**Ecole Nationale Polytechnique**

**Département du Génie Industriel**

**Option : Data Science & Intelligence Artificielle**

**2ème année**

Rapport de Projet

**Business Intelligence (BI)**



# Analyzing Job Market Data

*Case Study*

# Réalisé par

* Sarah BOUARABA
* Sofiane SADAT
* Yousra LAIB

## Sommaire

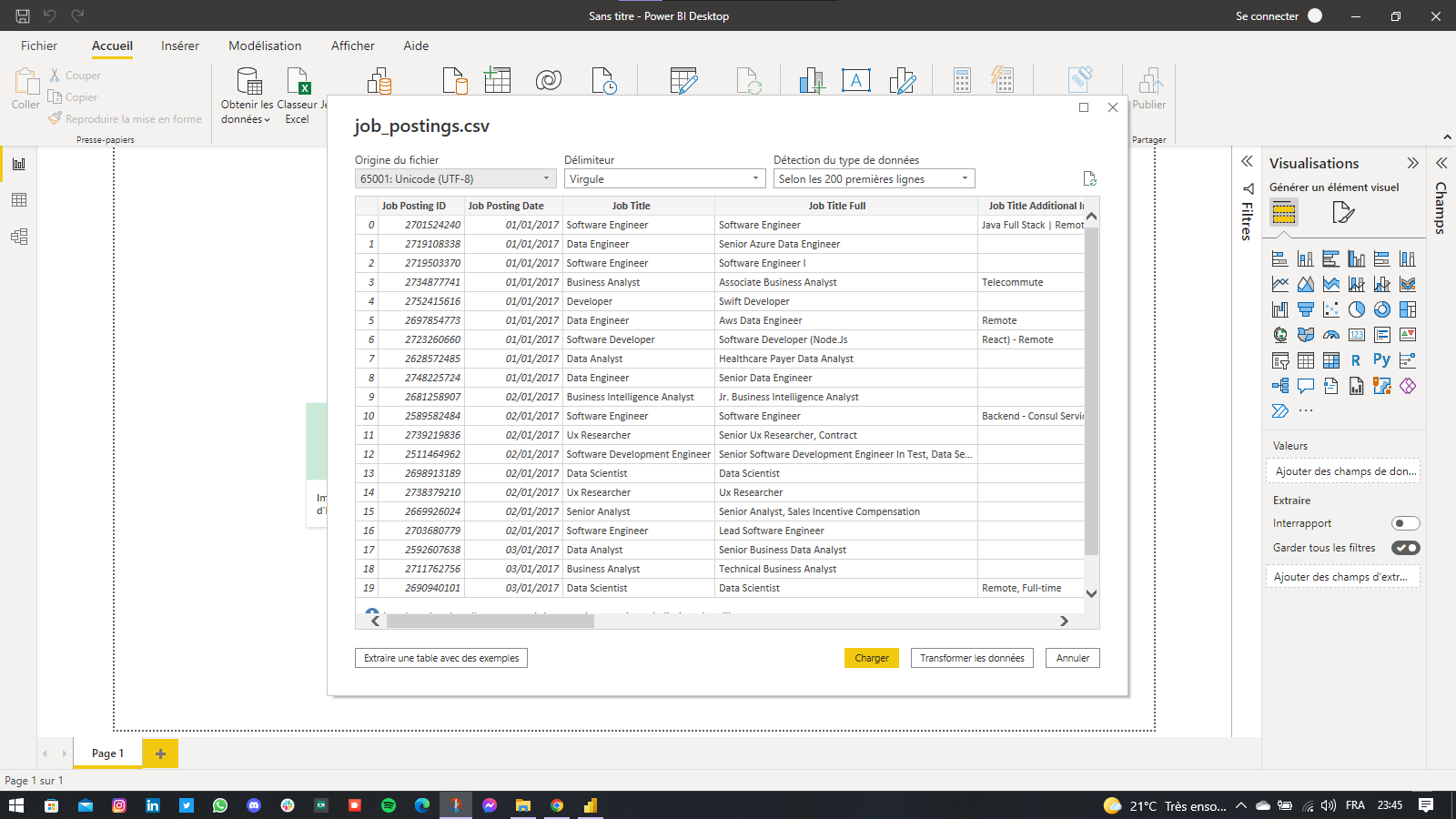
1. **Exploratory Data Analysis (EDA)**
2. Analyse globale initiale (Statistiques de base)
3. Analyse des tendances des offres d’emploi
4. Nettoyage de données pour les “Job Titles”
5. Effet des années d’expérience sur le salaire
6. Les tendances trouvées avec l’EDA (Résultats de l’EDA)
7. **Analyse des tendances du marché**
8. Nettoyage et analyse des “Skills”
9. Probabilité des “Skills” dans l’offre d’emploi
10. Tendances des “Skills” au fil du temps
11. Principales descriptions du poste
12. Autres recommandations d’emploi
13. Tendances issues de l’analyse (Interprétations)
14. **Création de tableau de bord**
15. Commencer les modèles de tableau de bord
16. Conception d’une mise en page de tableau de bord
17. Explorer les meilleurs skills pour les emplois
18. Jauge salariale pour les entreprises
19. Ajouter de l’interactivité avec des signets
20. **Exploratory Data Analysis**

Faire une analyse exploratoire de données (EDA) peut être difficile ou ambigu. Pour cette raison, avant de commencer, il est essentiel d’identifier l’étendue de notre analyse. On peut par exemple se poser les questions suivantes :

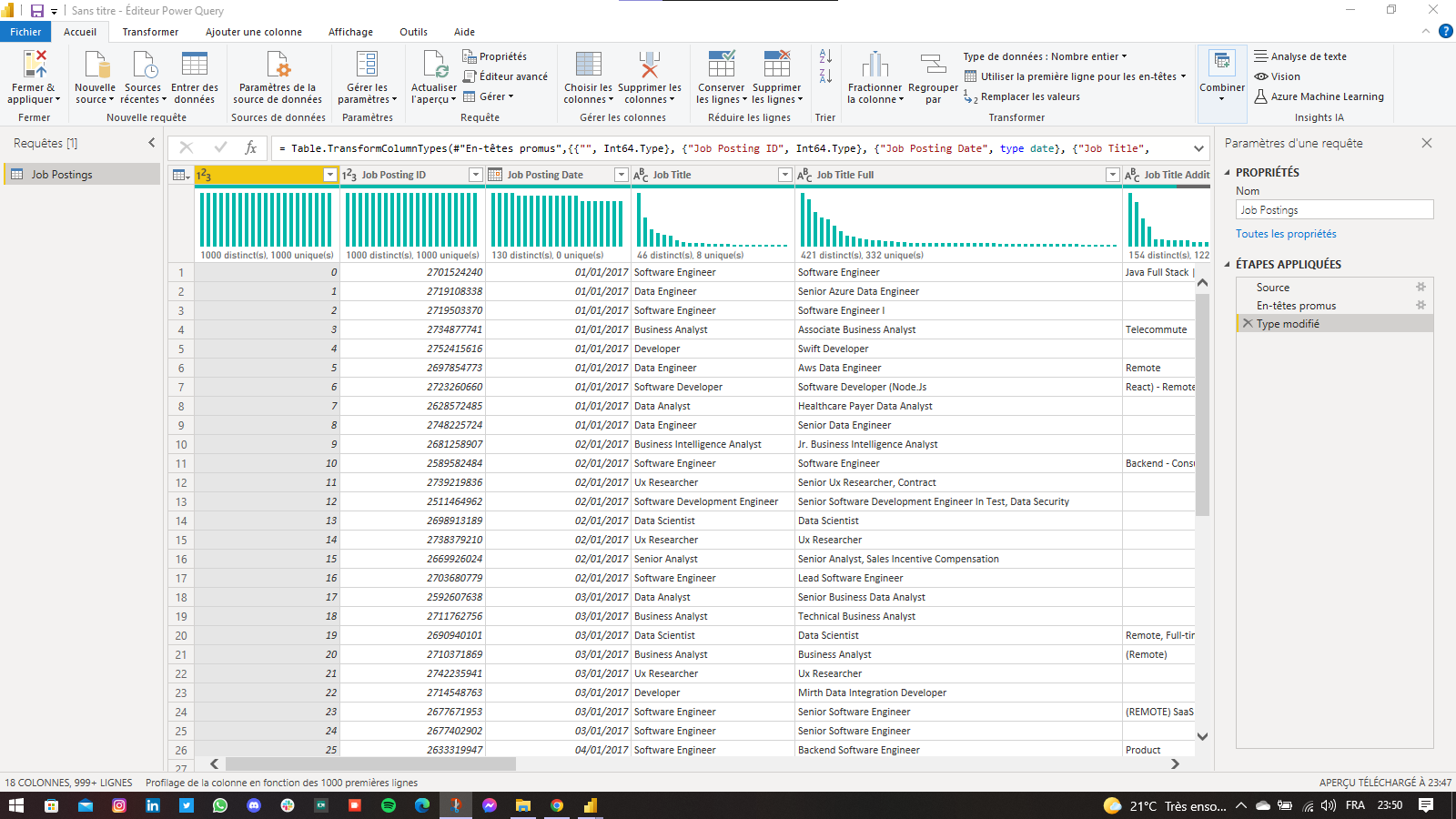
* Quelles compétences en Data Science sont en corrélation avec les offres d’emploi ?
* Quels sont les métiers les plus recherchés en Data Science ?
* Quel est le salaire moyen des postes en Data Science ?

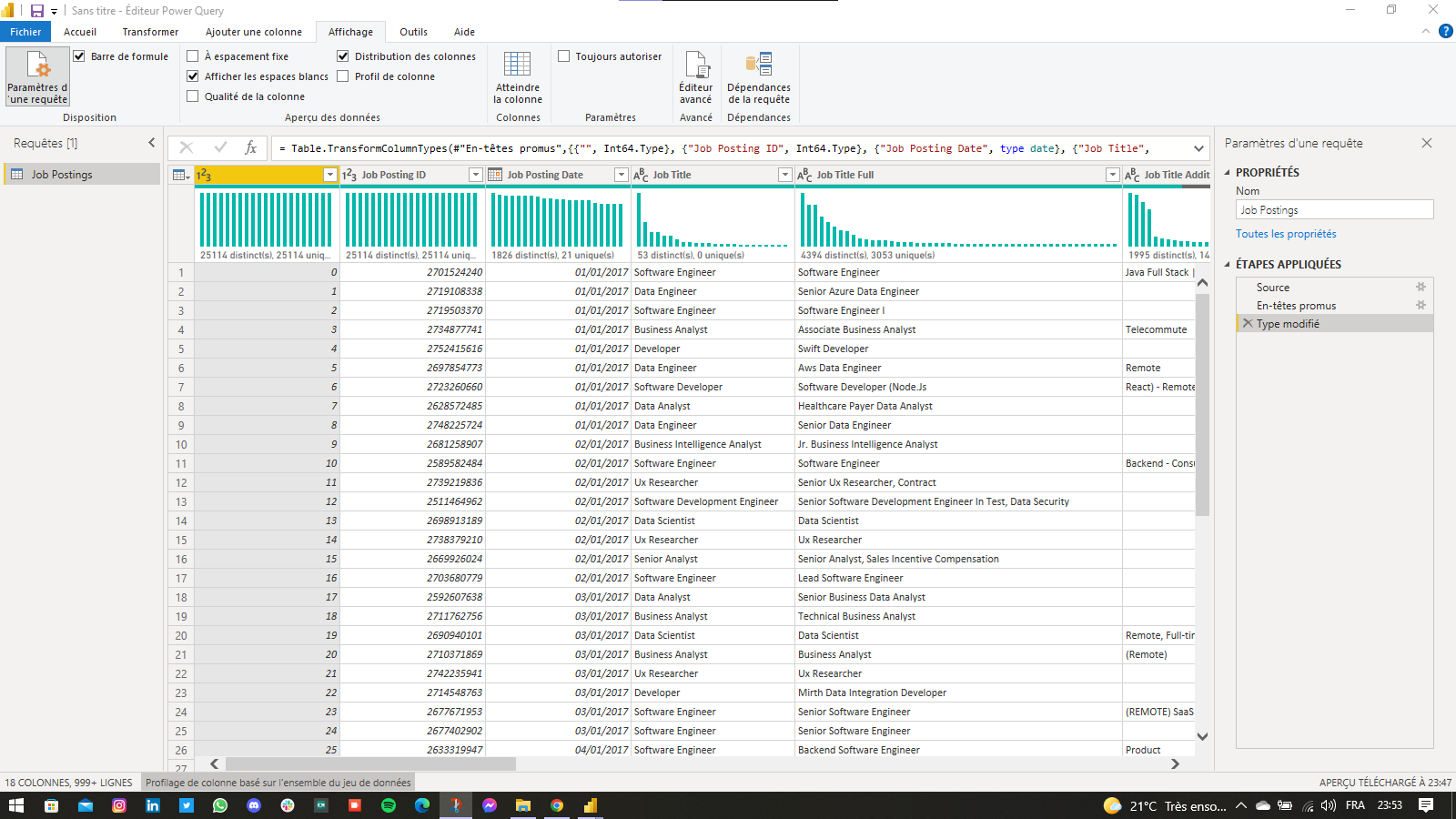
1. **Analyse globale initiale (Statistiques de base)**

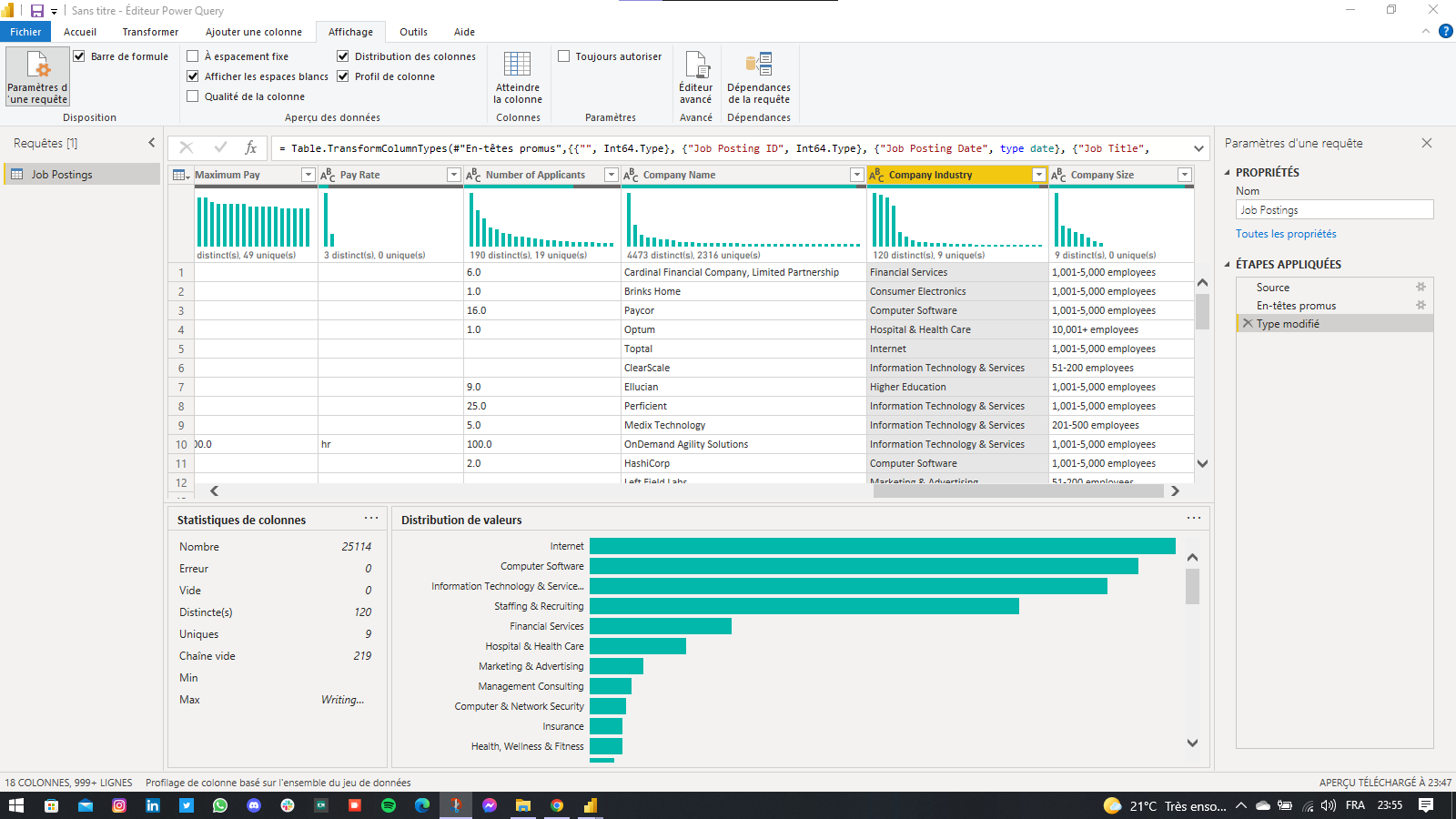
* Importation des données :



* Accès à l'éditeur Power Query pour rechercher les valeurs manquantes :

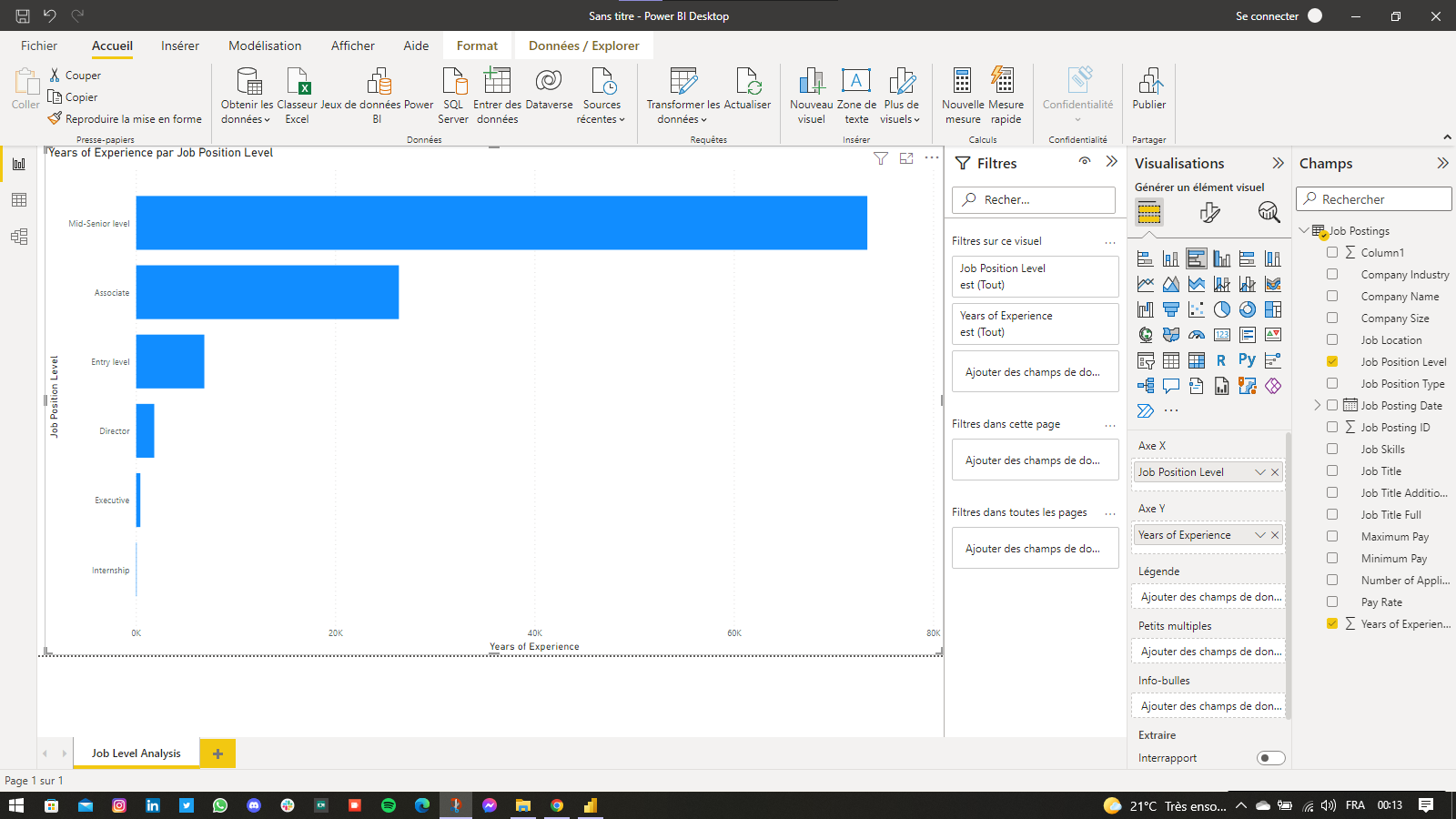


* Modification le profilage des colonnes à partie des 1000 premières lignes pour qu'il soit basé sur l'intégralité de l'ensemble de données :
* Statistiques des colonnes et la distribution des valeurs (Column Profiles) :

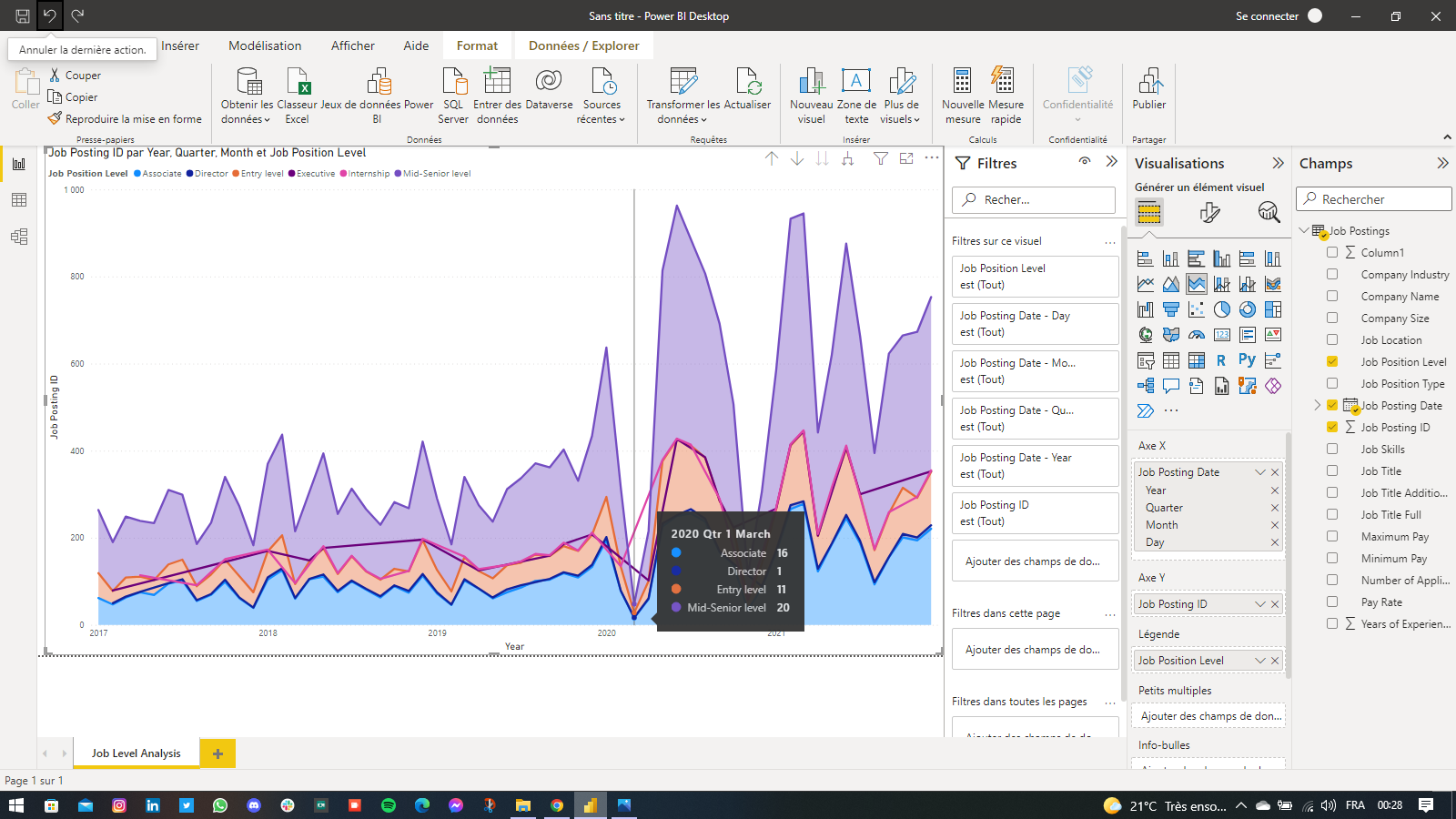


1. **Analyse des tendances des offres d’emploi**

* On veut examiner comment les années d’expérience se rapportent à un poste particulier, c-à-d : la relation entre “Years Experience” et “Job Position level” :



* Pour avoir un meilleur aperçu, visualisons l'évolution du nombre d'offres d'emploi (Job Posting ID) au fil du temps à l'aide de colonne "Job Posting Date" en utilisant la colonne “Job Position Level” pour différencier visuellement ces offres :

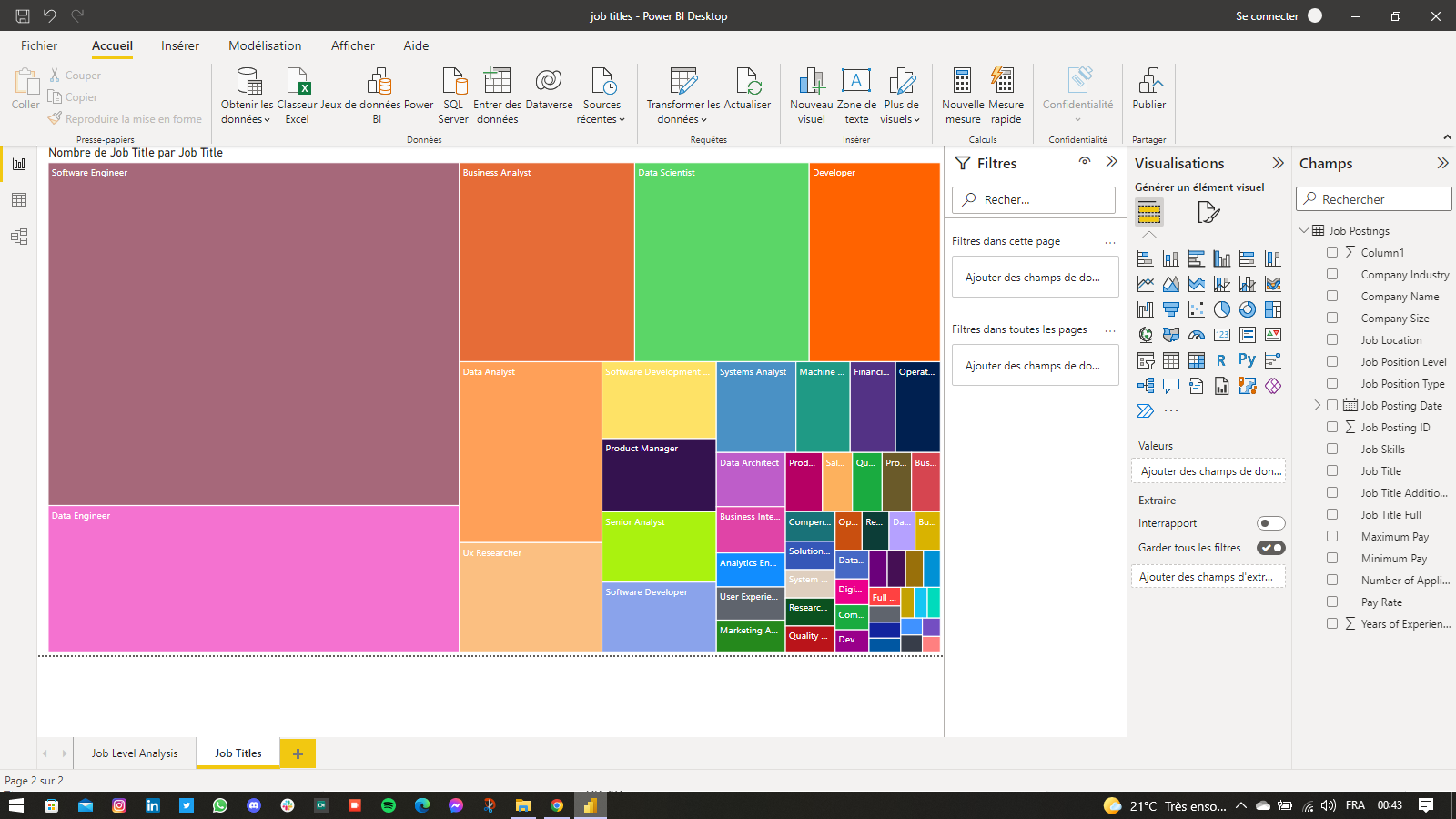


* On remarque ici que l’offre d’emploi est au minimum en Mars 2020, ce qui est logique, car durant cette période la pandémie a commencé, ce qui a engendré une diminution du nombre d'emplois.

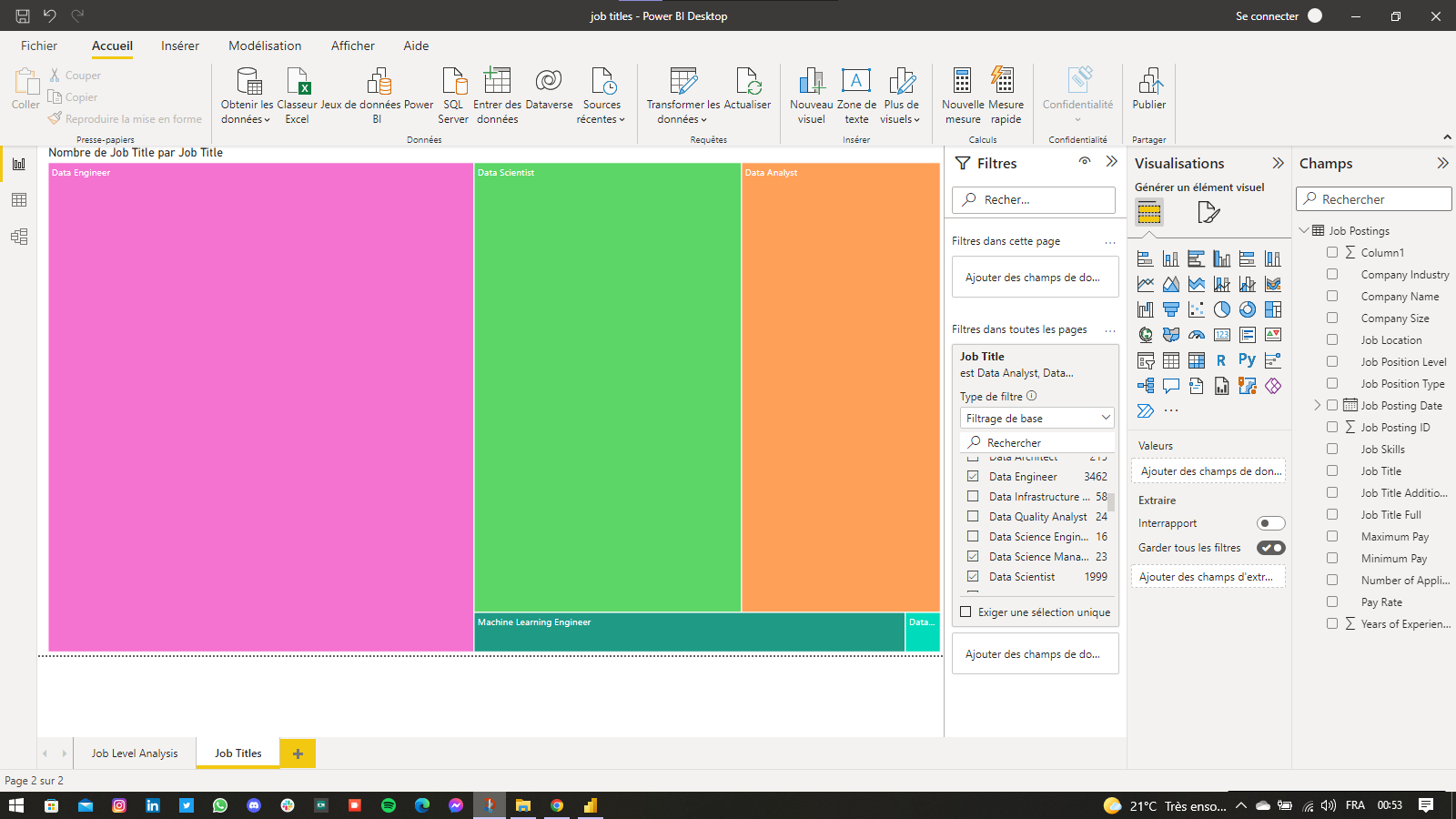
1. **Nettoyage de données pour les “Job Titles”**

On veut maintenant explorer les 5 métiers suivants (Les 5 meilleurs emplois en Data Science) : Data Scientist, Data Engineer, Data Analyst, Machine Learning Engineer et Data Science Manager, il faut donc faire un nettoyage de données.

* Visualisons le nombre de tous les différents “Job Titles” les uns par rapport aux autres à l’aide de l’outil “Treemap” :



* Faisons un filtrage pour toutes les pages afin de se concentrer sur les 5 métiers cités auparavant : Data Scientist, Data Engineer, Data Analyst, Machine Learning Engineer et Data Science Manager.



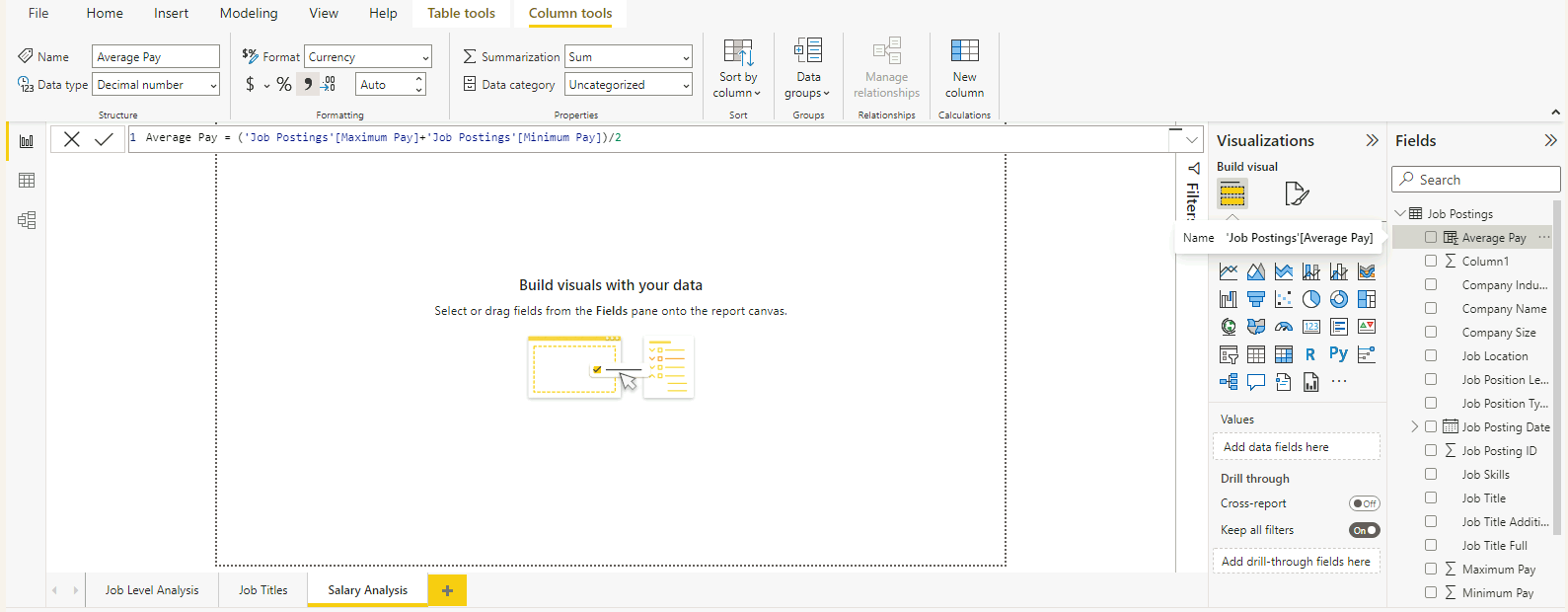
* On peut lire sur la visualisation le nombre de chaque “Job Title” :
* Data Engineer : 3462
* Data Scientist : 1999
* Data Analyst : 1485
* Machine Learning Engineer : 283
* Data Science Manager : 23

1. **Effet des années d’expérience sur le salaire**

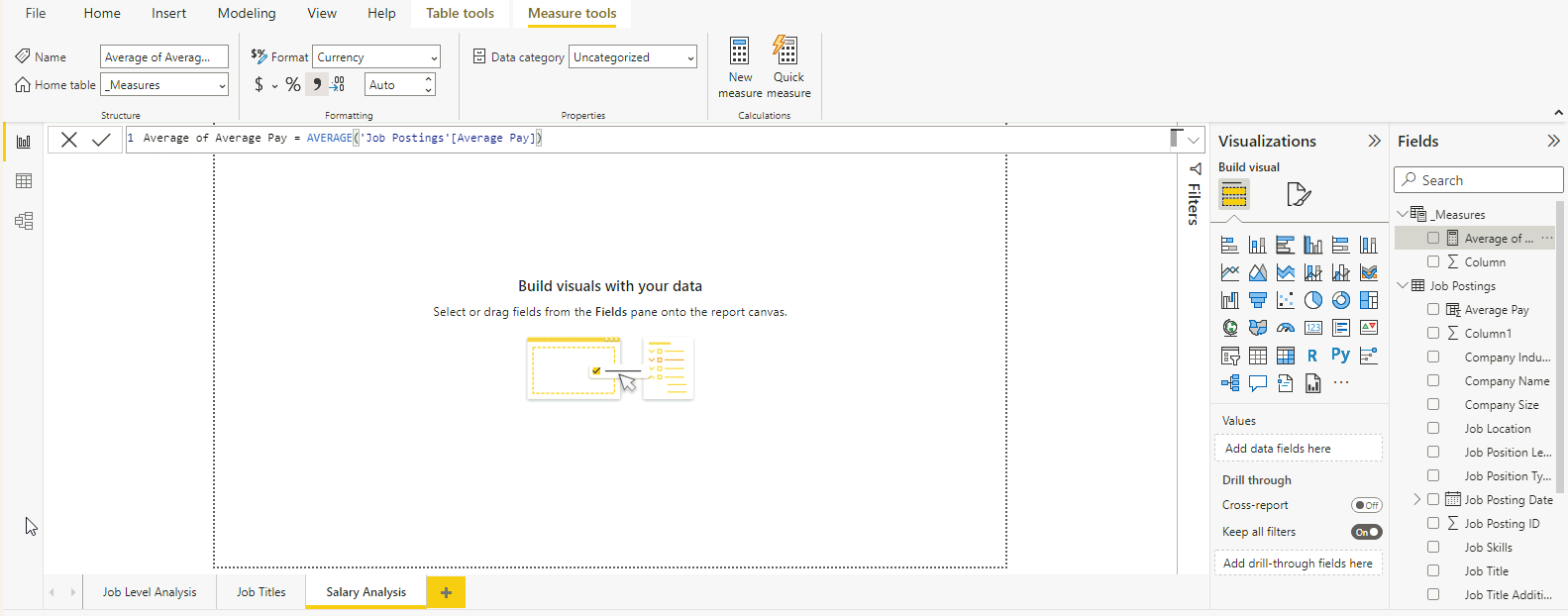
On veut maintenant voir ce qui influence les tendances du marché. Pour celà, examinons les salaires des différents emplois et analysons comment ils évoluent avec les années d’expérience. ( La relation entre “Salary” & “Years of Experience” )

Remarque : Etant donné que le dataset ne fournit que des valeurs minimales et maximales pour les salaires, nous devons d’abord effectuer un nettoyage de données.

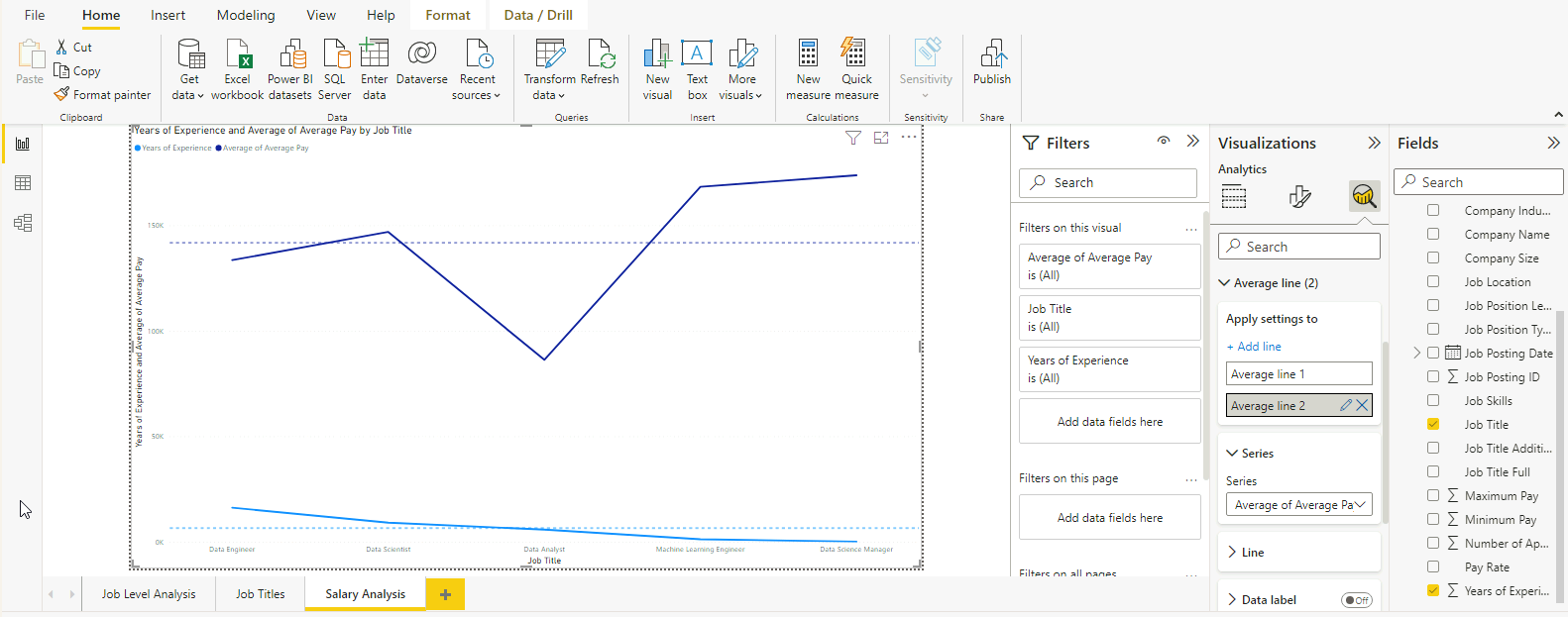
* Estimation du salaire moyen en créant une colonne calculée “Average Pay” (Format = Devise)



* On veut créer des mesures avec DAX, pour celà créons une nouvelle table intitulée “\_Measures”.
* Créons une nouvelle mesure pour la moyenne de “Average Pay” dans la table “\_Measures” intitulée “Average of Average Pay” (Format=devise) :



* Visualisation de la nouvelle mesure du salaire moyen par rapport aux années d’expérience pour chaque Job Title :



* On remarque que le poste d’un Data Analyst a le salaire moyen le plus bas en fonction des années d’expérience.

1. **Les tendances trouvées avec l’EDA (Résultats de l’EDA)**

* Il y a une corrélation entre le niveau de poste (Job Level Position) et les années d’expérience (Years Experience) pour un poste.
* Les offres d’emploi des 5 dernières années sont à la hausse.
* À mesure que les années d’expérience augmentent pour un emploi, le salaire associé augmente également.

1. **Analyse des Tendances du Marché**

Il est maintenant temps de se lancer dans la recherche d’informations en vue de la création de notre livrable

final à la société de recrutement.

Bien que nous avons pu effectuer une analyse de toutes les données précédemment, nous approfondirons

certaines des colonnes présentées pour découvrir d’autres informations en créant des visualisations

attrayantes.

Étudions ce qui suit :

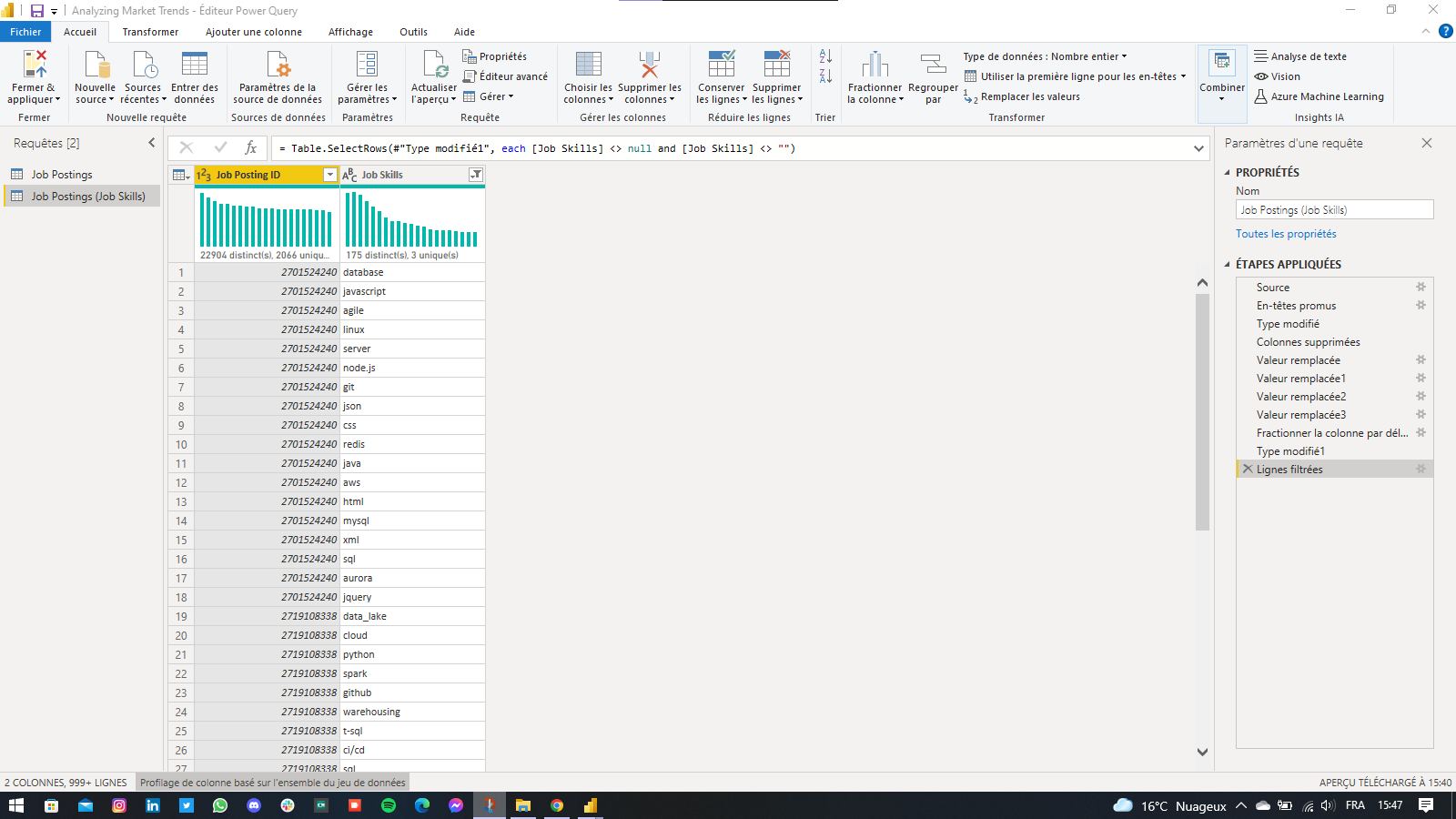
* La corrélation entre les “Skills” et “Job Titles”.
* Les principaux secteurs et entreprises qui demandent ces rôles.
* Recommandations potentielles : garder un œil sur les informations clés que nous souhaitons partager avec notre client.

1. **Nettoyage et analyse des “Skills”**

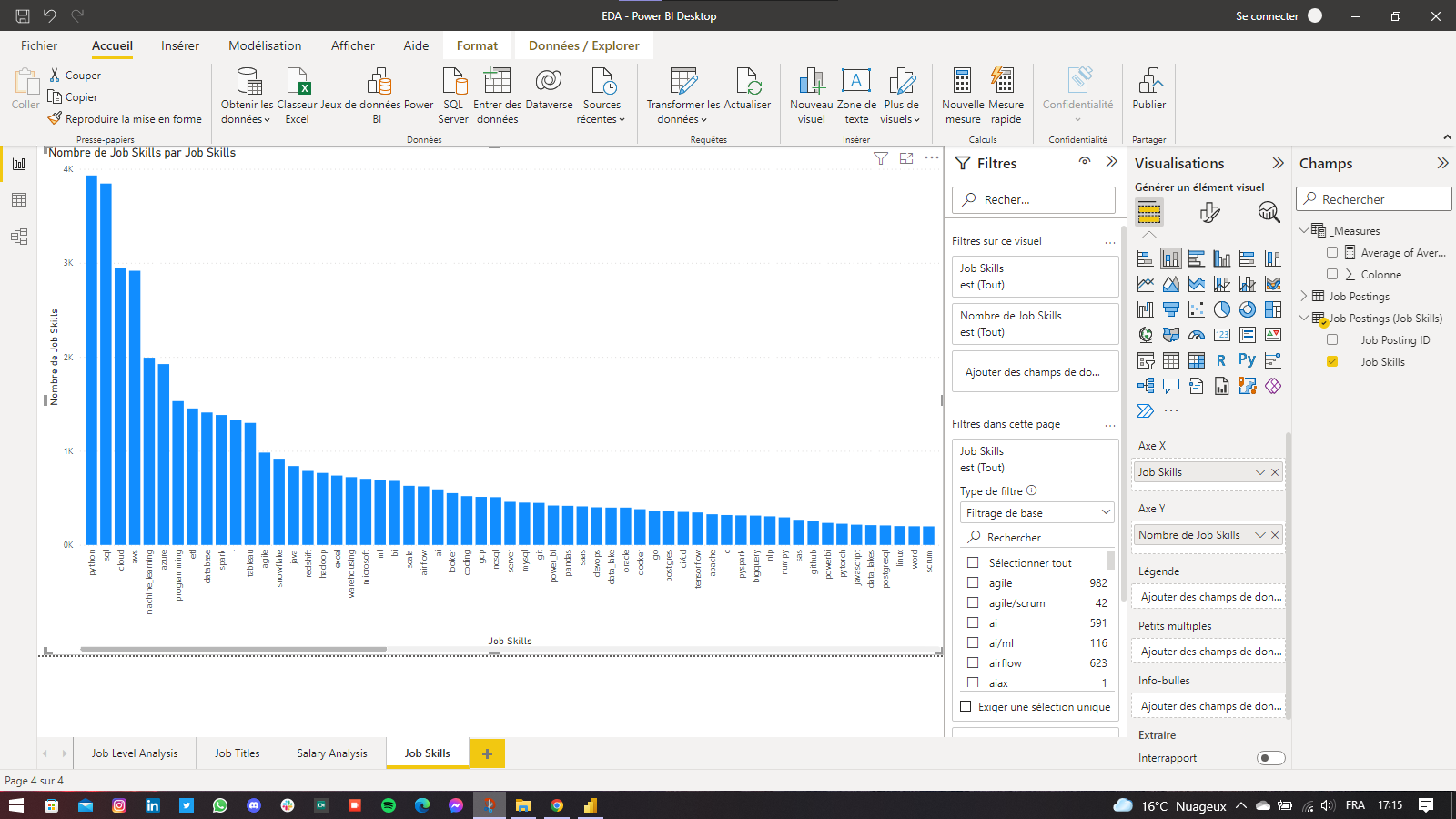
Passons à l’analyse des compétences. La colonne “Job Skills” contient des valeurs dans une liste

( ['database', 'javascript', …] ), nous devons donc nettoyer celà avant une analyse plus approfondie.

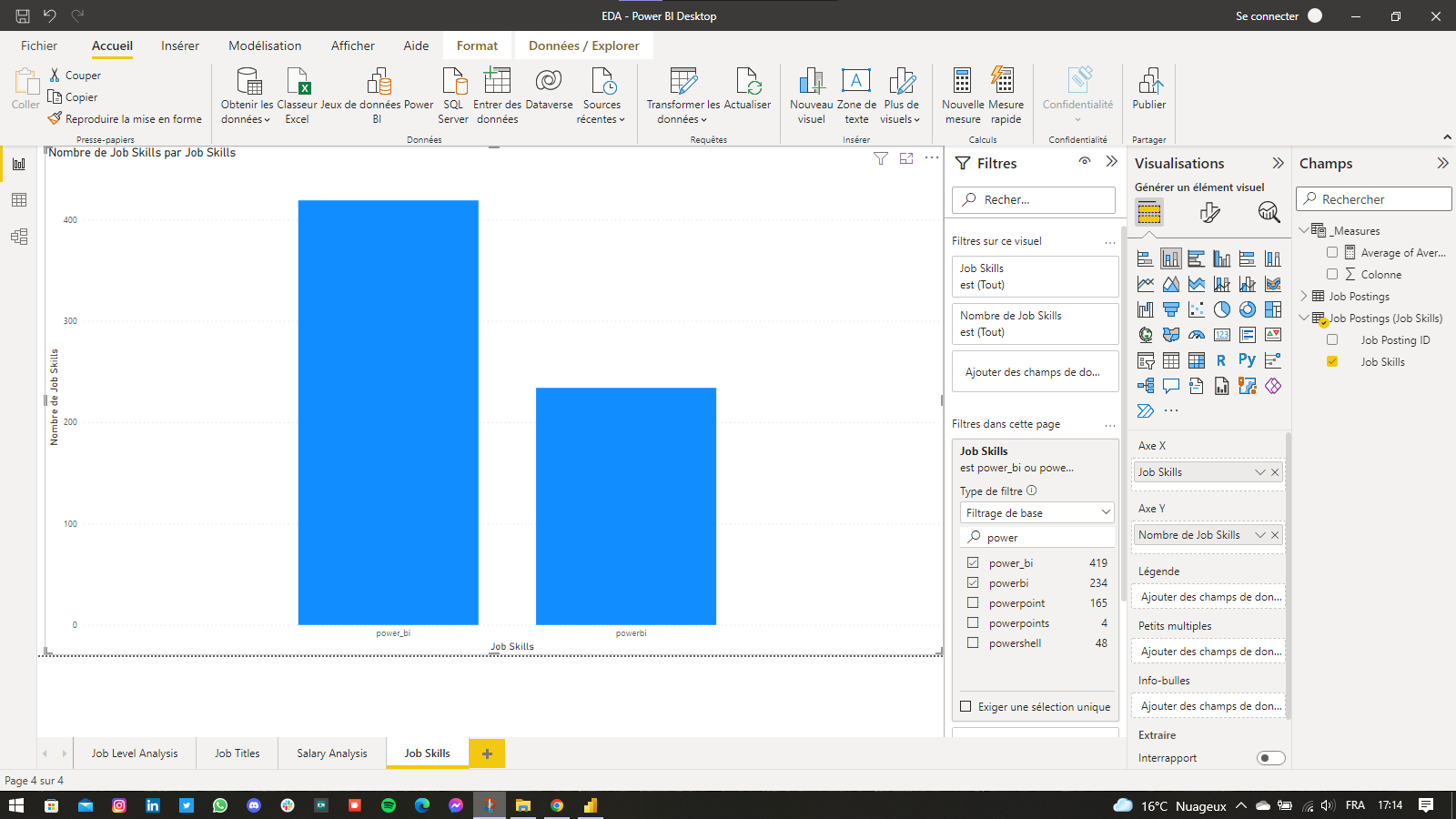
* Création d’une nouvelle requête avec un doublon de notre dataset “Job Postings” appelée “Job Postings (Job Skills)” qui contient les 2 colonnes : “Job Posting ID” & “Job Skills”.
* Nettoyons la colonne “Job Skills” pour supprimer les crochets, les guillemets et les espaces ( [ ],’ ‘,les espaces).
* Division de la colonne “Job Skills” par le délimiteur pour créer une ligne pour chaque compétence avec le fichier “Job Posting ID”.
* Suppression des lignes qui ont une valeur vide dans la colonne “Job Skills”.



* Création d’une visualisation montrant le nombre de tous les différents Skills :



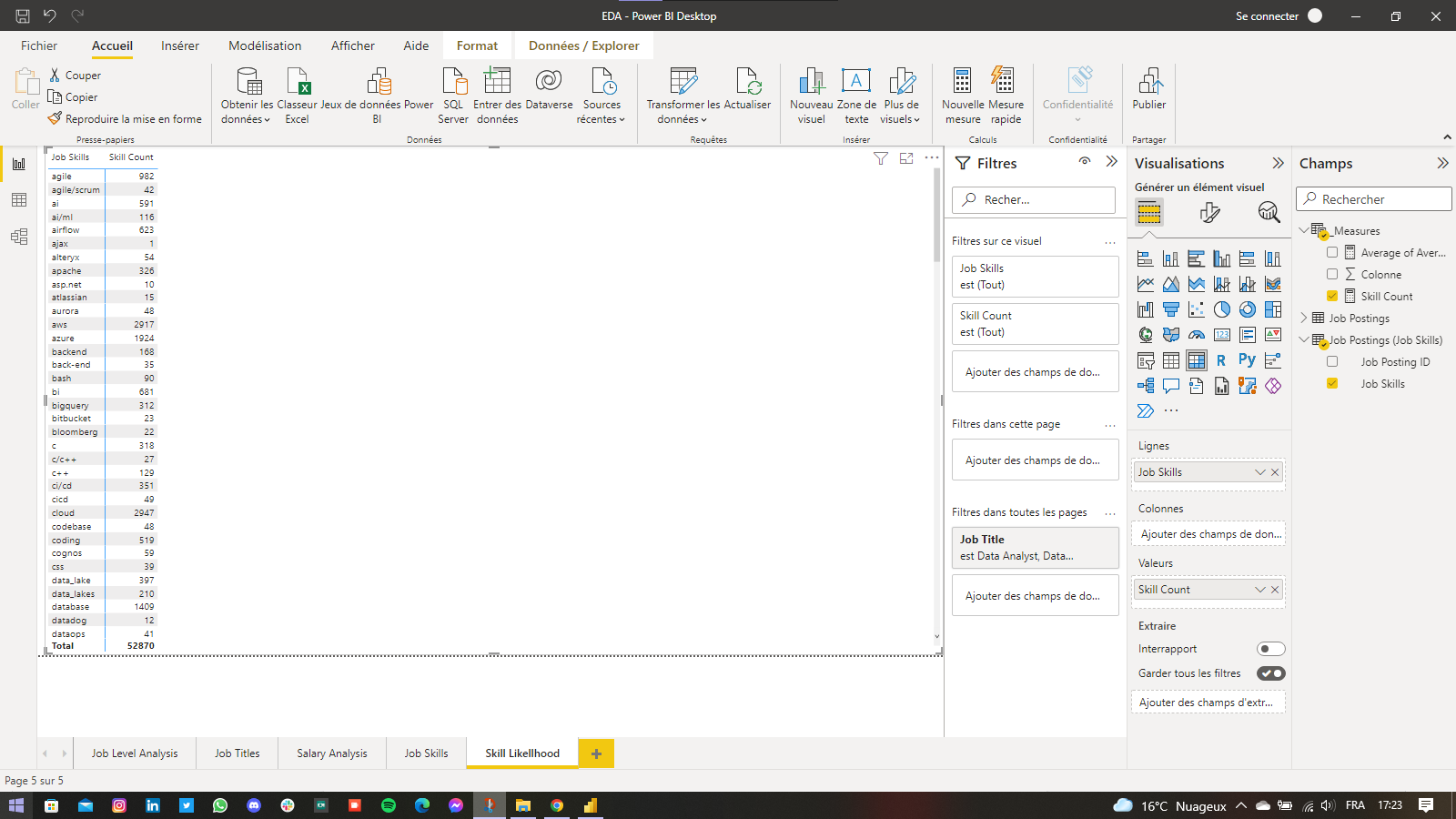
* On peut lire sur la visualisation combien de fois apparaît Power BI (powerbi, power\_bi) comme compétence (653) dans les offres d’emploi en ajoutant un filtre :



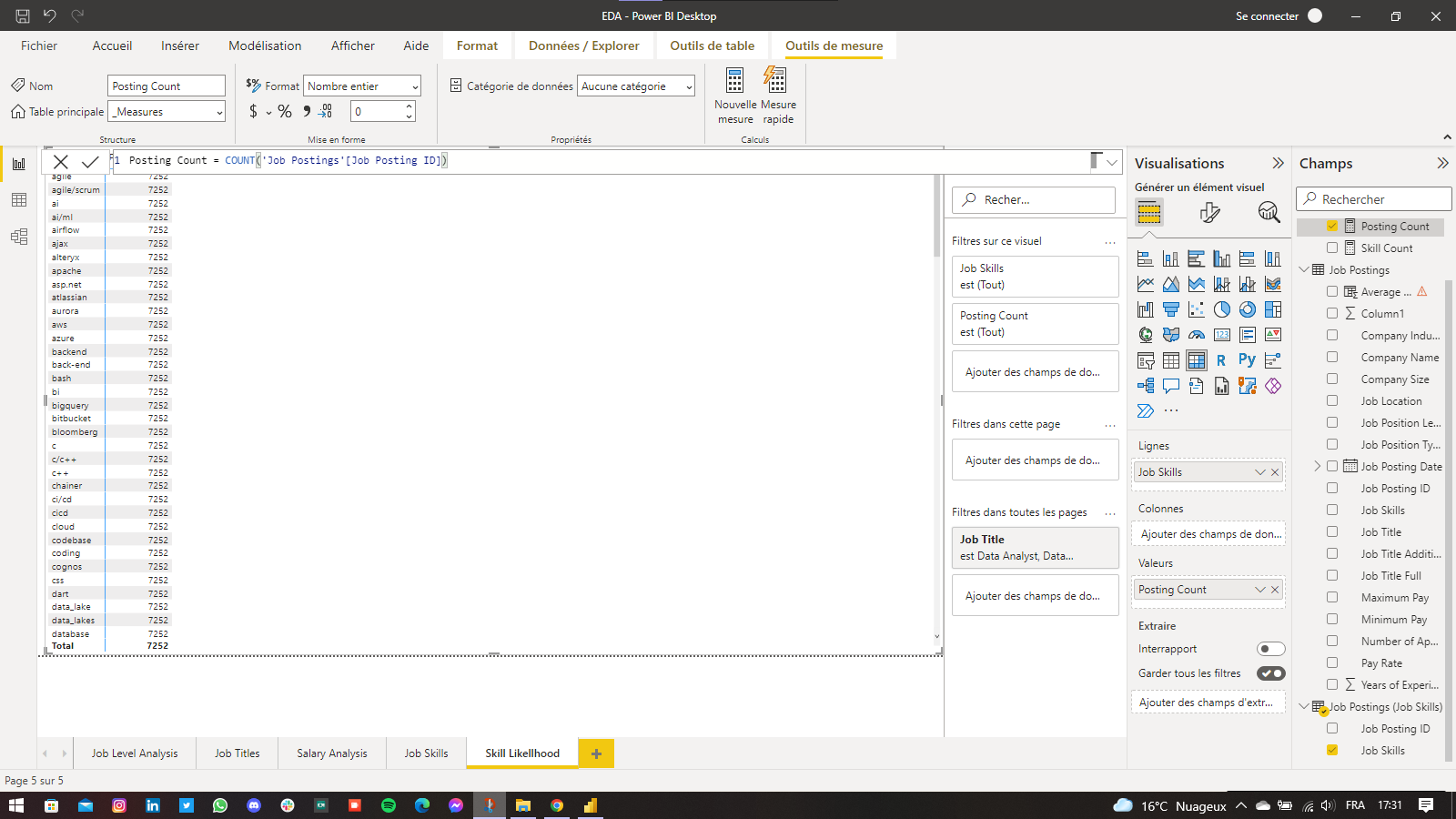
1. **Probabilité des “Skills” dans l’offre d’emploi**

On veut maintenant comprendre quel est le pourcentage de chance qu’un certain Skill soit répertorié dans une offre d’emploi

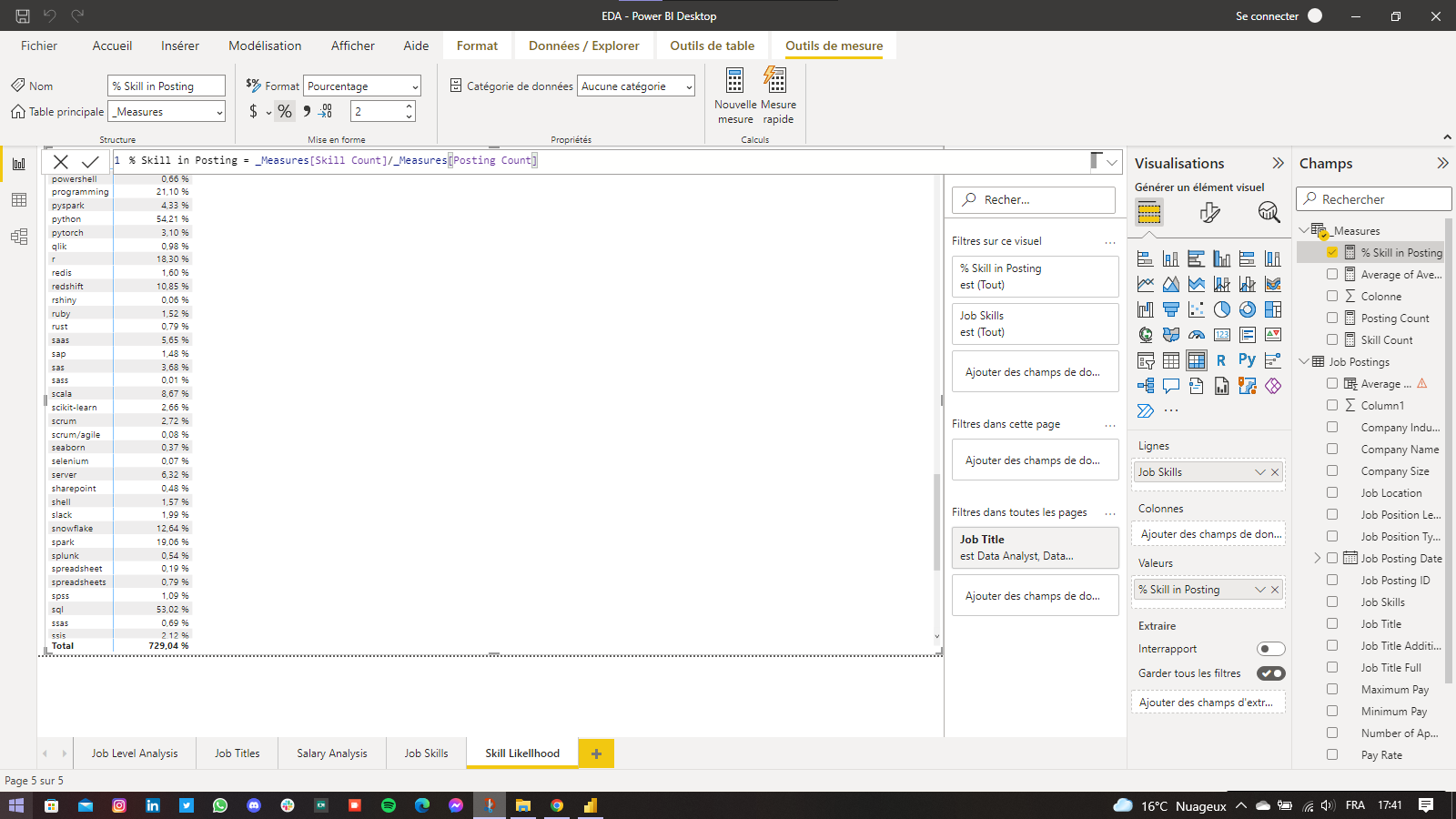
* Classons “powerbi” et “power\_bi” sous le même skill “power\_bi”.
* Création d’une mesure appelée “Skill Count” qui compte les Skills dans la table “Job Skills”.
* Création d’une matrice pour visualiser cette nouvelle mesure.



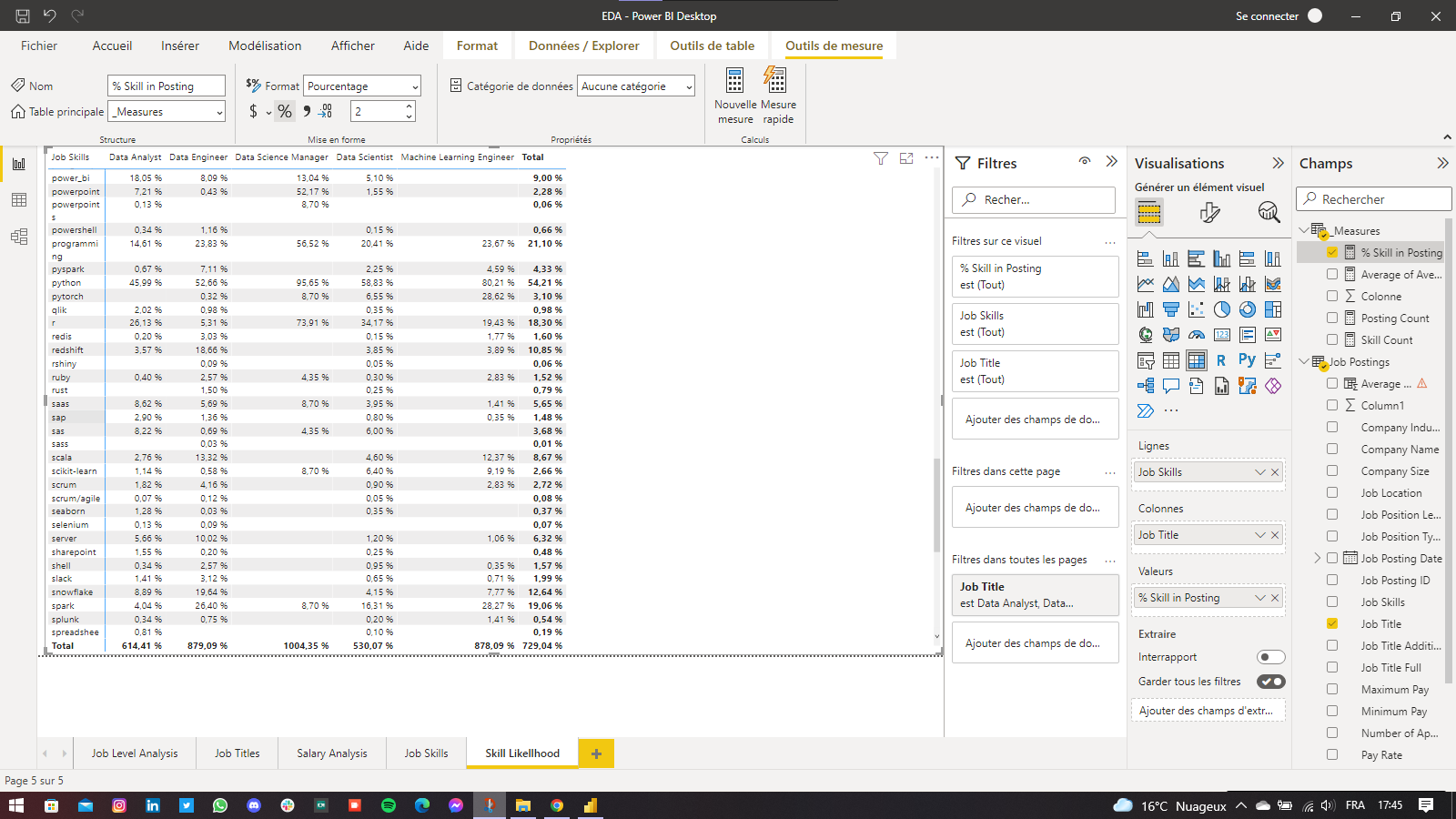
* Création d’une mesure appelée “Postings Count” pour le nombre de toutes les offres d’emploi “Job Postings”.



* Création d’une mesure finale appelée “%Skill in Posting” où nous voyons la probabilité qu’un skill se trouve dans un posting.



* Création de colonnes pour différents fichiers “Job Titles” :

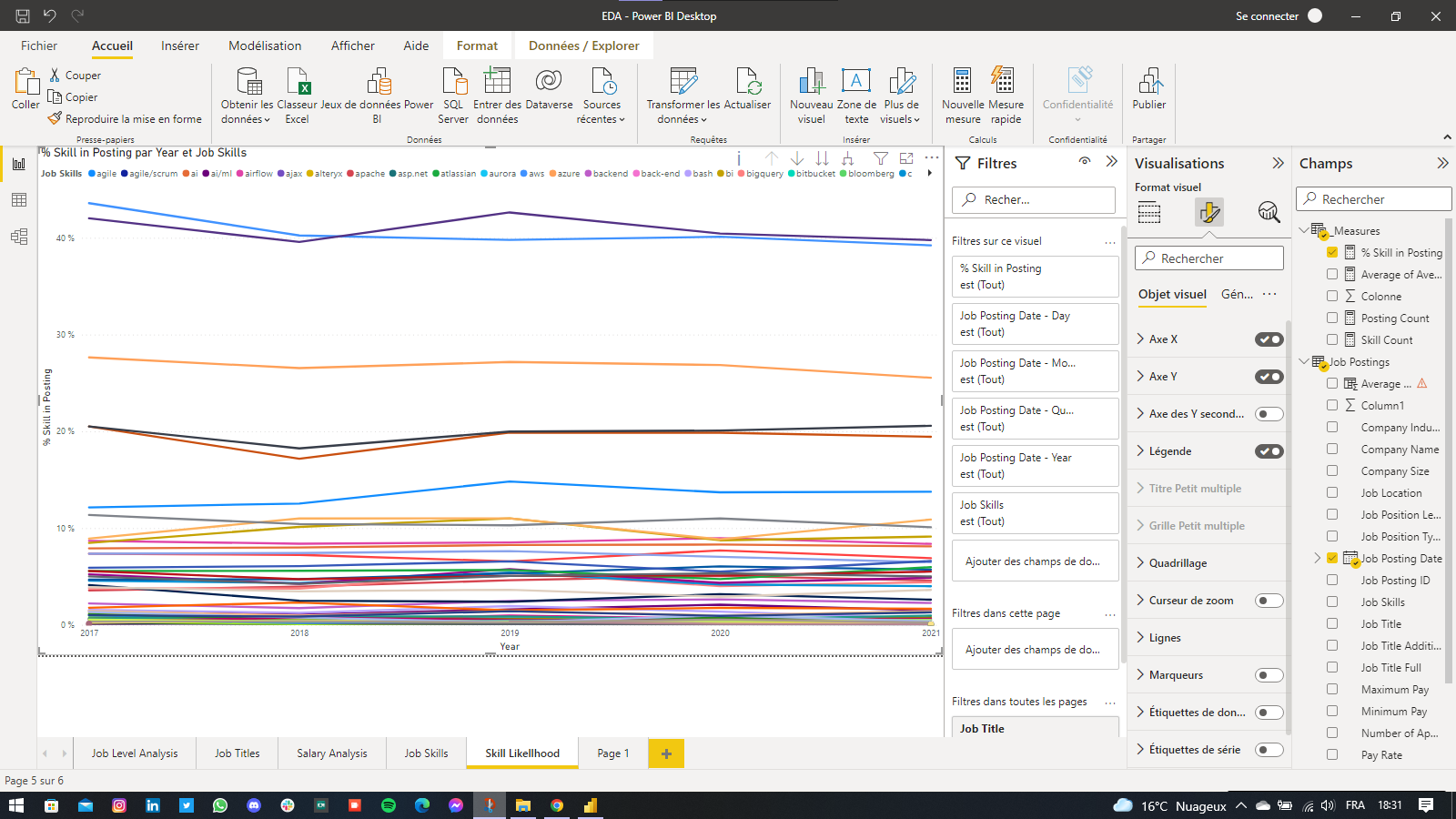


* On peut lire, par exemple, le pourcentage de probabilité que Power BI soit répertorié comme un skill dans un job posting de Data Scientist : 5.1 %

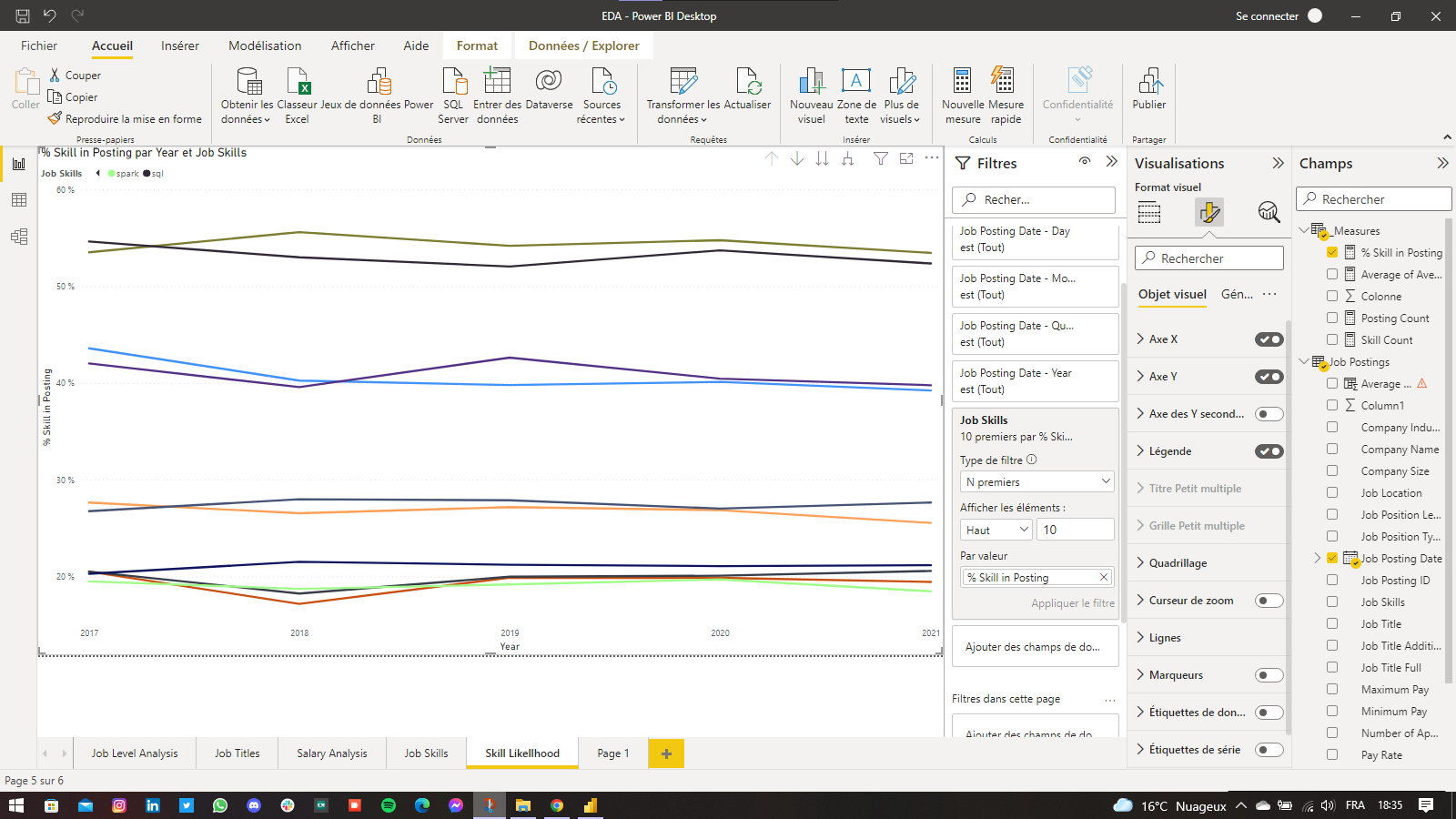
1. **Tendance des “Skills” au fil du temps**

On veut utiliser la nouvelle mesure en explorant l’évolution des Skills dans les offres d’emploi au fil du temps

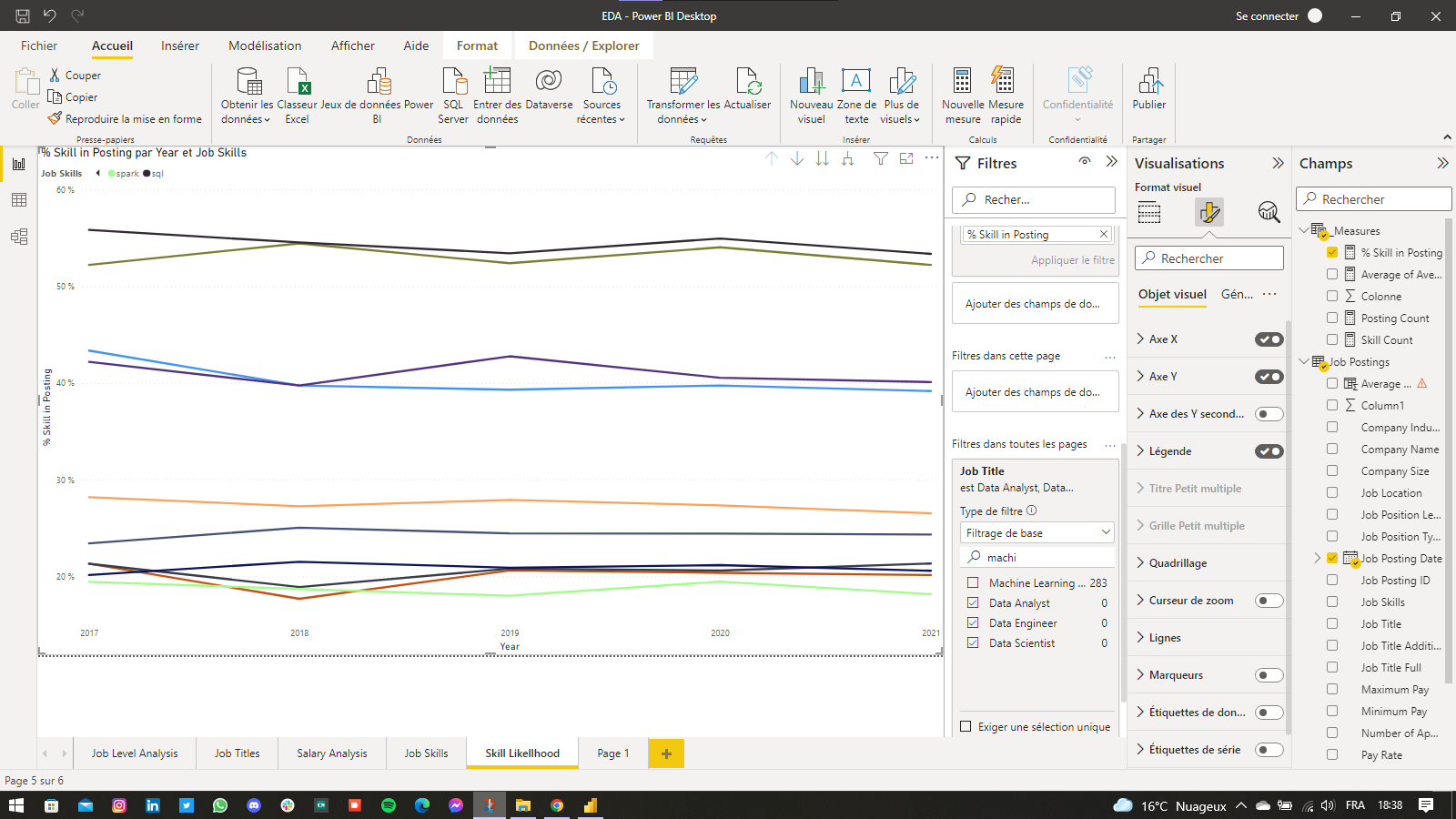
* Nettoyons la matrice de la dernière étape. Désactivons les sous-totaux des lignes.
* Création d’un graphique de série chronologique qui montre le “% Skill in Posting” au fil du temps :



* Cette visualisation est encombrée. Filtrons ce graphique pour n’afficher que les 10 meilleurs résultats pour “% Skill in Posting” :



* Pour les 3 principaux postes (Data Engineer, Data Analyst, Data Scientist), l’évolution des Skills les plus importants au fil du temps est maintenue à une demande constante :

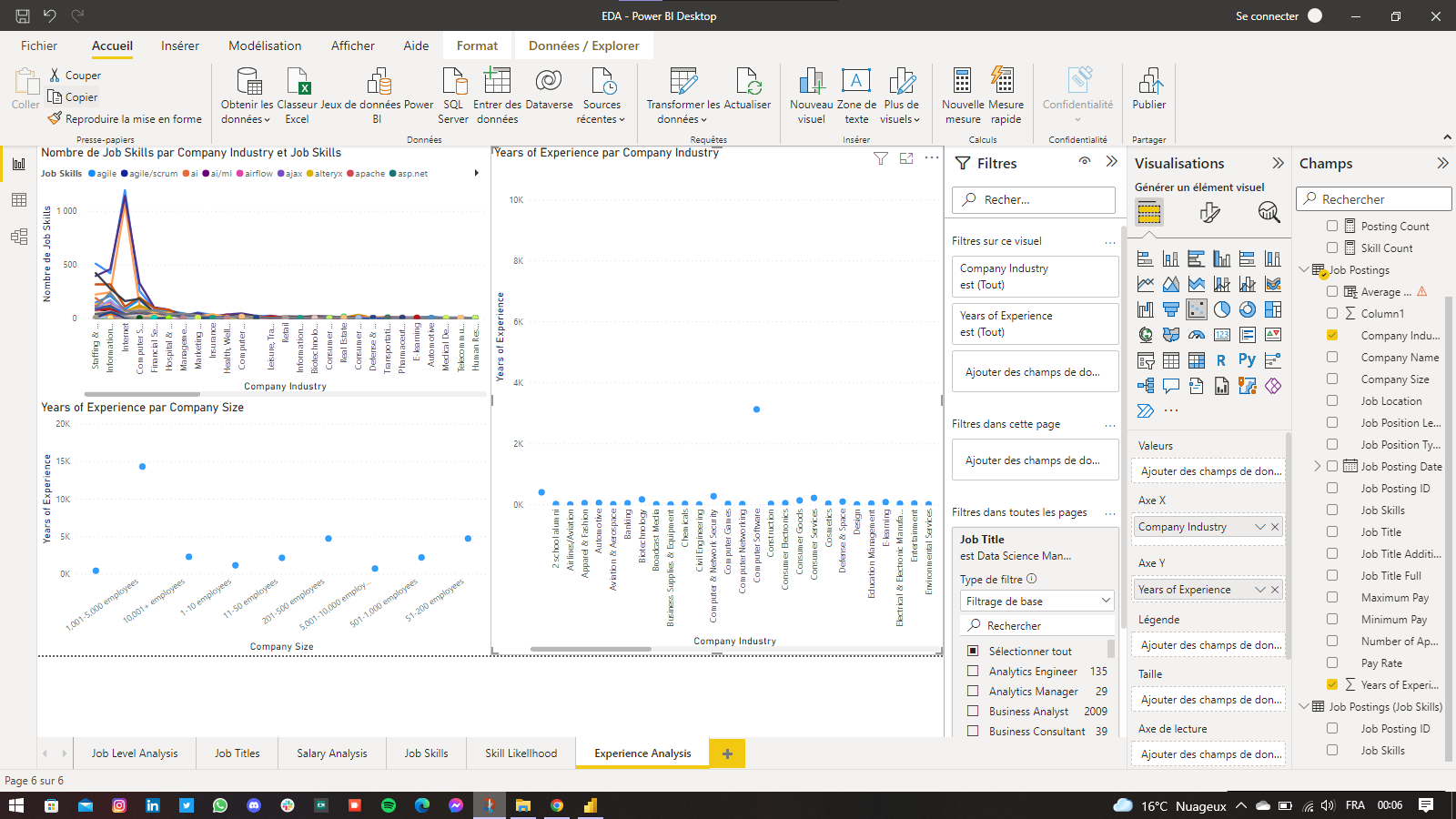


1. **Principales descriptions du poste**

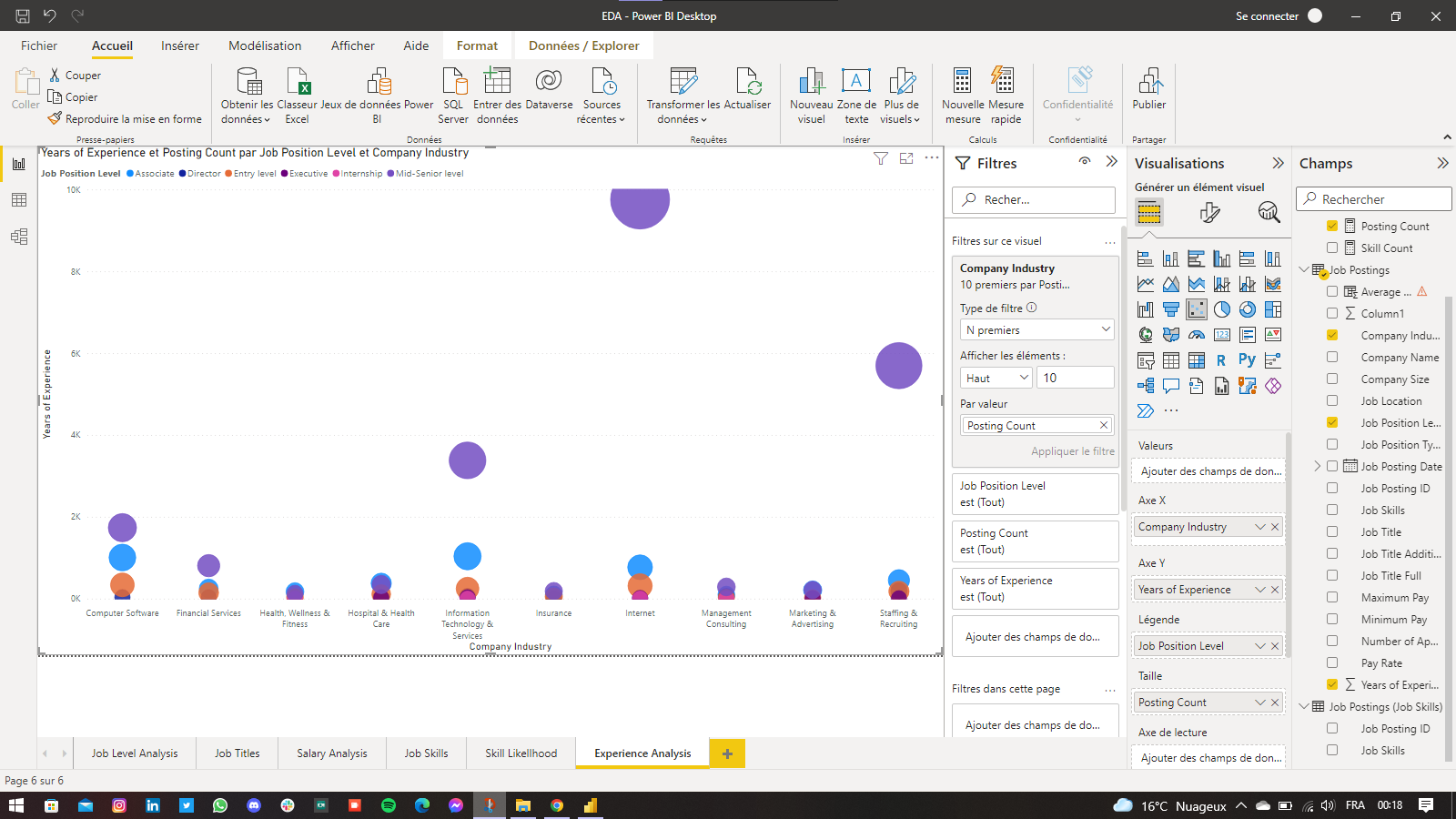
Nous voulons maintenant explorer les principales entreprises et industries à la recherche de postes en Data

Science.

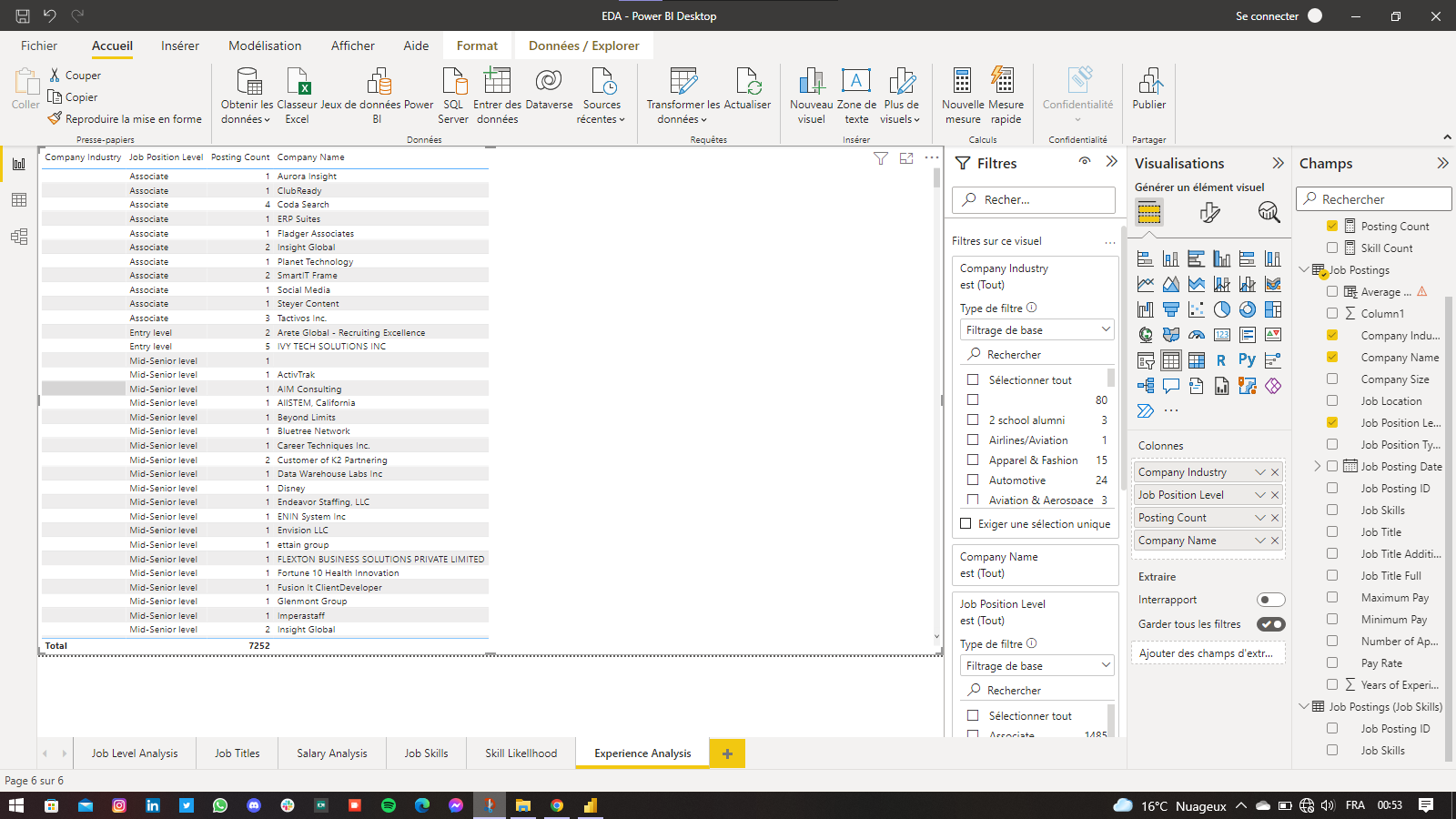
* Création de quelques visualisations :



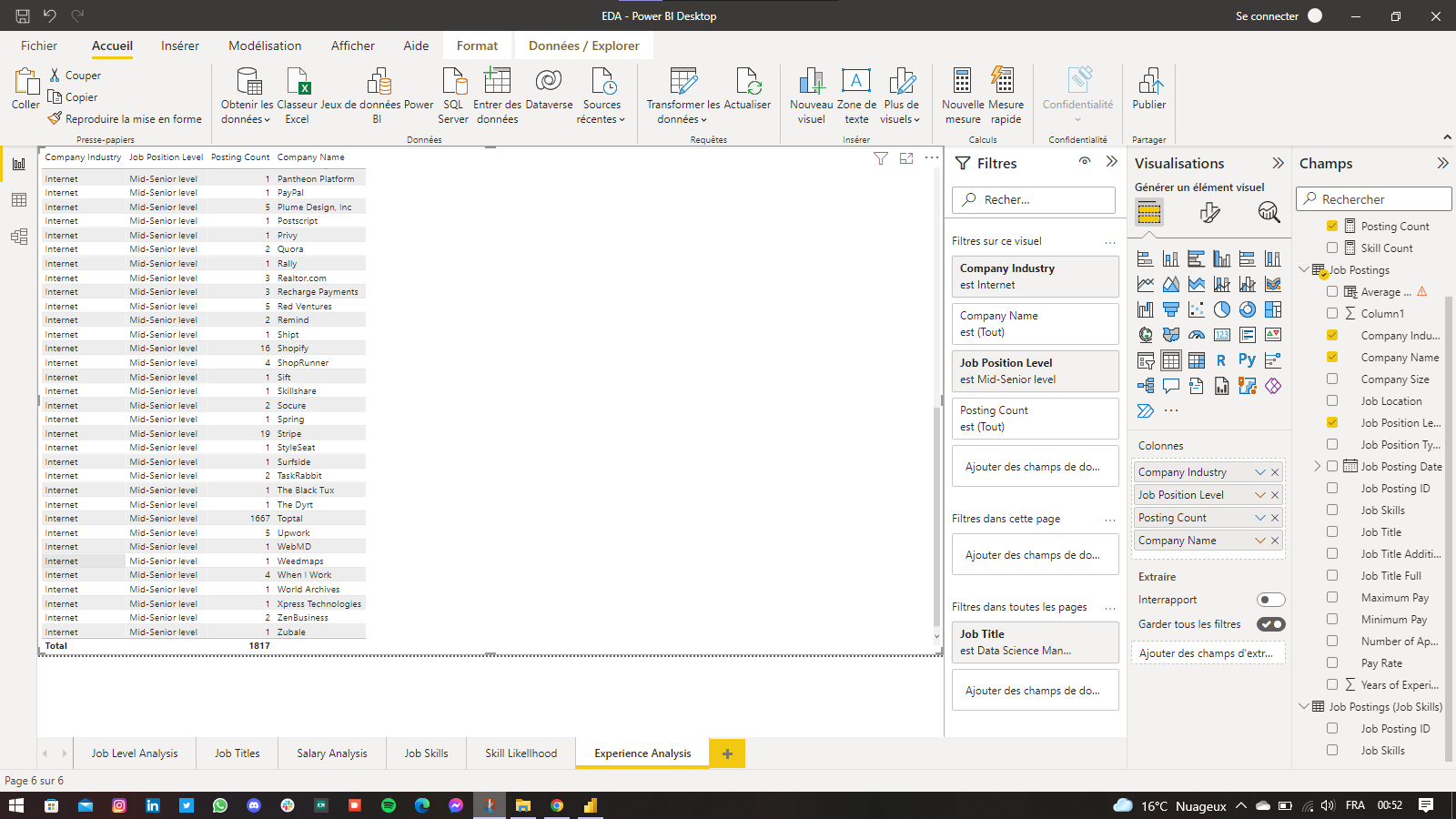
* Affinement du 3ème graphe davantage en présentant à la fois “Posting Count” et “Job Position Level” dans la visualisation (On filtre jusqu’au 10 principaux secteurs d’activité en fonction du nombre d’offres d’emploi) :



* Création d’un tableau pour compléter ce visuel en affichant les différents “Job Titles”, les entreprises, le nombre de “Job Postings” qu’elles ont et le “Company Name” :



* On peut appliquer un filtre, par exemple, pour voir quelle est l’entreprise qui recrute pour les postes “Mid-Senior level” dans l’industrie “Internet” :



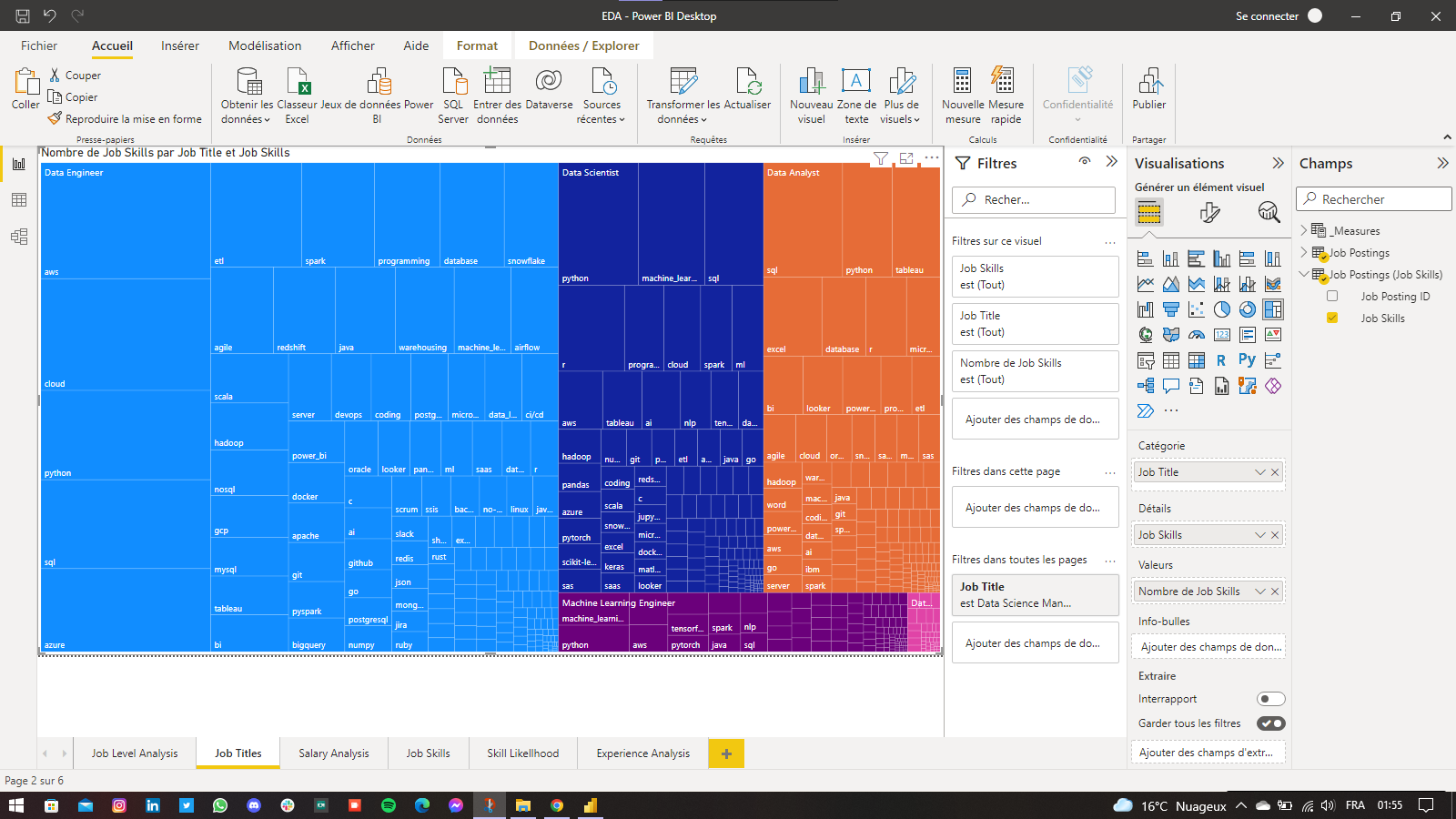
1. **Autres recommandations d’emploi**

Nous avons donc effectué une analyse très approfondie de nos 5 meilleurs “Job Titles” (Data Scientist, Data

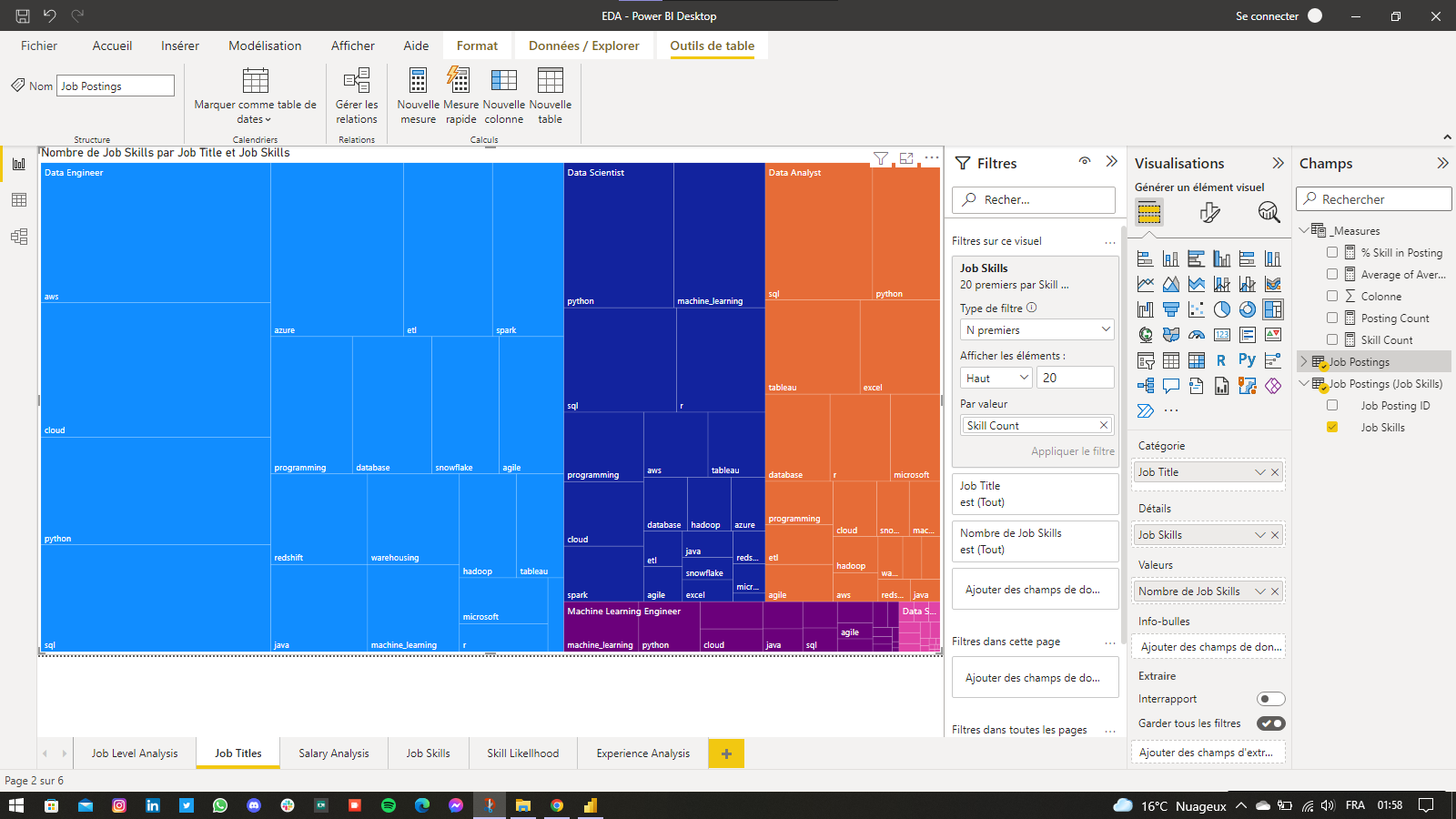
Engineer, Data Analyst, Data Science Manager, Machine Learning Engineer). Mais y a t-il d’autres “Job Titles”

que nous devrions également considérer et qui ont des exigences d’expérience similaires ?

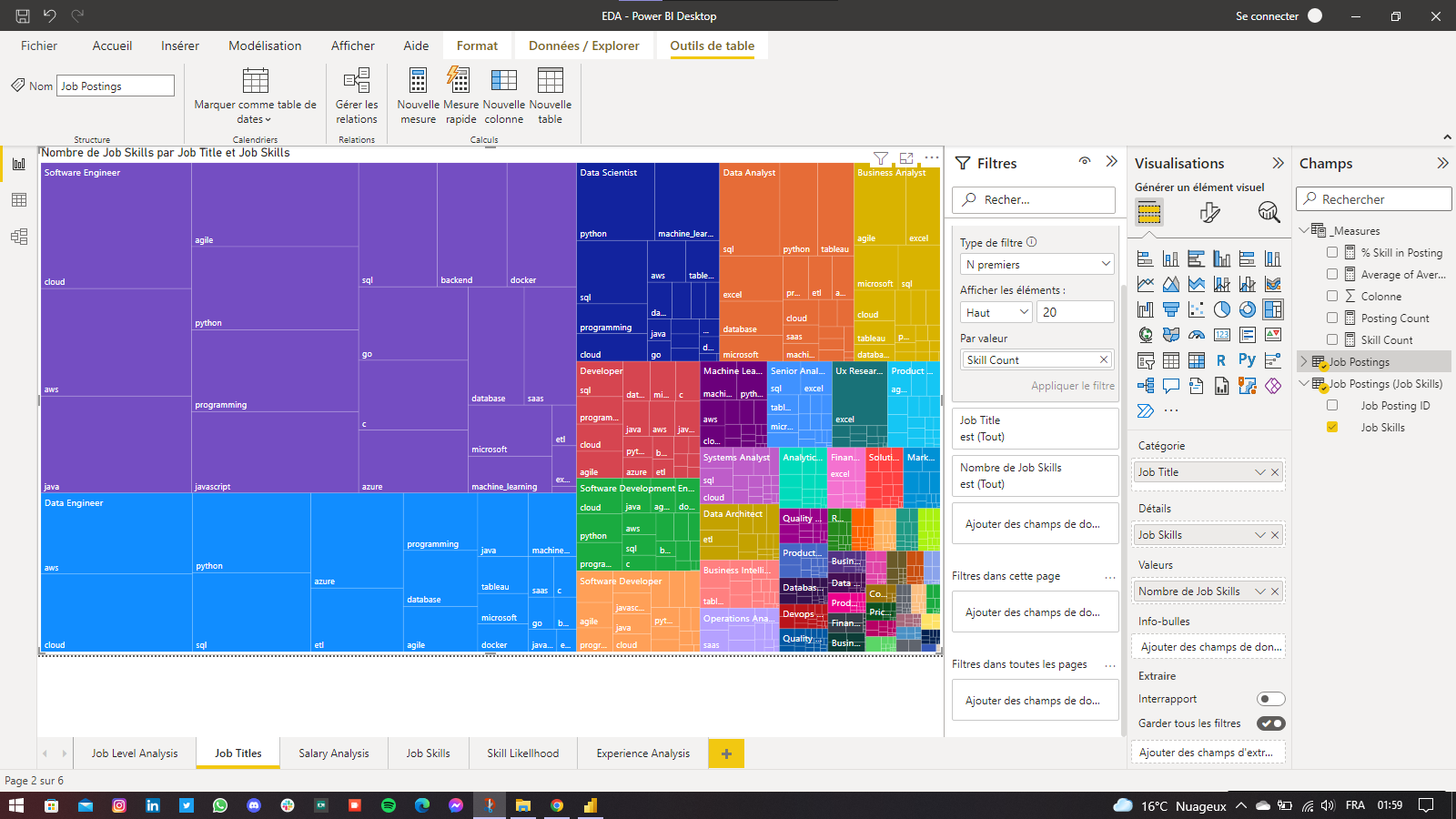
* On revient à la page “Job Titles” et mettre à jour la visualisation Treemap afin qu’elle affiche le nombre de compétences requises pour chacun des “Job Titles” :



* Ce n’est pas tous les Skills qui apparaissent sur ce Treemap. Filtrons cette visualisation afin d’afficher les 20 meilleurs Skills par “Skill Count” :



* Suppression des filtres sur toutes les pages afin que nous puissions rechercher d’autres “Job Titles” :



* On remarque, par exemple, que le Job qui ressemble le plus à un “Data Analyst” en ce qui concerne le type et le nombre de Skills requis est celui d’un “Business Analyst”. Ces 2 Jobs nécessitent des Skills en SQL, Excel, Tableau, Database et Cloud.

1. **Tendances issues de l’analyse (Interprétations)**

* Les “Software Engineers” et les “Data Engineers” ont des compétences très similaires.
* SQL est un Skill requis pour tous les Jobs en Data Science.
* Les industries basées sur la technologie avaient une demande plus élevée pour les postes en Data Science.
* La plupart des offres d’emploi ne provenaient pas de grandes entreprises.

1. **Création de Tableau de Bord**

Il est temps maintenant de rassembler toutes ces informations dans une histoire cohérente pour l’entreprise de recrutement. Commençons alors la création de notre Dashboard !

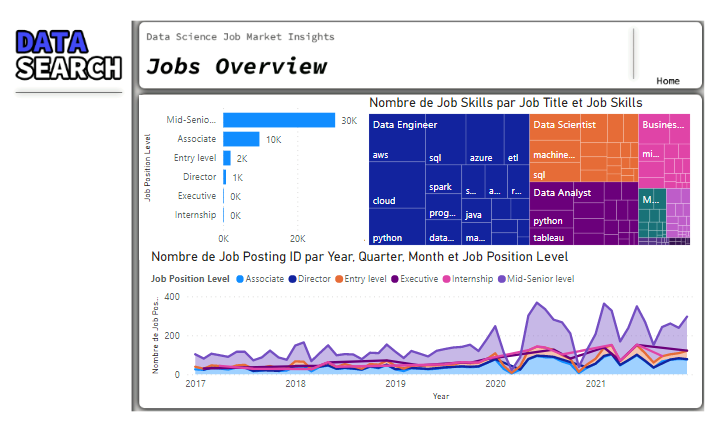
* Les recruteurs connaissent très bien ce domaine, donc ils comprendront toutes les mesures précédentes que nous avons utilisées et découvertes.
* L’objectif du recruteur est de mettre en relation ceux qui cherchent des postes avec des employeurs qui correspondent. Plus nous pouvons identifier facilement les domaines potentiels opportuns pour les candidats qualifiés, plus il sera facile pour un recruteur d’accomplir son travail.
* Nous devons transformer les informations de ces données en un formulaire utilisable pour les recruteurs. Ce sera le principal domaine d'intérêt lors de la création de ce Dashboard.

1. **Commencer les modèles de tableau de bord**

Créons les modèles de pages pour nos 2 différents tableaux de bord et la page d'accueil, nous

commencerons ensuite à ajouter des visuels à notre tableau de bord “Job”.

* Ajouter à nouveau le filtre de “Job Titles” sur toutes les pages pour sélectionner nos 5 métiers précédents ainsi que quelques autres : Data Scientist, Data Analyst, Data Engineer, Data Science Manager, Data Architect, Machine Learning Engineer, Business Analyst, Research Scientist.
* Créons des pages distinctes : Home, Jobs, Skills, Company en vue de la création de nos tableaux de bord.
* Ajout d’un background pour chaque page (Les images de background sont un excellent moyen de mieux organiser nos tableaux de bord).
* Copions les visuels graphiques de la page “Job Level Analysis” et “Job Titles” et collons-les dans la page “Jobs”.



* Quel “Job Position Level” pour les “Business Analyst” demande le plus “Excel” en tant que compétence ?



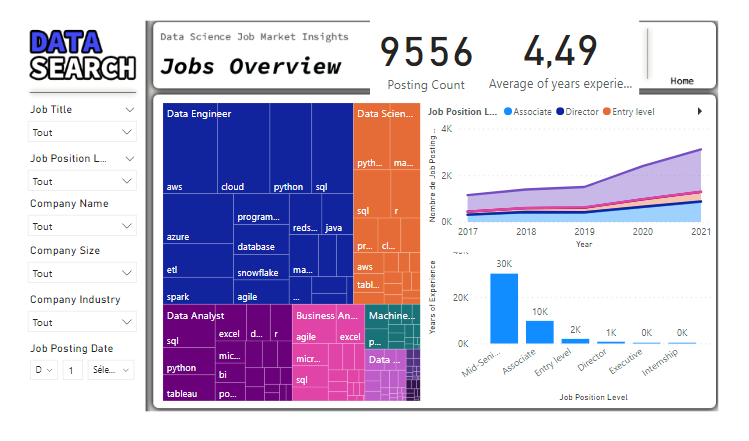
Réponse : “Associate” .

1. **Conception d’une mise en page de tableau de bord**

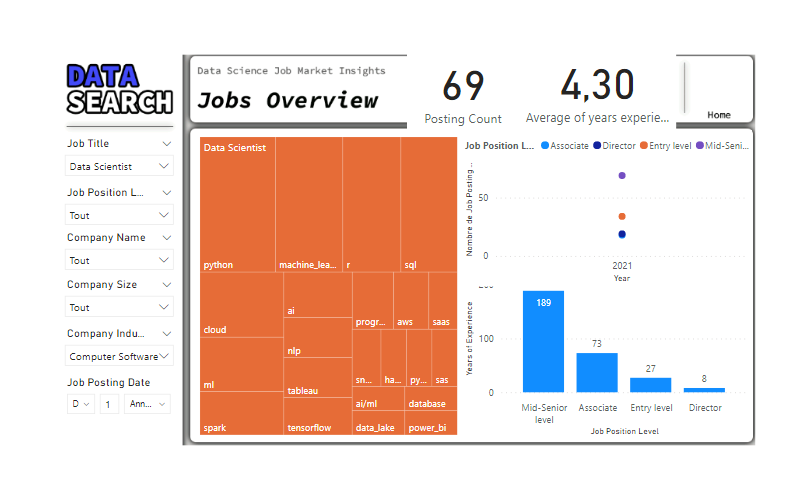
Términos la construction du tableau de bord “Jobs” en ajoutant des segments et des visuels de carte pour

les points de données clés.

* Sur la page “Jobs”, ajout des segments suivants dans la colonne de gauche.
* Copions ces segments et collons-les dans les pages “Skills” et “Company” avec synchronisation.
* Ajout de 2 cartes dans la case la plus en haut de la page “Jobs” affichant le nombre de job postings et la moyenne des années d’expérience.



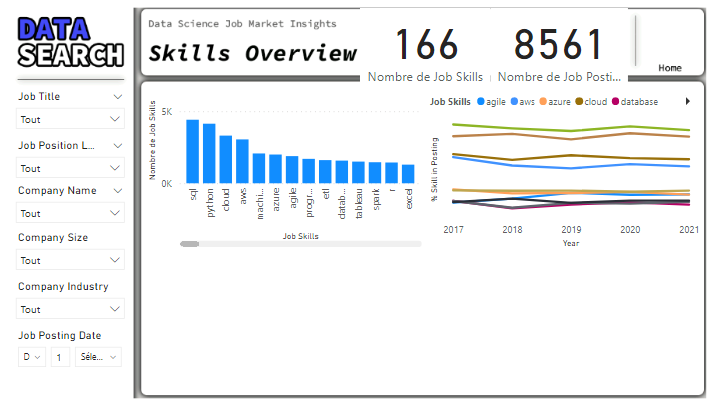
* En **“2021”**, par exemple, la moyenne d’années d’expérience des **“Data Scientists”** dans l’industrie des **“Computer Software”** est : 4.3



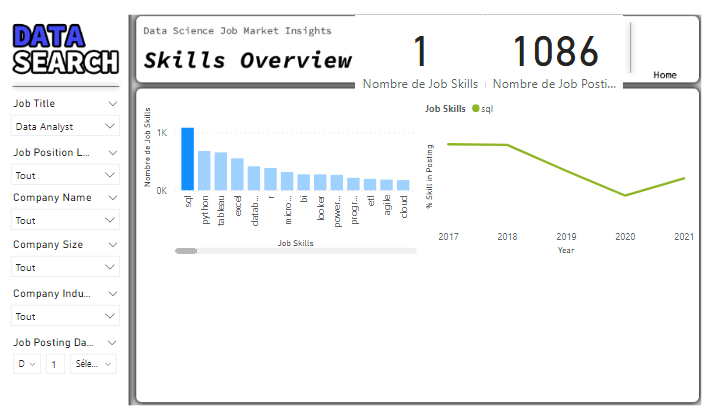
1. **Explorer les meilleurs Skills pour les emplois**

Construisons maintenant le tableau de bord “Skills”, nous utiliserons les données de la page “Job Skills”.

* Sur la page “Skills”, ajoutons les visuels graphiques des 2 pages “Job Skills” et “Skills Likelihood".
* Ajout d’une carte dans la case la plus haute pour un décompte distinct des Skills de la table “Job Skills”.
* Ajout d’une autre carte dans la case la plus haute pour un décompte distinct d’offres d’emploi qui ont un Skill correspondant.



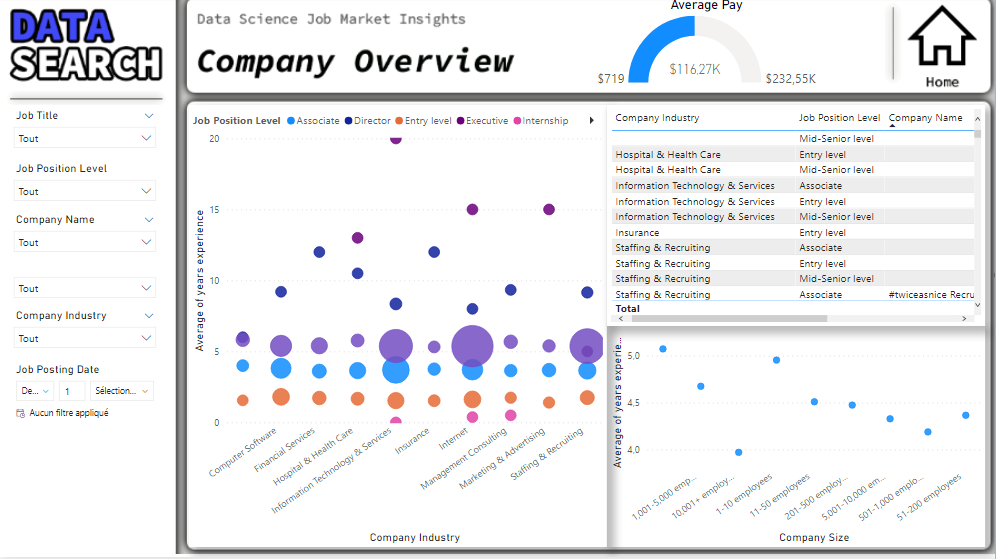
* Pour les **“Data Analyst”**, le nombre d’offres d’emploi qui incluent la compétence **“SQL”** est de : 1086



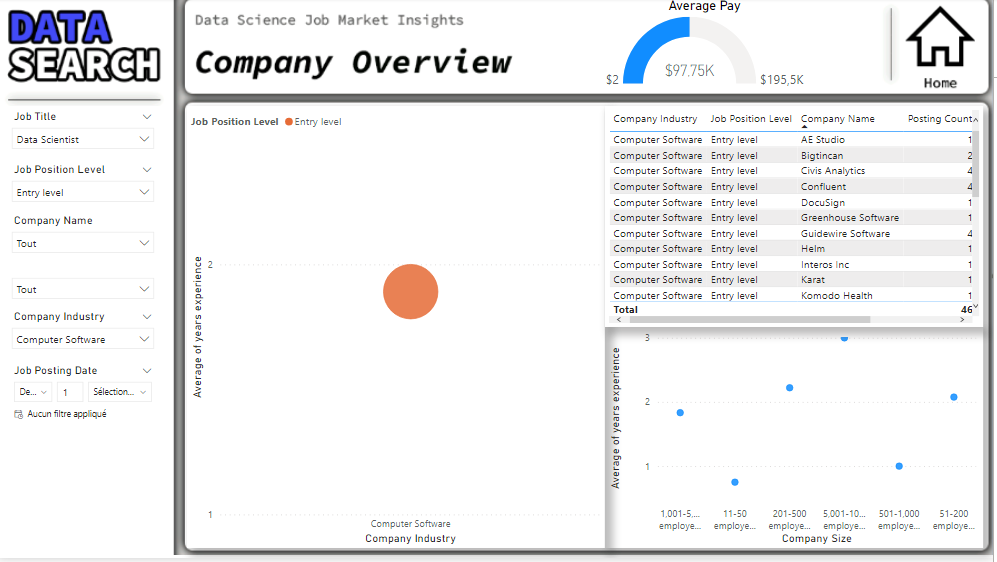
1. **Jauge salariale pour les entreprises**

Constituons notre troisième tableau de bord pour les informations sur l’entreprise.

* Sur la page “Company”, ajout des visuels graphiques de la page “Experience Analysis”.
* Ajout d’une Gauge dans la case plus haute pour afficher le salaire moyen.
* Formatage du visuel de Gauge pour inclure le salaire minimum et maximum



* Pour les **“Data Scientist”** débutants **(Entry Level)** dans l’industrie des **“Computer Software”**, leur salaire moyen est de : $97.75K



1. **Ajouter de l’interactivité avec des signets**

Enfin, ajoutons des icônes cliquables à notre page “Home” pour permettre à notre public cible de naviguer

plus rapidement.

