

Une enquête sur les profils « data »

Antoine GIRARD

Data analyst freelance

06 33 72 87 12

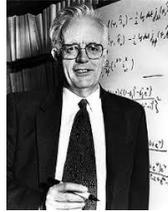
Antoine.girard@gamestatistics.fr

Juin 2022

Introduction

Statistiques

Statisticien / chargé d'études



2010



Data analyse / data science

Data analyst / data scientist

2022

Régression / modélisation



Analyse graphique

Data Viz



Machine learning

Qu'est-ce qui **caractérise** aujourd'hui un *data analyst* ou un *data scientist* ?

Qu'est-ce qui les **différencie** ?

Quelle est la réalité de leurs **profils**, **missions** et **outils** respectifs ?

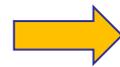
Méthodologie

- I. **Méthodologie**
- II. **Résultats**
- III. **Synthèse & conclusion**

Une enquête quantitative par sondage

indépendante et sans parti pris

Cible



métiers **opérationnels** de la **data**

- *Data analyst, data scientists, statisticiens, chargés d'études...*
- Professionnels en activité

Méthode



sondage **empirique**

- Répondants sur base volontaire

Diffusion du questionnaire

Diffusion du questionnaire exclusivement via **LinkedIn**

Novembre / Décembre 2021



Biais de sélection

de par l'algorithme de LinkedIn

Ciblage « manuel » de profils data

Invitations « aléatoires »

Questionnaire

Un questionnaire court et simple

Objectif : favoriser le taux de réponse



Profil

Age – genre

Niveau études



Outils

Connaissance – expertise

Fréquence d'utilisation



Missions

Missions effectives

Missions préférées

Structure d'échantillon

Recueil d'un échantillon de 150 réponses complètes

- *Diversité de profils et de secteurs* -

Axes d'analyses possibles

Data analyst vs data scientists

Par genre et tranche d'âge

Regroupements

- ✓ Chargés d'études associés aux *data analyst*
- ✓ Statisticiens associés aux *data scientist*

Exclusions

- ✓ Profils « informatique » ou « atypiques »
- ✓ Etudiants (hors alternance / stage) et inactifs

Résultats

I. Méthodologie

II. Résultats

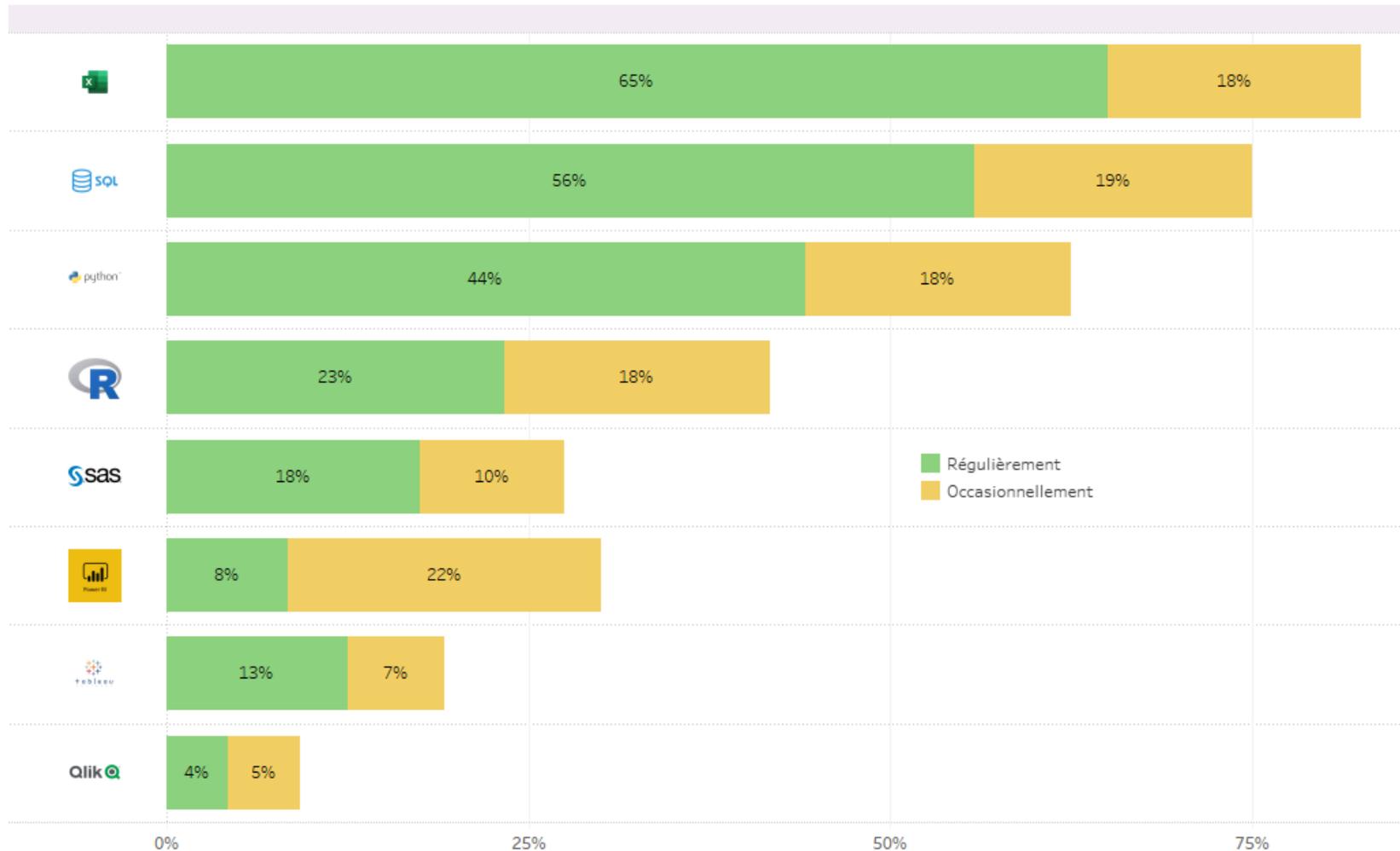
III. Synthèse & conclusion

Outils : SQL et Excel, deux indémodables

Fréquence utilisation outils

N=120

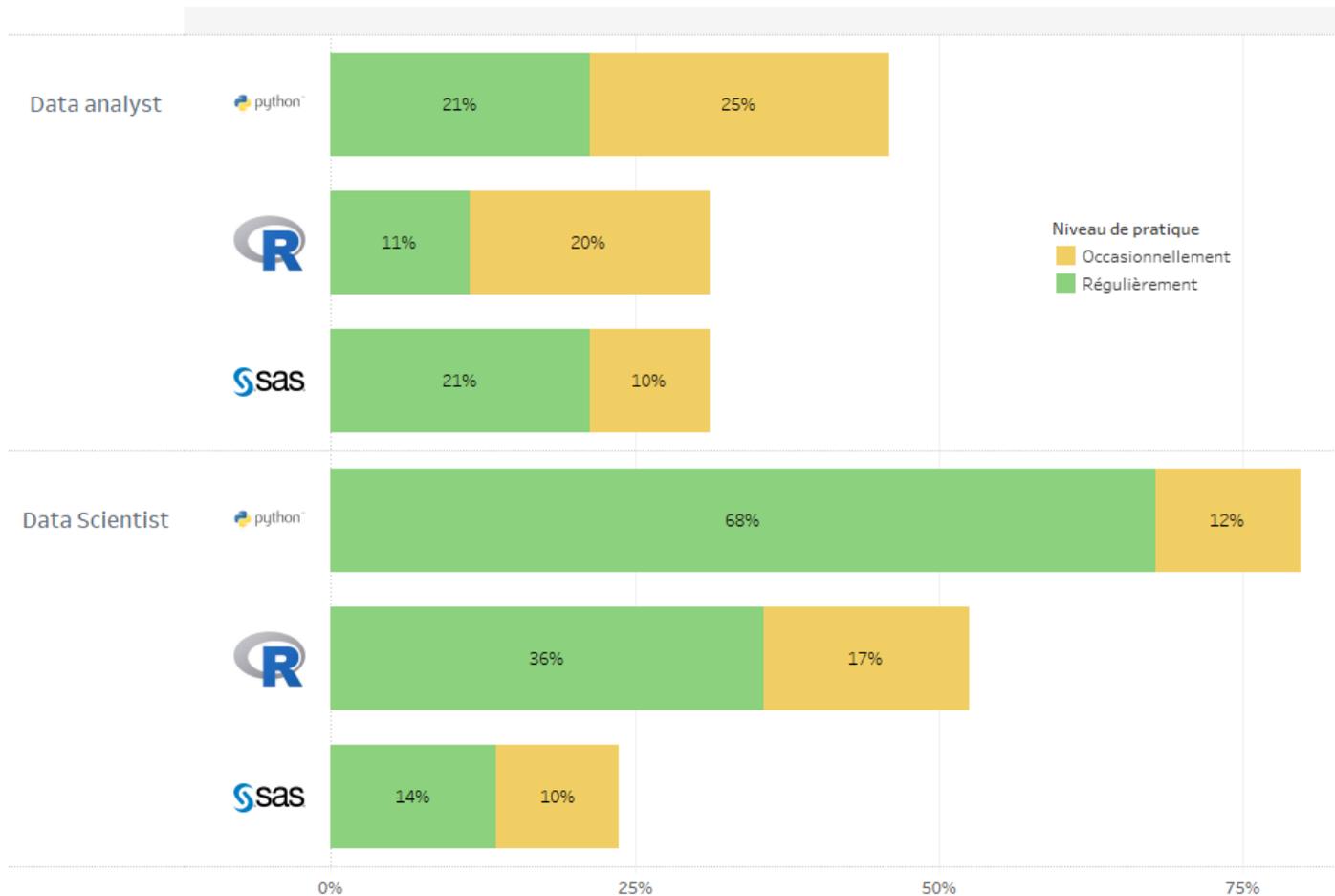
Pour chacun de ces outils que vous avez déjà utilisés, indiquez désormais si vous l'utilisez régulièrement, occasionnellement ou jamais dans votre poste actuel.



Programmation : Python prisé des data scientists

Fréquence d'utilisation des langages de programmation

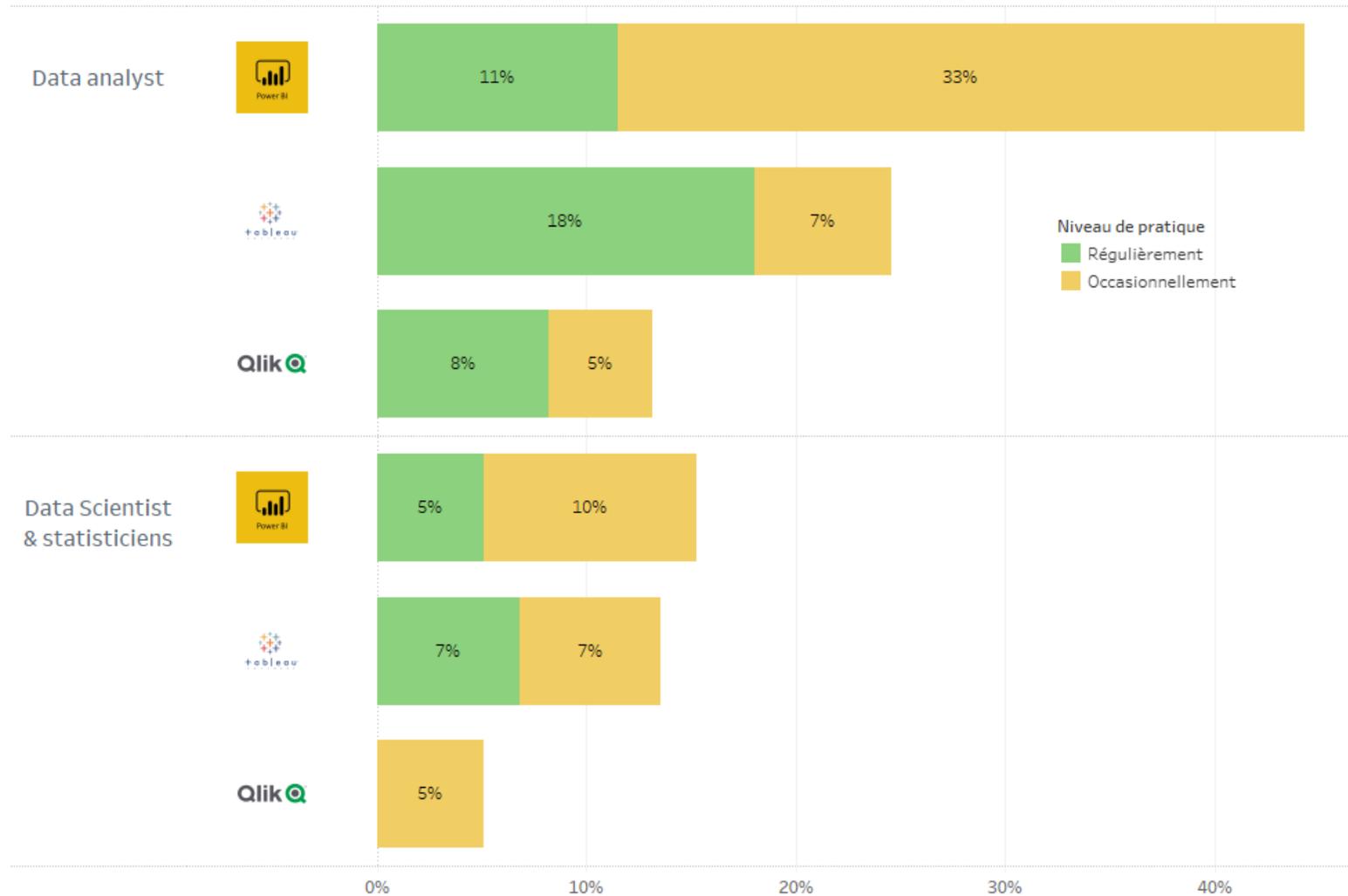
Pour chacun de ces outils que vous avez déjà utilisés, indiquez désormais si vous l'utilisez régulièrement, occasionnellement ou jamais dans votre poste actuel.



DataViz : des outils du data analyst

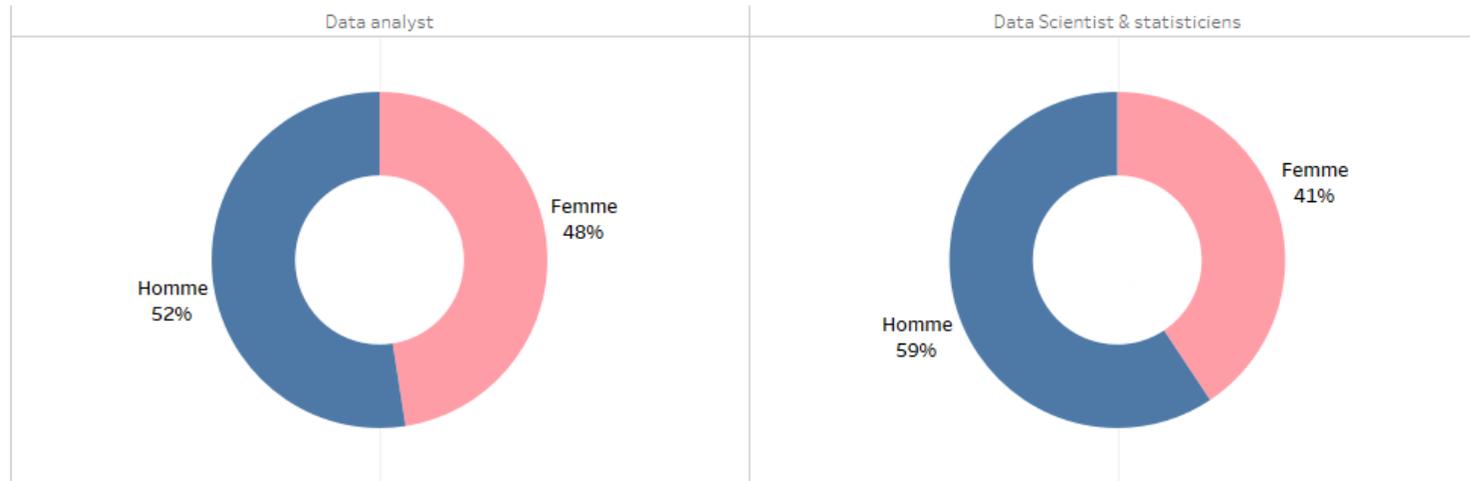
Fréquence d'utilisation des outils de dataviz

Pour chacun de ces outils que vous avez déjà utilisés, indiquez désormais si vous l'utilisez régulièrement, occasionnellement ou jamais dans votre poste actuel.

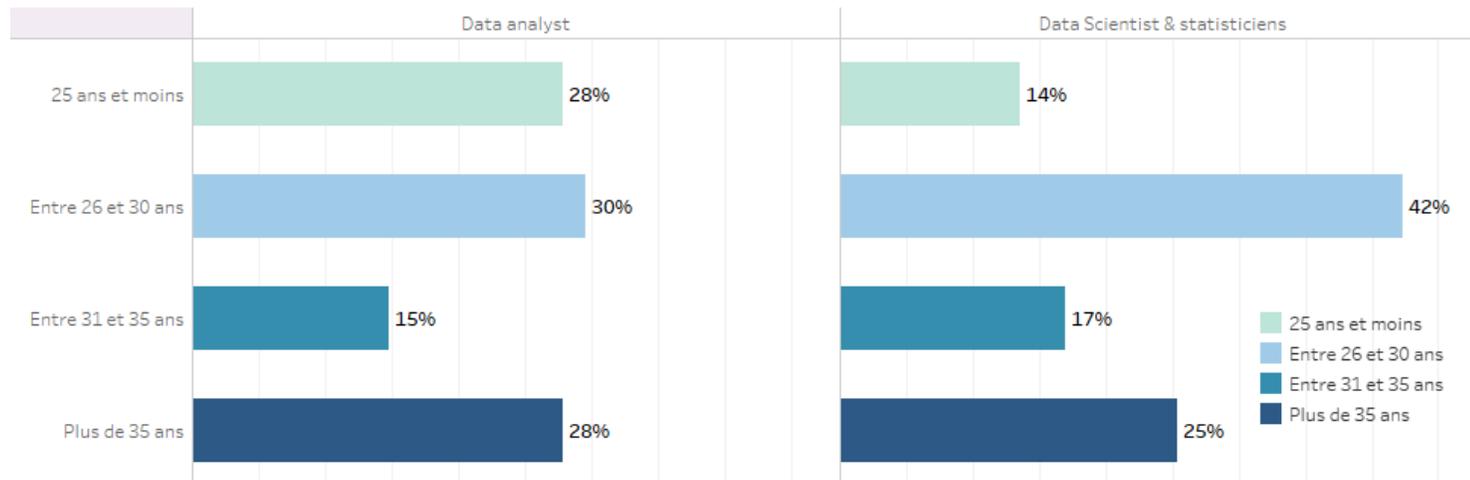


Des profils démographiques similaires

Répartition hommes / femmes



Répartition par tranche d'âge

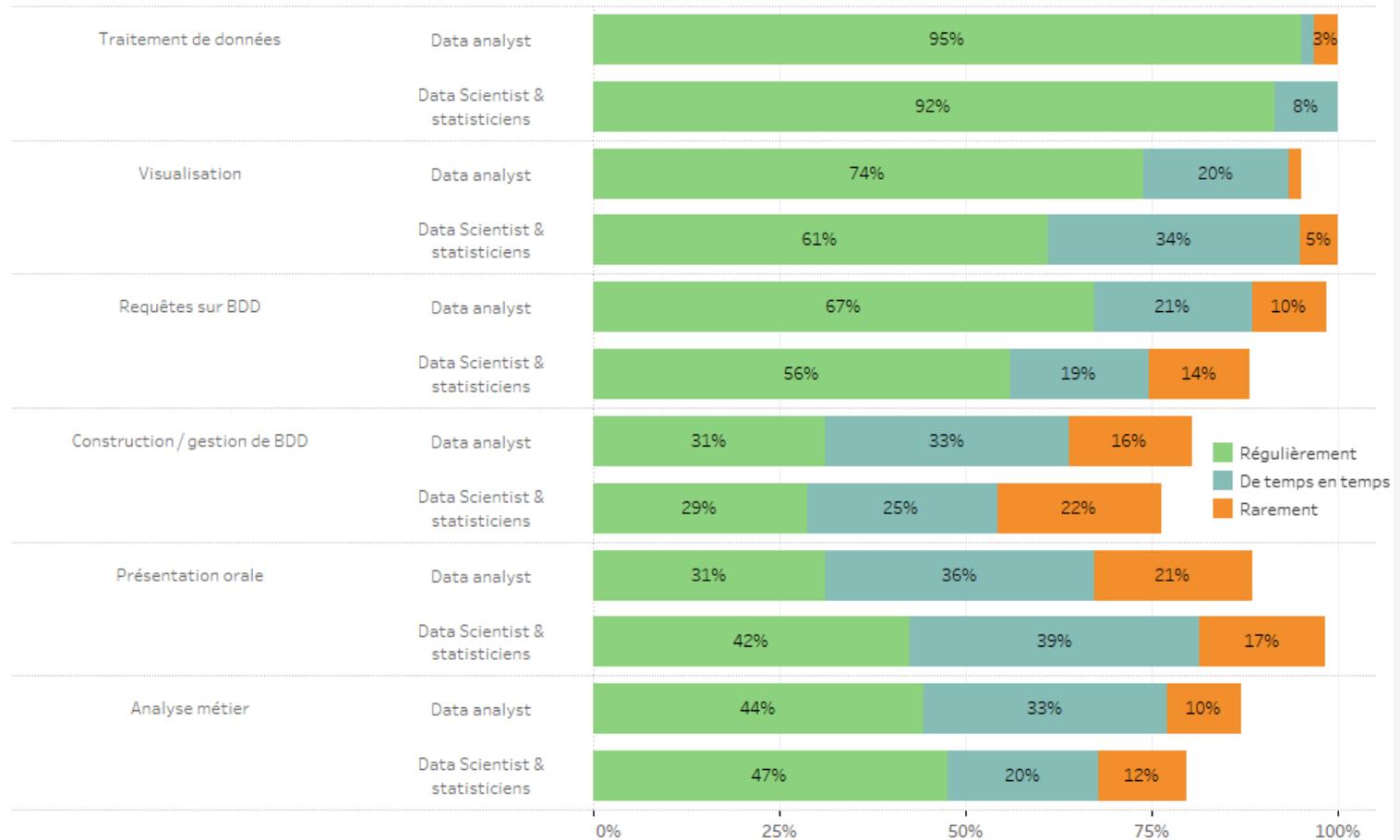


Enquête quantitative par enquête menée en Novembre 2021 par [Antoine Girard](#), sur un échantillon de 150 répondants volontaires "data" contactés via la plateforme Linked.

Beaucoup de missions communes DA vs DS ...

Tâches similaires entre data analyst et data scientists

Pour chacune de ces types de tâches liées à des profils « data », veuillez indiquer s'il s'agit de quelque chose que vous faites régulièrement, de temps en temps, rarement ou jamais dans votre poste actuel.

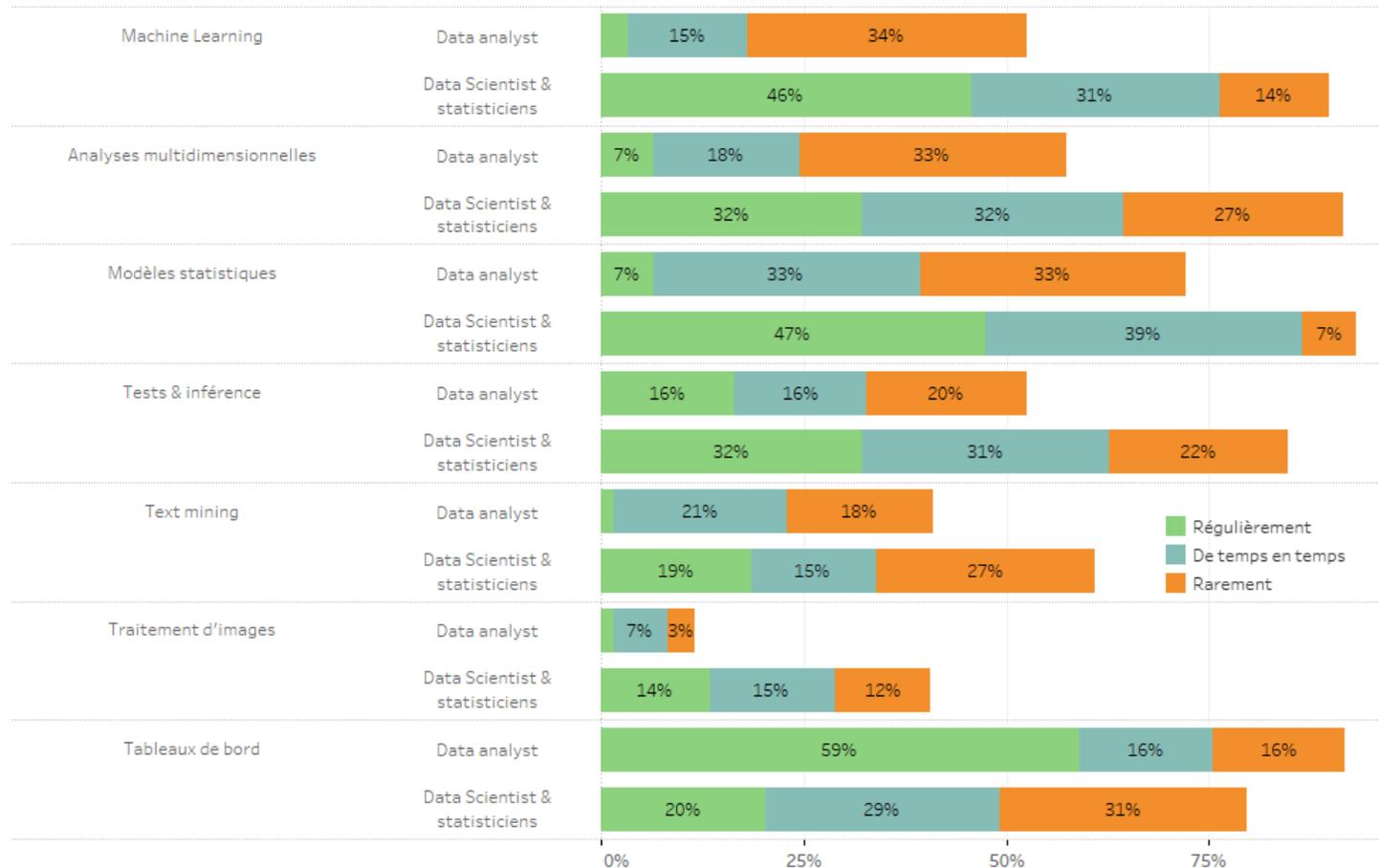


Enquête quantitative par enquête menée en Novembre 2021 par [Antoine Girard](#), sur un échantillon de 150 répondants volontaires "data" contactés via la plateforme Linked.

... mais des spécificités propres aux DS et DA

Tâches différenciées entre data analyst et data scientist

Pour chacune de ces types de tâches liées à des profils « data », veuillez indiquer s'il s'agit de quelque chose que vous faites régulièrement, de temps en temps, rarement ou jamais dans votre poste actuel.

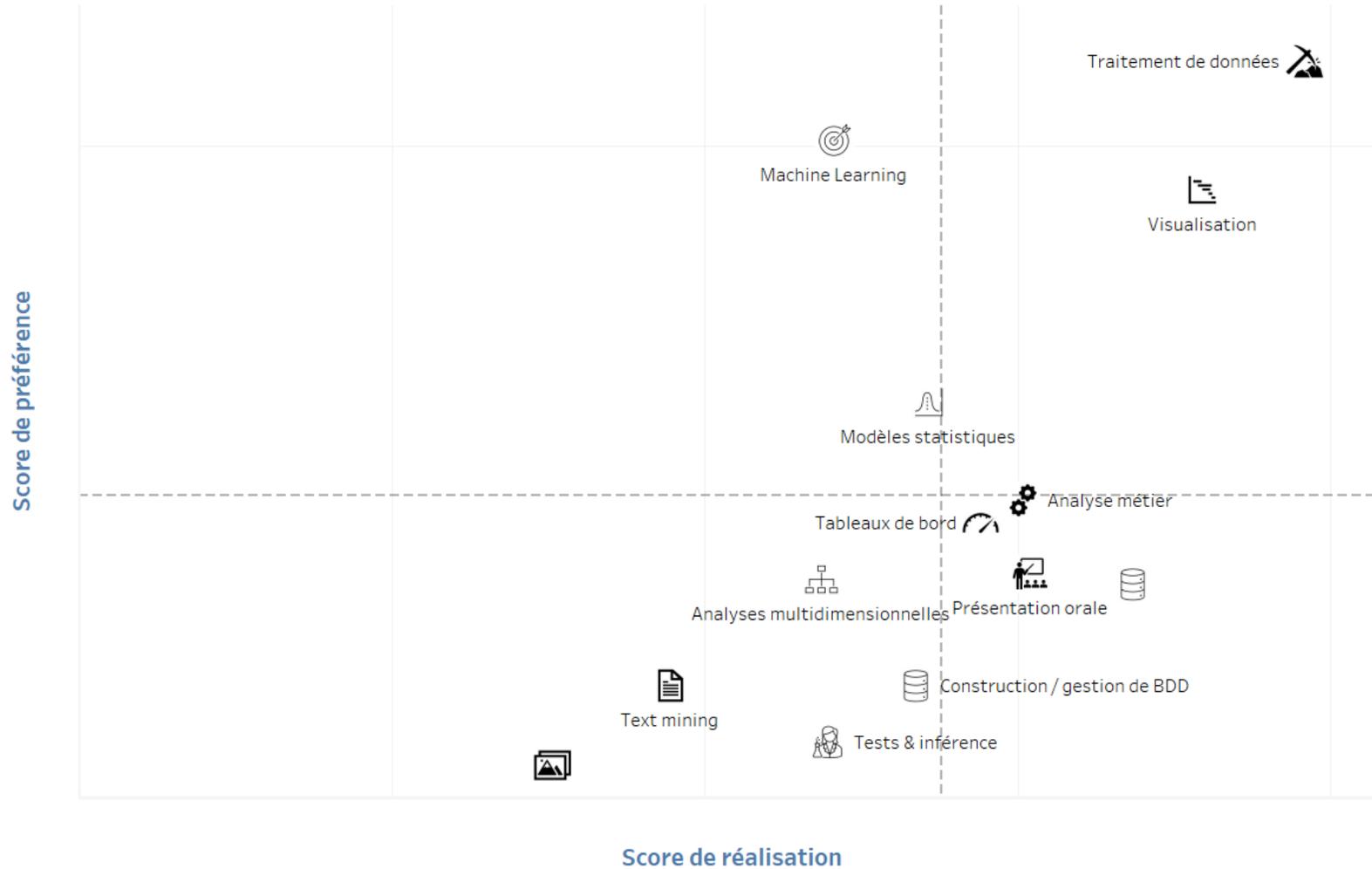


Enquête quantitative par enquête menée en Novembre 2021 par [Antoine Girard](#), sur un échantillon de 150 répondants volontaires "data" contactés via la plateforme Linked.

Synthèse tâches préférence / effectives

Taches préférées vs réalisées

Tout - Tout - 30 ans et moins et Plus de 30 ans -



Synthèse & conclusion

I. Méthodologie

II. Résultats

III. Synthèse & conclusion

Synthèse profil



→ Des métiers jeunes, plutôt paritaires

→ Peu de différence démographique entre *analyst* et *scientist*
Des *scientist* un peu masculins, des différences d'âge non-significatives

→ Une très large majorité de diplômés Bac +5
... et une minorité de doctorants pour les *scientists*

Synthèse outils



Excel et SQL, deux outils « indémoudables »

Très largement utilisés au quotidien tout autant par les *scientists* que les *analysts*



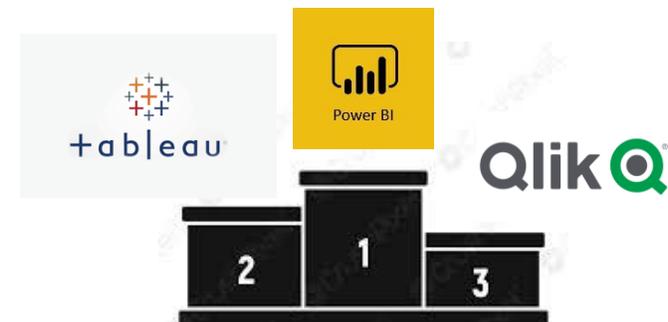
Python devance R et SAS, qui restent très utilisés

Un écart surtout marqué chez les *scientists*, où Python est l'outil de travail prédominant



Des outils de dataviz / BI très répandus également

... mais assez peu prisés des *scientists*, qui ont quelques outils spécifiques minoritaires



Synthèse missions



→ Manipuler, visualiser et extraire des données: les **tâches systématiques**

Analyst comme *scientists*, les mêmes combats du quotidien

→ Une nette différenciation de missions spécifiques ***analyst vs scientists***

Analysts : spécificité des tableaux de bord, et (un peu) plus de visualisation et de requêtes

Scientists : nette prédominance pour ce qui a trait aux algorithmes et aux statistiques

→ Des spécificités qui se retrouvent dans les préférences de missions

Les *analyst* aiment visualiser, les *scientist* prédire et modéliser...

Un intérêt plus modéré pour les problématiques métier (?) et liées aux bases de données

Synthèse globale

Qu'est-ce qui **caractérise** aujourd'hui un *data analyst* ou un *data scientist* ?

Qu'est-ce qui les **différencie** ?

Quelle est la réalité de leurs **profils**, **missions** et **outils** respectifs ?

→ Des profils démographiques proches

→ Une large étendue de missions et outils communs

→ Une différenciation des *scientists* marquée sur les statistiques et algorithmes, et l'usage plus généralisé de Python

Mise à disposition des résultats

- **Données brutes et traitements** :

- Fichiers csv et scripts R de traitements de données
- Lien : https://github.com/antoine1833/Enquete_data



- **Visualisation / exploration des résultats** :

- « Histoire » publiée sous Tableau Public
- Lien : <https://public.tableau.com/app/profile/antoine.girard/viz/Synthseenquedata2022/Synthseenquedata2022>



Une suggestion de pistes à creuser ? Je prends !

(ou vous faites...)

Une enquête à renouveler ?

- Une enquête à **renouveler à l'avenir** ?
- Méthodologie et questionnaire à **faire évoluer**
 - Thématiques : salaires ? Satisfaction au quotidien ? Perspectives d'évolution.... ?
 - Méthodologie : autres sources de diffusion ?

Des suggestions ou critiques d'ordre méthodologique / analytique ?

Je prends note !