

Une enquête auprès des métiers de la « data » : quelle place pour R et ses utilisateurs ?

Antoine GIRARD, *data analyst* indépendant

Antoine.girard@datag.fr

Quelle place pour R en 2024, dans le large univers de la « data » ? Peut-on dire que son usage soit en progression ou en déclin, parmi tous ces professionnels de la donnée ? Quelles sont les spécificités des utilisateurs de R (secteurs d'activité, missions, caractéristiques socio-démographiques) par rapport aux autres outils d'exploitation des données ? Une enquête menée en 2024 – et deuxième édition après une première réalisée en 2021 – tente d'apporter quelques éclairages quantitatifs à ces questions.

Mots-clefs (3 à 5) : Data enquête outils

Développement

Dans le large univers de la « data », les outils et missions évoluent vite : les *data analyst*, *data scientist* ou *data engineer* ont à leur disposition une large gamme d'outils, d'Excel à Python en passant par SAS et Power BI. Il n'est toutefois pas évident de savoir dans quelle mesure chacun de ces outils sont utilisés, et par qui. Des chiffres sont parfois partagés par des plateformes sur les utilisateurs (comme *Kaggle* ou *Pylote*), mais ces plateformes ont aussi de très fort biais liés à leur positionnement. Même remarque pour les classements de popularité des langages de programmation, qui concernent l'informatique au sens large mais pas forcément le créneau plus spécifique des analystes, scientifiques et ingénieurs de la donnée.

C'est dans ce but qu'une enquête quantitative a été menée par ma modeste personne. Diffusée principalement via le réseau social professionnel *Linked* en Mars 2024 (diffusion du questionnaire auprès de professionnels de la data), elle se base sur plus de 300 réponses de profils *data* en tous genre. Cet échantillon ne peut pas être considéré comme parfaitement représentatif, dans la mesure où les répondants ont été contactés de façon « pseudo aléatoire » via la plateforme *Linked*. Hors, rien ne prouve que *Linked* soit représentatif de l'ensemble de la population visée. Tout en étant conscient de ce biais inhérent à la méthodologie, cette enquête a été menée dans le souci d'une représentation la plus fidèle possible de la réalité de ces métiers : ainsi des profils choisis « aléatoirement » sur la plateforme ont été contactés, et non pas uniquement via mon réseau personnel, forcément biaisé.

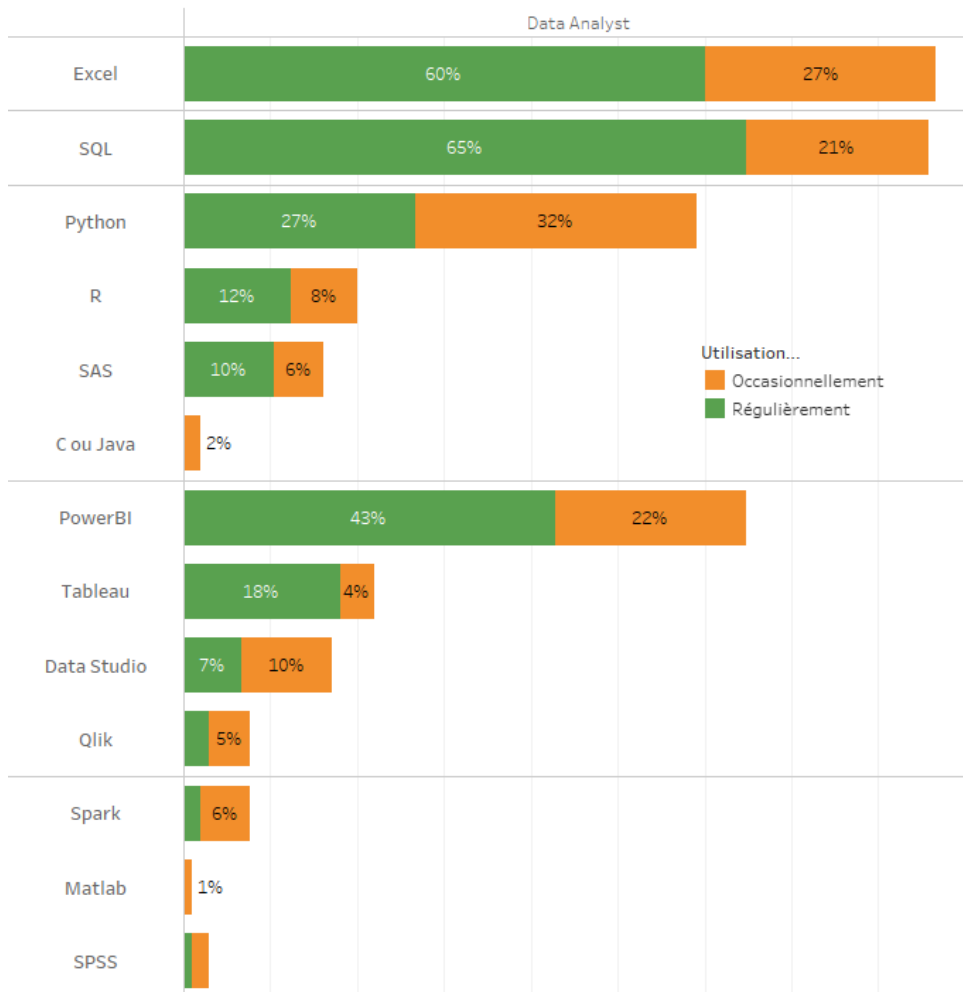
Les thématiques abordées par le questionnaire se limitent à trois grands domaines : les outils (connaissance et expertise), les missions (effectives et préférées) et le profil général (secteur d'activité, intitulé du poste, âge et genre). La volonté d'avoir un questionnaire court et simple (mais avec un bon taux de retours) explique ce nombre limité de thématiques.

Dans le cadre de cette présentation, les résultats seront évidemment orientés autour des utilisateurs du logiciel R : quel « poids » représentent-ils, notamment par rapport aux autres langages de programmation que sont SAS et Python ? Avec quels autres outils est-il le plus souvent utilisé ? Quelle évolution constate-t-on par rapport à la précédente enquête menée en 2021 ? Quelles sont les particularités de leurs profils, que ce soit sur leur mission ou leurs caractéristiques

*antoine.girard@datag.fr

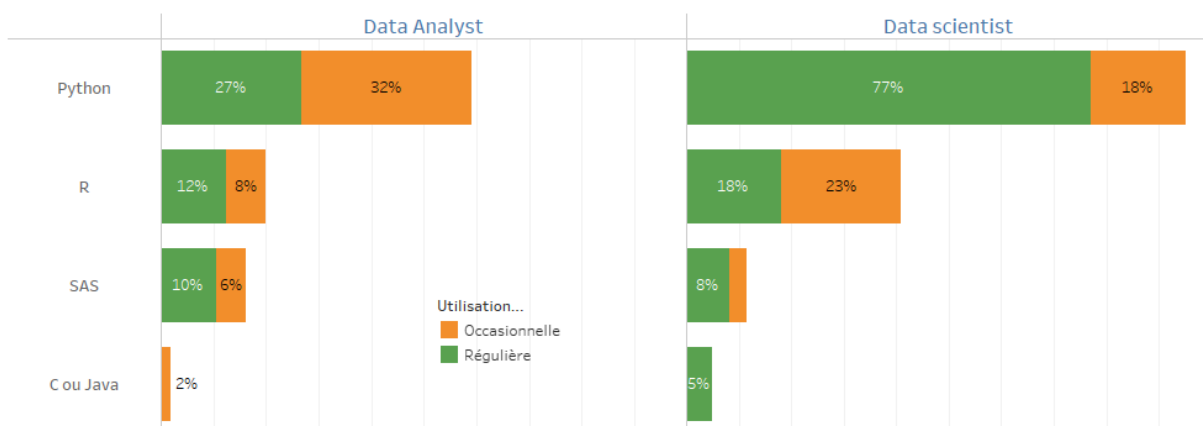
démographiques ? Autant de questions auxquelles cette enquête tâche d'apporter quelques réponses.

Les outils des *data analyst*



Enquête menée par Antoine Girard en Février 2024 auprès de 300 professionnels de la data

Outils de programmation utilisés par les *data analyst* et *scientists*



Enquête menée par Antoine Girard en Février 2024 auprès de 300 professionnels de la data

Références

Document de synthèse sur les principaux résultats de la précédente édition de l'enquête :

<https://public.tableau.com/app/profile/antoine.girard/viz/Synthseenquatedata2022/Synthseenquatedata2022>

Technologies utilisées par les freelances en informatique sur la plateforme *Pylote* :

<https://pylote.io/comparatif-competences-freelances-tech>

Les langages de programmation les plus populaires en 2023 selon l'*IEEE Computer society* :

<https://spectrum.ieee.org/the-top-programming-languages-2023>