi : 30 avril - 30 mai 2023
- Collecte des profils : 30 avril - 30 mai 2023
- Analyse des compétences demandées : 15 mai - 2 juin 2023
- Comparaison avec le programme de formation de l'ENSIAS : 22 mai - 5 juin 2023
- Analyse des profils des lauréats : 18 mai - 2 juin 2023
- Analyse comparative et recommandations : 22 mai - 5 juin 2023
- Réalisation du rapport : 30 mai 2023
- Présentation des résultats : 9 juin 2023

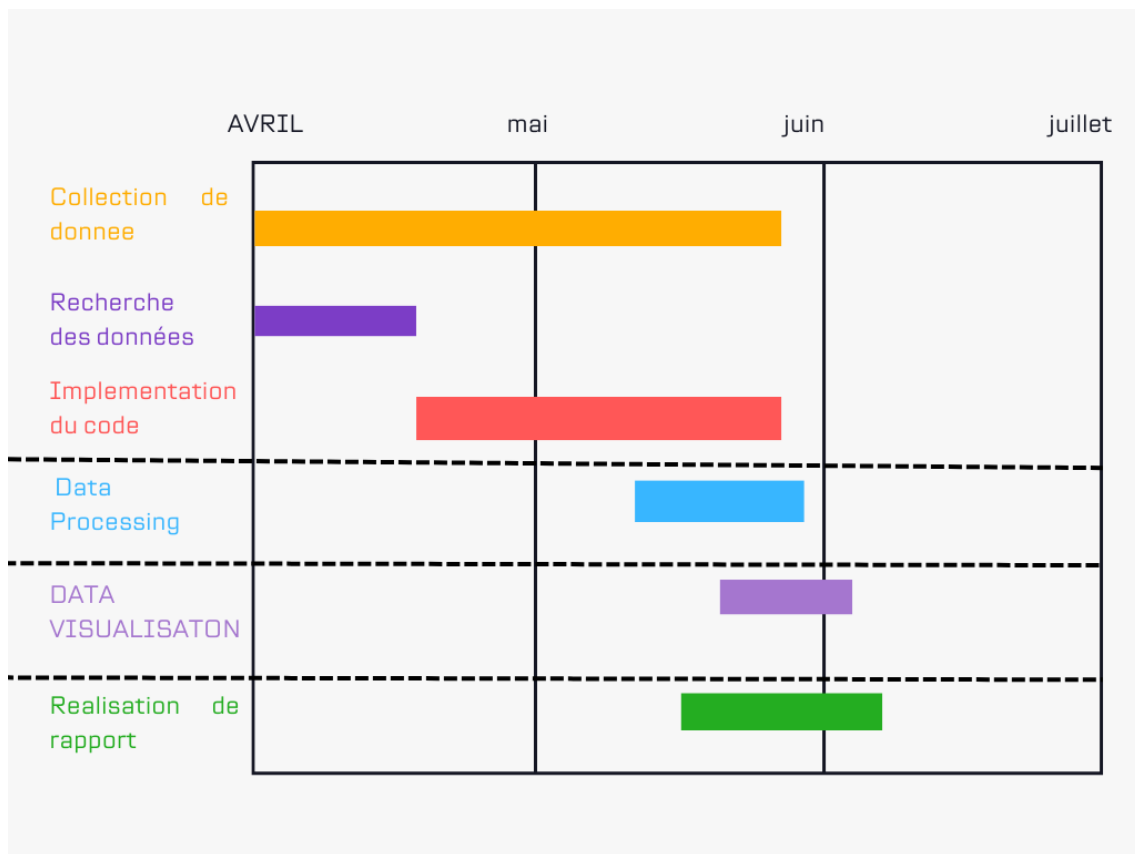


FIGURE 1.3 – Diagramme de Gantt

Synthèse :

En résumé, ce chapitre a permis de poser les bases de notre projet en présentant le contexte général et en tracant la voie à suivre.

Extraction des données

Ce chapitre est consacré à la présentation détaillée des étapes suivies lors de l'extraction des données, mettant en évidence la démarche adoptée , tout en précisant les outils utilisés.

2.1 Outils d'extraction des données

Dans notre projet, nous avons utilisé plusieurs outils d'extraction de données qui ont joué un rôle essentiel dans le processus.

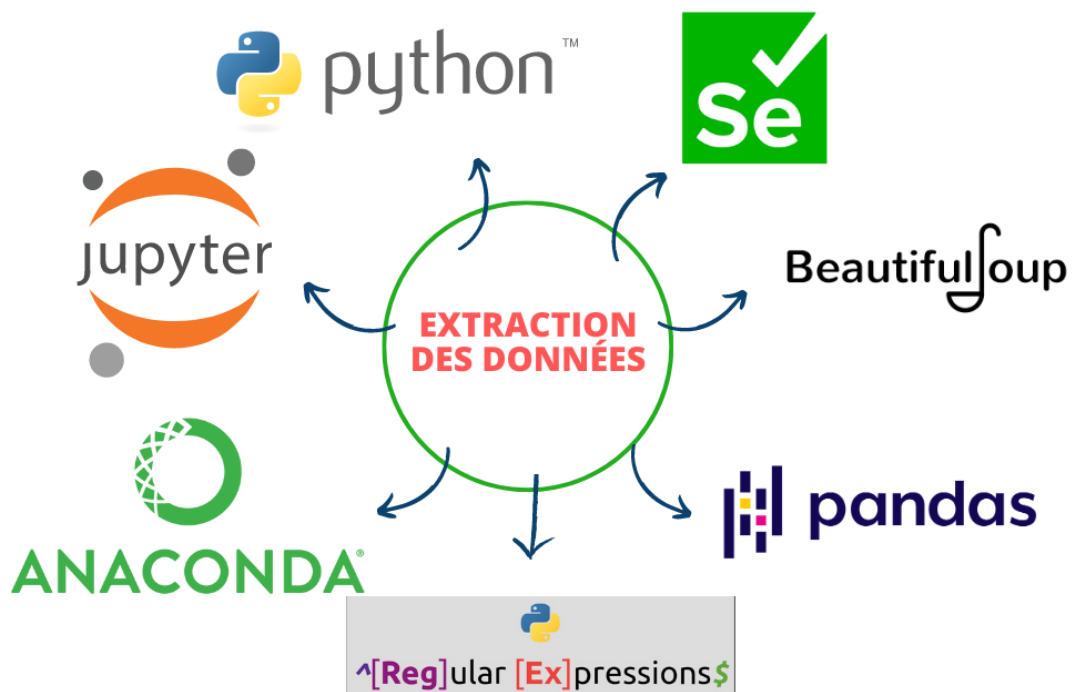


FIGURE 2.1 – Outils d'extraction des données

Voici comment nous avons utilisé chaque outil :

Anaconda : Nous avons utilisé Anaconda comme environnement de développement pour gérer nos packages Python et créer un environnement virtuel. Cela nous a permis de maintenir une configuration cohérente et d'installer facilement les bibliothèques nécessaires.

Jupyter : Nous avons utilisé Jupyter Notebook, qui est inclus dans Anaconda, pour développer et exécuter notre code Python de manière interactive. Jupyter nous a permis d'écrire du code, d'exécuter des cellules individuelles et de visualiser les résultats à mesure que nous progressions dans notre projet.

Python : Nous avons utilisé le langage de programmation Python comme principal outil pour le scraping des données et l'analyse. Python offre une syntaxe claire et concise, ainsi qu'une multitude de bibliothèques et de modules puissants qui facilitent l'extraction et le traitement des données.

Pandas : Nous avons utilisé la bibliothèque Pandas pour manipuler et analyser les données extraites. Pandas offre des structures de données flexibles, telles que les DataFrames, qui nous ont permis de nettoyer, transformer et organiser les données de manière efficace.

Regular Expression : Nous avons utilisé les expressions régulières (regular expressions) pour le traitement des descriptions des offres d'emploi. Les expressions régulières nous ont permis de rechercher et d'extraire des motifs spécifiques dans les textes des descriptions, tels que des compétences techniques. Cela nous a permis de structurer les informations pertinentes pour l'analyse ultérieure.

Beautiful Soup : Nous avons utilisé la bibliothèque Beautiful Soup lors du scraping des profils des lauréats. En analysant le code HTML des pages web des profils, Beautiful Soup nous a aidés à extraire les données spécifiques que nous recherchions, telles que les certificats obtenus, les formations suivies, ou la mention d'une double diplomation. Nous avons pu ainsi obtenir une vision globale des profils des lauréats.

Selenium : Lors du scraping des offres d'emploi, nous avons utilisé Selenium, un framework d'automatisation de navigateur web. Selenium nous a permis de simuler des actions humaines sur les sites web, notamment le remplissage de formulaires, le clic sur des boutons et le défilement des pages, afin d'accéder aux informations des offres d'emploi qui nécessitaient une interaction dynamique. Cela nous a permis de collecter de manière exhaustive les données des offres d'emploi.

2.2 Collecte des données

Dans ce chapitre, nous aborderons l'architecture technique mise en place pour réaliser notre projet de scraping des offres d'emploi LinkedIn et des profils des lauréats dans le domaine de la Business Intelligence et de l'analytique. Nous avons utilisé le langage de programmation Python pour développer nos solutions. En ce qui concerne le web scraping, nous avons utilisé les bibliothèques Selenium et BeautifulSoup en combinaison avec le navigateur Chromedriver pour automatiser la navigation sur LinkedIn et extraire les informations pertinentes des offres d'emploi. Selenium nous a permis de remplir les formulaires de connexion, de parcourir les pages de résultats et de récupérer les détails de chaque offre. En utilisant les mots-clés "Business Intelligence" et en définissant la localisation au Maroc, nous avons effectué une recherche ciblée d'offres d'emploi. Après un processus approfondi, nous avons réussi à extraire un total de 150 offres correspondantes.

sur une période allant du 30 mai. En plus de l'extraction des offres d'emploi, nous avons également extrait les informations des lauréats de l'ENSIAS. Nous avons pu collecter des données sur les profils des diplômés de l'ENSIAS, y compris leurs compétences, leurs certifications voire leur formation. Cette extraction d'informations supplémentaires nous a permis d'enrichir notre base de données et d'obtenir des insights plus précis sur les profils recherchés dans le domaine de la Business Intelligence et de l'analytique. Pour réaliser cette extraction, nous avons collecté les liens des profils des anciens lauréats de la filière Business Intelligence (BI) sur la page officielle de l'ENSIAS sur LinkedIn. Dans cette page, nous avons identifié la rubrique dédiée aux anciens élèves et avons effectué une recherche en filtrant les résultats avec le mot-clé spécifique : "Business Intelligence And Analytics". En intégrant les informations des lauréats de l'ENSIAS à notre projet de scraping des offres d'emploi LinkedIn, nous avons pu fournir une vision plus complète du marché de l'emploi dans ce domaine, en mettant en évidence les compétences et les profils les plus demandés par les entreprises.

Pour ce faire on a utilisé Jupyter Notebook comme un environnement de développement interactif avec Python. En prenant en considération les caractéristiques de LinkedIn, Nous avons adopté la structure suivante pour le code utilisé dans le scraping des offres d'emploi :

Après l'initialisation du pilote Chromedriver on procède au chargement de la page de connexion LinkedIn en utilisant la méthode `get()` du pilote et en spécifiant l'URL. On définit le nom d'utilisateur et le mot de passe pour finalement accéder au compte LinkedIn.

Chargement de la page des offres d'emploi spécifiques à la BI et à l'analytique au Maroc en utilisant la méthode `get()` avec l'URL appropriée.

Trouver toutes les offres d'emploi dans le bloc en utilisant la méthode `find-elements()`. Pour chaque offre d'emploi, trouver tous les éléments `<a>` et vérifier si l'attribut `href` commence par `"https://www.linkedin.com/jobs/view"` et si le lien n'est pas déjà dans la liste `links`. Si c'est le cas, ajouter le lien à la liste.

Faire défiler la page pour afficher l'élément de l'offre d'emploi en utilisant la méthode `execute-script()`. Le code visite chaque lien un par un et recherche ensuite les éléments HTML appropriés pour extraire les informations pertinentes suivantes :

jobtitles : Titres des offres d'emploi extrait avec la balise `<h1>`

companynames : Noms des entreprises de la classe CSS `".jobs-unified-top-card-company-name"`

Lieu de l'entreprise : la localisation des offres contenue dans la classe CSS `".jobs-unified-top-card-bullet"`

Date de publication : la date où l'offre est publiée, de la classe CSS `".jobs-unified-top-card-posted-date"`

Description de l'emploi : il s'agit d'un texte descriptif de chaque offre qui explique les missions et les compétences requises. Ces éléments sont généralement identifiés par la classe CSS `".jobs-description-content"`. La méthode `'find-elements'` retourne une liste de

tous les éléments correspondants. Ensuite, une boucle est utilisée pour parcourir chaque élément de la liste et extraire le texte de la description de l'emploi à l'aide de la méthode 'text'.

Stockage des informations extraites : Les informations (jobtitles, companynames, location, date,description) sont stockées dans des listes respectives.

Ajout des informations extraites à la liste de données , puis création d'un DataFrame à partir de cette liste et sauvegarde en tant que fichier EXCEL à l'aide de la bibliothèque pandas.

De même, pour le code de scraping des lauréats, voici la démarche suivie :

Vérification de la connexion à mon compte, puis cliquer sur le bouton de connexion une fois à l'intérieur.on donne une liste contenant tous les liens des profils linkedIn des lauréats de la BI

Parcours des liens de profil : Une boucle est utilisée pour parcourir tous les liens de profil donné . Pour chaque lien, les étapes suivantes sont effectuées :

Accès au profil LinkedIn : Le pilote Selenium accède au lien du profil.

Extraction du nom : La balise '<h1>' avec la classe "text-heading-xlarge inline t-24 v-align-middle break-words" contient le nom de l'utilisateur.

Extraction de la localisation : La balise '' avec la classe "text-body-small inline t-black-light break-words" contient la localisation de l'utilisateur.

Extraction de l'entreprise : La balise '<div>' avec la classe "inline-show-more-text" contient le nom de l'entreprise de l'utilisateur.

Extraction de la double diplomation : Nous extrayons les formations en utilisant la balise '' avec la classe "mr1 hoverable-link-text t-bold". Ensuite, nous vérifions si l'ENSIAS n'est pas la dernière étape dans le parcours scolaire de chaque individu.

Accès aux certifications : Le pilote Selenium accède à la section des certifications du profil. La balise '<a>' avec la classe "optional-action-target-wrapper" et l'ID "navigation-index-see-all-licenses-and-certifications" est recherchée, et son attribut "href" est utilisé pour accéder à la section des certifications.

Extraction des certifications : La balise '<div>' avec la classe "display-flex flex-column" contient les certifications de l'utilisateur. Les balises '' à l'intérieur de cette balise représentent chaque certification. Conversion de la liste des certifications en une chaîne séparée par des virgules.

Stockage des informations extraites : Les informations (nom, entreprise, localisation, certifications,double diplomation) sont stockées dans des listes respectives.

Ajout des informations extraites (nom, certifications, localisation, entreprise,double diplomation) à la liste de données (data), puis création d'un DataFrame à partir de cette liste et sauvegarde en tant que fichier EXCEL à l'aide de la bibliothèque pandas.

Finalement, nous disposons de deux bases de données distinctes. La première base de données concerne les offres d'emploi et contient les colonnes suivantes : nom de l'entreprise,

poste demandé, date de publication de l'offre, localisation, lien vers l'offre, description de l'offre et compétences requises.

La deuxième base de données concerne les profils et comprend les colonnes suivantes : nom, entreprise actuelle, localisation, certifications obtenues et formation. Dans la colonne de la formation, nous utilisons les codes "DD" pour indiquer que le lauréat a effectué une double diplomation, et "NO" pour indiquer qu'il ne l'a pas fait.

2.3 Traitement des données

Le prétraitement des données implique un processus visant à nettoyer et organiser les données pour les rendre structurées et cohérentes. Une fois que les données sont extraites, on obtient des données brutes qui sont les données collectées à partir de différentes sources, dans leur état d'origine. Cependant, ces données brutes ne sont généralement pas dans le format le plus adapté pour l'analyse ou la modélisation des données.

Le traitement des données englobe l'ensemble du processus de collecte, de transformation (tel que le nettoyage ou la mise en forme des données dans un état utilisable) . Les données propres désignent les données obtenues après le traitement des données brutes, c'est-à-dire des données prêtes à être analysées. Elles ont été transformées dans un format utilisable, et les données incorrectes, incohérentes ou manquantes ont été corrigées ou supprimées autant que possible.

Il existe plusieurs raisons pour lesquelles nous devons appliquer des opérations de traitement des données aux données brutes. Voici comment nous avons choisi de standardiser le contenu du dataframe :

- Suppression des doublons.
- Normalisation des dates.
- Uniformisation des emplacements.
- Nettoyage de la description.

Ces étapes permettent d'obtenir des données propres et cohérentes, prêtes à être analysées et utilisées dans des analyses ultérieures.

2.3.1 Exploration des données

Nous avons créé un DataFrame en utilisant les listes existantes contenant les titres des offres, les noms des entreprises, les emplacements, les dates de publication, les liens et les descriptions des emplois. Chaque liste a été convertie en une colonne distincte dans le DataFrame, ce qui nous a permis d'organiser les données de manière tabulaire. Et après le traitement des descriptions de chaque offre les résultats sont enregistrés dans des colonnes spécifiques du DataFrame, qui représentent les différentes catégories d'informations extraites. Enfin, le DataFrame modifié est enregistré dans un fichier Excel, remplaçant ainsi le fichier d'origine. Cette structure facilite l'accès et la manipulation des données, car nous pouvons maintenant effectuer des opérations et des analyses sur des colonnes spécifiques. Après avoir extrait les données des offres d'emploi, voici un extrait du fichier Excel obtenu.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Company	Job Title	Post Date	Location	Job Link	Job Description							
2	IBMMQ	Développ	il y a 11 h	Casablanca et périphérie	https://w	À propos de l'offre d'emploiNous sommes à la recherche d'un développeur talen							
3	Appen	Data Coll	il y a 18 h	Meknès et périphérie	https://w	À propos de l'offre d'emploiData Collection Project for Arabic Speakers in Moroc							
4	Scientiae	Analyste	il y a 1 ser	Rabat, Rabat-Salé-Kénitra, Maroc	https://w	À propos de l'offre d'emploiEntrepriseScientiae est fondée par des experts en cy							
5	Société Gi	Operatio	il y a 14 h	Casablanca, Casablanca-Settat, Maroc	https://w	À propos de l'offre d'emploiMissionsDans le cadre du département Front office,							
6	Société Gi	DATA AN	il y a 1 mo	Maroc	https://w	À propos de l'offre d'emploiVos missions au quotidienLe Data Analyst aura pour							
7	Société Gi	Chef de p	il y a 4 ser	Casablanca, Casablanca-Settat, Maroc	https://w	À propos de l'offre d'emploiVos missions au quotidienL'objectif principal d'un d							
8	ABL Aviat	Technical	il y a 1 jou	Casablanca, Casablanca-Settat, Maroc	https://w	À propos de l'offre d'emploiABL Aviation // Jr Technical Analyst• Location: Casab							
9	Devotean	Développ	il y a 3 jou	Dar Chaffai, Casablanca-Settat, Maroc	https://w	À propos de l'offre d'emploiDescription de l'entrepriseAvec plus de 8000 collabo							
10	take a che	Guest exp	il y a 14 h	Tarfaya, Laâyoune-Sakia El Hamra, Maroc	https://w	À propos de l'offre d'emploiAbout Take a ChefFounded in 2012, Take a Chef bega							
11	Deloitte	Data Anal	il y a 1 ser	Méchouar de Casablanca, Casablanca-Settat, Maroc	https://w	À propos de l'offre d'emploi« Tous nos postes sont ouverts au télétravail »Rejoir							
12	TalentKor	Junior Daf	il y a 1 ser	Essaouira, Marrakech-Safi, Maroc	https://w	À propos de l'offre d'emploiLOCATION: Remote WorkSCHEDULE: Part-timeAre yc							
13	Stellantis	Bi Data Ar	il y a 3 ser	Casablanca, Casablanca-Settat, Maroc	https://w	À propos de l'offre d'emploiLe Pilote PFM a La Responsabilité Du Pilotage Et Réal							
14	Leyton	Senior Bui	il y a 1 ser	Casablanca et périphérie	https://w	À propos de l'offre d'emploiExperienced Automation Software Engineer WANTEI							
15	American	Informati	il y a 2 ser	Casablanca-Settat, Maroc	https://w	À propos de l'offre d'emploiThe appropriate candidates need to have been instr							
16	Cnexia	DevOps Ei	il y a 3 ser	Fès, Fès-Meknès, Maroc	https://w	À propos de l'offre d'emploiJoining Cnexia is choosing to be part of an ambitious							
17	OMNISHC	Ingénieur	il y a 8 h	Casablanca et périphérie	https://w	À propos de l'offre d'emploiNous sommes à la recherche d'un développeur talen							
18	Société Gi	Business	il y a 1 mo	Casablanca, Casablanca-Settat, Maroc	https://w	À propos de l'offre d'emploiVos missions au quotidienLa mission du BA pour le c							
19	BCG Platir	Senior IT	(il y a 1 ser	Casablanca, Casablanca-Settat, Maroc	https://w	À propos de l'offre d'emploiWHO WE AREAt BCG Platinion we're keen to engage							
20	Arrow Ele	EMEA Bus	il y a 1 ser	Casablanca, Casablanca-Settat, Maroc	https://w	À propos de l'offre d'emploiPosition:EMEA Business AnalystJob Description:Our							

FIGURE 2.2 – Resultat des offres avant traitement

Et après avoir extrait les données des lauréats, voici le résultat final obtenu :

1	Name	Company	Location	Certifications	Formations
2	Abdelhalim Asfiri Eng.	IBM	Rabat, Rabat-Salé-Kénitra	SAP Certified Application Associate - SAP Analytics Cloud, SAP Certified Application Associate - Reporting, Modeling and Data Acquisiti	No
3	Oumayma Touji	Orange Business	Rabat, Rabat-Salé-Kénitra	The First Steps in Negotiating, QlikView: Getting Started with QlikView for Data Visualization, Certified SAFE® 5 Practitioner	No
4	Imane CHAOUI	OCF SA	Maroc	Data Science and Big Data Analytics : Making Data-Driven Decisions	No
5	Abdessamad Azrou	INTEL CIA	Maroc		No
6	Ziyad HARROUD	Business & Decision	Clermont-Ferrand, Auvergne-Rhône-Alpes, France		DD
7	Imane Benchgroun	OMNISHORE Groupe	Casablanca et périphérie		No
8	Samaha Eddallil	Orange Business Serv	Rabat, Rabat-Salé-Kénitra, Maroc		No
9	Mouad Assaital	IBM	Rabat, Rabat-Salé-Kénitra	A Crash Course in Data Science	No
10	Mariam Janati Idrissi	Ecole Nationale Supér	Luxembourg	POWER BI DATA ANALYST, Statistics, Data Science & Analytics Intro	No
11	Anas Saoudi	Ecole Nationale Supér	Casablanca et périphérie	Enterprise Systems, Python Data Structures	No
12	Ahmed Elhami	Pharma Pilot (part of	Rabat-Salé-Kénitra, Maroc		No
13	Reda Menaoui	Umanis	Mohammedia, Casablanca-Settat, Maroc		No
14	Monia EL ADIB	CGI	Montpellier, Occitanie, Fr	Microsoft Certified: Data Analyst Associate	No
15	Badr-Eddine Aboulkar	Orange Business Mar	Rabat, Rabat-Salé-Kénitra, Maroc		No
16	Fatima AL hammouti	CGI	Rabat et périphérie	Neural Networks and Deep Learning, What is data science ?, Autopsy Training	No
17	Abdelhak ELBIARI	EURAFRIC INFORMATI	Mohammedia, Casablanc	Google Cloud Big Data and Machine Learning Fundamentals, Improving Deep Neural Networks: Hyperparameter Tuning, Regularizatio	No
18	Yassine Toumi	Plus Company	Montréal, Québec, Cana	Spécialisation Deep Learning, Machine Learning, HCIA-AI	DD
19	REBBAJ Chamsseddin	Société Générale Mar	Rabat, Rabat-Salé-Kénitra, Maroc		No
20	Mounia Katih	CGI	Maroc	Big Data Engineer - Explorer Award 2021	No
21	Mehdi BELCHITI	Deloitte	Maroc	Applied Data Science with Python - Level 2, Data Analysis Using Python, Data Visualization Using Python, Database Management Esse	No
22	CHOUIEKH Adil	IBM	Rabat et périphérie	MDM Essentials, Microsoft Certified: Azure Data Fundamentals, IBM SPSS Modeler V18.2.x Essentials, Enterprise Design Thinking Co-I	No
23	Yassine BOURICHI	IBM	Casablanca-Settat, Mar	License to Lead - First-Line Manager, Technical Specialist Profession Certification - Level 1, Software Application Implementation Speci	No
24	Bassma EL GHAAZAZ	IBM	Rabat, Rabat-Salé-Kénitra	Consumer Products Industry Jumpstart, Enterprise Design Thinking Practitioner, IBM Agile Explorer, Method Essentials	No
25	Asmae AMNAS	IBM	Casablanca-Settat, Mar	IBM Agile Explorer, IBM Consulting Way, Enterprise Design Thinking Practitioner, Google Cloud Certified Professional Data Engineer, NDD	No
26	Moncef Bender	IBM	Rabat et périphérie	AWS Certified Data Analytics - Specialty, AWS Certified Cloud Practitioner, Forward Program, Foundations of Strategic Business Analy	No

FIGURE 2.3 – Resultat des profils avant traitement

2.3.2 Uniformation de la date

En utilisant les bibliothèques pandas, datetime, locale et re, nous avons pu effectuer un traitement sur la colonne de dates de notre fichier Excel. Nous avons converti les durées de publication en dates de publication spécifiques. Pour ce faire, nous avons extrait les quantités et les unités de temps des durées à l'aide des expressions régulières. En fonction de l'unité de temps (heures, semaines, mois), nous avons soustrait la quantité correspondante à partir de la date ultérieure, qui représente la date à laquelle nous avons effectué le scraping des offres. Cela nous a permis de calculer les dates de publication réelles. Finalement, nous avons ajouté une nouvelle colonne contenant ces dates de publication au dataframe et enregistré le dataframe modifié dans un nouveau fichier Excel.

Voici le resultat obtenu :

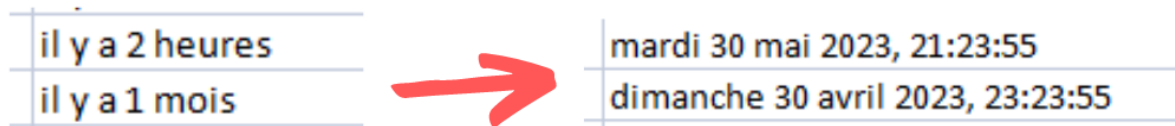


FIGURE 2.4 – Resultat après traitement de la date

2.3.3 Uniformation de la localisation

L'uniformation de la localisation consistait notamment à regrouper différentes variations de noms de villes pour les rendre plus cohérentes et compréhensibles. Par exemple, nous avons transformé des villes telles que "Casablanca, Casablanca-Settat, Maroc" en "Casablanca Périphérie". De cette manière, nous avons pu harmoniser les données et faciliter leur analyse ultérieure. Cette uniformisation de la localisation nous a permis d'obtenir une vue plus claire et homogène des différentes offres d'emploi et d'effectuer des analyses plus précises sur les zones géographiques. Voici le resultat après uniformation des location :

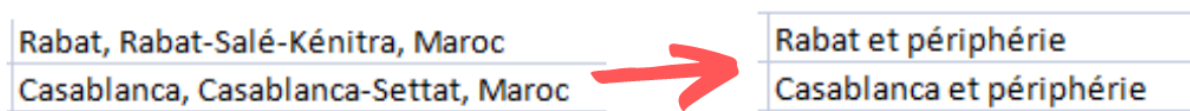


FIGURE 2.5 – Resultat des offres après traitement de la localisation

2.3.4 Traitement de la description

En utilisant le module "re" de Python, nous avons pu effectuer un nouveau traitement sur la description des offres d'emploi. Ce traitement consiste à extraire des informations pertinentes en utilisant des expressions régulières pour identifier des phrases contenant des mots-clés spécifiques et des compétences individuelles.

Nous avons défini une liste de mots-clés qui représentent des compétences recherchées dans les offres d'emploi, telles que Python, Pandas, Numpy, SQL, Data Science, etc. En parcourant chaque mot-clé, nous recherchons s'il est présent dans le texte de la description de l'offre d'emploi. Si un mot-clé est trouvé, il est ajouté à une liste de compétences. Cette liste de compétences est ensuite organisée en une chaîne de caractères avec des virgules pour chaque offre d'emploi. Enfin, le dataframe modifié est enregistré dans un nouveau fichier Excel.

Le résultat après ce traitement :

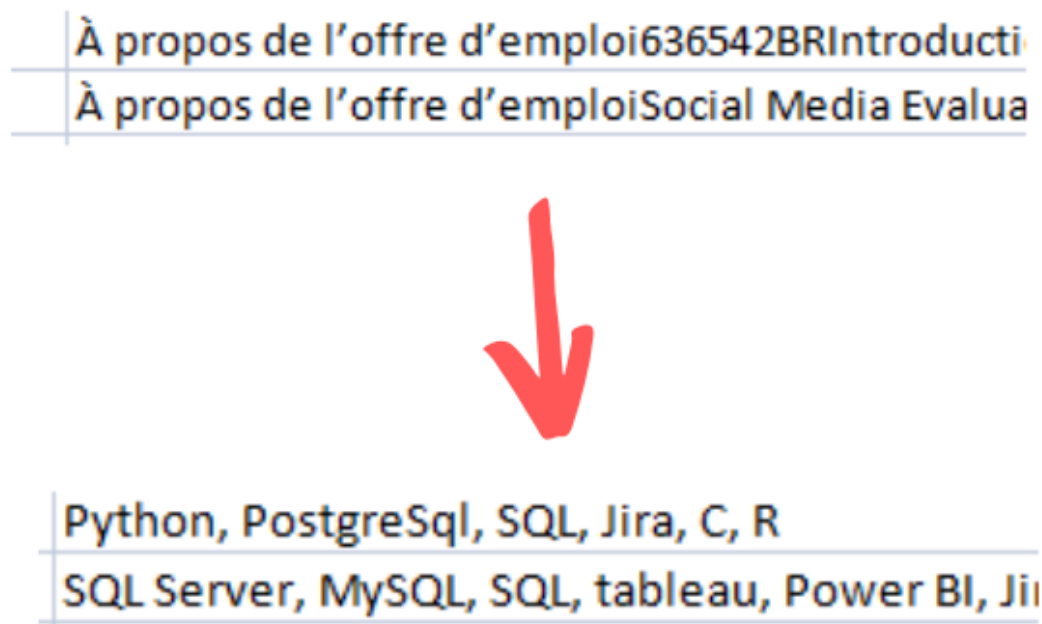


FIGURE 2.6 – Resultat des profils après traitement de la description

2.3.5 Traitement des certificats

Lors du scraping des certificats de chaque profil, le résultat était une liste contenant les certificats pour chacun, délimitée par des virgules. Afin de tirer profit de ces données lors de l'analyse, nous avons procédé au traitement de cette colonne dans l'éditeur Power Query. Nous avons utilisé la fonction de fractionnement de colonne en utilisant la virgule comme délimiteur. Cela nous a permis de créer une colonne distincte pour chaque certificat, où chaque certificat occupe sa propre colonne.

2.3.6 Suppression des duplications

Lors du scraping des données LinkedIn, il peut y avoir des redondances lors de l'extraction des profils des lauréats. Pour résoudre ce problème, l'éditeur Power Query offre une fonctionnalité de suppression des doublons. Cette option permet de filtrer les enregistrements en éliminant les entrées en double, garantissant ainsi une liste de profils uniques. Cela facilite l'analyse ultérieure des données. En utilisant cette fonctionnalité, les doublons peuvent être éliminés efficacement, améliorant la qualité des données obtenues.

2.4 Indicateurs et axes d'analyse

La première dataset obtenue est formée par les colonnes suivantes :

- Company : Le nom de l'entreprise pour les offres d'emploi.
- Job Title : Le titre de l'offre.
- Date de publication : La date de publication de l'offre.
- Location : La ville où se situe l'entreprise pour les offres.
- Job Description : La description de l'offre.

- Link : Le lien de l'offre.
- Compétences : Le niveau de diplôme requis.

La deuxième dataset obtenue est formée par les colonnes suivantes :

- Name : Le nom de l'entreprise pour les profils des lauréats.
- Location : La ville où se situe l'entreprise pour les lauréats.
- Certifications : Les certifications des lauréats.
- Company : Le nom de l'entreprise pour les lauréats.
- Formation : Indique si le lauréat en question a fait une double diplomation ou non.

Synthèse :

Ce chapitre est consacré à la collecte des données sur LinkedIn, où nous avons exposé les outils utilisés et la méthodologie suivie. Nous avons également effectué le traitement des données afin d'en tirer le meilleur parti.

Analyse des données

Après avoir collecté les données, nous avons entrepris leur processus de nettoyage. Dans ce chapitre, nous nous concentrerons sur la visualisation de ces données.

3.1 L'outil d'analyse :

Power BI :

Power BI est un outil d'analyse et de visualisation de données développé par Microsoft. Il permet aux utilisateurs de se connecter à différentes sources de données, de les transformer et de les modéliser. Avec Power BI, il est facile de créer des tableaux de bord interactifs et des rapports dynamiques. L'outil offre des fonctionnalités avancées telles que la création de mesures personnalisées et le partage de données. Power BI facilite l'analyse approfondie des données et la prise de décisions éclairées. Cet outils est utilisé pour supprimer les doublons pour les profils et pour visualiser les données



FIGURE 3.1 – Power BI

3.2 Conception des visualizations

Une fois que nous avons nettoyé et préparé nos données, nous pouvons passer à l'étape d'exploration et de visualisation. Dans ce sens nous proposons les visuels suivant :

1. Effectuer une recherche sur les compétences les plus recherchées sur le marché du travail au Maroc pour le domaine de la Business Intelligence (BI).
2. Les entreprises qui proposent le plus grand nombre d'offres d'emploi dans le domaine de la BI.

3. Les zones géographiques qui affichent un nombre plus élevé d'offres d'emploi dans le domaine de la BI.
4. Analyser la répartition géographique des diplômés de l'ENSIAS en filière BI.
5. Identifier les entreprises qui recrutent le plus de diplômés en BI.
6. Le pourcentage de personnes n'ayant pas obtenu de certifications supplémentaires et identification de leurs employeurs.
7. Examiner les domaines les plus fréquemment choisis par les diplômés pour leurs certifications.
8. Recenser le nombre de personnes ayant effectuer une double diplomation.

3.3 Visualisations

3.3.1 Graphes relatifs aux offres

Combien d'offres chaque société propose-t-elle ?

Company	NombreOffre
ABL Aviation	1
Agoda	10
ALTEN	2
ALTEN MAROC	2
Americana Restaurants	1
Appen	5
Apside	1
ARAGO Consulting	1
Arrow Electronics	6
Artefact	1
Attijariwafa bank	1
AVL	1
Azentio Software	1
BCG Platinion	1
BDSI Maroc	1
BMCI Groupe BNP Paribas	2
BNP Paribas	1
Boston Consulting Group (BCG)	5
Capgemini Engineering	3
CGI	7
Cnexia	3
Total	150

FIGURE 3.2 – Nombre d'offre par chaque entreprise

La table présentée ci-dessus affiche le nombre d'offres d'emploi fournies par chaque société, avec les colonnes correspondant au nom de la société et au nombre d'offres. En analysant les données, nous constatons que la société Agoda propose le plus grand nombre de postes de travail liés à la Business Intelligence, avec un chiffre de 10 offres. Ensuite, nous avons la société CGI qui offre un nombre équivalent de 7 postes.

Combien d'offres d'emploi sont disponibles dans chaque zone géographique ?

Location	NombreOffre
Casablanca et périphérie	95
Essaouira, Marrakech-Safi, Maroc	2
Fès, Fès-Meknès, Maroc	2
Maroc	9
Marrakech, Marrakech-Safi, Maroc	11
Meknès et périphérie	3
Rabat et périphérie	25
Tanger-Tétouan-Al Hoceïma, Maroc	1
Tarfaya, Laâyoune-Sakia El Hamra, Maroc	1
Zaouit Irs, Souss-Massa, Maroc	1
Total	150

FIGURE 3.3 – Répartition géographique des offres d'emploi

La table ci-dessus présente la répartition géographique des offres d'emploi, avec les colonnes indiquant la localisation et le nombre d'offres correspondant pour chaque région. L'analyse des données révèle que la région de Casablanca offre un total de 95 offres, soit 63% de l'ensemble des 150 offres répertoriées. Ensuite, nous avons la région de Rabat qui propose 25 offres.

Combien d'offres d'emploi exigent cette compétence spécifique ?

Pour analyser les compétences demandées par chaque offre d'emploi, nous avons regroupé des compétences précises pour chaque visualisation.

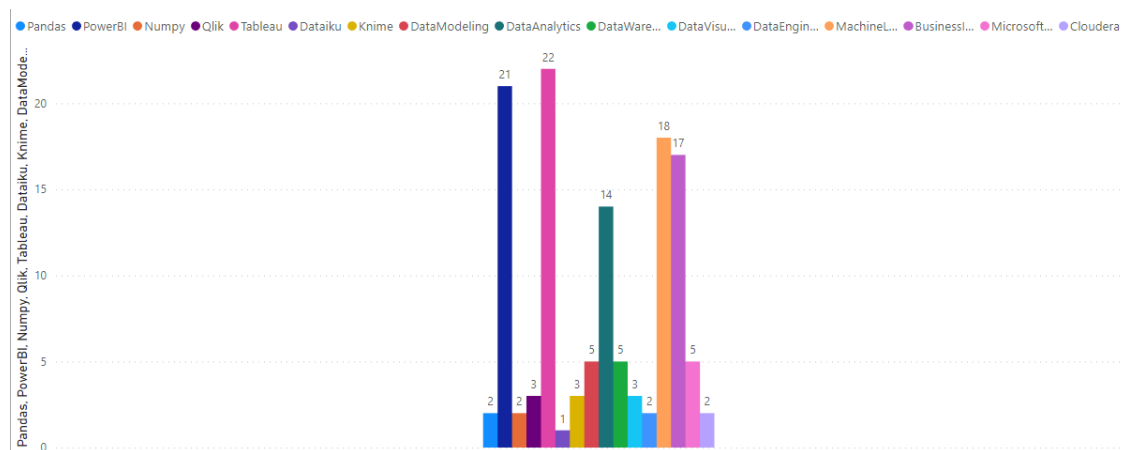


FIGURE 3.4 – nombre d'offre demandant les outils d'analyse et de visualisation des données

L'histogramme de groupe illustre le nombre de compétences requises dans les 150 offres d'emploi que nous avons collectées grâce au scraping. Toutes ces offres sont axées sur les outils d'analyse de données et de visualisation. Les compétences les plus demandées, dans l'ordre décroissant de leur occurrence, sont :

- Tableau : 22 offres, représentant 14,67% du total.
- Power BI : 21 offres, représentant 14% du total.
- Machine Learning : 18 offres, représentant 12% du total.
- Business Intelligence : 17 offres, représentant 11,33% du total.

- Data Analytics : 14 offres, représentant 9,33% du total.

Par ailleurs, nous avons observé que Data Modelling, Data Warehouse et Microsoft Excel sont mentionnés dans 5 offres sur les 150, représentant 3,33% du total. En revanche, les autres outils tels que Pandas, Numpy, Qlik, Dataiku, Cloudera, Knime et Data Visualisation sont moins demandés, avec seulement 2 à 3 occurrences parmi les 150 offres.

MYSQL, NOSQL, POSTGRESQL, PLSQL, SQL, SQLSERVER ET SGBD

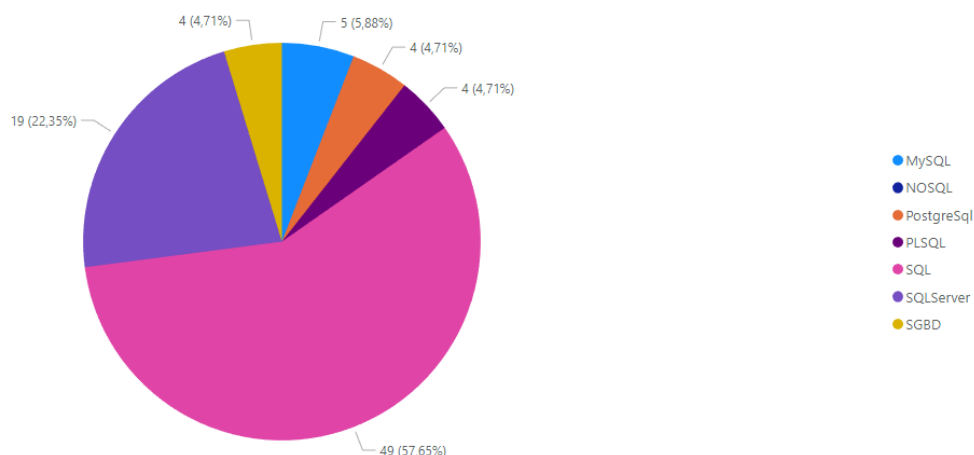


FIGURE 3.5 – nombre d'offre demandant langages de requête et de manipulation de données

Dans ce graphique en secteurs représentant les compétences liées aux langages de requête et de manipulation de données, nous pouvons conclure que SQL est la compétence la plus demandée, avec un pourcentage de 57,65%. Ensuite, nous trouvons SQL Server avec un pourcentage de 22,35%. La compétence MySQL représente 5,88% des demandes, tandis que PostgreSQL, SGBD et PL/SQL sont chacun demandés à hauteur de 4,71%. Il est à noter que la compétence NoSQL ne figure pas parmi les demandes.

Cependant, il est important de noter que la compétence NoSQL n'est demandée dans aucune des offres analysées.

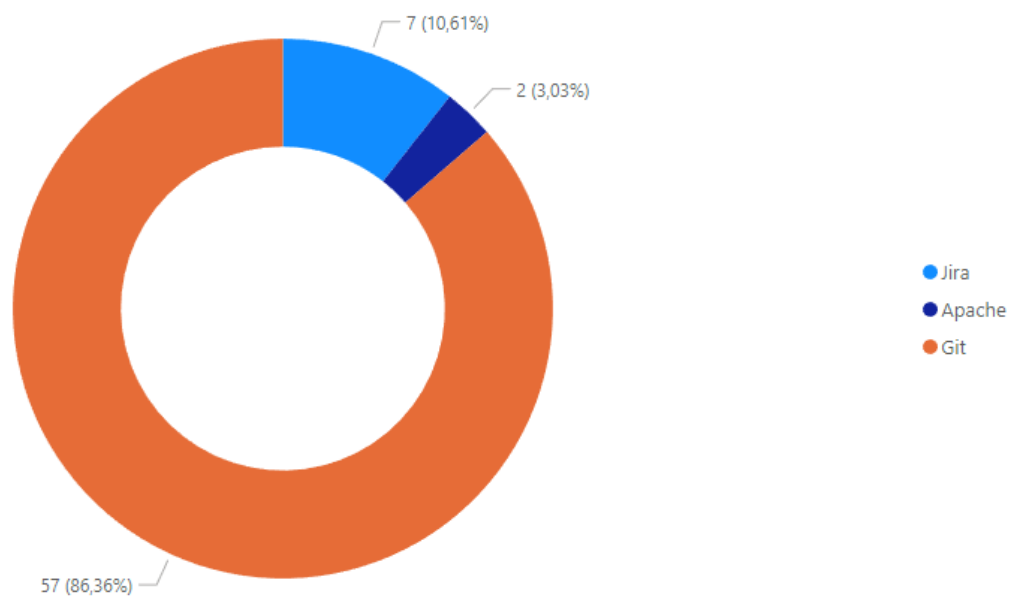


FIGURE 3.6 – nombre d'offre demandant Outils de développement et de gestion de versions

Ce graphique en anneau représente les outils de développement et de gestion de versions. À partir de celui-ci, nous pouvons conclure que GIT est la compétence la plus demandée, avec un pourcentage de 38%. En revanche, JIRA est demandé dans un pourcentage de 4% et Apache dans un pourcentage de 1%.

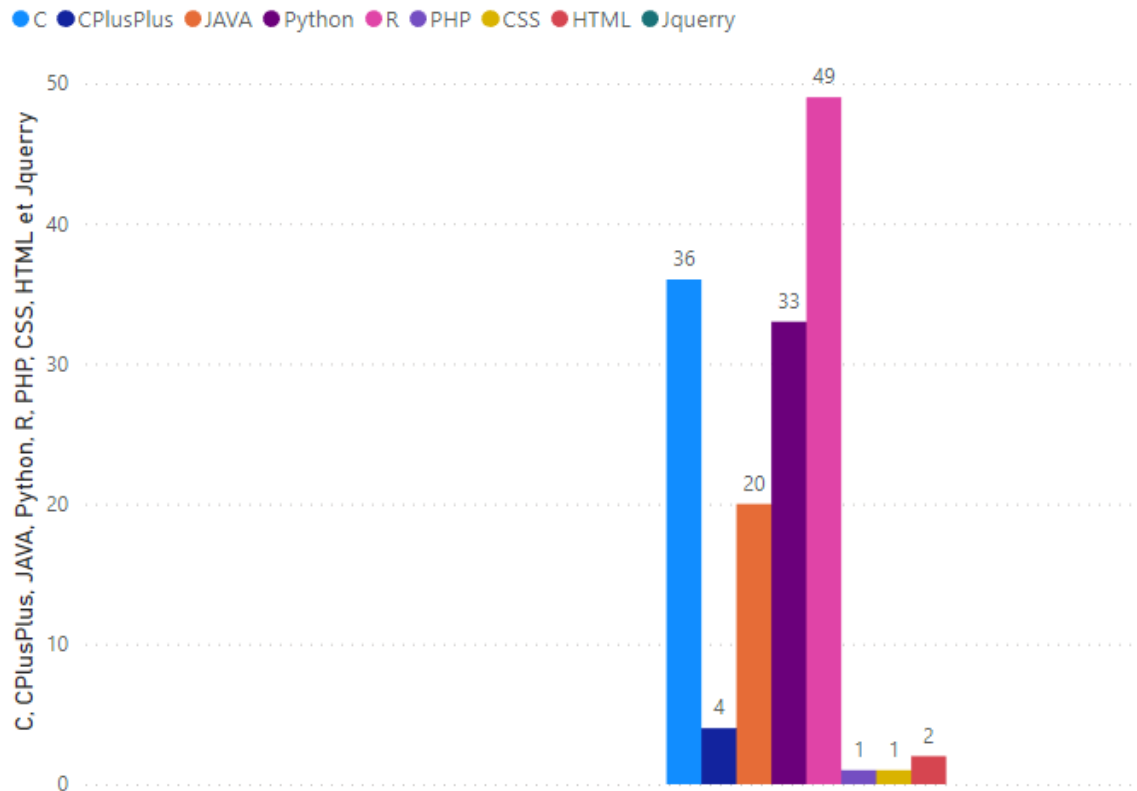


FIGURE 3.7 – nombre d'offre demandant langages de développement et de programmation :

D'après cet histogramme de groupe représentant les langages de développement et de programmation les plus demandés dans 150 offres, nous pouvons conclure ce qui suit :

- R est le langage le plus demandé, apparaissant dans 49 offres sur 150. Cela représente un pourcentage de 32,67%.
- Ensuite, nous avons C, qui est également très demandé, avec un pourcentage de 24% .
- Python est également populaire, apparaissant dans 33 offres sur 150, soit un pourcentage de 22%.
- Java est présent dans 20 offres sur 150, ce qui représente un pourcentage de 13,33%.
- En revanche, les langages C++, PHP, CSS et HTML ont un pourcentage inférieur à 2,67%.

Il est important de noter que JQuery n'est demandé dans aucune des offres analysées.

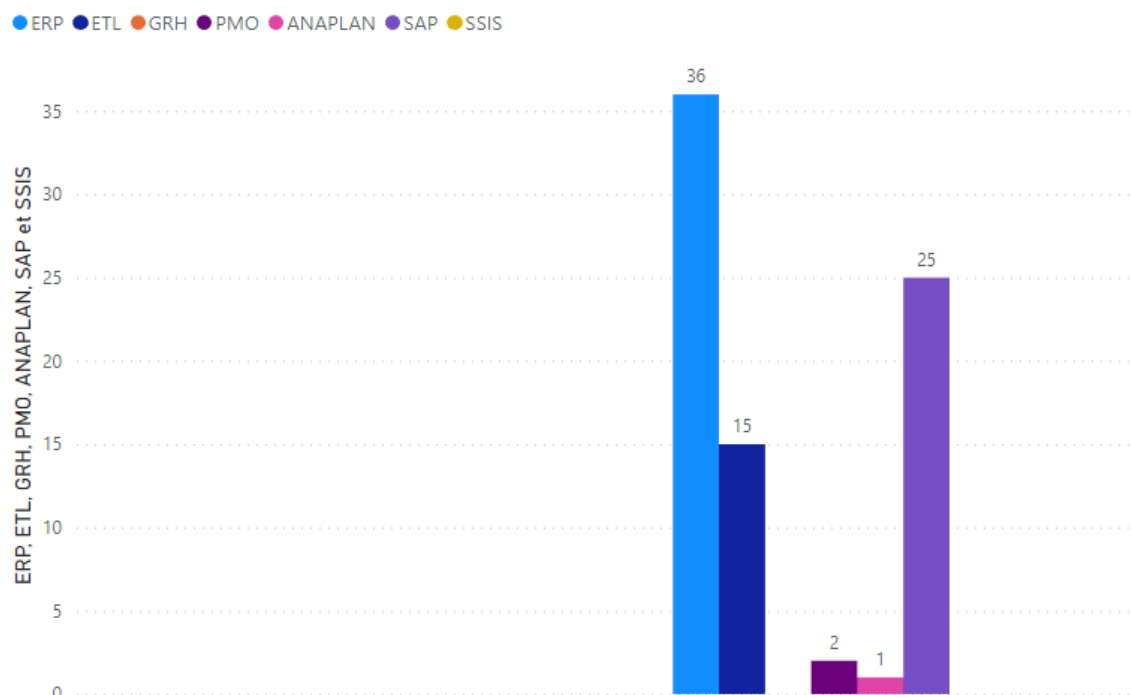


FIGURE 3.8 – nombre d'offre demandant outils de gestion d'entreprise et de projets

Dans cet histogramme représentant les outils de gestion d'entreprise et de projets, nous pouvons observer les tendances suivantes :

- ERP, qui fait référence à un système de planification des ressources de l'entreprise, est le plus demandé, apparaissant dans 36 offres sur 150. Cela représente un pourcentage de 24%
- Ensuite, nous trouvons SAP, qui est un système intégré de gestion d'entreprise, avec un pourcentage de 16,67%.
- L'ETL, représentant l'outil d'extraction, de transformation et de chargement de données, est mentionné dans 15 offres sur 150, soit un pourcentage de 10%.
- En revanche, ANAPLAN et PMO, qui signifient respectivement un outil de planification et de gestion de projets, sont demandés seulement 1 à 2 fois parmi les 150 offres.

Il est important de noter que les compétences SSIS (SQL Server Integration Services) et GRH (Gestion des Ressources Humaines) sont absentes dans toutes les offres analysées.

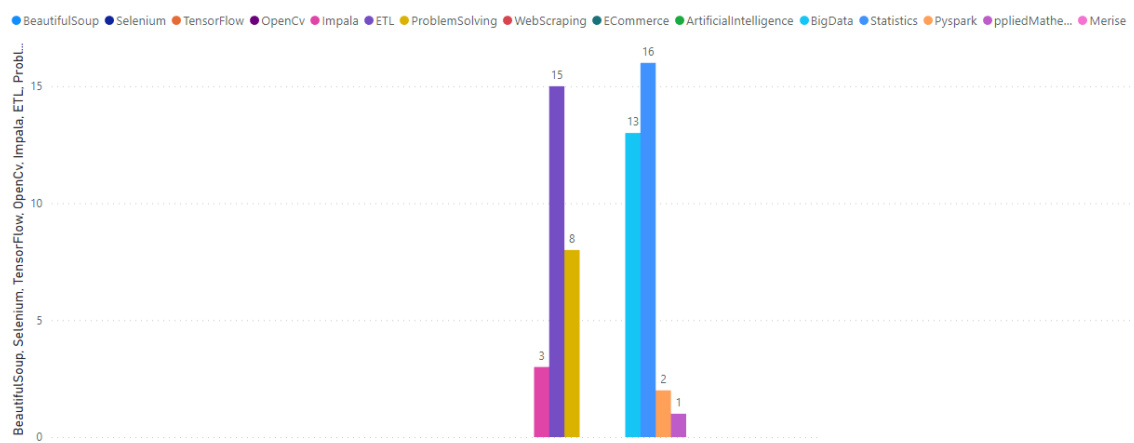


FIGURE 3.9 – nombre d'offre demandant ces autres Outils

Cet histogramme met en évidence les compétences et les outils spécifiques requis pour des tâches et des domaines particuliers. Une analyse des données révèle les conclusions suivantes :

- Les compétences en statistiques sont les plus recherchées dans cette catégorie, apparaissant dans 16 offres sur 150. Cela correspond à un pourcentage de 10,67%.
- L'ETL (Extract, Transform, Load) est également en demande, avec 15 offres sur 150, soit un pourcentage de 10%.
- Le Big Data est mentionné dans 13 offres sur 150, représentant un pourcentage de 8,67%.
- Le problem solving est une compétence recherchée dans 8 offres sur 150, soit un pourcentage de 5,33%.

En revanche, des compétences telles que Impala, PySpark et les mathématiques appliquées sont moins demandées, avec un pourcentage assez bas allant de 1 à 3 sur les 150 offres. De plus, des compétences telles que Beautiful Soup, Selenium, TensorFlow, OpenCV, le web scraping, l'intelligence artificielle, le e-commerce et Merise ne sont pas du tout demandées dans les offres analysées.

3.3.2 Graphes relatifs aux profils

Comment se répartissent géographiquement les lauréats ?



FIGURE 3.10 – Carte de répartition des lauréats

Le tableau ci-dessus présente la répartition des lauréats de la filière BI&A en fonction de leur localisation. Une observation initiale montre que ces lauréats sont répartis dans plusieurs pays autres que le Maroc. Plus précisément, 6 personnes ont choisi de s'installer au Canada, tandis qu'un nombre significatif de 22 personnes a opté pour la France. On note également la présence d'une personne travaillant en Angleterre, une autre en Finlande et une autre au Luxembourg.

En ce qui concerne le Maroc, la majorité des lauréats se trouvent à Casablanca, où résident 39 personnes. Par ailleurs, 23 lauréats ont choisi Rabat comme lieu de résidence, tandis que quelques-uns se sont installés à Fès, khouribga et Tanger, avec seulement une personne dans chaque ville.

La carte adjacente est une représentation réelle de la répartition géographique des lauréats de l'Ensias. Cette carte géographique illustre la distribution de ces derniers en fonction des pays. La taille et la couleur de chaque bulle reflètent l'importance du nombre de lauréats.

Combien de lauréats compte chaque entreprise ?

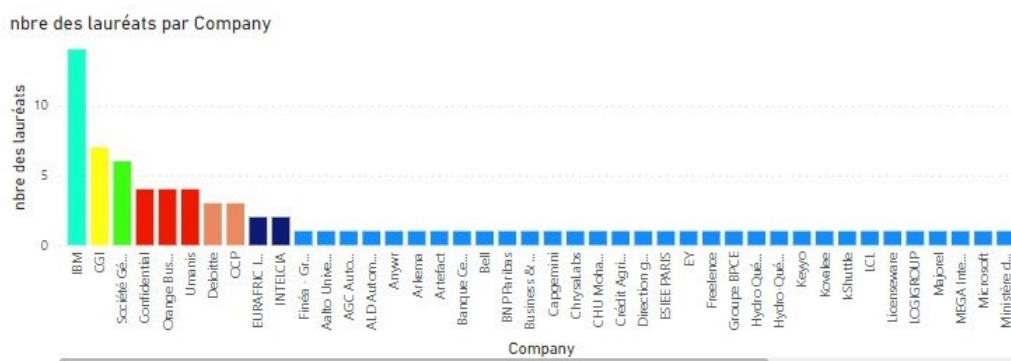


FIGURE 3.11 – Répartition des lauréats par entreprise

Le diagramme à barres ci-dessus représente le nombre de lauréats dans chaque entreprise. En examinant ce visuel, on remarque qu'IBM est l'entreprise qui recrute le plus d'étudiants en BI, avec un effectif de 14 personnes parmi les 96 lauréats. En deuxième position, nous retrouvons CGI avec un nombre de 7 personnes, suivi de près par Société Générale qui a recruté 6 personnes. Ensuite, les entreprises Confidential, Orange Business et Umanis se trouvent à égalité avec 4 lauréats chacune.

Pour analyser cette répartition en fonction de chaque pays, nous pouvons ajouter un filtre basé sur la localisation. Cela nous permettra d'évaluer la répartition des lauréats dans les différentes entreprises pour chaque pays.

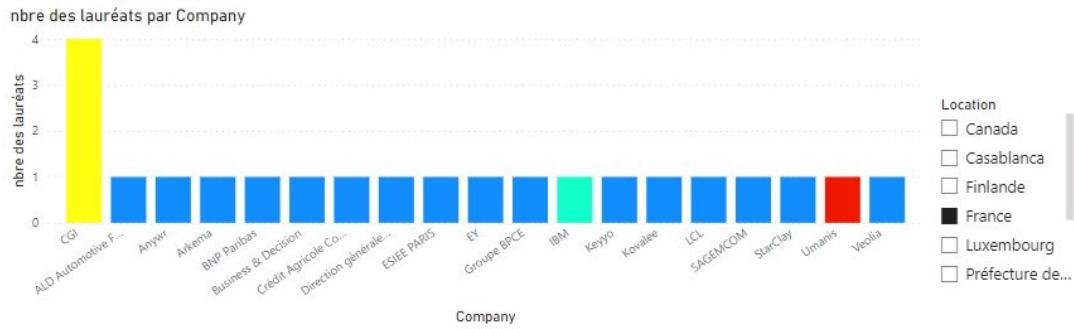


FIGURE 3.12 – Répartition des lauréats par entreprise dans la France

À titre informatif, ce diagramme illustre la répartition des diplômés présents en France en fonction de leur entreprise d'emploi. On peut observer que CGI est l'entreprise où les spécialistes en Business Intelligence travaillent le plus.

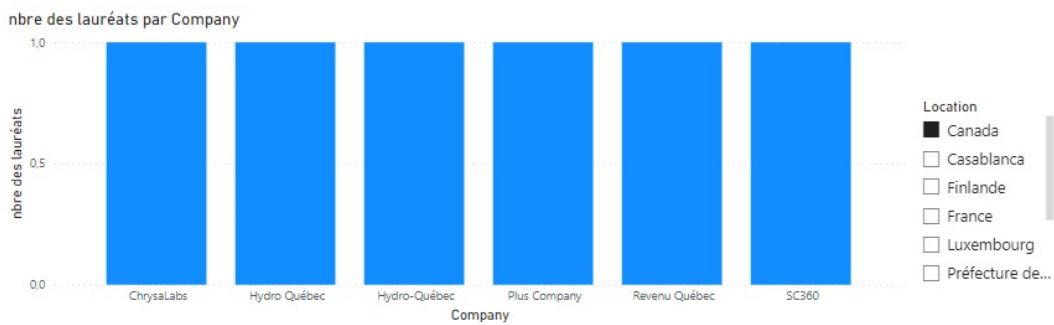


FIGURE 3.13 – Répartition des lauréats par entreprise dans Canada

Ce diagramme illustre la répartition des diplômés au Canada parmi diverses entreprises. Chaque entreprise compte un lauréat de l'Ensias parmi ses effectifs.

Combien de lauréats n'ont aucune certification et quelle est leur entreprise respective ?

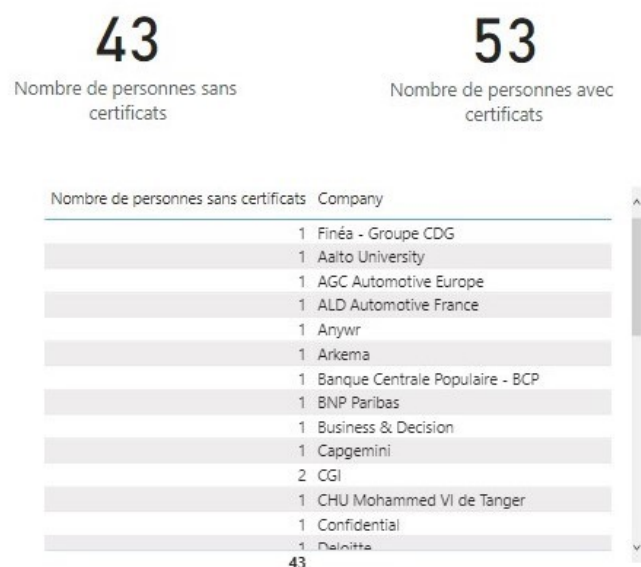


FIGURE 3.14 – Tableau des entreprises des lauréats sans certifications

Après la collecte de données établies pour les 96 lauréats de la filière, nous avons pu déterminer le nombre de personnes possédant des certifications. En effet, la première carte présente les résultats d'une mesure réalisée, révélant que parmi ces 96 lauréats, 43 personnes n'ont pas procédé à l'acquisition de certifications, ce nombre représente un pourcentage de 44,79%.

La deuxième carte met en évidence que 53 personnes possèdent des certifications, une valeur qui constitue 55,21%. D'après les mesures représentées dans les cartes, on peut conclure que le nombre de lauréats qui ont choisi de faire des certifications et ceux qui ne l'ont pas fait diffèrent légèrement.

Le tableau ci-dessous représente la répartition des lauréats sans certifications. On remarque que ces derniers occupent des postes au sein d'entreprises reconnues telles que IBM, CGI, Orange Business Services, et d'autres.

Quel est le nombre total de certifications par personne ?

Nombre de Cases Remplies 2	Name
6	Abdelhak ELBIARI
6	Abdelhalim Asfiri Eng.
8	Abdelkarim AZOUZI
2	Abdellah Rami
3	Abdenmour EL MALKI
0	Abderahmane Outaleb
0	Abdessamad Azrour
4	Achraf BOUGADIR
1	Afaf SIDALI
0	Ahlam NARHAM
0	Ahmed El Kourri
0	Ahmed Elhami
0	Ahmed Tijane
0	Amal EL BRIKI
18	AMINE HAMAQUI
5	Amine Hamdouchi
238	

FIGURE 3.15 – Certifications par personnes

Le tableau ci-dessus représente le nombre de certifications pour chaque personne dans notre base de données. Après le traitement effectué sur les certificats, chaque colonne contient une certification. Ainsi le chiffre maximum de certifications obtenues par une personne est de 19. Pour calculer le nombre de certificats pour chaque enregistrement, nous avons calculé la somme des colonnes remplies pour chaque lauréat.

Quels sont les domaines les plus couramment choisis par les lauréats pour leurs certifications ?

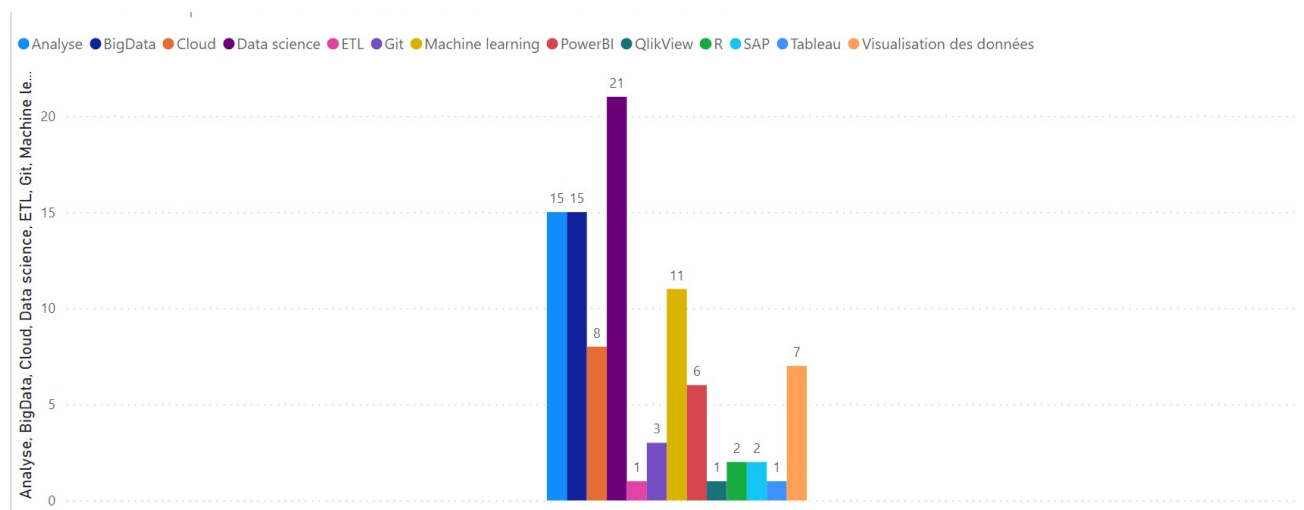


FIGURE 3.16 – Domaine de certifications

Afin d'analyser les certifications en termes de contenu, nous avons créé les cartes suivantes qui représentent les domaines les plus envisagés par les lauréats de la BI. Dans ce contexte, nous pouvons constater que de nombreuses personnes sont certifiées dans les domaines suivants : Data science, Machine learning, Analyse de données, Big data et Cloud. Ces domaines de certifications témoignent de l'intérêt et de l'engagement des lauréats de la BI dans des disciplines clés liées à l'analyse et à la gestion des données.

Il est important de souligner que le nombre de personnes envisageant de se certifier en power BI est assez faible. Cela indique que la formation académique dans le domaine de la visualisation des données est suffisante pour développer des compétences solides dans ce domaine. De plus, on constate que l'outil de visualisation Qlickview n'est pas envisagé par les lauréats dans leurs certifications .

Combien de lauréats ont choisi de faire une doubles diplômes ?



FIGURE 3.17 – Double diplomation

La formation proposée par l'ENSIAS offre à ses étudiants la possibilité de réaliser une double diplomation dans différents pays. Nous avons étudié le cas de nos 96 lauréats en Business Intelligence (BI) pour déterminer s'ils ont choisi de faire une double diplomation ou non.

Les résultats sont présentés dans les tableaux suivants. Il est important de souligner qu'un nombre intéressant de 32 lauréats ont opté pour une double diplomation à l'étranger, ce nombre représente un pourcentage de 33,33% , tandis que 64 lauréats ont poursuivi leurs études au sein de l'ENSIAS constituant donc une proportion de 66,67% .

Parmi ces 32 lauréats, une écrasante majorité de 19 personnes a choisi de poursuivre leurs études en France, plaçant ainsi ce pays en tête de liste. Le Canada se classe en deuxième position avec 5 personnes, tandis que la Finlande et l'Angleterre comptent chacune une personne.

Il est également à noter que certains lauréats sont revenus au Maroc après leur double

diplomation, avec un total de 6 personnes qui reflète un pourcent de 18,75%.

Conclusion :

Pour s'intégrer aisément dans le marché du travail marocain, les diplômés de la filière Business Intelligence et Analytics de l'ENSIAS sont encouragés à suivre les tendances et les technologies utilisées par les entreprises. Cette étude met en évidence les compétences requises en les comparant à celles étudiées à l'ENSIAS. Dans ce contexte, de nombreux lauréats choisissent de passer des certifications afin d'approfondir leurs connaissances dans un domaine spécifique et de le maîtriser davantage. De plus, de nombreuses offres d'emploi exigent une maîtrise solide des langues et des compétences en communication, ce qui renforce le rôle essentiel d'un data analyst qui doit être en mesure de transmettre les résultats au responsable concerné.

Discussion

4.1 Résultats importants

L'analyse des résultats obtenus à partir du scraping des offres d'emploi de LinkedIn pour les postes en Business Intelligence et des profils des diplômés de l'ENSIAS révèle plusieurs tendances intéressantes.

- Il apparaît que les sociétés Agoda et CGI sont parmi celles qui proposent le plus grand nombre d'offres d'emploi dans ce domaine, ce qui peut indiquer une demande accrue de professionnels de la Business Intelligence dans ces entreprises.
- En ce qui concerne la localisation géographique, Casablanca et Rabat se distinguent en tant que zones où le nombre d'offres d'emploi en Business Intelligence est élevé. Cela suggère que ces villes sont des pôles d'activité importants dans le domaine de la Business Intelligence au Maroc.

Les compétences demandées et enseignées	Les compétences demandées mais non enseignées	Les compétences enseignées mais moins demandées
<div> <div> <div>Power BI</div> <div>C</div> <div>Java</div> <div>Machine Learning</div> <div>SQL</div> <div>Business Intelligence</div> <div>Data Analytics</div> <div>ERP</div> <div>Statistique</div> <div>Big Data</div> </div> <div> </div> </div>	<div> <div>Tableau</div> <div>R</div> <div>Git</div> <div>ETL</div> <div>SAP</div> </div>	<div> <div>Data Warehouse</div> <div>Data Modelling</div> <div>SGBD</div> <div>PL/SQL</div> <div>HTML</div> <div>CSS</div> <div>JQuery</div> <div>GRH (Gestion des Ressources Humaines)</div> <div>PMO (Project Management Office)</div> </div>

- En ce qui concerne les employeurs qui recrutent le plus, il est intéressant de noter qu'IBM et CGI sont en tête de liste, suggérant qu'ils offrent de nombreuses opportunités d'emploi dans le domaine de la Business Intelligence.
- En ce qui concerne les diplômés de l'ENSIAS, il est important de mentionner que

96 d'entre eux ont été recensés. Parmi ces diplômés, 43 ont des certifications, tandis que 53 n'en ont pas. Le diplômé avec le plus grand nombre de certifications enregistrées compte 19 certificats différents.

- En termes de domaines de spécialisation, 21 diplômés possèdent une certification en Data Science, 15 en Analyse, 15 en Big Data et 11 en Machine Learning.
- Il est également intéressant de noter que parmi les 96 diplômés, 32 ont suivi une double diplomation, dont 6 ont poursuivi leurs carrières au Maroc et 26 à l'étranger, indiquant une mobilité internationale relativement élevée parmi les diplômés de l'ENSIAS.

4.2 Recommandations

Pour améliorer la formation des profils BI&A à l'ENSIAS, nous proposons les recommandations suivantes :

- Intégrer les compétences demandées mais non enseignées : L'ENSIAS devrait envisager d'introduire des modules spécifiques couvrant des compétences telles que Tableau, R, Git, ETL et SAP dans son programme de formation. Cela permettra aux étudiants d'acquérir des compétences recherchées par les employeurs et de répondre à leurs besoins actuels.
- Renforcer l'accent sur les compétences enseignées mais moins demandées : Bien que ces compétences puissent être moins demandées sur le marché du travail, elles restent importantes dans le domaine de la Business Intelligence et de l'Analytics. L'ENSIAS devrait envisager de fournir des opportunités supplémentaires aux étudiants pour se familiariser avec des sujets tels que Data Warehouse, Data Modelling, SGBD, PL/SQL, HTML, CSS, JQuery, GRH et PMO. Cela peut être réalisé en intégrant des projets pratiques, des études de cas ou des modules optionnels axés sur ces compétences.
- Encourager les certifications complémentaires : Pour les compétences demandées mais non enseignées, l'ENSIAS pourrait encourager les étudiants à obtenir des certifications spécifiques, telles que la certification Tableau, la certification R ou la certification Git. Ces certifications complémentaires renforceront leur profil professionnel et les rendront plus compétitifs sur le marché du travail.
- Renforcer la collaboration avec l'industrie : L'ENSIAS devrait développer des partenariats solides avec des entreprises et des professionnels du secteur de la Business Intelligence et de l'Analytics. Cela permettra d'identifier les compétences les plus recherchées et de les intégrer dans le programme de formation. Les stages en entreprise, les projets conjoints et les interventions de professionnels du secteur peuvent également offrir aux étudiants une expérience pratique et une meilleure compréhension des exigences de l'industrie.

- Mettre à jour le programme de formation en fonction des évolutions technologiques : Le domaine de la Business Intelligence et de l'Analytics évolue rapidement. Il est donc essentiel que l'ENSIAS suive ces évolutions et mette régulièrement à jour son programme de formation pour refléter les nouvelles tendances, les outils émergents et les meilleures pratiques de l'industrie.

En suivant ces recommandations, l'ENSIAS sera en mesure de mieux préparer ses étudiants aux exigences du marché du travail et de renforcer leur employabilité dans le domaine de la Business Intelligence et de l'Analytics. Cela permettra aux diplômés de l'ENSIAS de se démarquer et de réussir dans leur future carrière professionnelle.

4.3 Limites

Le projet présente certaines limites qui doivent être prises en compte :

- Échantillonnage limité des offres d'emploi : Le scraping des offres d'emploi à partir d'une seule plateforme, telle que LinkedIn, peut limiter la représentativité des résultats. Il est possible que certaines opportunités d'emploi pertinentes n'aient pas été incluses dans l'échantillon, ce qui pourrait influencer les conclusions.
- Variabilité des compétences demandées : Les compétences demandées dans les offres d'emploi peuvent varier en fonction de divers facteurs tels que la taille de l'entreprise, l'industrie et la localisation géographique. Par conséquent, les résultats obtenus peuvent ne pas être généralisables à l'ensemble du marché du travail.
- Manque de données sur les lauréats : Le scraping des profils des lauréats peut également présenter des limites, notamment en termes de disponibilité des données. Il est possible que certains profils ne soient pas accessibles ou que certaines informations clés, telles que les certificats obtenus, ne soient pas mentionnées sur les profils.
- Absence d'évaluation des compétences pratiques : Le projet se concentre principalement sur l'analyse des compétences mentionnées dans les offres d'emploi et les profils des lauréats. Cependant, il ne tient pas compte des compétences pratiques réelles acquises par les étudiants tout au long de leur formation, telles que les projets réalisés ou les stages effectués. Cela peut limiter la compréhension globale des compétences des diplômés.
- Évolution rapide du domaine : La Business Intelligence et l'Analytics sont des domaines en constante évolution, avec l'émergence de nouvelles technologies, méthodologies et compétences. Les résultats obtenus à un moment donné peuvent ne pas refléter les besoins et les demandes du marché du travail à long terme.

Il est important de reconnaître ces limites pour interpréter les résultats du projet avec prudence et prendre en compte d'autres facteurs pertinents lors de la prise de décisions concernant l'amélioration de la formation et l'adéquation des compétences.

Conclusion générale

En conclusion, ce projet de fin d'année a permis de mettre en évidence des résultats importants quant à l'adéquation entre la formation dispensée par l'ENSIAS dans le domaine de la Business Intelligence et les compétences demandées par le marché du travail. Les résultats ont révélé à la fois des correspondances encourageantes et des écarts à combler.

Il est encourageant de constater que la majorité des compétences enseignées sont en adéquation avec les compétences recherchées par les employeurs. Cela témoigne de la qualité de la formation offerte par l'ENSIAS et de la pertinence de son programme académique. Cependant, les compétences spécifiques telles que Tableau, R, Git, ETL et SAP, qui sont demandées par les employeurs mais non enseignées, doivent être prises en considération pour mieux préparer les étudiants à répondre aux exigences du marché.

D'autre part, les compétences enseignées mais moins demandées sur le marché du travail, telles que Data Warehouse, Data Modelling, SGBD, PL/SQL, HTML, CSS, JQuery, GRH et PMO, doivent être réévaluées pour mieux répondre aux besoins des employeurs. Une révision du programme de formation, en accordant une plus grande importance aux compétences les plus recherchées, pourrait améliorer l'employabilité des étudiants et renforcer leur adaptation aux exigences de l'industrie.

Il convient également de noter les limites de ce projet, telles que l'échantillonnage limité des offres d'emploi, la variabilité des compétences demandées sur le marché, le manque de données sur les profils des lauréats et l'absence d'évaluation des compétences pratiques. Ces limites soulignent l'importance de considérer ces résultats avec prudence et de les compléter par d'autres informations et études pour une compréhension plus complète.

Dans l'optique d'améliorer la formation en Business Intelligence à l'ENSIAS, nous recommandons de prendre en compte les compétences demandées mais non enseignées, de renforcer les compétences enseignées mais moins demandées, d'encourager les certifications complémentaires, de favoriser la collaboration avec l'industrie et de maintenir le programme de formation à jour. En adoptant ces recommandations, l'ENSIAS pourra mieux préparer les étudiants à relever les défis du marché du travail dans le domaine de la Business Intelligence et de l'Analytics.

Bibliographie

[1] Documentation Pandas.

Librairie. <https://pandas.pydata.org/docs/>

[2] Documentation BeautifulSoup.

Librairie. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

[3] Documentation Selenium.

Librairie. <https://www.selenium.dev/documentation/en/>

[4] Documentation des expressions régulières.

Librairie. <https://docs.python.org/fr/3/library/re.html>

[5] Documentation Power BI.

Outil de Business Intelligence. <https://powerbi.microsoft.com/fr-fr/what-is-power-bi/>

[6] Documentation Canva.

Outil de construction de figure. <https://www.canva.com/design/DAFk6do2Xkk/Jof8U79IQ86Lkqck1zf318-47d7-8818-e7dcec2c47fa>