**Comprendre les Tendances du Marché de l'Emploi en IA, Data Science et Big Data : Optimisation du Recrutement et de la Gestion des Talents**



**Réalise par : Taha Srhayar**

**Encadré par : Moad Tajmouati**

**INTRODUCTION**

Dans le cadre de ce projet, en tant que développeur Data, nous avons entrepris une analyse approfondie du marché du travail en Intelligence Artificielle (IA), Data Science et Big Data. Nous avons commencé par traiter et nettoyer les données pertinentes contenues dans un fichier CSV, afin de garantir leur qualité et leur fiabilité pour nos analyses ultérieures.

En utilisant des outils d'analyse de données avancés, nous avons identifié les tendances clés du marché de l'emploi dans ces domaines émergents. Nous avons également mis en évidence les compétences les plus demandées par les employeurs et les titres de poste les plus courants pour les professionnels de l'IA, de la Data Science et du Big Data.

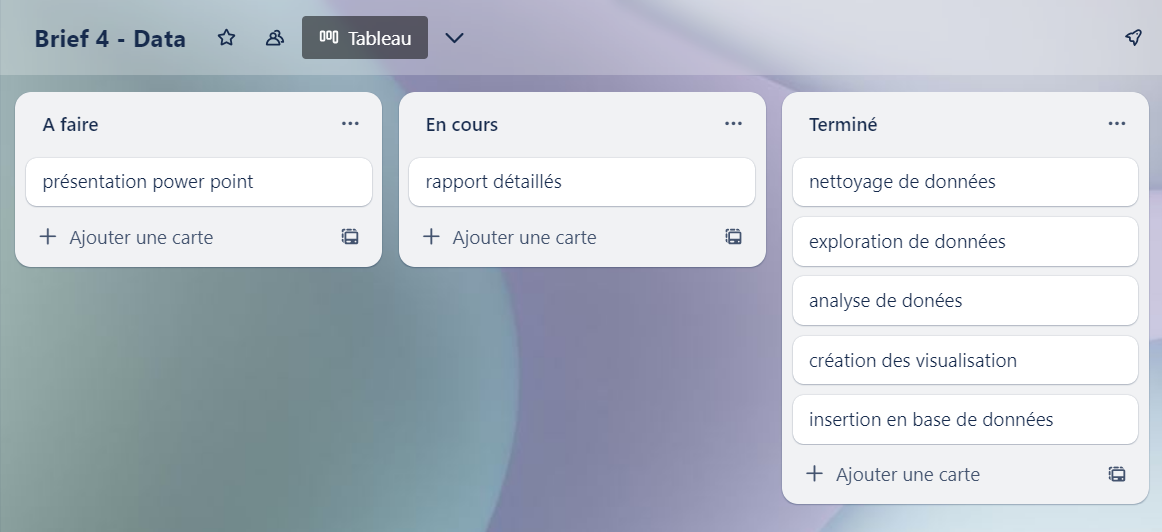
Afin de faciliter la compréhension et la communication des résultats, nous avons créé des visualisations claires et informatives. Ces visualisations ont permis aux parties prenantes de mieux saisir les conclusions tirées de notre analyse.

Pour assurer une gestion optimale des données tout au long du projet, nous avons mis en place un système de stockage dans une base de données MySQL. Cela nous a permis d'accéder facilement aux informations collectées et d'assurer l'intégrité des données.

Pour planifier et organiser notre projet de manière efficace, nous avons utilisé l'outil Trello. Cela nous a permis de suivre les étapes clés du processus, de gérer les tâches et de maintenir une cohérence dans nos actions.

En résumé, notre analyse approfondie du marché du travail en IA, Data Science et Big Data nous a fourni des informations essentielles pour optimiser nos stratégies de recrutement, de gestion des talents et de développement de compétences. Grâce au traitement des données, à leur visualisation et à leur stockage, ainsi qu'à une planification rigoureuse du projet, nous sommes en mesure de prendre des décisions éclairées pour rester compétitifs dans ce monde numérique en constante évolution.

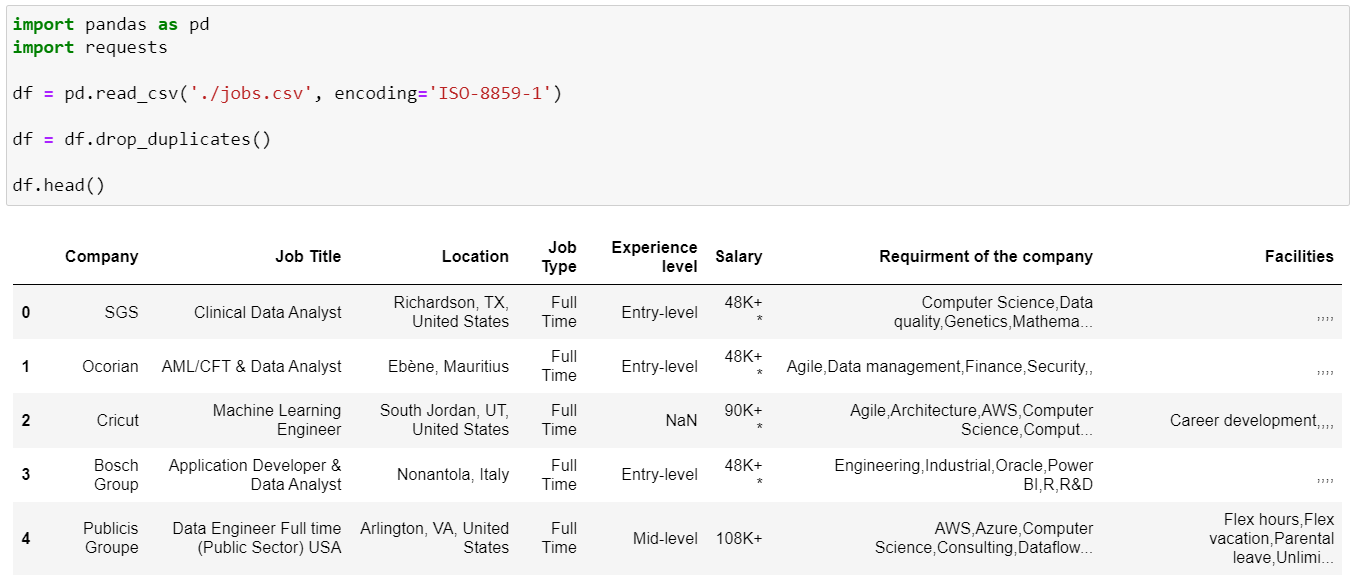
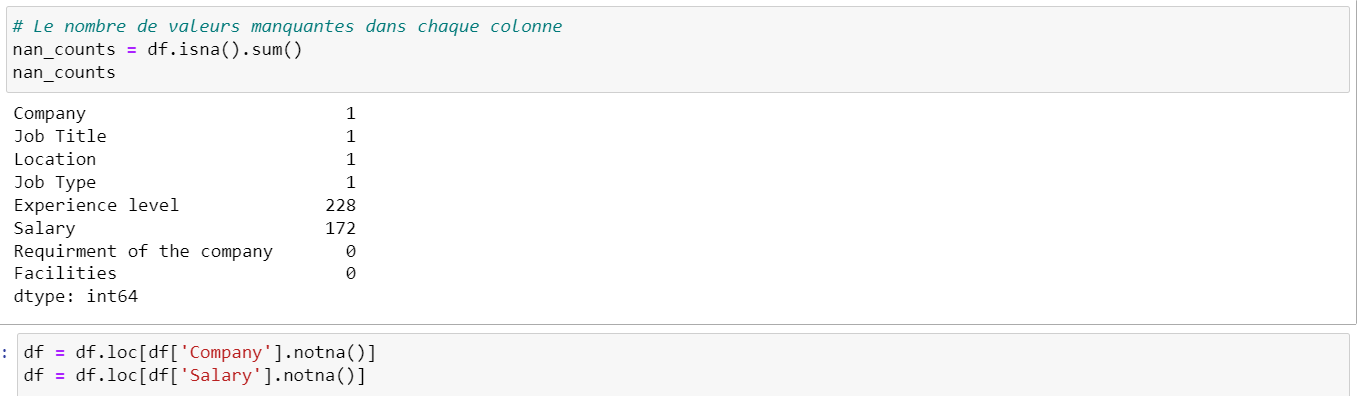
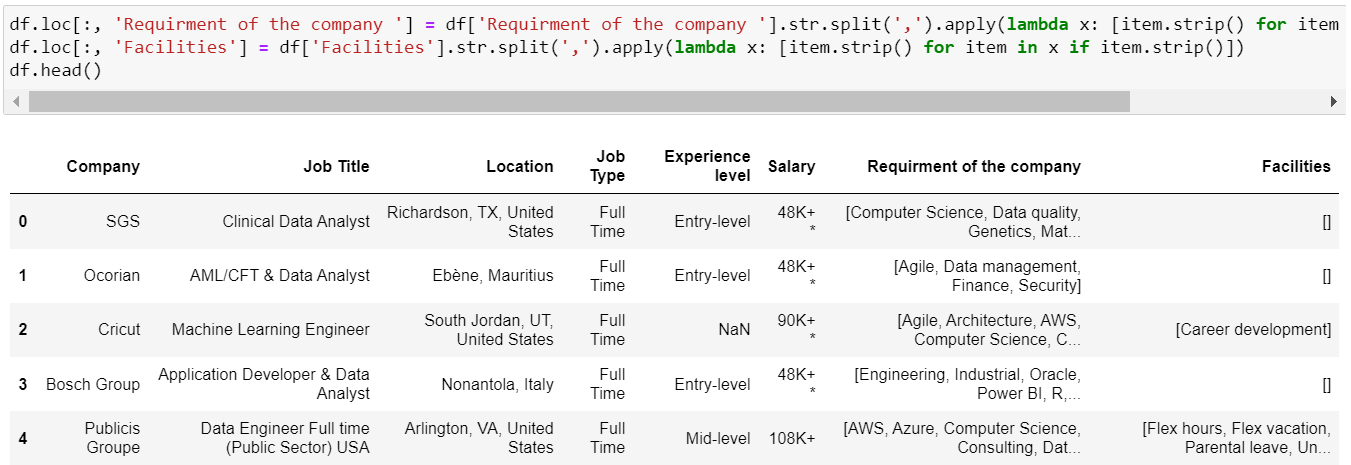
**Gestion Efficace des Tâches avec Trello**



En utilisant la plateforme de gestion de projet Trello, j'ai divisé efficacement mes tâches en trois divisions : "**À faire**", "**En cours**" et "**Terminé**". Trello m'a offert une interface intuitive pour créer des listes dédiées à chaque étape du processus. Dans la liste "À faire", j'ai regroupé toutes les tâches en attente, me permettant ainsi de prioriser mes actions. Une fois que je commence à travailler sur une tâche, je la déplace vers la liste "En cours", signalant ainsi clairement ce sur quoi je suis concentré. Et lorsque j'ai terminé une tâche, Trello me permet de la déplacer facilement vers la liste "Terminé", me donnant une vision instantanée de mes réalisations. Cette approche, facilitée par Trello, a considérablement amélioré l'efficacité de ma gestion des tâches, me permettant de rester organisé, de suivre les progrès du projet et de célébrer mes succès au fur et à mesure de leur accomplissement.

**Préparation des Données pour l'Analyse : Nettoyage et Transformation**

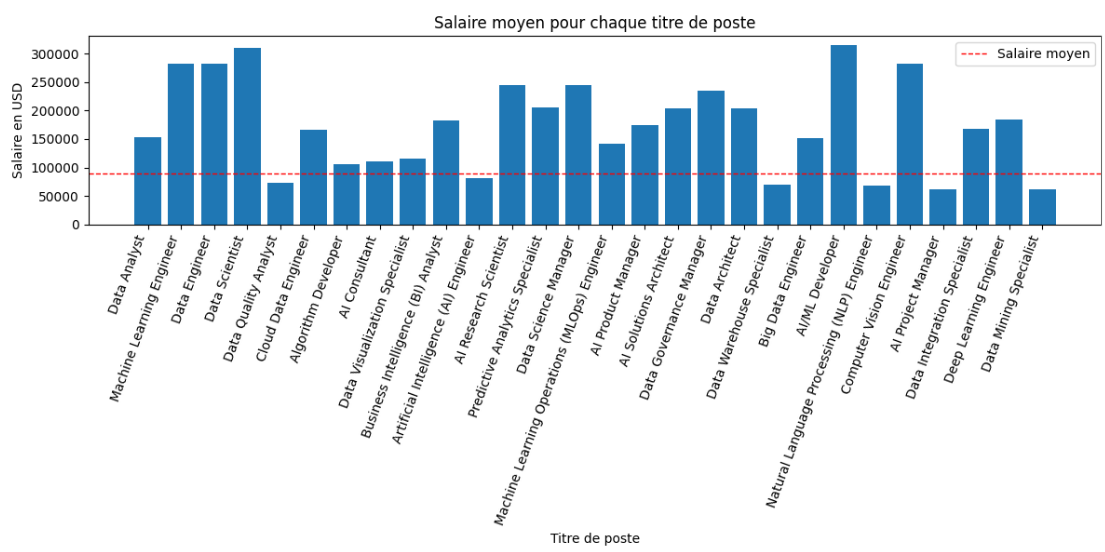
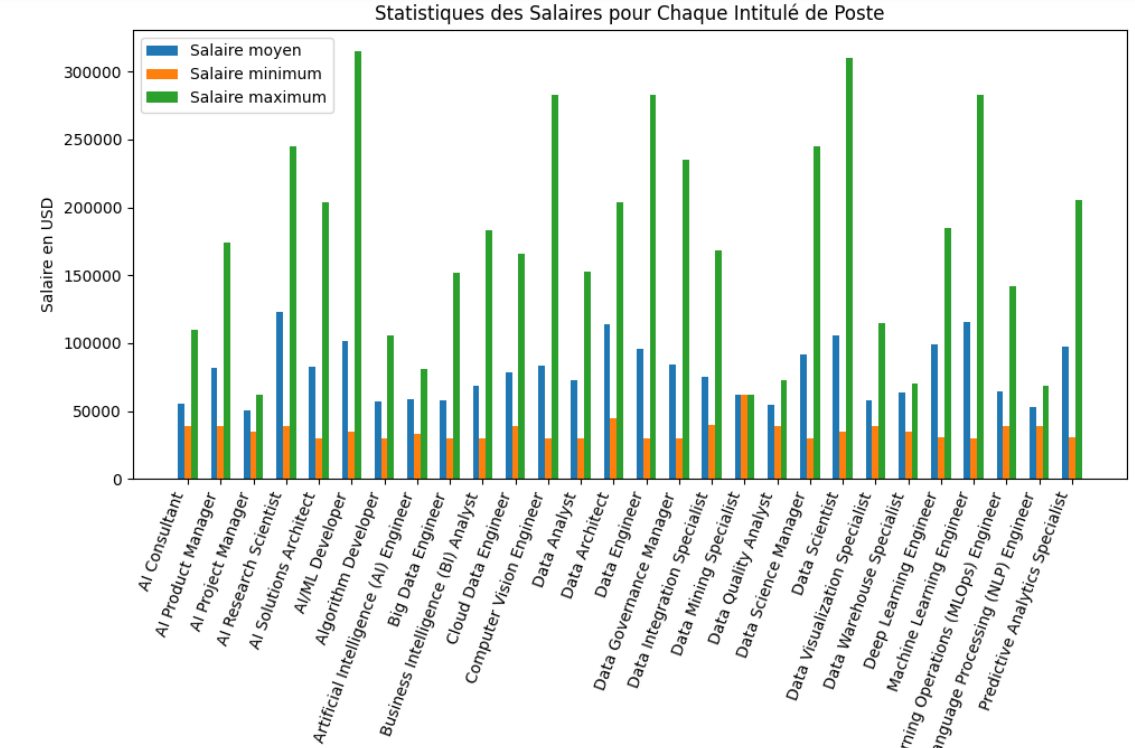
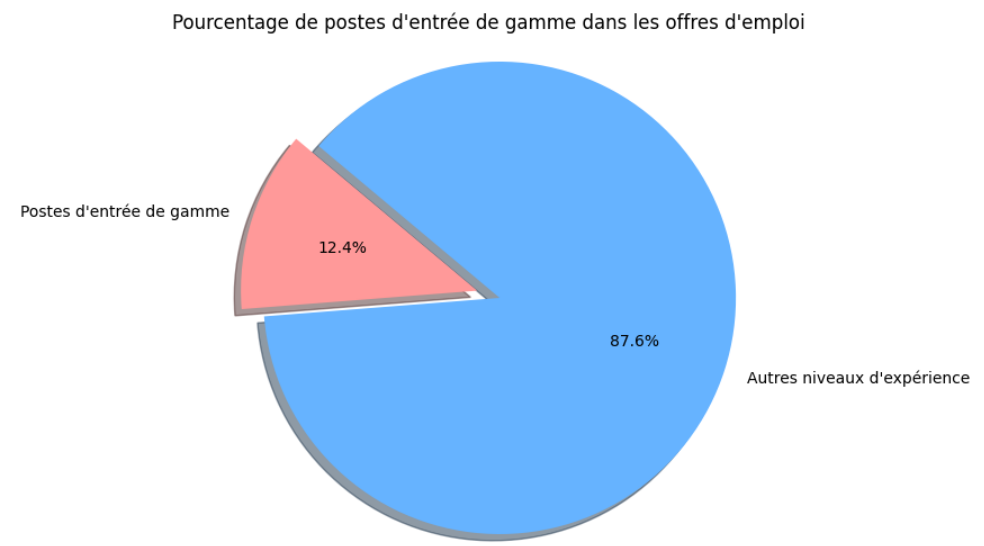
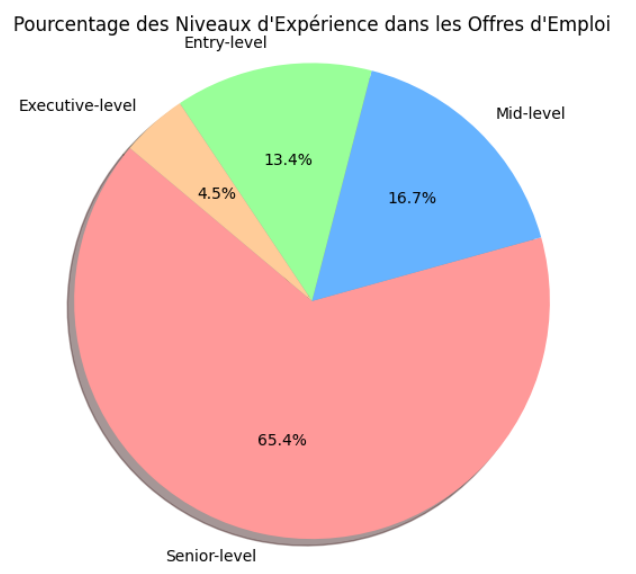
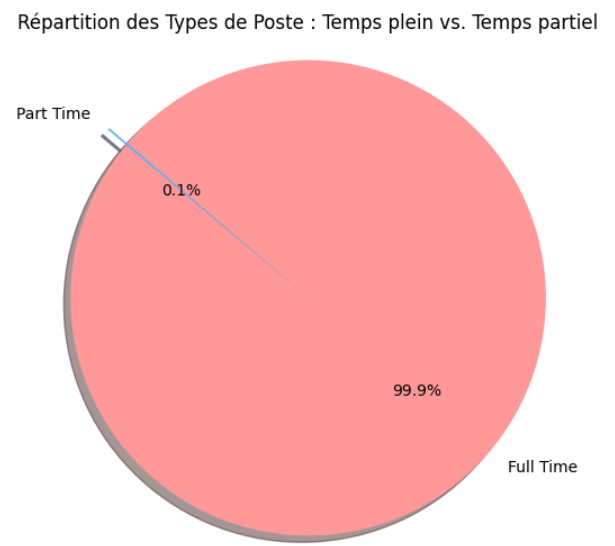
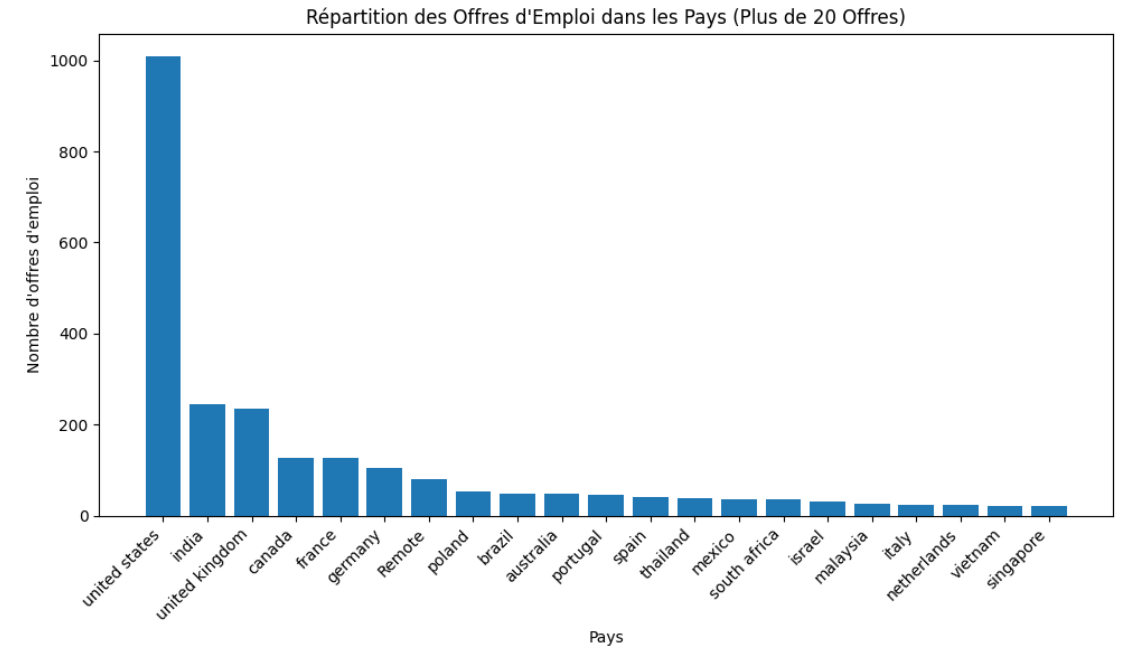
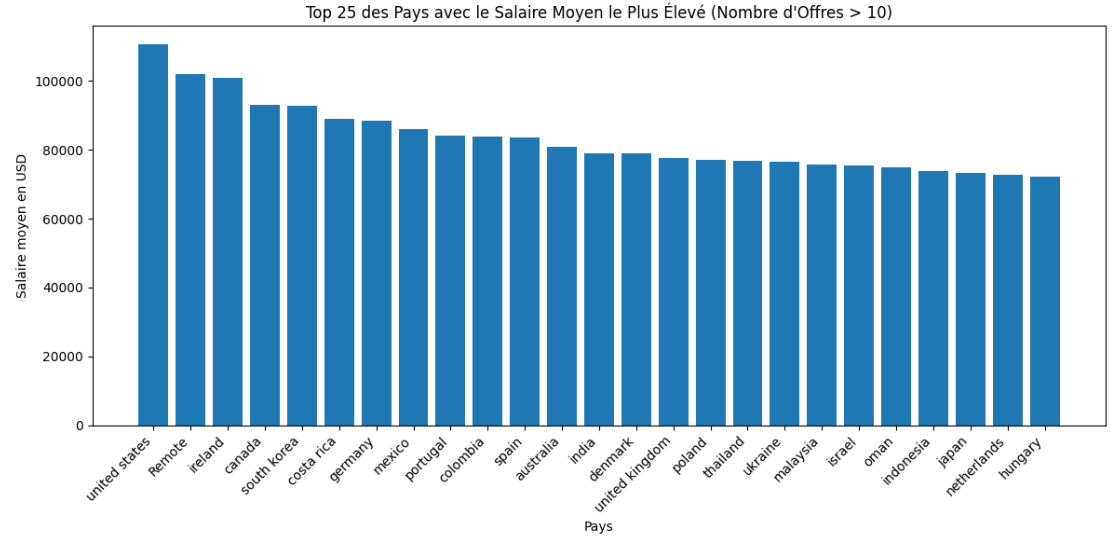
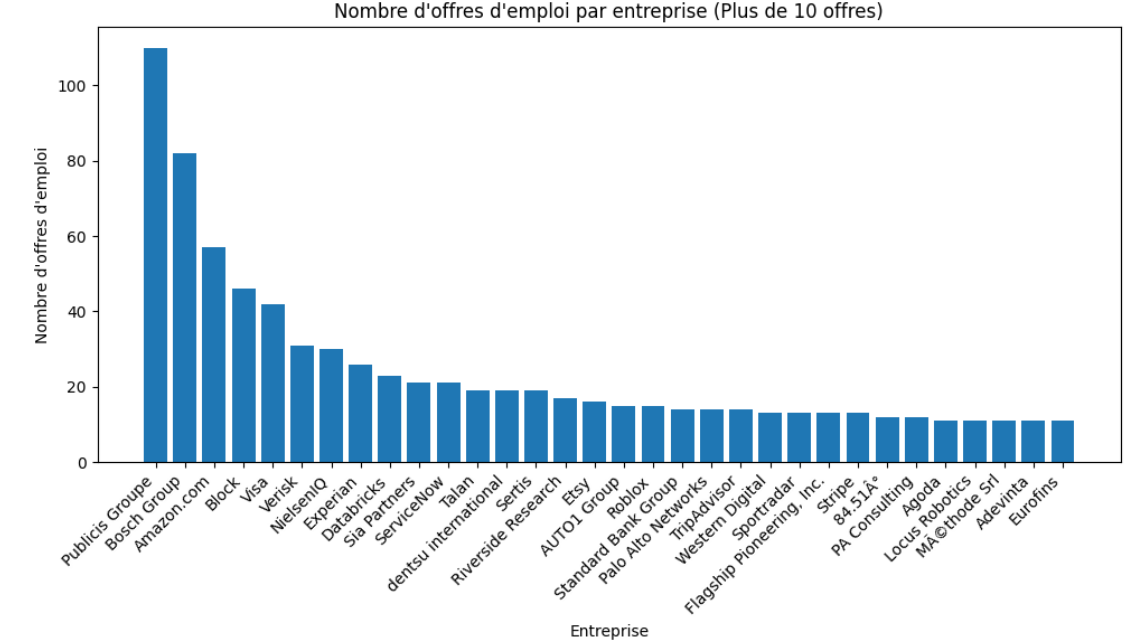
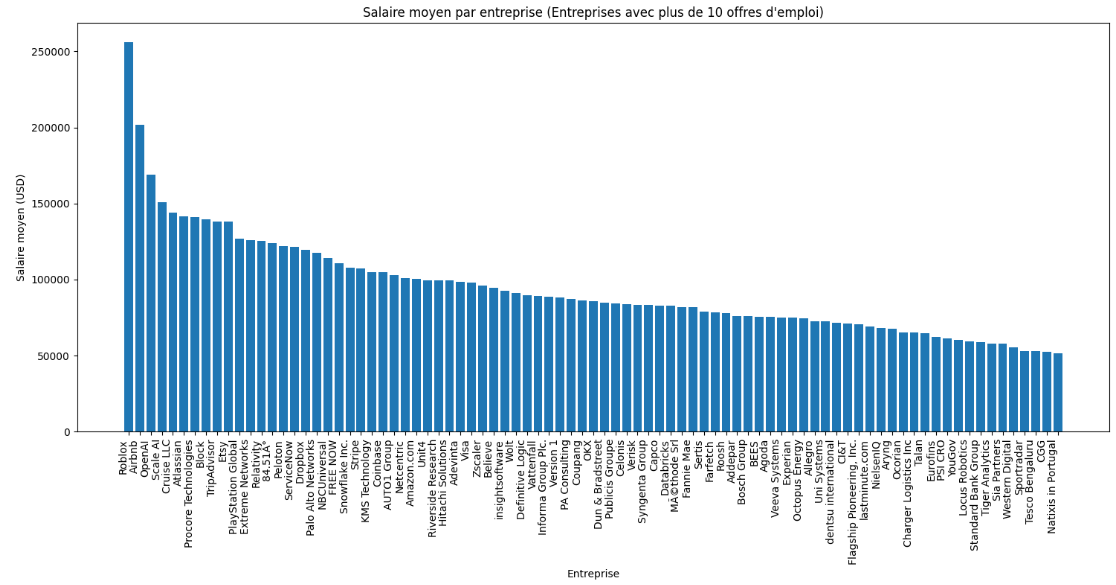
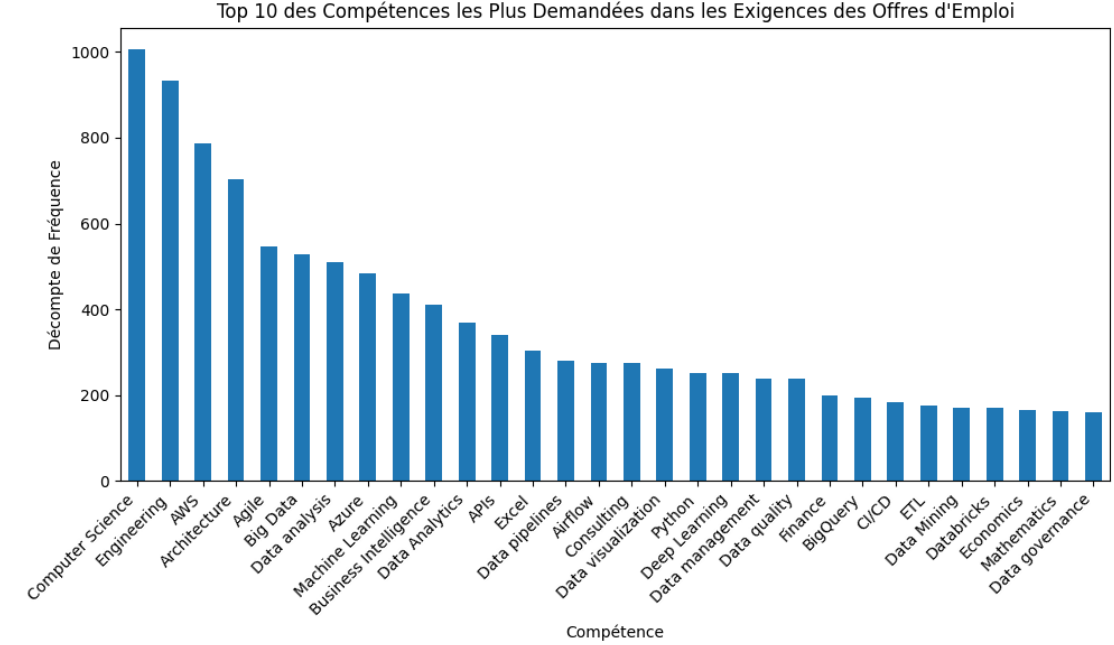
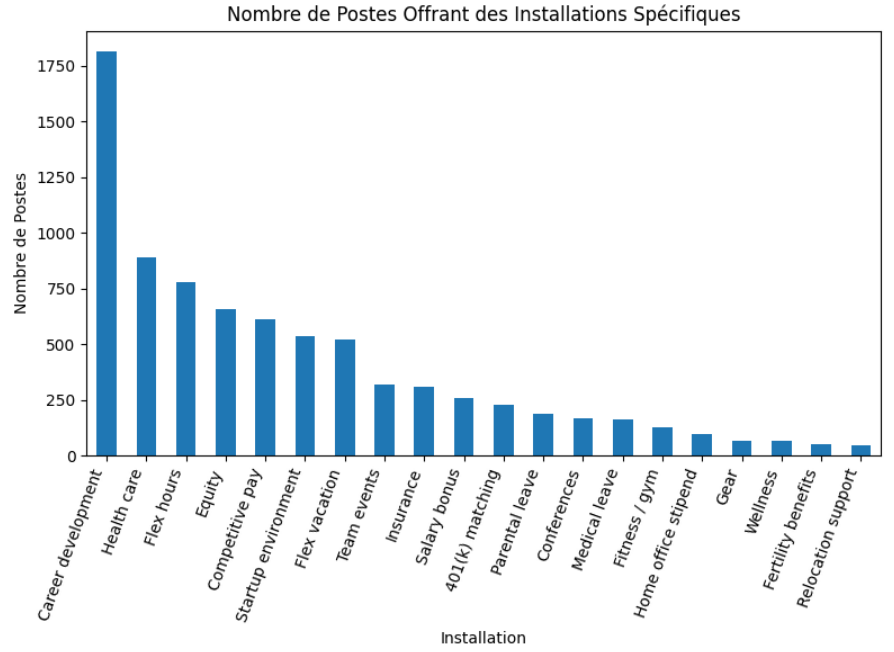
Pour nettoyer les données, nous avons effectué plusieurs tâches essentielles dans notre ensemble de données (df).

* Tout d'abord, nous avons éliminé les doublons en utilisant la fonction df.drop\_duplicates(), garantissant ainsi l'unicité des enregistrements. 
* Ensuite, nous avons examiné le nombre de valeurs manquantes dans chaque colonne en utilisant des méthodes telles que df.isnull().sum(). Pour assurer la qualité de notre analyse, nous avons supprimé les lignes contenant des valeurs manquantes dans les colonnes "Company" et "Salary" en utilisant df = df.loc[df['Company'].notna()] et df = df.loc[df['Salary'].notna()].
* Afin de mieux analyser les informations contenues dans les colonnes "Requirement of the company" et "Facilities", nous avons divisé les chaînes de caractères en listes à l'aide de df['column\_name'].str.split(',') et appliqué une fonction lambda pour supprimer les espaces inutiles avec [item.strip() for item in x if item.strip()].
* Une autre étape cruciale a été de convertir les salaires en valeurs numériques et de les convertir en USD pour faciliter les comparaisons. Nous avons utilisé des méthodes telles que pd.to\_numeric() et des taux de change pour effectuer cette transformation. 
* Pour obtenir une vue plus complète de nos données, nous avons créé la colonne "country" en se basant sur les informations existantes. De même, nous avons utilisé l'algorithme de traitement naturel du langage (NLP) de la bibliothèque scikit-learn (sklearn) pour créer la colonne "jobtitle", en utilisant des techniques de classification pour extraire les titres de poste pertinents à partir des descriptions fournies. 

En résumé, en effectuant ces différentes tâches de nettoyage, nous avons préparé nos données de manière optimale pour l'analyse ultérieure. Ces étapes ont permis d'éliminer les erreurs potentielles, d'assurer la cohérence des données et de transformer les informations brutes en un ensemble de données exploitable pour nos analyses avancées.

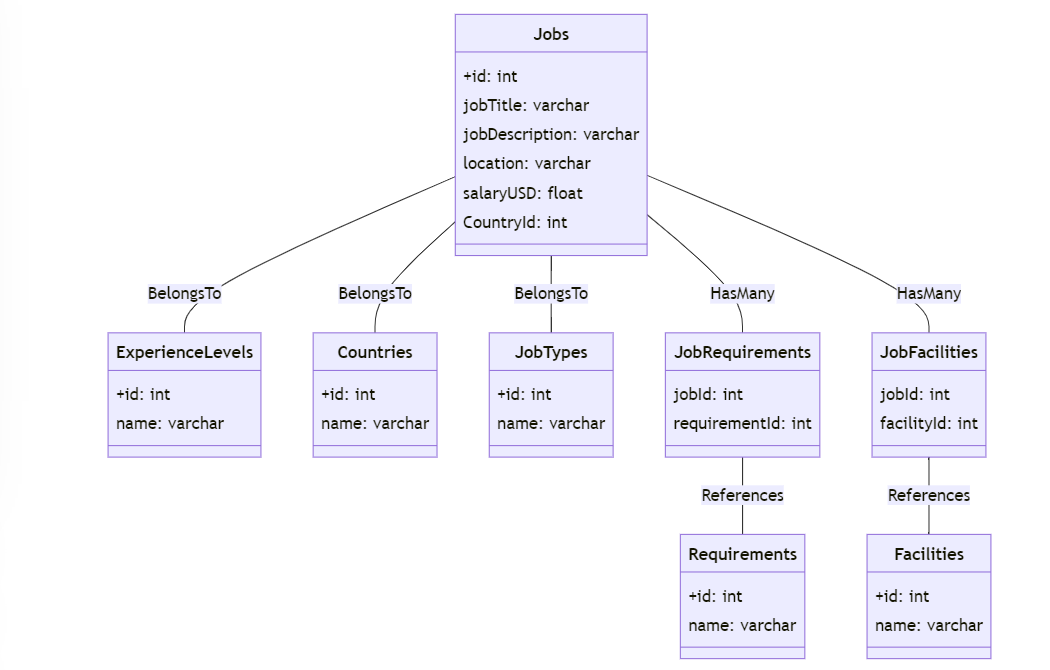
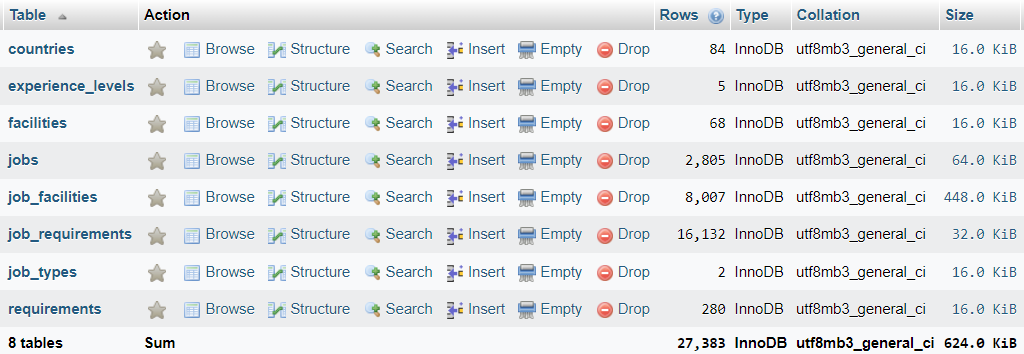
**Visualisation du Marché du Travail en IA, Data Science et Big Data avec Matplotlib**

Pour visualiser de manière efficace les résultats de notre analyse du marché du travail en IA, Data Science et Big Data, nous avons utilisé Matplotlib, une bibliothèque de visualisation en Python. Nous avons créé des graphiques et des diagrammes pertinents pour représenter les statistiques sur les salaires, la répartition du niveau d'expérience, du type de poste, par pays et par entreprise, ainsi que l'analyse des compétences et des exigences. Ces visualisations ont permis de présenter clairement le salaire moyen, le salaire minimum et le salaire maximum, de visualiser la proportion de postes d'entrée de gamme par rapport aux autres niveaux d'expérience, et de mettre en évidence les compétences les plus demandées. Grâce à Matplotlib, nous avons pu communiquer efficacement les tendances et les opportunités du marché du travail à nos parties prenantes, les aidant ainsi à prendre des décisions éclairées en matière de recrutement, de gestion des talents et de développement de compétences.

* **Graphique des Salaires Moyens par Titre de Poste**
* **Statistiques des Salaires pour Chaque Titre de Poste** ****
* **Répartition des Postes d'Entrée de Gamme dans les Offres d'Emploi** ****
* **Répartition des Niveaux d'Expérience dans les Offres d'Emploi** ****
* **Répartition des Types de Poste : Temps plein vs. Temps partiel**
* **Répartition des Offres d'Emploi dans les Pays (Plus de 20 Offres)** ****
* **Top 25 des Pays avec le Salaire Moyen le Plus Élevé (Nombre d'Offres > 10)** ****
* **Nombre d'Offres d'Emploi par Entreprise (Plus de 10 Offres)** ****
* **Salaire Moyen par Entreprise (Entreprises avec plus de 10 Offres d'Emploi)** ****
* **Top 10 des Compétences les Plus Demandées dans les Exigences des Offres d'Emploi** ****
* **Nuage de Mots des Compétences Demandées dans les Exigences des Offres d'Emploi** ****
* **Nombre de Postes Offrant des Installations Spécifiques** ****
* **Nuage de mots des Installations Offertes dans les Offres d'Emploi** ****

**Facilité d'Utilisation grâce à l'Insertion des Données dans une Base de Données (MySQL)**

L'insertion des données issues de notre analyse approfondie du marché du travail en IA, Data Science et Big Data dans une base de données MySQL représente une étape clé pour faciliter leur utilisation dans des applications web, mobiles et des API. Cette base de données offre une structure organisée et sécurisée pour stocker les informations essentielles telles que les salaires, les titres de poste, les compétences demandées et les caractéristiques des entreprises. Grâce à cette infrastructure solide, les développeurs pourront créer des applications interactives, des tableaux de bord visuels et des services personnalisés, offrant aux utilisateurs une expérience fluide et permettant aux parties prenantes de prendre des décisions éclairées en matière de recrutement, de développement de carrière et de planification stratégique. L'intégration de cette base de données avec des API ouvrira également de nouvelles possibilités d'innovation et d'analyse des tendances sur le marché de l'emploi, favorisant ainsi un environnement professionnel dynamique et interconnecté.

* **Diagramme de Classes : Gestion des Offres d'Emploi et des Exigences**
* **Les tables dans la base de données**

**Conclusion :**

L'analyse approfondie du marché du travail dans les domaines émergents de l'IA, de la Data Science et du Big Data a révélé des tendances prometteuses pour l'avenir. Les résultats montrent clairement que le domaine de la data est en plein essor, avec une demande croissante chaque jour. Les compétences les plus demandées et les titres de poste les plus courants dans ces domaines ont été identifiés, permettant ainsi aux entreprises de mieux cibler leurs initiatives de recrutement, d'acquisition de talents et de développement de compétences.

Les visualisations claires et informatives ont rendu les conclusions de l'analyse facilement compréhensibles par toutes les parties prenantes. Les pourcentages de postes d'entrée de gamme et les répartitions par niveau d'expérience ont offert des aperçus précieux sur le paysage de l'emploi dans ces domaines en pleine expansion.

La pertinence des questions clés identifiées pour l'analyse des données a permis de mettre en évidence les opportunités de croissance et les domaines où des investissements supplémentaires peuvent être bénéfiques. Le traitement des données, leur nettoyage, et leur stockage dans une base de données MySQL ont facilité l'utilisation ultérieure dans des applications web, mobiles et API, assurant ainsi une meilleure prise de décision dans le domaine de la data.

Dans l'ensemble, cette analyse approfondie confirme que le domaine de la data possède un avenir prometteur et un potentiel d'évolution significatif. Les entreprises doivent prendre en compte ces conclusions pour s'orienter de manière stratégique et saisir les opportunités offertes par le marché du travail en constante évolution dans le secteur de la data. En capitalisant sur les compétences clés et en répondant aux demandes croissantes du marché, elles seront bien positionnées pour prospérer dans cet environnement dynamique et compétitif.