

Une enquête sur les profils « data »

GIRARD Antoine, antoine.girard@gamestatistics.fr, travailleur indépendant, Bordeaux

Thématique **Analyse des données**

Résumé : Derrière ces mots désormais si répandus de « *data analyst* » et « *data scientist* », quelle est la réalité des missions et outils associés à ces postes ? Une enquête menée auprès de profils « data » en tous genres – *analyst*, *scientist* ou encore « statisticiens » - permet d'apporter un éclairage sur leurs profils respectifs, les outils utilisés et le type de missions effectuées.

Mots clés : *enquête sondage outils missions profils dataviz*

1. Introduction

Il y a une quinzaine d'années seulement, les termes de « *data analyst* » et « *data scientist* » étaient encore inconnus : il y avait le statisticien et le chargé d'études, et d'ailleurs le terme même de « data » était assez peu utilisé. Si l'on peut arguer que les fondamentaux de l'analyse de données n'ont pas vraiment changé, des nouveaux outils ont émergé, et d'autres ont décliné : technologies et missions en analyse de données évoluent vite, au point qu'il est parfois difficile de comprendre ce qui peut se cacher derrière ces nouveaux termes. En 2021, quel est le profil moyen de ces métiers opérationnels de la *data* ? Quels outils utilisent-ils ? Quelles missions effectuent-ils ? Qu'est-ce qui différencie, au-delà du titre, le *data analyst* du *data scientist* ?

2. Méthodologie

Pour apporter des éléments de réponse à cette question, une enquête quantitative a été menée par ma modeste personne. Diffusée principalement via *Linked* en novembre 2021 (diffusion du questionnaire auprès de professionnels de la data, plus ou moins aléatoirement choisis), elle se base sur environ 150 réponses complètes de profils data en tous genre. Tout en étant conscient des biais inhérents à la méthodologie – influence de l'algorithme de *Linked* dans le choix des répondants, surreprésentation probable de certains types de profils (bordelais notamment !) – cette enquête a été menée dans le souci d'une représentation la plus fidèle possible de la réalité de ces métiers.

Les thématiques abordées par le questionnaire se limitent à trois grands domaines : les outils (connaissance / expertise / utilisation), les missions (effectives et préférées) et le profil général (secteur d'activité, type de poste, type d'emploi, 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.

L'analyse de ces données suit la méthodologie employée par les instituts d'étude : tris simples et tris croisés, tests statistiques et analyse multidimensionnelle. En particulier, ce type de données et la problématique centrale (qu'est-ce qui différencie le *data analyst* du *data scientist*) se prêtent particulièrement bien à la problématique de la méthodologie statistique dite de l'AFM (Analyse Factorielle Multiple), mise en œuvre dans le traitement des résultats.

Une présentation synthétique des principaux résultats de cette enquête sera proposée au cours de la présentation, après avoir détaillé sa méthodologie (y compris ses limites) et ses objectifs.

3. Originalité / perspective

Il y a à ma connaissance peu d'enquêtes ciblant spécifiquement les métiers opérationnels de l'analyse de données. Les enquêtes effectuées par le site *Kaggle* (dont les résultats de la dernière édition viennent d'être publiés) permettent de mettre en évidence les outils utilisés par les *data scientist* - avec un nombre très important de participants – mais il s'agit à mon sens d'une communauté assez pointue et spécialisée dans le *Machine Learning*. Cette enquête se veut plus généraliste et touchant tous types de profils, du chargé d'études en alternance travaillant sur Excel au *data scientist* expérimenté et spécialisé dans le traitement d'images en *deep learning*.

Par ailleurs, l'expérimentation de la méthode de l'AFM – moins connue que ses grands frères de l'ACP et AFC / ACM – pourrait intéresser le public ne connaissant pas forcément cette méthodologie spécifique.

Enfin, si les retours sur cette enquête et ces résultats sont positifs, alors je tenterai très certainement de la relancer l'année prochaine, en visant une amélioration de la qualité de réponse et des problématiques plus diversifiées.

Références

L'enquête en question est encore toujours ouverte même si les réponses ne sont plus comptabilisées, aussi le questionnaire utilisé peut-il être consulté via ce lien :

<http://www.gamestatistics.fr/limesurvey/index.php/526253?newtest=Y&lang=fr>

Une présentation de l'Analyse Factorielle Multiple, méthodologie mise au point par l'Agrocampus de Rennes, peut être consultée ici :

<http://factominer.free.fr/more/PagesAFM.pdf>