

Une enquête sur les profils « data »

Décembre 2024



Antoine GIRARD

Data analyst freelance

06 33 72 87 12

Antoine.girard@datag.fr

Introduction



- **Data Analyst *freelance***

- **Indépendant** (*même si utilisateur de R et Tableau essentiellement !*)

- Un parcours essentiellement porté sur l'**analyse de données** et les **statistiques** dans le marketing, avec une **spécialité « enquêtes »**



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. **Principaux résultats**
- III. **Pour aller plus loin**

Une enquête quantitative par sondage

indépendante et sans parti pris...

Cible



Métiers **opérationnels** de la **data**

- *Data Analyst, data Scientists, data Engineer*
- Professionnels en activité

Méthode



Sondage **empirique**

- 311 réponses complètes récoltées

Diffusion du questionnaire

Diffusion du questionnaire exclusivement via **LinkedIn**

Mars / Avril 2024



Biais de sélection

de par l'algorithme de LinkedIn

Ciblage « manuel » de profils data

Invitations « aléatoires »

La deuxième édition d'une enquête déjà effectuée en 2021

Objectif : suivre les évolutions de tendance

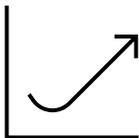
Enquête data 2021

(~150 répondants)



Enquête data 2024

(~ 310 répondants)



- Méthodologie identique
- Questionnaire quasi-similaire
- Plus de réponses en 2024, permettant de mieux cibler par métier / secteur

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

+ points spécifiques enquête 2024 :

- *rapport au télétravail*
- *satisfaction professionnelle*
- *utilisation de l'IA*

Structure d'échantillon (1)

Recueil d'un échantillon de 310 réponses complètes

- *Diversité de profils et de secteurs* -

Axes d'analyse

- **Evolution** vs enquête 2021
- Comparatifs par **métier** : DE / DA / DS

Exclusions

- ❖ Profils data très « atypiques »
- ❖ Etudiants (hors alternance / stage) et inactifs

Acronymes utilisés :

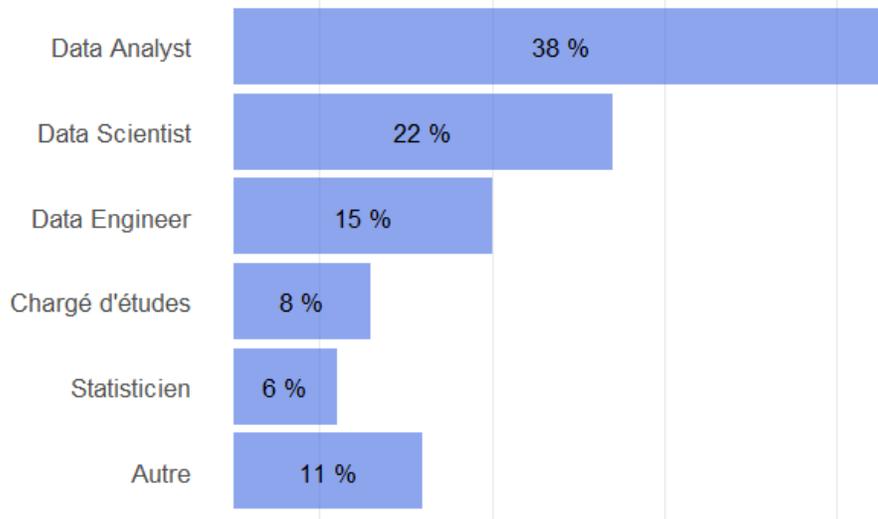
DA : *Data Analyst*

DS : *Data Scientist*

DE : *Data Engineer*

Structure d'échantillon (2)

Répartition des répondants par métier



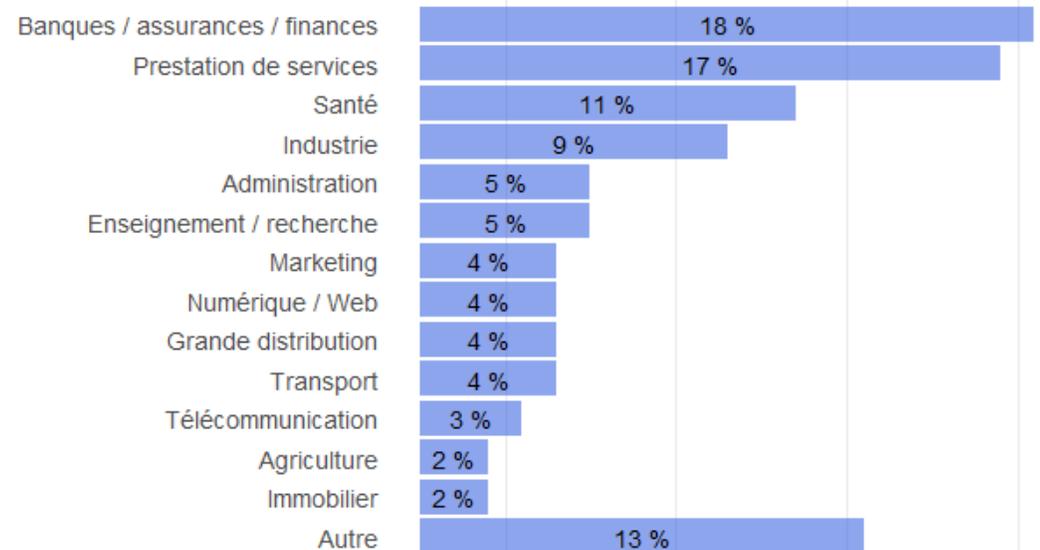
Une majorité de *data analyst* dans l'échantillon

... mais aussi une quantité non-négligeable de DS / DE

Une grande diversité de secteurs

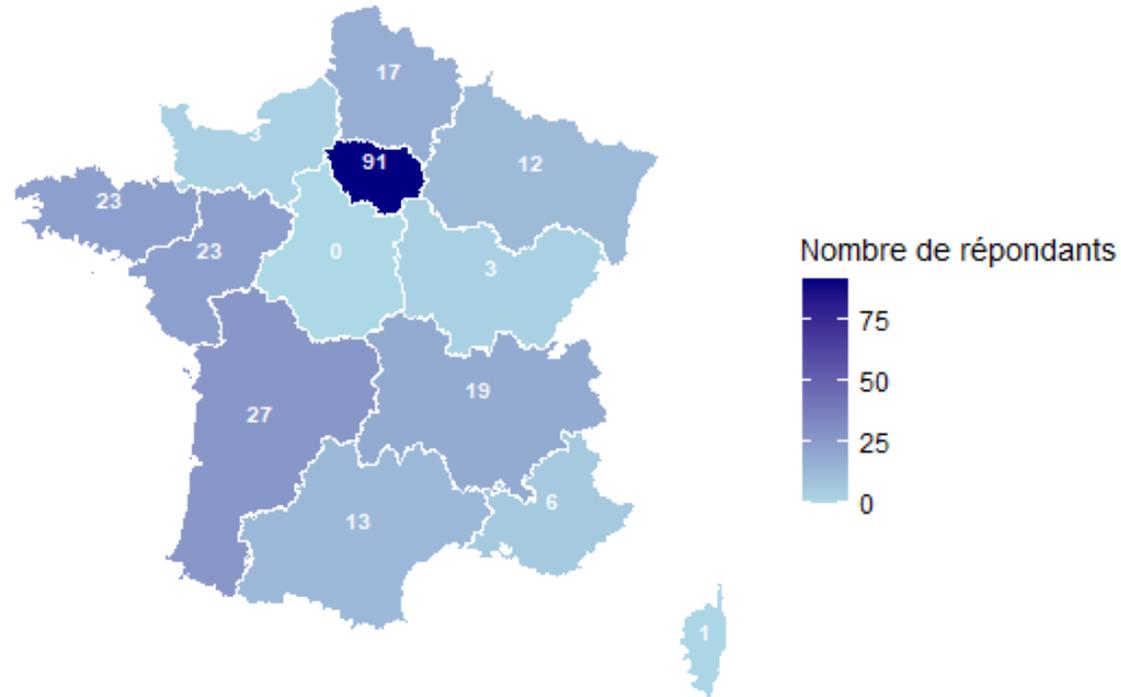
Banque / assurance et prestation de services parmi les plus représentés

Répartition des répondants par secteur



Structure d'échantillon (3)

Répartition des répondants par région



Une **grande diversité géographique** également

La surreprésentation de l'île de France est assez « logique » dans ces métiers

Principaux résultats

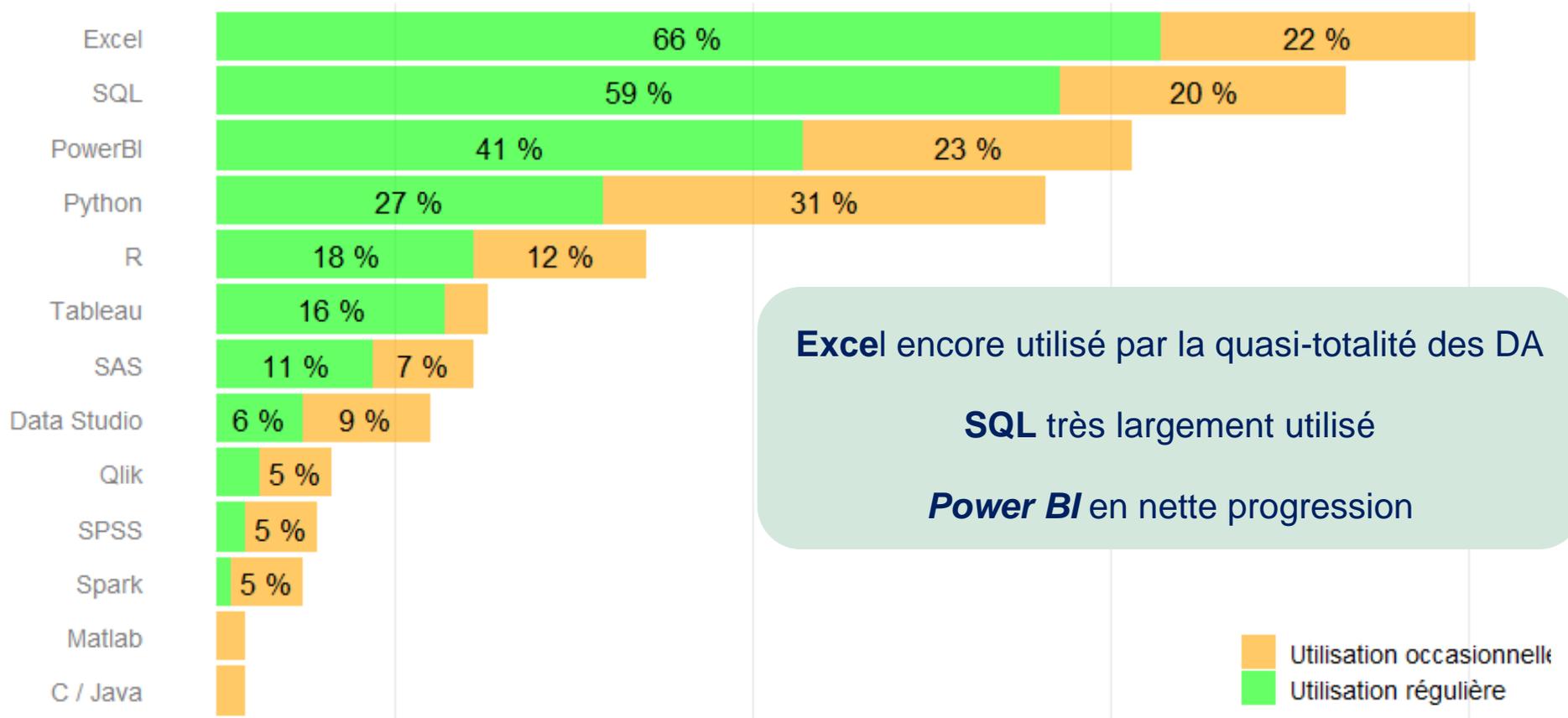
I. Méthodologie

II. Principaux résultats

III. Annexes

Les outils du Data Analyst

Fréquence d'utilisation des outils data *Data analyst / chargés d'études uniquement*

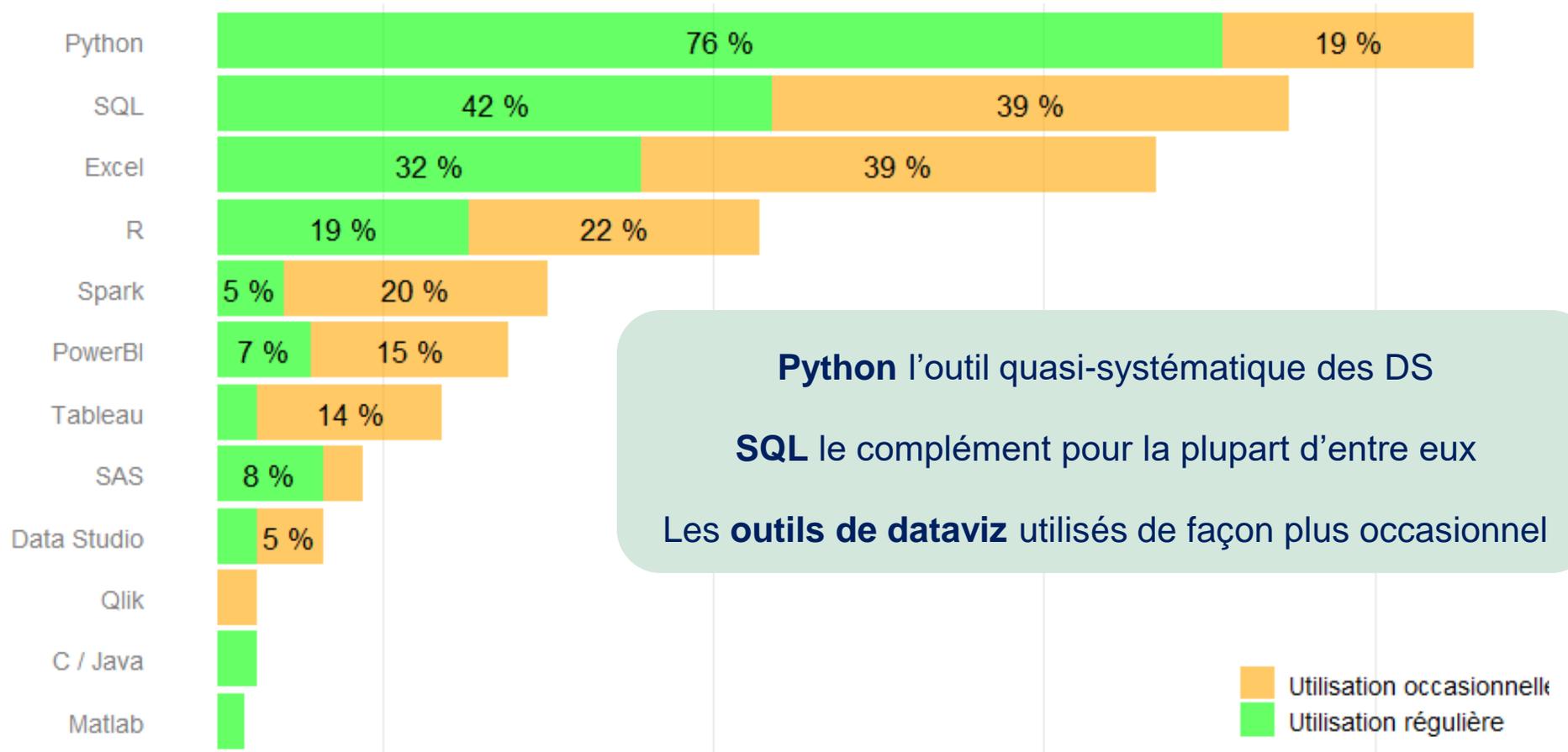


Question posée : pour chacun de ces outils que vous avez déjà utilisés, indiquez désormais si vous l'utilisez régulièrement, occasionnellement ou jamais dans votre poste actuel.

Les outils du Data Scientist

Fréquence d'utilisation des outils data

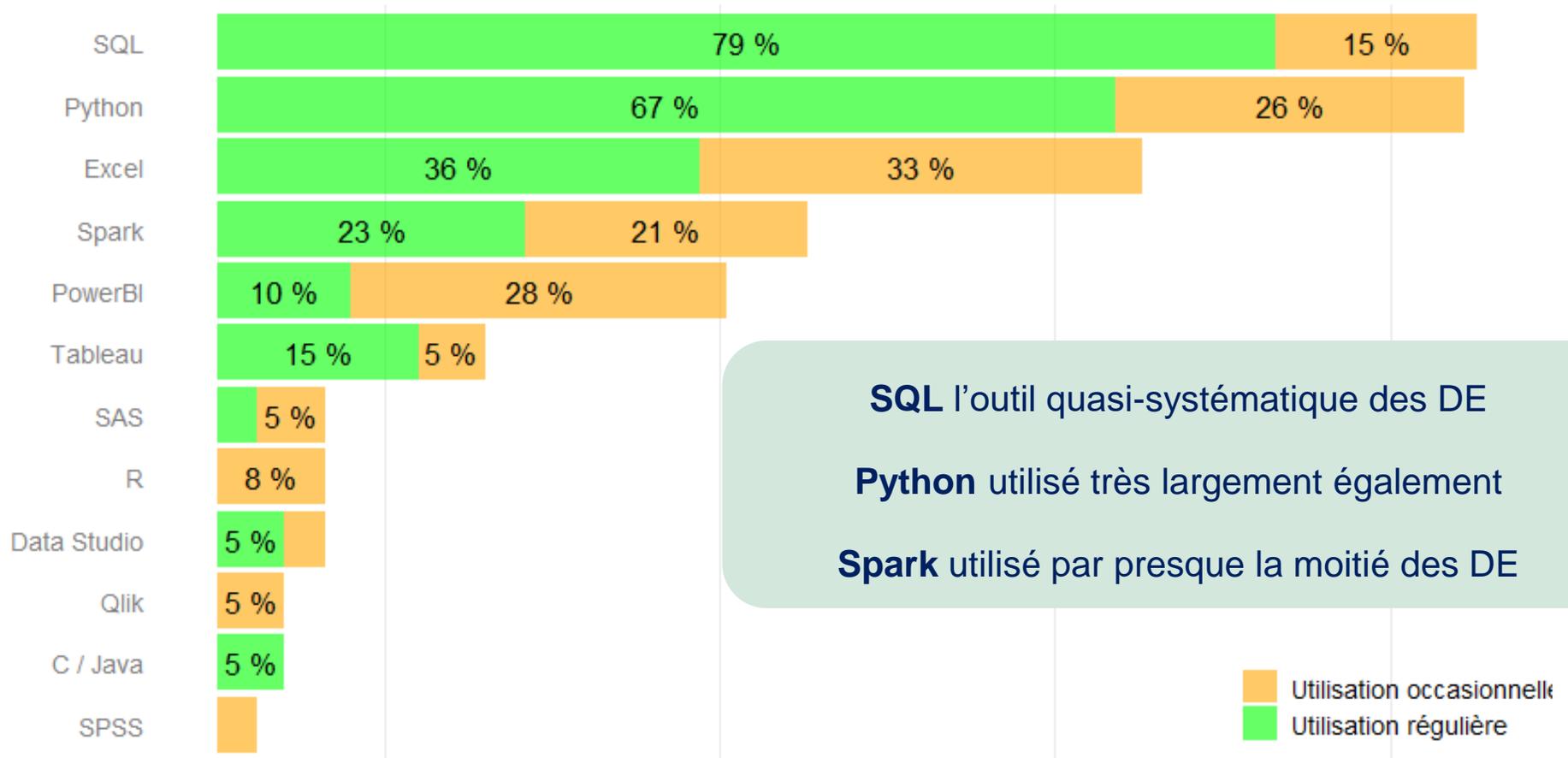
Data scientist uniquement



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

Les outils du Data Engineer

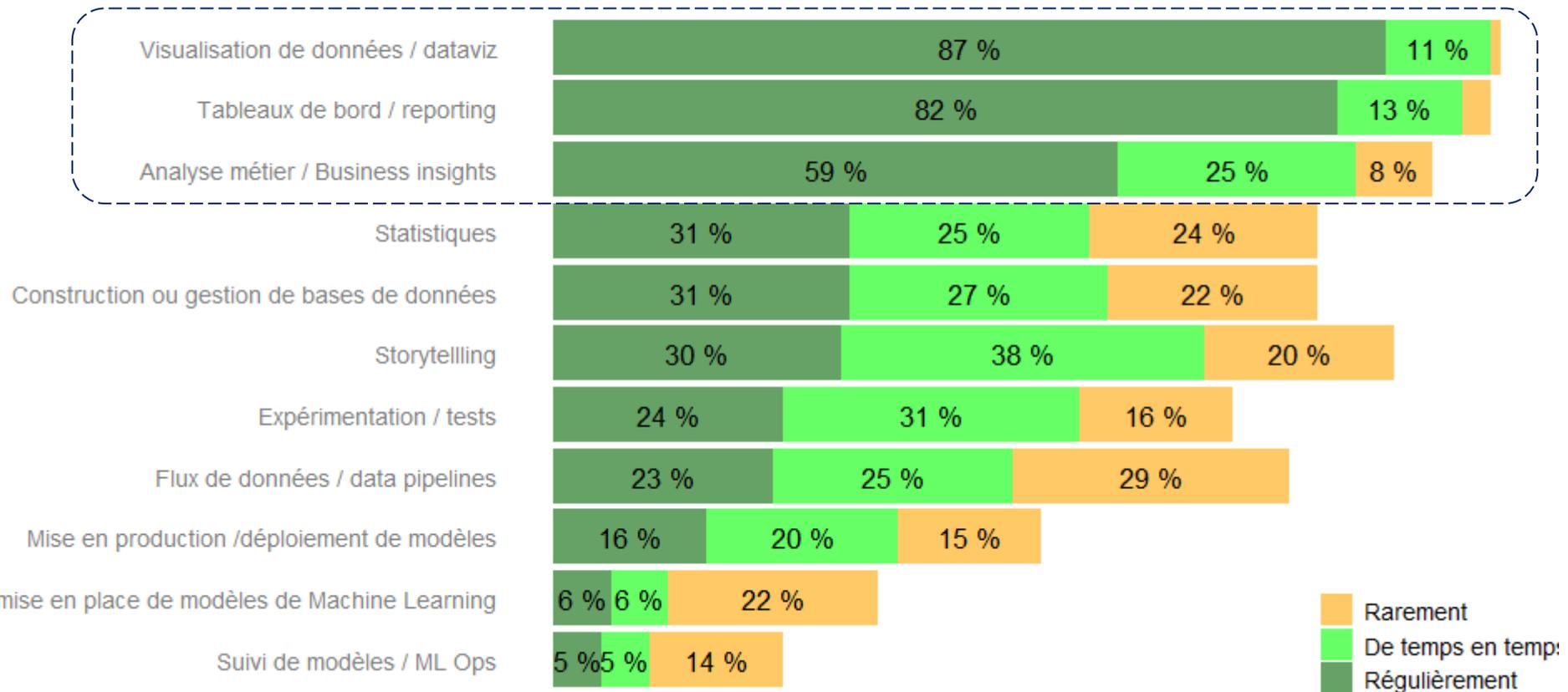
Fréquence d'utilisation des outils data *Data Engineer uniquement*



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

Les missions du Data Analyst

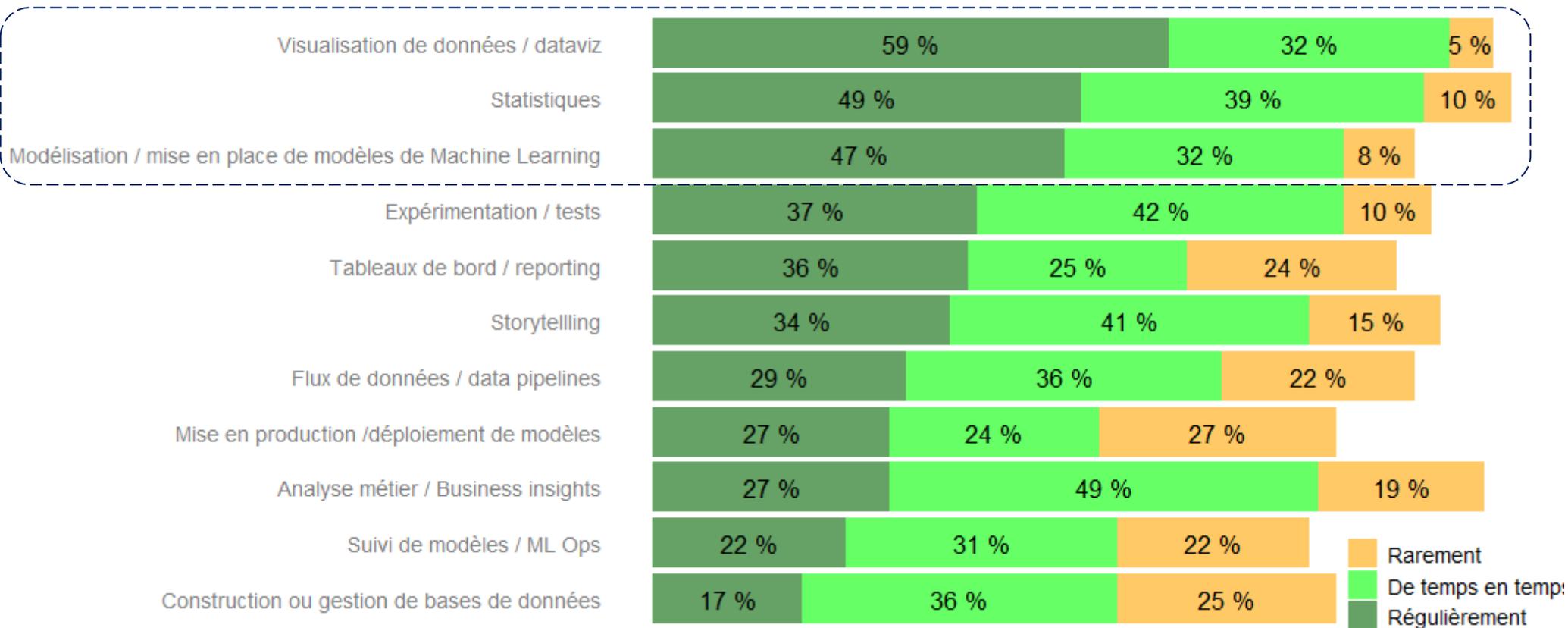
Fréquence de réalisation des tâches *Data analysts / chargés d'étude uniquement*



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.

Les missions du Data Scientist

Fréquence de réalisation des tâches *Data Scientist uniquement*

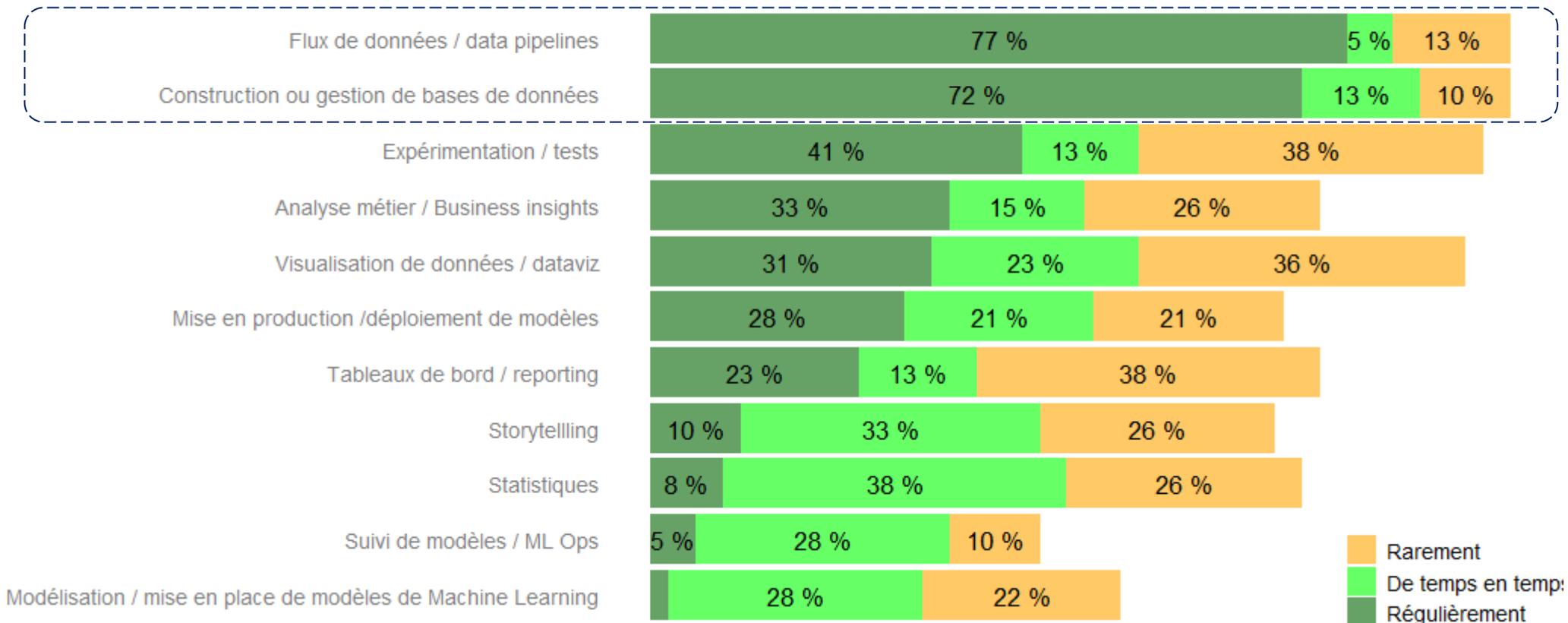


Pour chacune de ces types de tâches liées à des profils « data », veuillez indiquer s'il s'agit de quelque chose que vous faites régulièrement, de temps en temps, rarement ou jamais dans votre poste actuel.

Les missions du Data Engineer

Fréquence de réalisation des tâches

Data Data Engineer uniquement

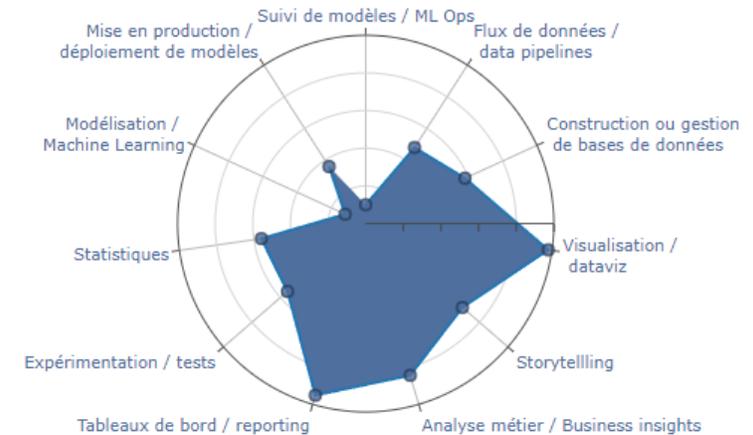


Pour chacune de ces types de tâches liées à des profils « data », veuillez indiquer s'il s'agit de quelque chose que vous faites régulièrement, de temps en temps, rarement ou jamais dans votre poste actuel.

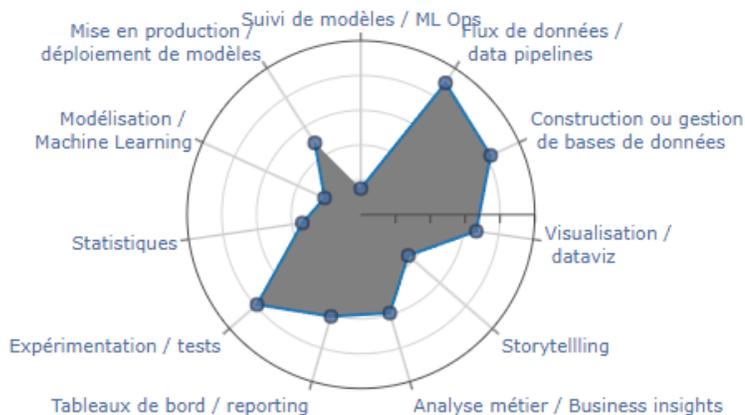
Résumé des missions par métier

- **Data Analyst** : une spécialisation sur la **dataviz**, le **reporting** et **l'analyse métier**
- **Data Scientist** : une certaine **polyvalence**, avec une part importante de **machine learning** et des **statistiques**
- **Data Engineer** : principalement la **gestion des données** et **bases de données**

Les missions du Data Analyst

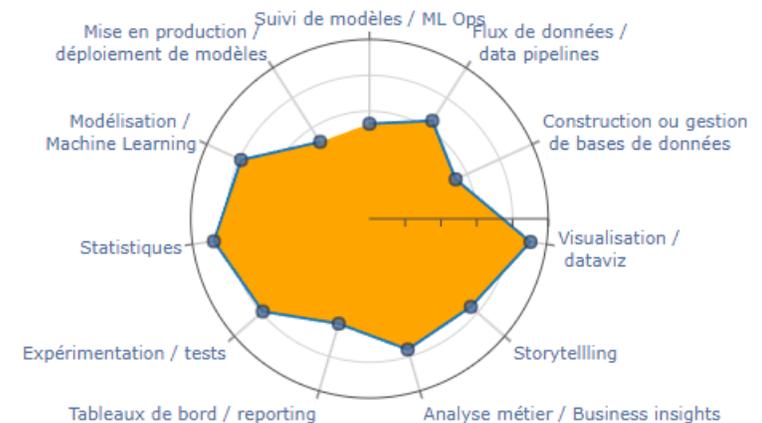


Les missions du Data Engineer



Note : reprise des célèbres visuels de **Kevin Rosamont Prombo**, mais adapté aux données issues de l'enquête

Les missions du Data Scientist



Pour aller plus loin

- I. Méthodologie
- II. Principaux résultats
- III. Pour aller plus loin



Excel reste un **outil très largement utilisé** par les DA

Une proportion non-négligeable des DS et DE l'utilisent également

Fréquence d'utilisation d'Excel

Selon le métier



Une **utilisation en recul** toutefois par rapport à la dernière enquête, là où les logiciels de Dataviz progressent

- Pour les DA : utilisation régulière passant de 80% en 2021 à 66% en 2024

Focus outils de dataviz



Les outils de Dataviz sont désormais massivement utilisés par les DA

les $\frac{3}{4}$ d'entre eux utilisant Power BI, Tableau ou Qlik



Fréquence d'utilisation des outils de dataviz

Selon le métier



Une utilisation en nette progression par rapport à la dernière enquête

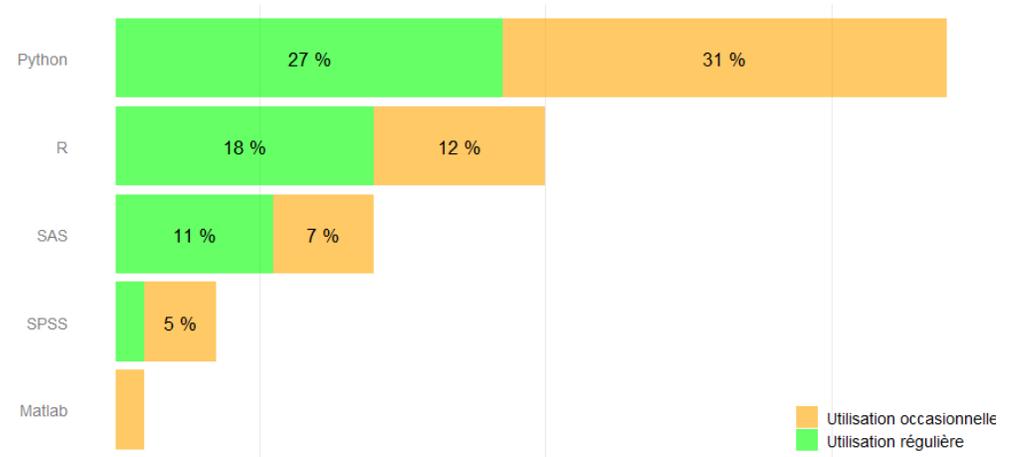
Focus SAS



SAS reste utilisé par une portion non-négligeable de DA, bien que largement distancé par Python et R.

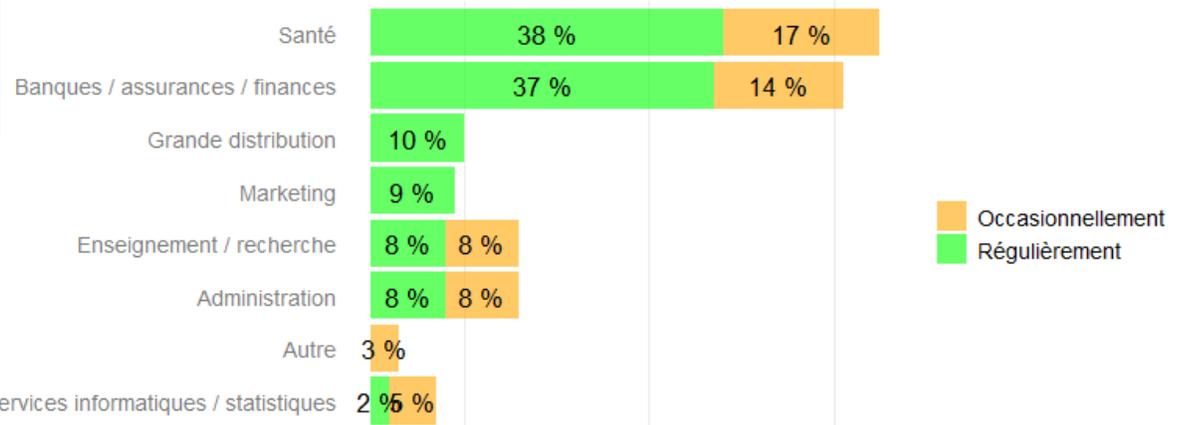
On retrouve ces utilisateurs essentiellement dans les secteurs de la **santé** et la **banque-assurance**

Fréquence d'utilisation des outils d'analyse de données
Data analyst / chargés d'études uniquement



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

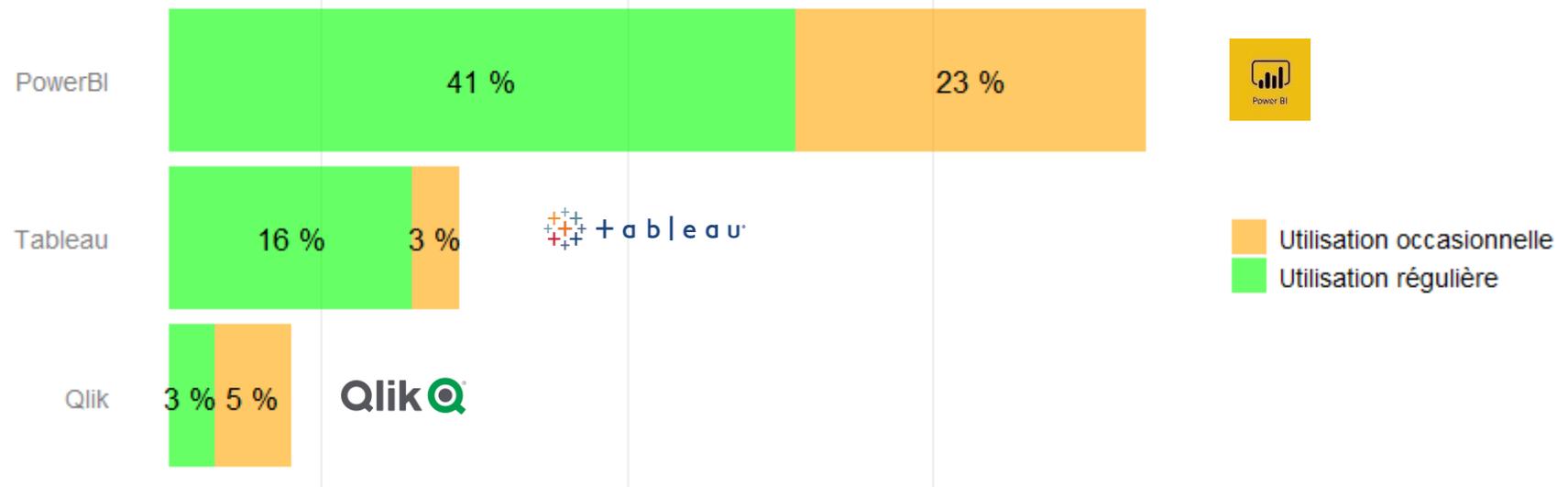
Fréquence d'utilisation de SAS par secteur
Tous métiers



Focus diversité des outils de Dataviz

Fréquence d'utilisation des outils de dataviz

Data analyst / chargés d'étude uniquement



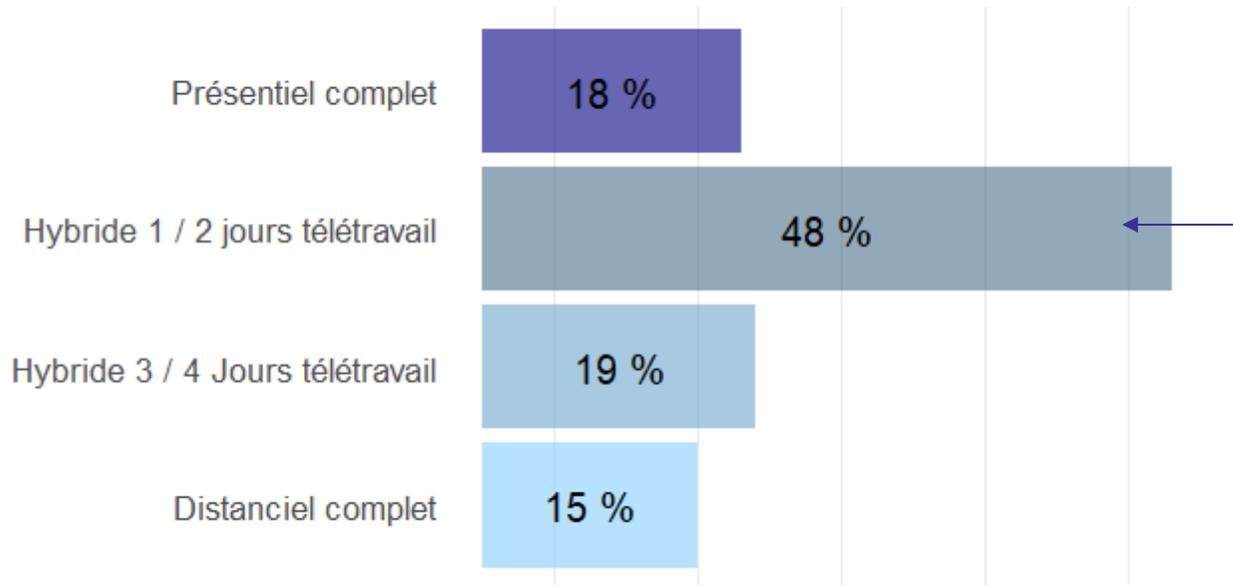
Power BI creuse l'écart avec ses concurrents

Tableau conserve sa place de « **solide second** »

Télétravail

Format de télétravail

Tous métiers / secteurs



Environ **la moitié** des professionnels de la data en télétravail un ou deux jours par semaine

Dans votre poste actuel, quelle est la part de télétravail / travail sur site ?

Les secteurs....

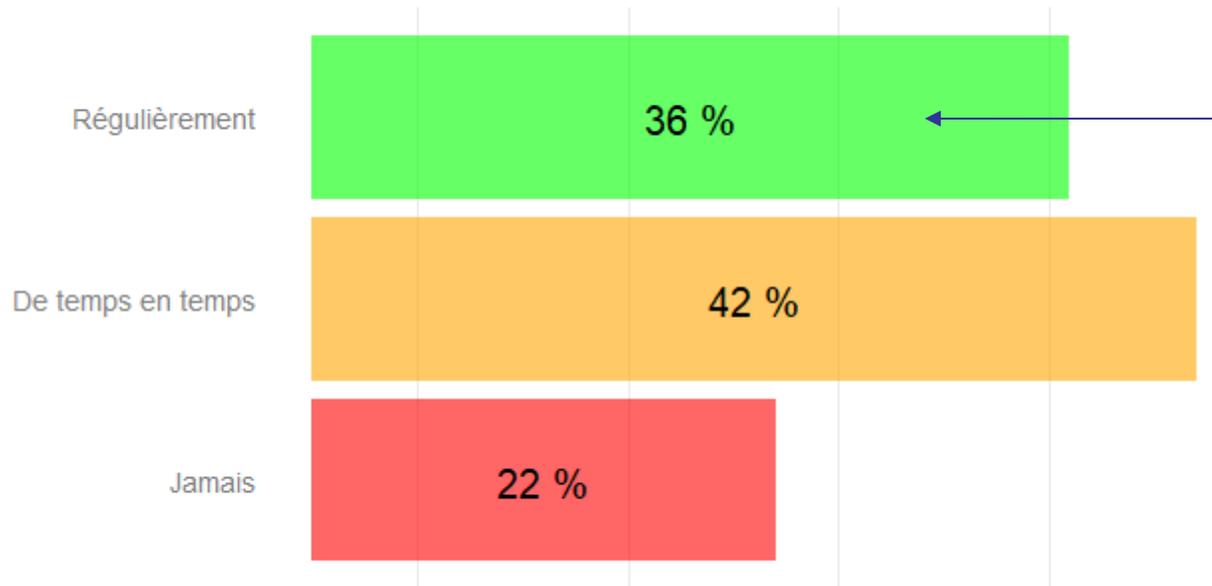
... les plus fréquemment associés au « ***pur distanciel*** » : prestation de service, Web et santé

... plus fréquemment associés au « ***pur présentiel*** » : enseignement, recherche, industrie et administration

Utilisation de l'IA au travail

Utilisation de l'Intelligence Artificielle au travail

Tous métiers / secteurs

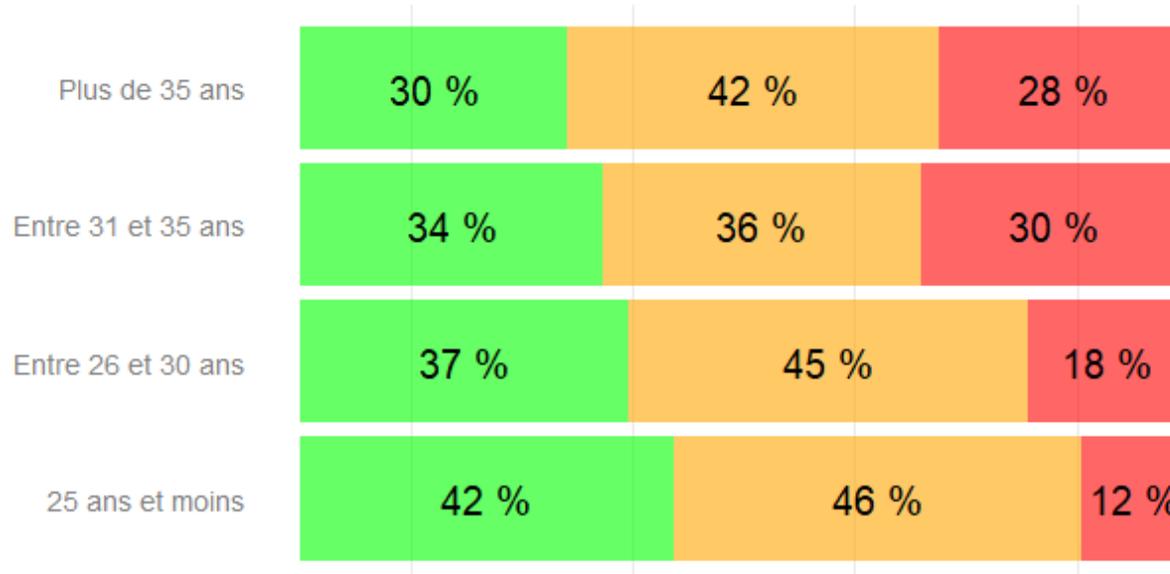


Plus d'un **1/3 des professionnels** de la data utilisant l'IA de façon régulière

Question posée : utilisez-vous l'Intelligence Artificielle type chatGPT dans votre travail ?

IA croisée par genre et tranche d'âge

Utilisation de l'IA au travail par tranche d'âge
Tous métiers / secteurs



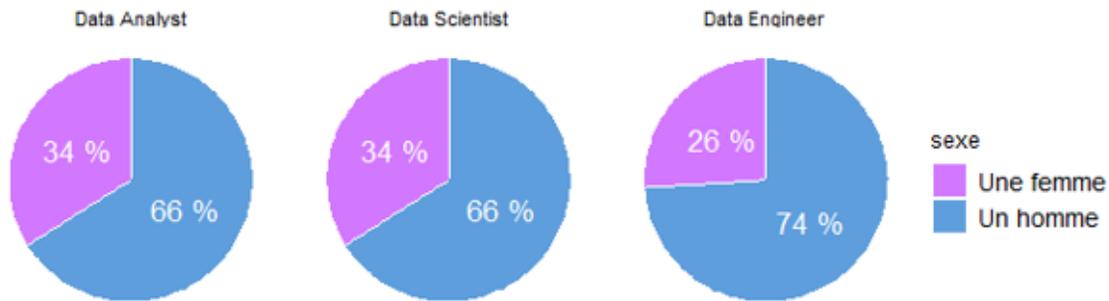
Question posée : utilisez-vous l'Intelligence Artificielle type chatGPT dans votre travail ?

L'utilisation de l'IA est très corrélée à l'âge, les plus jeunes l'utilisant plus fréquemment

- L'écart reste toutefois mesuré, on ne peut pas parler de « choc générationnel »

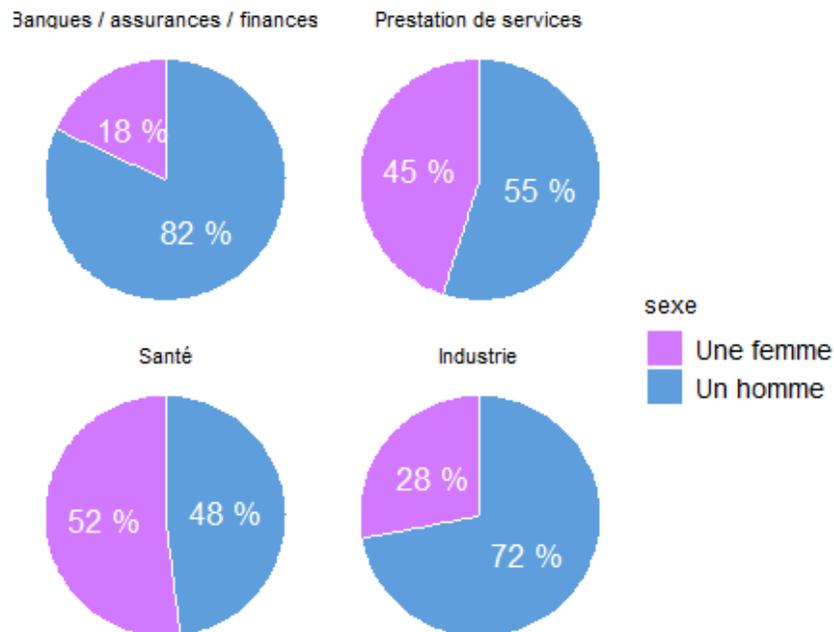
Genre

Répartition par sexe des principaux métiers de la data



Des ratios **hommes / femmes assez proches** d'un **métier** à l'autre

Répartition par sexe sur les principaux secteurs data



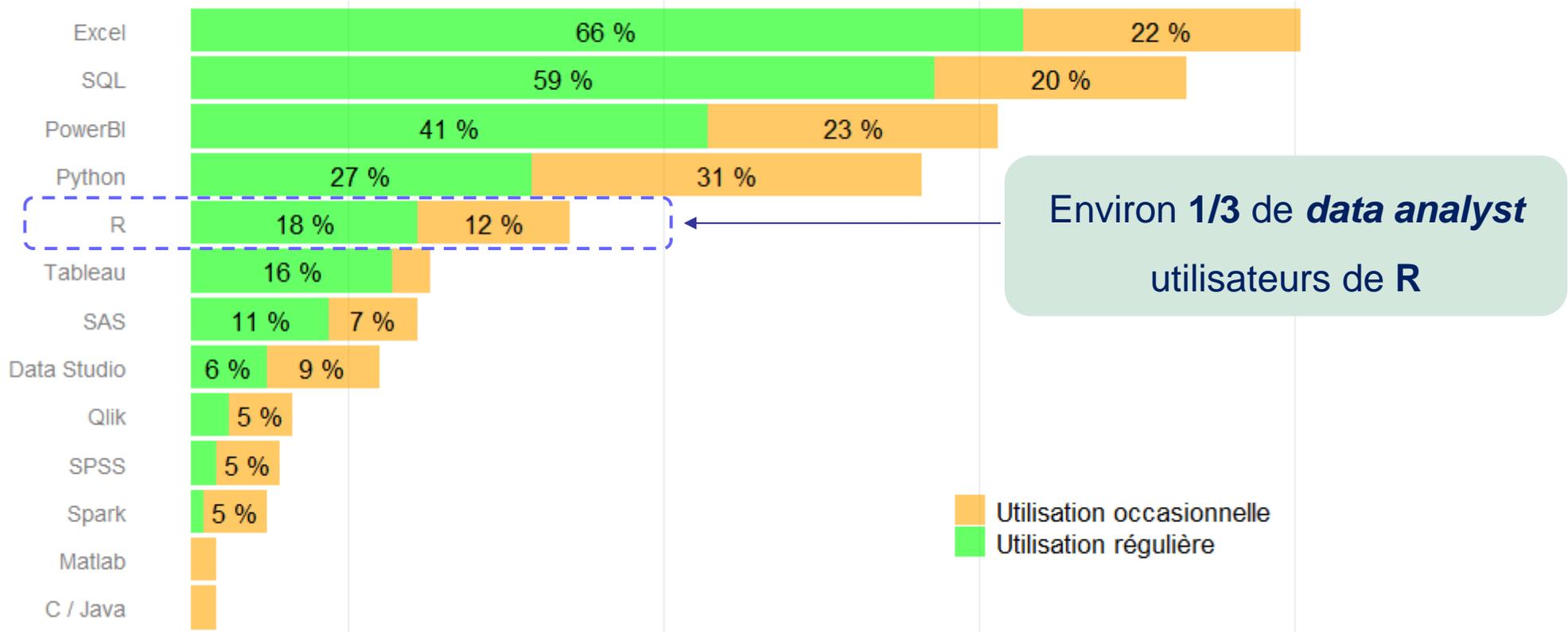
Des différences plus marquées **par secteur**

- Des **femmes plus nombreuses** dans la **prestation de service** et la **santé**

Focus R : utilisation



Fréquence d'utilisation des outils data *Data analyst / chargés d'études uniquement*



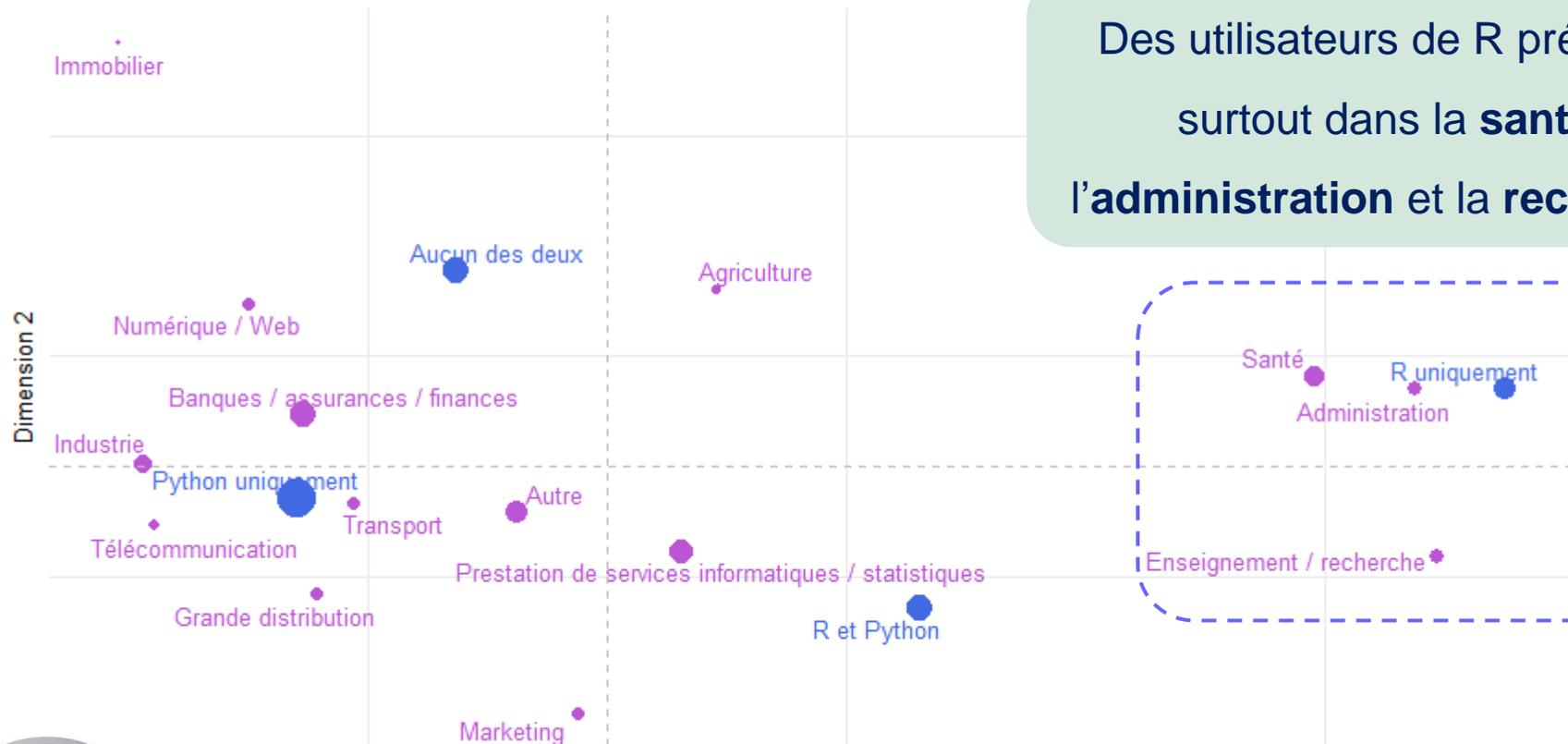
Question posée : pour chacun de ces outils que vous avez déjà utilisés, indiquez désormais si vous l'utilisez régulièrement, occasionnellement ou jamais dans votre poste actuel.

Data scientists (base faible) : 40% d'utilisateur, mais tous utilisateurs de Python également

Focus R : secteur d'activité

Liens entre le secteur d'activité et l'utilisation de R / Python

Tous profils



Des utilisateurs de R présents surtout dans la **santé**, l'**administration** et la **recherche**



Méthode utilisée : analyse des correspondances entre secteur d'activité et utilisation conjointe de R / Python

Pas de différences significatives sur l'âge ou le sexe R vs Python

Focus R : utilisation



Corrélations missions / outils

Tous profils

	Data Studio	Excel	PowerBI	Python	Qlik	R	SAS	Spark	SQL	Tableau
Visualisation de données / dataviz	-0.22	0.17	0.26	0.01	0.01	0.03	-0.13	-0.18	0.1	0.21
Tableaux de bord / reporting	-0.16	0.3	0.35	-0.15	0.08	-0.1	0	-0.2	0.13	0.19
Suivi de modèles / ML Ops	0.03	-0.08	-0.09	0.27	-0.01	-0.01	-0.08	0.2	0.02	-0.04
Statistiques	-0.1	0.05	-0.16	0.01	0.02	0.42	0.16	-0.07	-0.16	0.08
Modélisation / mise en place de modèles de Machine Learning	0.03	-0.12	-0.11	0.35	-0.04	0.25	-0.05	0.2	-0.16	0
Mise en production /déploiement de modèles	0.02	-0.09	-0.02	0.12	0.05	-0.03	-0.12	0.14	0.11	0
Flux de données / data pipelines	0.03	-0.13	0.03	0.28	0.16	-0.26	-0.14	0.24	0.32	0.08
Expérimentation / tests	0.03	-0.09	-0.03	0.22	-0.02	0.1	-0.09	0.18	0.04	0
Construction ou gestion de bases de données	0.09	0.06	-0.05	0.01	0	-0.04	0.09	0.04	0.18	0.12
Analyse métier / Business insights	-0.18	0.14	0.35	0.13	0.1	-0.19	-0.11	-0.13	0.28	0.22

Des utilisateurs de R qui se démarquent des autres outils par l'importance des **statistiques**

Mise à disposition des données

- **Données brutes et traitements** :

- **Fichiers csv de données** et scripts R / quarto de traitements de données
- Lien : https://gitlab.com/angirard1/enquete_data_2024/-/tree/main/

Un autre point / croisement à explorer ?

Une question sur cette enquête ?

N'hésitez pas à me contacter.

Antoine.girard@datag.fr