

A complex network diagram with numerous nodes of varying sizes (blue, grey, and dark brown) connected by thin lines, creating a web-like structure across the entire slide background.

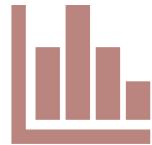
# Analyse d'un sondage mené auprès des professionnels de la data : Insights et tendances à l'aide de Power BI

- ESSOMBA LENJI Chris Manuel
- 04/06/2023

# SOMMAIRE



Introduction



Méthodologie



Résultats



Conclusion

# INTRODUCTION

- Passionné par le monde de la Data, j'ai longtemps été déconcerté par le choix de carrière à suivre dans ce domaine. Malgré mes recherches approfondies sur chaque métier, j'étais toujours perplexe. J'ai donc décidé de demander l'avis des professionnels exerçant ces métiers. Comment? En exploitant des données réelles d'un sondage mené auprès de 630 professionnels de la Data.
- Préparez-vous à plonger dans les résultats captivants de cette analyse. Ensemble, nous dévoilerons les métiers les plus rémunérateurs et les langages de programmation les plus populaires dans le monde de la Data.

# MÉTHODOLOGIE

## **Etape 0 : Importation des données sur Kaggle**

- Téléchargement du jeu de données sur les décisions de crédit depuis Kaggle.

## **Etape 1 : Nettoyage des données**

- Traitement des valeurs manquantes, valeurs mal saisies, valeurs sous le mauvais format.

## **Etape 2 : Visualisation sous Power BI**

- Identifier les visualisations clés nécessaires pour répondre aux questions posées.
- Identifier les visualisations clés nécessaires pour répondre aux questions posées.
- Utiliser les fonctionnalités de Power BI pour créer des graphiques, des tableaux de bord et des rapports interactifs en fonction des variables d'intérêt
- Ajouter des filtres, des segments et des interactions pour permettre une exploration approfondie des données.
- Ajouter des filtres, des segments et des interactions pour permettre une exploration approfondie des données.



A hand holding a spray bottle, spraying a fine mist of water against a dark background. The mist is captured in motion, creating a soft, ethereal glow. The hand is positioned on the left side of the frame, and the spray extends towards the right.

# ETAPE 1 NETTOYAGE DES DONNEES EN UTILISANT POWER QUERY

Rapport réalisé par Chris ESSOMBA

# TRAITEMENT DES DONNÉES MANQUANTES NUMERIQUES: REMPLACEMENT PAR LA MOYENNE

- Nous avons utilisé la méthode *List.Average* afin de remplacer les valeurs manquantes par la moyenne dans chaque colonne qui contient des valeurs numériques.

## Remplacer les valeurs

Remplacez une valeur dans les colonnes sélectionnées par une autre.

Valeur à rechercher

Remplacer par

> Options avancées

OK

Annuler

# TRAITEMENT DES DONNÉES MANQUANTES QUALITATIVES: REMPLACEMENT PAR LE MODE

- Nous utilisons la commande *List.Mode* pour remplacer les valeurs nulles dans chaque colonne contenant des valeurs qualitatives.

## Remplacer les valeurs

Remplacez une valeur dans les colonnes sélectionnées par une autre.

Valeur à rechercher

Remplacer par

> Options avancées

OK

Annuler

# TRAITEMENT DES DONNÉES MAL SAISIES

- L'objectif de cette section était de rectifier les erreurs d'orthographe dans les noms de métiers (illustrés dans l'image de gauche) et de les réécrire sans le préfixe '*Other (Please Specify)*'.

Q1 - Which Title Best Fits your Current Role?

Tri croissant  
Tri décroissant  
Effacer le tri  
Effacer le filtre  
Supprimer les éléments vides

Filtres textuels

Rechercher

- ☒ Data Scientist
- ☒ Database Developer
- ☒ Other (Please Specify)
- ☒ Other (Please Specify):Account manager
- ☒ Other (Please Specify):Ads operations
- ☒ Other (Please Specify):Analyst
- ☒ Other (Please Specify):Analyst Primary Market I
- ☒ Other (Please Specify):Analytics Consultant
- ☒ Other (Please Specify):Analytics Engineer
- ☒ Other (Please Specify):Analytics Manager
- ☒ Other (Please Specify):BI consultant
- ☒ Other (Please Specify):BI Developer
- ☒ Other (Please Specify):BI Manager
- ☒ Other (Please Specify):Billing analyst
- ☒ Other (Please Specify):Business Analys
- ☒ Other (Please Specify):Business Analyst
- ☒ Other (Please Specify):Business analyst

La liste peut être incomplète. [Charger plus](#)

OK Annuler

Tri croissant  
Tri décroissant  
Effacer le tri  
Effacer le filtre  
Supprimer les éléments vides

Filtres textuels

Rechercher

- ☒ Analytics Manager
- ☒ Business Analyst
- ☒ Business Analyst
- ☒ Business Intelligence Consultant
- ☒ Business Intelligence Developer
- ☒ Business Intelligence Manager
- ☒ Business Intelligence Analyst
- ☒ Business Intelligence Consultant
- ☒ Business Intelligence Developer
- ☒ Business Intelligence Engineer
- ☒ Consultant
- ☒ Continuous Quality Improvement Specialist
- ☒ Data Analyst
- ☒ Data Architect
- ☒ Data Coordinator
- ☒ Data Engineer

La liste peut être incomplète. [Charger plus](#)

OK Annuler



# TRANSFORMATION DES VALEURS DONNÉES SOUS FORMES EN MOYENNE NUMÉRIQUE

- L'objectif était de déterminer la moyenne des données salariales qui étaient présentées sous forme d'intervalles, afin de simplifier leur analyse.

Q3 - Current Yearly Salary (in USD)	Average Salary
106k-125k	115,5
41k-65k	53
0-40k	20
150k-225k	187,5
41k-65k	53
0-40k	20
0-40k	20
125k-150k	137,5
86k-105k	95,5
41k-65k	53
66k-85k	75,5
0-40k	20
0-40k	20
0-40k	20
41k-65k	53
41k-65k	53
0-40k	20
0-40k	20
41k-65k	53
0-40k	20
41k-65k	53
106k-125k	115,5
0-40k	20
0-40k	20
0-40k	20



# VISUALISATION SUR POWER BI

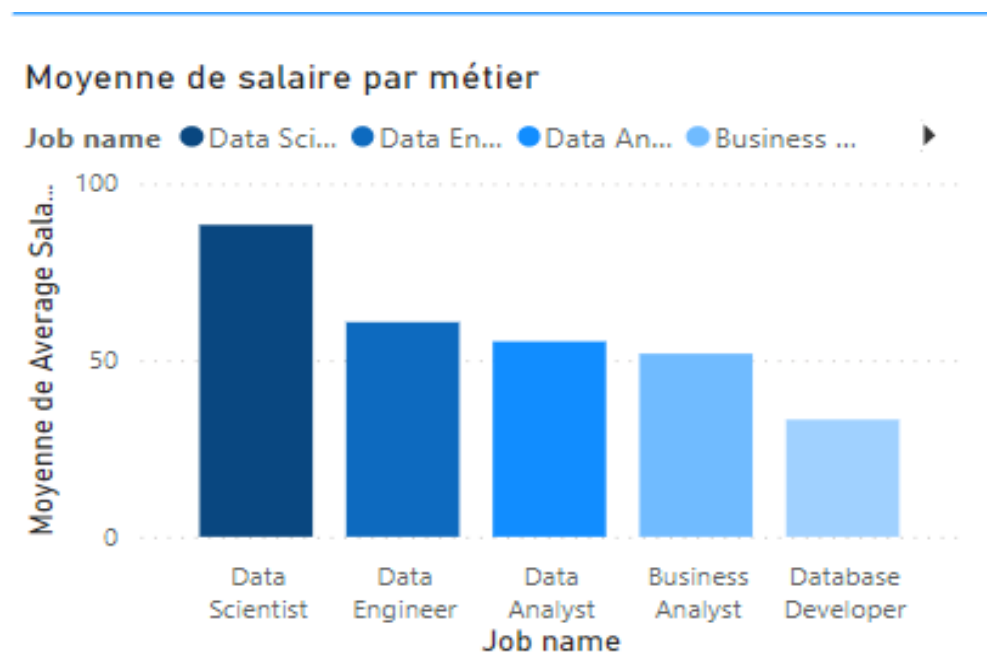


# PARTIE 1 : ANALYSE DES VARIABLES CLÉES



# ANALYSE DES MÉTIERS LES MIEUX RÉMUNÉRÉS

- D'après les résultats de ce sondage, le métier le mieux rémunéré est le Data Scientist, avec une moyenne de 88 000 \$ par an. Cependant, au fur et à mesure de notre analyse approfondie, nous remarquerons que ces chiffres sont influencés par divers facteurs.

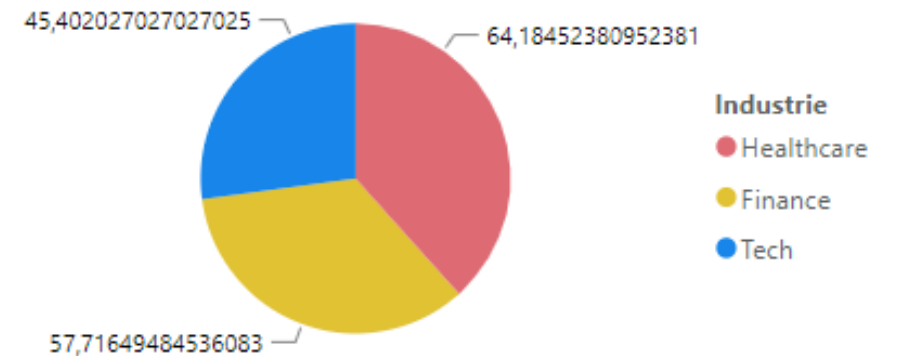




# ANALYSE DES SECTEURS D'ACTIVITÉS RÉÉNUMÉRANT LE MIEUX

- D'après les résultats de ce sondage, il est remarquable de constater que le secteur de la santé offre la meilleure rémunération aux professionnels de la Data, avec un salaire moyen dépassant les 64 000 \$ par an.

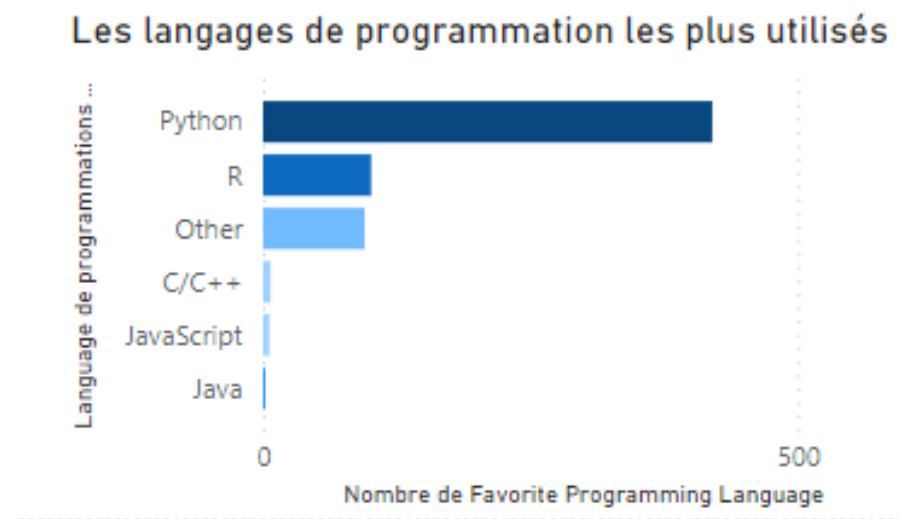
Les secteurs d'activités les plus rémunérateur





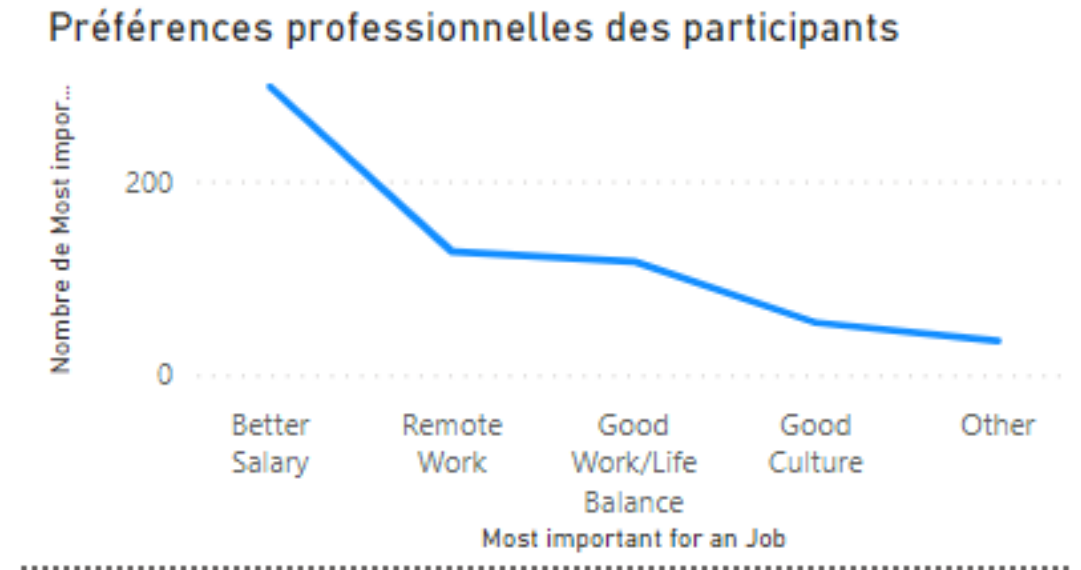
# ANALYSE DES LANGAGES DE PROGRAMMATIONS LES PLUS UTILES

- En se basant sur les résultats de ce sondage, une conclusion s'impose : Python est le langage de programmation le plus populaire auprès des professionnels de la Data.



# ANALYSE DES PRÉFÉRENCES PROFESSIONNELLES

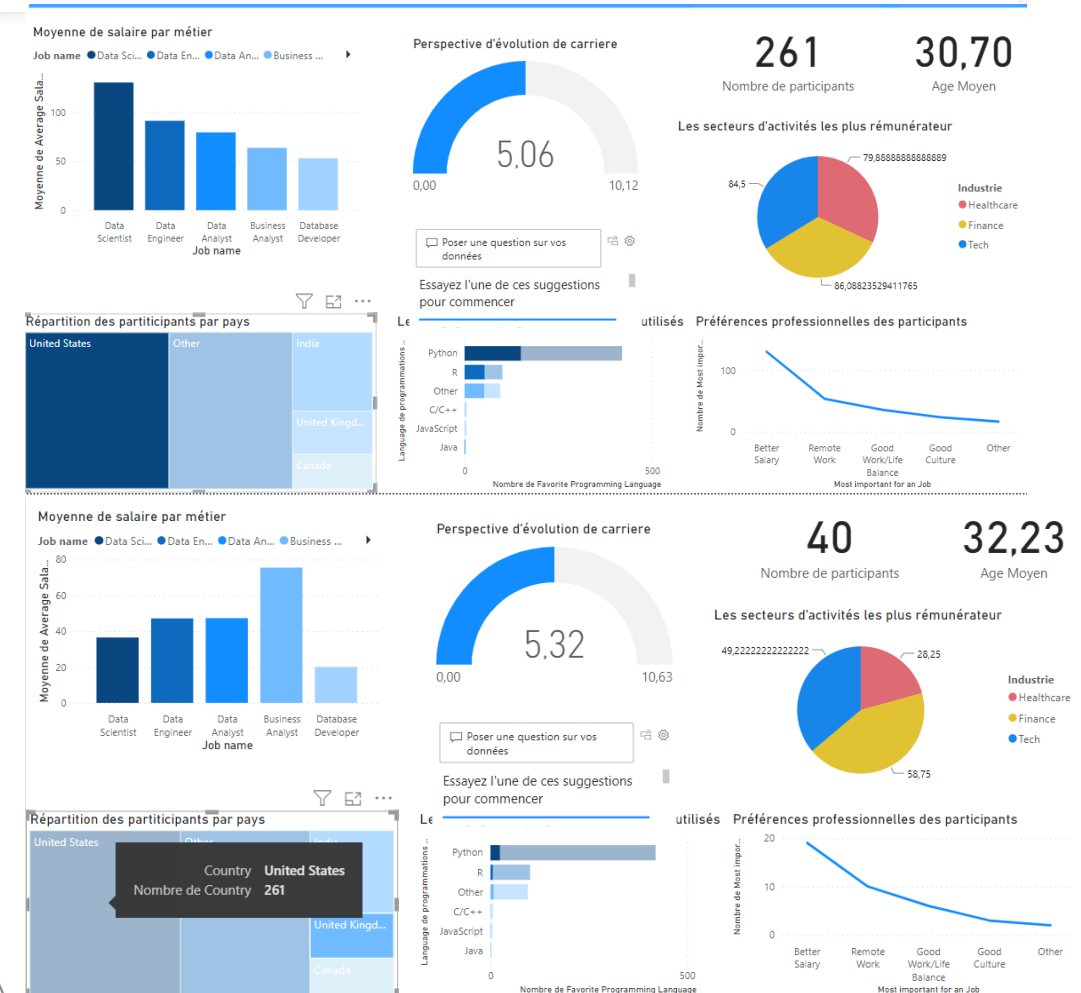
- L'analyse des préférences professionnelles de ce groupe de participants met en évidence un constat clair : le salaire est de loin le critère le plus apprécié. En deuxième position, la possibilité de télétravail est également très importante, tandis que l'importance de la culture de l'entreprise n'est pas à négliger.



## PARTIE 2 : EXPLORATION DES VARIABLES QUI IMPACTENT LES SALAIRES DES PROFESSIONNELS DE LA DATA.

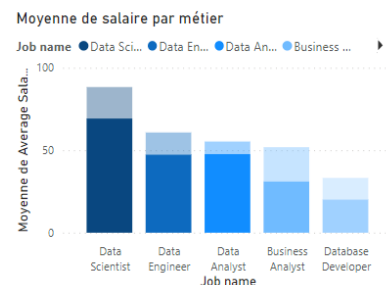
# IMPACT DU PAYS DE TRAVAIL : Etats-Unis VS Royaume-Unis

- Cette analyse démontre clairement l'impact significatif du pays de travail sur les salaires. Un exemple frappant est celui des Data Scientists travaillant aux États-Unis, qui perçoivent plus de trois fois le salaire de leurs pairs travaillant au Royaume-Uni. Les Data Scientists américains gagnent en moyenne 130 000 \$ par an, alors que leurs collègues britanniques touchent en moyenne 36 000 \$ par an.

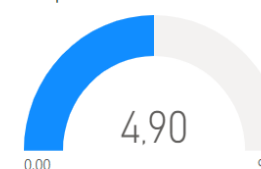


# IMPACT DES SECTEURS D'ACTIVITÉ: FINANCE VS TECHNOLOGIE.

- L'analyse met en évidence l'impact significatif des secteurs d'activité sur les salaires des professionnels de la data. Dans l'ensemble, il est observé que les salaires sont plus élevés dans le secteur de la finance par rapport à celui de la technologie. De manière intéressante, les Business Analysts travaillant dans la finance gagnent en moyenne le double de leurs collègues travaillant dans la technologie.

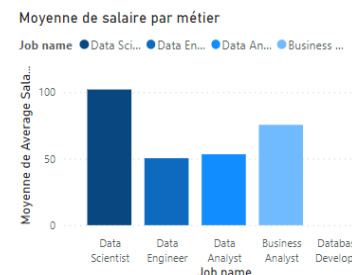
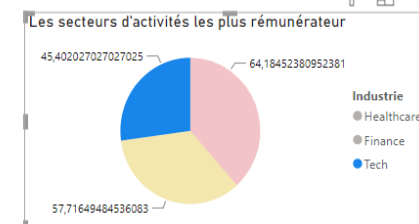


Perspective d'évolution de carrière

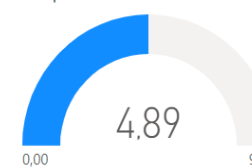


150 28,57

Nombre de participants Age Moyen

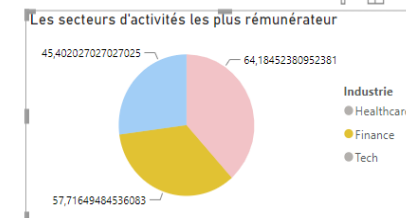


Perspective d'évolution de carrière



97 30,21

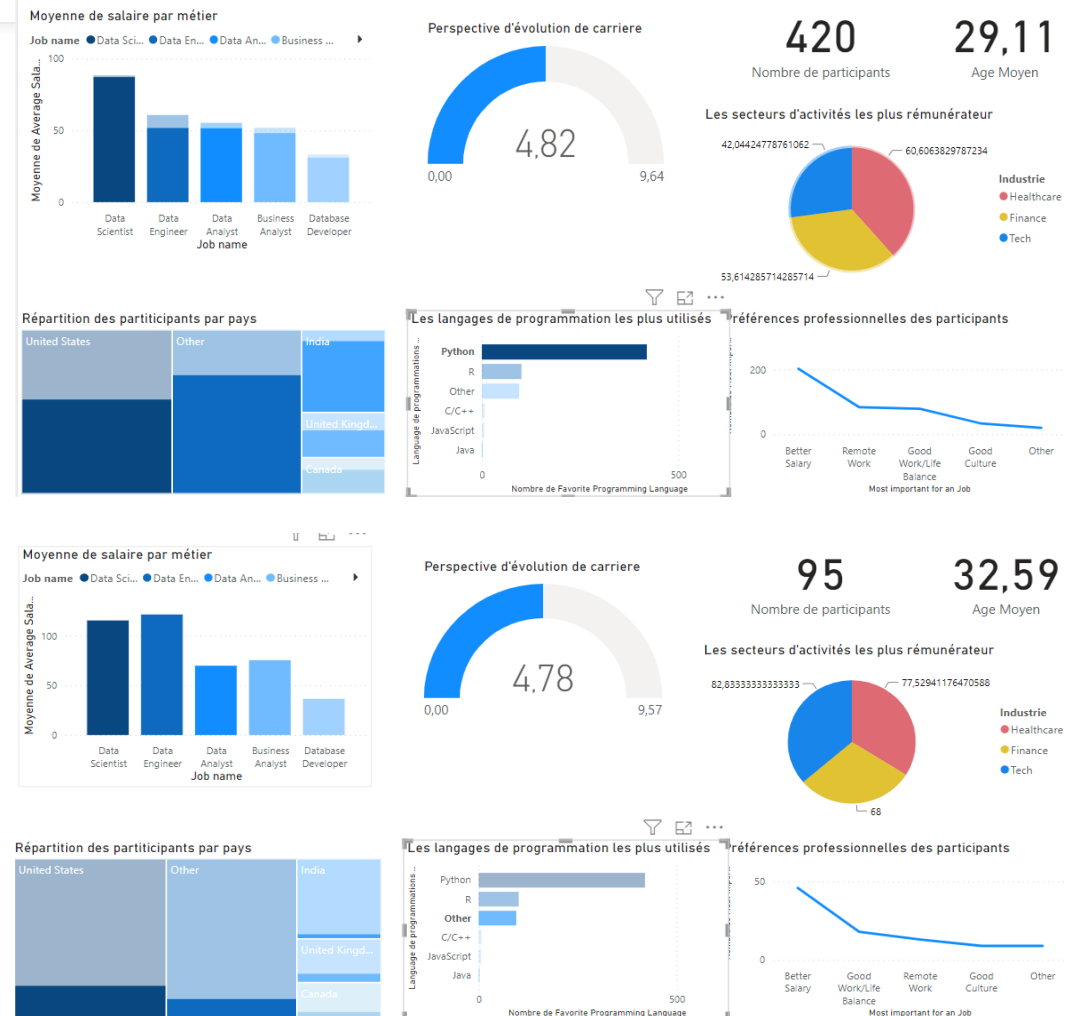
Nombre de participants Age Moyen





# IMPACT DE L'UTILISATION DE PYTHON.

- Cette analyse révèle que les professionnels utilisant majoritairement Python dans leur travail (au-dessus) gagnent moins que ceux utilisant d'autres. Ceci peut s'expliquer par le fait que ce langage de programmation soit un langage très populaire et plutôt facile à apprendre entraînant ainsi une très grande offre de personnes possédant cette compétence comparée à la demande des entreprises, donc ce qui peut justifier une baisse des salaires proposés.



# CONCLUSION

- En conclusion, l'analyse approfondie du sondage mené auprès des professionnels de la data a permis de tirer les conclusions suivantes :
- Le métier de Data Scientist est le mieux rémunéré, avec une moyenne de 88 000 \$ par an.
- Le pays de travail joue un rôle déterminant dans les salaires des professionnels de la data, avec des écarts significatifs entre les pays.
- Le secteur d'activité a également une influence sur les salaires, avec des salaires généralement plus élevés dans le domaine de la finance.
- Python est le langage de programmation le plus essentiel pour les professionnels de la data, bien qu'il puisse être associé à des salaires légèrement inférieurs en raison de sa popularité et de sa facilité d'apprentissage.
- Ces conclusions soulignent l'importance de prendre en compte ces facteurs lorsqu'on envisage une carrière dans le domaine de la data et lors de l'établissement de politiques de rémunération. Elles fournissent des informations précieuses pour les professionnels de la data et les entreprises afin de prendre des décisions éclairées.

# REFERENCE

- La base de données utilisée dans notre projet a été téléchargée à partir de Kaggle, une plateforme en ligne populaire pour les scientifiques des données et les praticiens du Machine Learning. Le jeu de données spécifique que nous avons utilisé est intitulé *Data Science survey* et il a été fourni par ALEX THE ANALYST sur Kaggle. Ce jeu de données regroupe les réponses à un sondage réalisé auprès d'une communauté de professionnels exerçant dans le domaine du Big Data.



Merci pour votre  
attention !