

Σ | STEPSWISE

DESSINE MOI UN DATA SCIENTIST

8 LEÇONS POUR LE RECRUTER



{ REMORA
Talent Advisory

Quantmetry



DESSINE MOI UN DATA SCIENTIST

8 LEÇONS POUR LE RECRUTER



Sommaire

0 - Édito.....	p6-p7	<i>Jeremy Harroch et Simon Majoulet</i>
1 - Compétences recherchées en Data science.....	p8-p10	<i>Nicolas Gibaud</i>
2 - Équipe type en Data science.....	p11-p13	<i>Xavier Vincelot</i>
3 -L'encadrement d'une équipe de Data scientists.....	p14-p16	<i>Karl Neuberger</i>
4 - Les profils types en Data science.....	p17-p21	<i>Isabelle Robin</i>
5 - Marché du recrutement.....	p22-p24	<i>Michèle Marchand</i>
6 - Processus de recrutement.....	p25-p26	<i>Olivier Kahn</i>
7 - On boarding en Data science.....	p27-p29	<i>Issam Ibnouhsein & Rony Rozas</i>
8 - Enjeux RH.....	p30-p32	<i>Matthieu Vautrot</i>
9 - Références.....	p33	

Pour un sujet encore neuf, le Big Data a déjà épuisé un bon nombre de métaphores, les plus communes ayant été la révolution, la ruée vers le pétrole ou l'or, l'exploration (la découverte de l'Eldorado), voire le nautisme (vague, tsunami, « fleuve Amazone de données »).

A tout prendre, l'image de la ruée vers l'or n'est pas si mauvaise ; il s'agit de rechercher des richesses dans de vastes contrées inexplorées, et de beaucoup piocher. Las, nous ne sommes plus, en 2015, dans la période faste et candide des premières découvertes et des emballements. Les prospecteurs sont désormais dans les montagnes du Yukon, il fait froid, les pépites sont bien enfouies, et tout le monde, soudain, se proclame vendeur de pioches. Et l'excitation peut laisser place à une certaine frustration.

Avec la multiplication de titres fantaisistes et des annonces d'emplois très libérées sémantiquement autour de la Data Science, il nous apparaît comme nécessaire de donner un peu plus de clarté aux profils et aux rôles qui émergent dans cette phase de transition et de structuration. Ce

travail de formalisation est le fruit de la collaboration entre une structure de conseil spécialisée dans la Data Science, Quantmetry, qui est l'un des tous premiers acteurs du domaine en France, et d'un cabinet de conseil en recrutement au niveau exécutif, Remora Talent Advisory.

Avec ou sans métaphore, l'émergence de la Data Science reste une excellente nouvelle, pour l'innovation, la création de valeur et l'emploi, avec déjà près de 4.000 créations de postes de « data scientists » attendus en France.

Bonne lecture...



Jeremy Harroch
Fondateur et PDG de
Quantmetry



Simon Majoulet
Co-fondateur et Directeur
Associé de Remora Talent
Advisory



Compétences recherchées en Data science

On observe aujourd'hui une certaine inflation autour du titre, souvent auto-attribué, de « data scientist », floraison qui n'est probablement pas sans rapport avec la nouvelle attractivité de ces compétences (« the sexiest job of the 21st century » d'après un récent article de la Harvard Business School). Cependant, rares sont ceux qui ont réellement le profil requis, parce qu'il repose sur des expertises pointues, et surtout parce que ces différentes expertises n'étaient jusqu'à présent que rarement combinées au sein d'un seul profil.

**THE SEXIEST JOB
OF THE CENTURY ?**



LES COMPÉTENCES DU DATA SCIENTIST EN MOUVEMENT

Les entreprises ont aujourd'hui une conscience de plus en plus précise du potentiel que recèlent leurs données, et recherchent les moyens de les exploiter et d'en tirer le maximum d'information utile. Pour les aider dans cette tâche, les data scientists (traduit littéralement : « scientifique de la donnée ») sont les personnes en charge de récupérer, stocker, organiser, traiter cette masse d'informations afin d'en tirer de la valeur.

Le data scientist est un profil d'un genre nouveau, issu de la convergence des statistiques et de l'informatique. La définition d'un data scientist reste encore floue. Ce qui le caractérise certainement le mieux, c'est la variété des compétences qu'il doit maîtriser. Il s'agit en effet d'un profil hybride, qui doit disposer d'un solide bagage en mathématiques, statistiques, maîtriser les outils informatiques et les infrastructures nécessaires à la gestion et au traitement des données, et avoir la curiosité et la soif de comprendre le métier et le secteur dans lequel il intervient. Voici un tour d'horizon des compétences qu'on attend d'un data scientist dans chacun de ces domaines.

Mathématiques, statistiques

La maîtrise des outils de base en statistique est indispensable à qui prétend savoir décrire et analyser des données. Savoir ce qu'est une distribution, comprendre le principe de l'analyse en composantes principales, connaître les grands tests statistiques ou encore savoir en quoi consiste un maximum a posteriori sont autant de points qu'il faut avoir assimilés. On ne demande pas nécessairement d'être un expert, mais d'avoir le recul nécessaire pour pouvoir dire si telle ou telle technique est une bonne approche dans un contexte défini.

Intimement lié aux statistiques, l'apprentissage automatique, ou machine learning, est un des fondamentaux de la data science. Il est indispensable d'en connaître les grands principes : la différence entre l'apprentissage supervisé et non supervisé, les notions d'over-fitting et de validation croisée. Une bonne connaissance des algorithmes classiques est également importante : arbres de décision, support vector machines, k-means, réseaux de neurones, etc. De nombreuses implémentations « clés en mains » existent pour ces algorithmes, par exemple dans des bibliothèques spécialisées en R (gbm, randomForest, nnet) ou en Python (scikit-learn). Aussi, ici encore, on ne demande pas d'avoir une connaissance détaillée des théories derrière chacun des algorithmes, mais d'avoir une compréhension générale et le recul nécessaire pour savoir quand et comment appliquer le bon algorithme.

De plus une bonne compréhension de ces domaines implique un minimum de connaissances en algèbre linéaire et analyse multivariée, qui constituent le cadre mathématique sur lequel elles reposent.

Informatique, architecture IT

On ne peut pas se prétendre data scientist sans un minimum de compétences en développement et en architecture IT. En particulier, la maîtrise d'un langage de script « haut-niveau », comme Python, R ou simplement BASH, est indispensable. En effet, le data scientist doit être capable de pouvoir développer, tester et valider ses modèles seul.

A cela s'ajoute la connaissance d'un panel d'outils IT de data mining et d'analyse de données. Il faut des compétences sur les outils IT de stockage et traitement de la donnée : savoir manipuler un

système de gestion de bases de données relationnelles (SQL), ou non relationnelles (technologies NoSQL comme Cassandra ou MongoDB), connaître les nouvelles technologies de calculs distribués et parallélisés, souvent indispensable pour un traitement efficace de larges volumes de données : MapReduce, écosystème Hadoop (Hive, Pig, etc) et les dernières technologies du Complex Event Processing (Storm), du calcul parallèle ou inmemory (Spark) et de l'indexation (ElasticSearch).

Le data scientist doit être au courant des technologies les plus récentes et en cours de développement, mais également des technologies naissantes. Il doit identifier les technologies d'avenir comme Vowpal Wabbit et d'autres bibliothèques d'online learning, libfm pour la factorisation, les packages qui améliorent la performance des algorithmes comme blaze, numba ou xgboost, ou encore les packages qui implémentent les algorithmes ou les méthodes les plus récentes, comme theano pour le deep learning. Il est aussi extrêmement important de se maintenir au courant des bonnes pratiques et des nouveaux usages des technologies. Le meilleur moyen est d'avoir une pratique régulière de la veille et de participer à des conférences.

Connaissance métier

L'objectif du data scientist, est de mettre en œuvre les meilleurs moyens pour répondre à une problématique métier précise. Comprendre un enjeu propre à un secteur, formaliser une problématique et développer la meilleure approche possible pour y répondre implique une connaissance pointue du secteur et une compréhension en profondeur des données que l'on manipule et une intuition sur les valeurs que l'on peut y trouver.

Communication, visualisation

Un data scientist doit savoir expliquer les outils qu'il a développé et les méthodes qu'il a mise en œuvre, en sachant adapter son discours à une audience qui n'a pas forcément un bagage technique avancé. C'est pourquoi des capacités de communication et de vulgarisation sont indispensables. Inutile de construire le modèle le plus sophistiqué du monde si vous n'arrivez à convaincre personne de sa pertinence! Savoir présenter ses résultats de manière à la fois claire, précise et esthétique est un immense atout. Les outils de visualisation sont une aide extrêmement précieuse et la maîtrise des outils de data visualization tels que Tableau, Qlikview, Spotfire et des bibliothèques comme D3.js, matplotlib ou ggplot2, shiny sont des atouts plus pour construire des représentations graphiques plus avancées.

.....Nicolas Gibaud.

Équipe type en Data science

La composition d'une équipe type en data science doit permettre de couvrir l'ensemble des phases du processus d'exploration. Elle se constitue de profils techniques et opérationnels puisant leurs connaissances dans différents domaines scientifiques, unis par une forte cohésion.

Faut-il recruter une équipe de Data scientists ou de data science ?

Le champ des compétences à mettre en œuvre par le data scientist est très large : mathématiques, informatique et business. Il apparaît clairement que tout miser sur la découverte d'un tel profil est périlleux, d'autant que la majorité d'entre eux ont moins de 3 ans d'expérience. Le parfait data scientist s'assimile plus à une illusion qu'à une réalité industrielle.



LA PIZZA TEAM : UNE ÉQUIPE DE SUPER HÉROS AU SERVICE DE LA DATA SCIENCE

Une équipe de data science :

Pour réunir l'ensemble de ces compétences, une équipe cohérente et complémentaire semble être un moyen plus sûr et plus efficace. Comment doit se constituer une telle équipe ? Quels en sont les rôles et la répartition ? En data science, ces questions sont neuves et les managers dépourvus de réponses. Les compétences de l'équipe de data science doivent s'articuler autour des activités du processus d'exploration. La data science est une discipline rationnelle qui suit un processus scientifique, permettant de garantir la qualité des travaux d'exploration de données.

On doit compter parmi ses membres les ressources nécessaires à chacune des phases du processus :

- *Collecte de données* : des profils fonctionnels et techniques capables de localiser les données à l'intérieur, la périphérie et à l'extérieur de l'organisation ;
- *Préparation des données* : des data-miners habitués à des manipulations de données pour les intégrer et les nettoyer : lecture, fouille systématique, formatage et feature-engineering, comprenant notamment le travail sur les champs plus complexes (texte, images);
- *Consolidation* : des spécialistes des bases de données travaillant en collaboration avec les équipes métier pour joindre des tables issues de systèmes informatiques différents afin de dédoublonner et d'identifier les références uniques
- *Modélisation* : des profils quantitatifs et mathématiciens pour modéliser et croiser les données et donner sens aux corrélations ;
- *Apprentissage des modèles prédictifs* : des profils statisticiens adeptes de machine learning, pour mettre en œuvre des algorithmes et garantir leur pertinence et leur robustesse.
- *Interprétation* : des experts de data-vizualisation ayant l'expérience de représenter la donnée, ainsi que des profils techniques pour la mise en place de dashboards fonctionnels.
- *Intégration dans les processus opérationnels* : Des profils opérationnels capables de comprendre simultanément les enjeux métiers et les difficultés techniques pour diffuser les transformations nécessaires.
- *Visualisation* : à mi-chemin entre user-expérience et créativité, les prototypes proposent une prise en main éloquente des résultats trouvés.

Rassembler ces compétences est un premier sujet. Créer l'harmonie entre ces différents rôles et favoriser leur collaboration en est un second.

La taille. Une équipe trop petite manquerait de ressources, d'énergie ou de créativité pour aborder des problèmes complexes. A contrario, une équipe trop grande perd en efficacité car les

membres communiquent plus difficilement, et des jeux de pouvoirs apparaissent. La théorie de la 2-pizzas-team donne un critère simple : Une équipe est trop grande si elle ne peut pas être nourrie par 2 pizzas. En pratique, on parle d'équipe entre 5 et 12 personnes.

Feature team. Il est tentant de rassembler ses équipes par compétences : les informaticiens, les développeurs, les analystes. Or une telle architecture non seulement peut présenter des goulots d'étranglement, mais nuit à la fluidité de la chaîne de transformation. Le découpage par fonction s'avère très souvent plus efficace : chaque équipe est orientée vers un produit, un projet pilote, pour lequel elle vit en autonomie et exploite tout son savoirfaire.

Alignement. Entre le mythe du généraliste polyexpert et l'association de peintures monotâche, il existe un juste milieu. Aussi petite soit une équipe, les membres doivent pouvoir parler le même langage et savoir s'exécuter un minimum dans le langage de l'autre. De même pour le choix des outils : passer de R à Python à SQL, pour faire ce dont chacun des trois individuellement est capable s'avère être une perte de temps considérable.

Le fonctionnement de l'équipe de data science incombe à son chef de projet. Son rôle va bien plus loin que la gestion traditionnelle de projets, car il s'agit de projets d'exploration. Son leadership, son intuition et sa finesse d'analyse permettront à l'équipe de data science de fonctionner.

Innovation. Quelle que soit sa taille, l'équipe doit avoir conscience de travailler sur des sujets nouveaux avec des méthodes nouvelles dans une culture agile, très orientée delivery, tout en étant conscient d'un certain niveau d'incertitude assumée par l'esprit test and learn. L'équipe doit savoir s'adapter avec tout le champ des possibles en informatique et algorithmique alternant entre des phases de hacking (en réadaptant des scripts existants) et de mash-up (en combinant plusieurs solutions provenant de plusieurs langages).

Industrialisation. Le grand enjeu, une fois la valeur trouvée et prouvée par les données, consiste à pouvoir assurer la capacité à rejouer les algorithmes sur de nouvelles données (update des scripts) et à les améliorer (upgrade) dans un système robuste, stable et respectant les critères de sécurité de la DSI. C'est alors que l'équipe se voit complétée par des développeurs informatiques, des administrateurs de bases de données et une assistance à la maîtrise d'ouvrage.

Partage de connaissance. L'équipe de data science garde à l'esprit de conserver un temps raisonnable au transfert de compétence afin que chaque membre soit capable de réaliser à nouveau, en autonomie des projets similaires.

.....Xavier Vincelot.

L'encadrement d'une équipe de Data scientists

Si le recrutement d'une équipe de data scientists est la première brique de la transformation digitale d'une entreprise par le quantitatif, la qualité de leur encadrement, notamment lors des phases de projets, en constitue la clé de voûte. En effet, pour délivrer de la valeur actionnable, les data scientists doivent travailler de concert avec le métier et être encadrés par des professionnels selon une vraie méthodologie projet.



LE PREMIER PROJET BIG DATA : UNE COURSE AUX CONNAISSANCES ET AU DELIVERY

Pour créer de la valeur au sein d'une entreprise grâce à la data science, recruter des data scientists n'est pas suffisant. En effet, ces derniers doivent être encadrés par des professionnels capables de les guider d'un point de vue business afin d'aligner la stratégie globale de l'entreprise et le travail d'analyse de données.

Manager une équipe d'experts diplômés des meilleures écoles d'ingénieur ou titulaires de thèses en science n'est cependant pas chose aisée. En effet, cela nécessite à la fois des compétences techniques afin d'avoir la crédibilité nécessaire, mais aussi une expertise business en gestion de projet et en

communication afin d'interagir avec les décideurs métier pour piloter l'avancée d'un projet et s'assurer du caractère opérationnel de la solution proposée.

Dans le cadre d'un projet de data science, la définition d'un objectif business clair constitue la première condition sine qua non. Pour ce faire, plusieurs parties prenantes doivent être impliquées dans la construction et le suivi d'un projet :

- Un sponsor soutenant le projet à haut niveau et guidant les grandes orientations stratégiques en terme business. Le sponsor est généralement un Directeur d'une entité avec un niveau hiérarchique suffisamment élevé pour assurer le soutien politique du projet
- Des responsables des différentes entités concernées par le projet : Direction des Systèmes d'Information, Direction Marketing et Connaissance Clients, Direction de la Transformation Digitale, Direction Stratégique, Direction Générale, etc.
- Un data scientist senior ou des leaders techniques peuvent également accompagner le chef de projet afin de l'aider à encadrer les équipes sur des points particuliers (statistiques, image processing, data design, etc.).

Du point de vue opérationnel, la méthodologie utilisée pour assurer le delivery d'un projet de data science est inspirée des méthodes agiles utilisées dans le développement Web. En effet, l'analyse de données étant par essence exploratoire, une approche itérative permet de confronter très régulièrement les résultats avec les différentes parties prenantes et ainsi d'affiner et de recentrer l'analyse à chaque étape. Au quotidien, cette méthodologie de continuous delivery vise à mettre en place plusieurs rituels entre les parties prenantes au projet sur la base de sprints (ou itérations) de deux semaines :

- Des stand-up meetings quotidiens (15 min) permettant de faire le point sur les objectifs de la journée entre le chef de projet et les équipes techniques.
- Des comités projets hebdomadaires ou bimensuels (1h à 2h) destinés à affiner les stratégies d'analyse à déployer dans l'itération et à planifier celle-ci (répartition de travaux, priorisation et suivi du backlog) entre les chefs de projet, les équipes techniques et plusieurs responsables des entités impliquées.
- Des comités de pilotage mensuels (2h) réunissant l'ensemble des parties prenantes au projet et le sponsor dans le but de présenter les résultats du mois et d'orienter la suite du projet.

Tout au long du projet, la répartition des travaux est suivie dans un backlog (Kanban) réunissant les tâches dont chaque Data Scientist est responsable sur chacun des chantiers identifiés.

A chaque itération, le rôle du chef de projet est d'apporter les réponses aux questions suivantes aux équipes techniques :

- Pourquoi : quel est le sens business de l'analyse de données en cours ?

- Comment : quelle méthodologie de data science utiliser ? Le chef de projet doit dans ce cadre faire profiter les équipes de ses expériences passées et de son expertise technique, tout en laissant aux data scientists la liberté nécessaire pour exprimer leur créativité.
- Combien : quelle équipe et quelles ressources pour quelles tâches ?

En dehors du mode projet, chaque data scientist dépend d'un manager expérimenté responsable d'assurer son suivi à long terme et de gérer le développement de sa carrière dans l'entreprise. A ce titre, le manager est responsable de la formation et de la montée en compétence d'un data scientist.

Par ailleurs, les managers ont aussi la tâche de garantir la motivation de leurs data scientists par la reconnaissance de leur travail (implication des data scientists lors des comités de pilotage ou mise en valeur du travail réalisé lors de séances de présentation à l'équipe notamment), mais aussi leur épanouissement au travail (temps « libre » dédié à la veille technologique ou scientifique par exemple (<http://fr.slideshare.net/joelhorwitz/agile-data-science-36258963>))

En définitive, l'encadrement d'une équipe de data scientists est donc une activité nouvelle nécessitant de connaître à la fois la data science, mais aussi le profil de ces nouveaux experts de la donnée. Une bonne formation scientifique, ainsi qu'une expérience managériale en gestion de projet et des qualités de communication sont donc des éléments indispensables au recrutement d'un chef de projet en data science.

L'encadrement d'une équipe de data scientists est une activité nouvelle qui, si elle est mal conduite, peut faire échouer un projet. Le choix des profils de managers et les méthodologies utilisées sont donc tout aussi importants que le choix des data scientists.

.....Karl Neuberger.

Les profils types en Data science

Les Data Scientists possédant un mélange de compétences à la fois éclectiques et complémentaires, il n'existe pas un unique profil-type de Data Scientist, mais une multitude ! Formations, parcours, décryptons les univers différents qui gravitent autour des métiers de la donnée...

Les profils junior :

Côté junior, il n'existe que peu de formations réellement spécialisées en data science et celles-ci sont très récentes. Les data scientists juniors viennent donc d'horizons très divers.

Historiquement, les analystes et data miners proviennent d'écoles spécialisées en statistiques. Ces écoles offrent aujourd'hui des cursus qui permettent d'appréhender les statistiques pour traiter des volumes de données massifs, à l'instar de l'option « data science » à l'ENSAE. À ces écoles, s'ajoutent des masters spécialisés destinés à donner une coloration « Big Data » à un parcours bac+5. Ces masters sont généralement proposés par des écoles de statistiques comme l'ENSAI, ou des écoles orientées vers les technologies de l'information comme Télécom ParisTech, Centrale Executive ou le master de l'Ecole Polytechnique.

Quant aux écoles d'ingénieurs généralistes, elles proposent pour la plupart, des options en informatique, mathématiques appliquées ou recherche opérationnelle qui permettent aux élèves d'avoir un premier socle de connaissances sur l'analyse de données. Moins formés aux statistiques que leurs confrères et consœurs d'écoles spécialisées, leur côté technophile et leurs acquis mathématiques leur permettront d'envisager les algorithmes avec un regard différent.

Au-delà des écoles d'ingénieurs, le monde universitaire est également très présent dans cette discipline qui met souvent au service des entreprises les avancées de la recherche. De nombreux docteurs deviennent donc data scientists après une thèse en machine learning. Mais il n'est également plus surprenant de découvrir des profils de docteurs en physique, qui ont des compétences pointues en statistiques.

Pour démarrer en tant que data scientist, des parcours plus atypiques sont également possibles car il s'agit d'un métier où l'auto-apprentissage est très valorisé. Cela permet d'apporter des compétences et visions différentes des profils classiques issus des grandes écoles.

Les profils seniors

Pour les seniors, les profils peuvent être encore plus variés mais les domaines ont souvent un point commun : une touche de mathématiques dans un univers technologique. Certains secteurs ont en effet connu plus tôt le besoin de traiter et exploiter les données collectées. Ce sont aujourd'hui de ces secteurs que proviennent la majorité des profils seniors, qui réinventent leur métier sous le jour de la data science.



ASSURANCE

BANQUE

ECOLE
STATS

INGENIEURS

STAT

BI

Physiques

Pig

MATHS
APPLIQUEES

Bio



BIG DATA

RESEARCH

DATA SCIENTIST

HADOOP

PYTHON

Recherche
opérationnelle

Les parcours du monde de la Business Intelligence (BI) sont les plus classiques et probablement les plus évidents. Un expert en BI aura souvent rencontré des problématiques métier très similaires à celles retrouvées dans la data science et sera sensible au besoin de valorisation des données qu'expriment les entreprises. Ils pourront également s'avérer être des atouts précieux pour la visualisation de données, qui demande à la fois la maîtrise d'outils particuliers et une certaine compréhension.

Après des années au sein de grands groupes d'assurance, les personnes provenant de ce secteur ont appris à gérer le risque à l'aide de modèles statistiques avancés. Aujourd'hui, ils peuvent mettre à profit leur expérience des données pour d'autres domaines, plus novices en la matière.

Au sein du domaine bancaire, ce sont souvent les spécialistes du trading algorithmique qui se réorientent vers le métier de data scientist. Leur maîtrise des cours de la Bourse pourra s'étendre à d'autres variables que les kilos-euros et leur connaissance de langages informatiques de scripting les avantage dans l'acquisition de la « boîte à outils » du data scientist.

Enfin, des personnes travaillant en biostatistique continuent d'apporter leur connaissance profonde des statistiques au domaine médical. Les besoins en analyse de données sont également existants dans le secteur agro-alimentaire et l'industrie pharmaceutique. Ce type de profil sera habitué aux processus de traitement des données sensibles.

Le renouveau de la data science a permis à de nombreux profils de se reconvertir pour tirer parti de leur expérience avec les données dans d'autres secteurs. Aujourd'hui, les entreprises s'arrachent ces candidats si rares et les frontières entre les domaines se sont estompées pour ces derniers, à la reconversion facilitée par la forte demande.

La différence entre un data scientist et un data miner.

Au delà du changement de nom, de mode et de génération, il y a des vraies différences entre ces deux profils compatibles avec leur culture propre. Les data miners formés sur des logiciels pendant les années 90 sont souvent dans une démarche d'utilisation de techniques classiques autour de la statistique descriptive, avec des interfaces graphiques très réussies qui limitent la créativité et qui accordent beaucoup d'importance à la lisibilité des calculs. Ils ont aussi acquis une compétence en échantillonnage, nécessaire pour assurer la performance de leur calculs en temps mais maintenant devenue obsolète avec les nouvelles technologies. Les data scientists sont plus bidouilleurs, s'autorisent à utiliser des algorithmes plus black box (dont on ne peut pas interpréter les résultats) comme certains algorithmes de machine learning, préfèrent mesurer leurs résultats sur la base de la précision de leurs calculs que sur la base du temps de convergence et ne sont pas gênés d'envisager de faire un full read (relecture complète) sur leur données. Les data scientists ont adoptés des comportements plus adaptés à la recherche de signaux faibles, les grands tendances sont toujours aussi bien interprétés par les data miners.

Les voies pour devenir data scientist sont aujourd’hui nombreuses car il s’agit d’un métier qui s’invente, un métier en pleine évolution qui nécessite des compétences très diverses. L’un des enjeux autour du métier de data scientist de demain est d’arriver à structurer l’évolution des formations tout en conservant cette riche diversité de profils.

.....Isabelle Robin.

Le Data scientist, où et comment le trouver ?

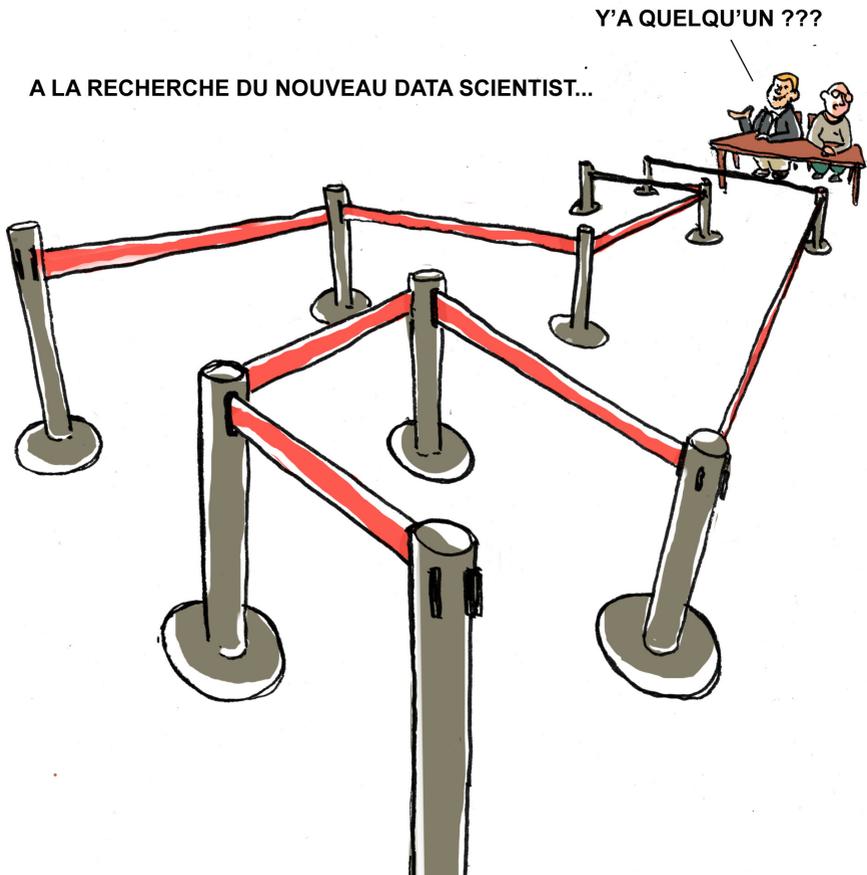
En France selon les DRH et les cabinets de recrutement, on serait autour de 2000 à 3000 profils recrutés par an. Mais la demande est en réalité loin d'être couverte et les RH sont aujourd'hui l'une des problématiques les plus importantes dans le domaine du Big Data. Le data scientist est un profil encore assez rare, avec des compétences très recherchées et les recruteurs doivent s'armer de patience afin de trouver la perle rare.

On compte aujourd'hui une dizaine de formations spécialisées mettant chacune sur le marché une promotion d'une trentaine d'étudiants. S'y ajoutent quelque 300 personnes participant à des cercles d'apprentissage en dehors des écoles et universités, et environ autant passent par l'autoformation. La demande étant plus forte que l'offre, le marché du recrutement des Data Scientists reste très tendu. De plus, le nombre de recrutements ne va cesser d'augmenter pendant encore plusieurs années avec la prise de conscience générale selon laquelle les data scientists constituent des ressources très importantes au sein des entreprises. Parmi elles, on peut traditionnellement compter les géants du web comme Facebook, Apple ou encore Google. Le secteur de la publicité en ligne, est également très demandeur sachant que les entreprises y sont en pleine croissance et que l'analyse de données constitue leur cœur de métier. De plus en plus, des entreprises du secteur bancaire, de l'assurance, de l'énergie, de la santé, des médias ou encore des sociétés de transports en pleine transformation digitale s'intéressent à leur données et à ce que la data science et le Big Data peuvent leur apporter. Elles cherchent avant tout à mieux connaître leurs clients, à optimiser plus finement leurs processus et à proposer des services innovants. Enfin, des sociétés de conseil spécialisées dans la data se positionnent aussi sur le marché, ainsi que des fabricants d'objets connectés ou des éditeurs d'applications liées à ces objets connectés. Les grands gagnants de demain seront donc ceux qui cherchent dès maintenant à cumuler expertises technologique, scientifique et métier et à les articuler au cœur d'équipes pluridisciplinaires qui sont au sein de la stratégie d'entreprise.

Les recruteurs doivent donc rentrer en contact avec les profils atypiques des data scientists, comprendre les nouvelles compétences exigées par ces nouveaux métiers et réussir à les évaluer. Il existe pour cela plusieurs moyens :

- Participer aux forums de recrutement orientés sur la data. DataJob est le plus grand salon de rencontre autour des métiers de la data. Cet événement est une très bonne occasion de connecter des entreprises avec les data scientists de demain. (C'est une journée consacrée au recrutement, au networking ainsi qu'à des conférences, des keynotes et des tables rondes autour des métiers de la Data.) Des grands comptes, des cabinets de conseil en statistiques, aussi bien que des PME ou encore des start-ups viennent à la rencontre de plus de 800 étudiants issus d'écoles d'ingénieurs et de statistiques telles que l'ENSAE, l'ENSAI, l'Ecole polytechnique ou encore Telecom ParisTech, ainsi que des universités.

- Connaître et se rapprocher des écoles proposant des formations d'ingénieur en data science et en Big Data. Ces formations délivrent des compétences de haut niveau et des profils suffisamment polyvalents pour répondre à une bonne partie des problématiques des entreprises.
- Faire appel à un cabinet de conseil en data science ou une entreprise spécialisée dans l'outsourcing de compétences en data science. Certaines sociétés n'ayant pas encore la possibilité de créer une structure en interne faute de moyens financiers, humains ou techniques, utilisent cette alternative au recrutement qui permet de ne pas engager de coûts excessifs dans un premier temps, ainsi que de répondre à un ensemble de problématiques data rapidement. Cela de manière ponctuelle et en fonction de ses besoins. Attention, de nombreux cabinets prétendent maîtriser la compétence en data science lors qu'il proviennent d'univers proches mais moins experts comme la business intelligence ou les SSII en développement informatique. Afin de distinguer leur expérience, un bon moyen consiste à évaluer la proportion de data scientists qui les composent. S'il y en a trop peu, la tentation sera trop forte pour les positionner sur des métiers traditionnels et leur montée en compétence ne sera pas au niveau d'attente. Les pure players assurent de concentrer sur leurs équipes des formations spécialisées sur les attentes des clients en data science.



- Passer par un cabinet de recrutement afin de dénicher des profils spécialisés en data science. C'est un investissement financier important, de 10 000 à plus de 20 000 euros pour certains profils seniors. Cependant c'est un bon moyen de gagner du temps et de définir avec précision les besoins afin de trouver le profil adéquat. La aussi, beaucoup d'agences RH prétendent avoir une expertise sur ce sujet sans avoir d'expérience notable. La combinaison avec un talent assessment (évaluation des compétences techniques) d'un cabinet expert et indépendant est un bon moyen pour cumuler une force de détection et une capacité de filtration. Voir à ce sujet, l'article 6 sur le processus de recrutement.
- Assurer sa visibilité et augmenter sa notoriété en communiquant auprès de médias de l'écosystème (Alliancy, Le Journal du Net, Maddynews etc.) afin d'être identifié par de potentiels candidats comme un acteur incontournable et attractif du secteur.
- Créer une marque employeur en éditant du contenu de qualité sur la data science est un bon moyen d'attirer spontanément des candidats : un blog, des livres blancs, partager des retours d'expérience, organiser des meetups sont des démarches très adaptées à ces profils à la recherche de connaissances.

Selon le plan Big Data, la France veut créer 130 000 emplois dans ce secteur et l'Europe aurait besoin de former 3 000 000 de Data scientists dans les années à venir. C'est donc pour cela que de nombreuses écoles créent des masters Big Data, sur le Machine Learning ou d'autres thématiques de la data science, afin de palier à cette pénurie et répondre au besoin de ce marché d'avenir.

.....Michèle Marchand.

Processus de recrutement

En data science, les profils sont rares et les compétences recherchées couvrent un large spectre de domaines. Ces deux particularités font du processus de recrutement un exercice complexe et exigeant. Les qualités à évaluer sont à la fois les savoir-faire techniques, principalement en mathématiques appliquées et en informatique, mais également les compétences humaines de communication et de pédagogie et une capacité à appréhender les enjeux business et économiques d'entreprises diverses.

Concernant les concepts mathématiques, il s'agit pour le recruteur d'évaluer tout d'abord les connaissances théoriques du candidat portant sur les différentes composantes de la data science et notamment les principaux modèles d'apprentissage automatique. Pour vérifier cela, il est nécessaire de demander au candidat d'expliquer comment fonctionne l'algorithme, de préciser les concepts mathématiques sous-jacents et de détailler les points d'attention dont il faut tenir compte pour l'utiliser. Le candidat doit aussi être capable de maîtriser les différentes notions de statistiques qu'il sera amené à utiliser au quotidien : tests d'hypothèses, sélection de variables et feature engineering,

**ON M'A DIT QUE VOUS
ÉTIEZ UN MOUTON À 5 PATTES**



sélection de modèles, statistique multivariée, etc. En complément des connaissances théoriques, le candidat doit être capable de démontrer un solide savoir-faire technique. Il est donc fortement recommandé de proposer un exercice, qui peut prendre la forme de la résolution d'un problème technique auquel l'entreprise a déjà pu être confrontée.

Un candidat au poste de data scientist doit également avoir de bonnes compétences en informatique. Les compétences à tester sont différentes de celles requises pour un poste de développeur, elles sont pour le cas du data scientist de deux natures : d'une part la capacité à programmer et à implémenter des algorithmes d'analyse de données et d'autre part la compréhension des principaux outils Big Data. Afin de contrôler cela, la première vérification à effectuer est celle du CV : l'objectif est de déterminer si le candidat mentionne bien ces éléments et s'il a déjà mené un ou plusieurs projets dans ce type d'environnement technique. Des questions techniques simples peuvent être posées pour vérifier que le candidat a bien le niveau attendu, et un exercice de programmation peut permettre d'apprécier plus objectivement l'expérience et le savoir-faire du candidat. Outre la programmation, le candidat doit également démontrer une connaissance des outils Big Data. Ainsi, le candidat doit être capable de présenter et décrire les principales technologies Big Data – notamment l'écosystème Hadoop et les bases de données NoSQL – et selon le poste proposé, être capable de développer en MapReduce, Spark ou Storm par exemple.

Le candidat doit également savoir exposer ses méthodes pour la veille technologique, sur les aspects mathématiques de la data science (nouvelles méthodes, nouveaux algorithmes, etc.), ainsi que sur les technologies (programmation, Big Data). Il doit démontrer sa capacité à monter rapidement en compétence et de façon autonome sur un nouveau sujet technique.

Les connaissances métiers sont également centrales pour un data scientist. Ainsi, pour un poste de data scientist expérimenté, il est fortement apprécié de pouvoir démontrer une compréhension économique d'un secteur d'activité. Pour vérifier ce type de compétences, il est très utile de proposer une étude de cas de type consulting au candidat. Les études de cas peuvent être fondées sur des projets réels sur lesquels l'entreprise a travaillé, l'objectif est de déterminer si le candidat est capable de proposer une méthodologie basée sur la data science pour résoudre un problème métier rencontré dans une entreprise. Une étude de cas n'attend pas spécifiquement de réponse juste ou fausse, le cadre de cet exercice est davantage une discussion ouverte, dont le but est d'échanger avec le candidat sur un projet réel. Une étude de cas est souvent un bon moyen de tester les compétences qualitatives du candidat : attitude, éthique, énergie, motivation et compétences interpersonnelles.

Finalement, si le métier de data scientist est complexe et multifacette, cela fait du recrutement de ce type de profil un exercice délicat, alliant entretiens techniques et entretiens métiers. Pour juger des diverses qualités nécessaires à ce poste, une approche intéressante peut être de faire rencontrer au candidat divers employés, tels que des développeurs, statisticiens, ou directions métiers par exemple, qui sont par ailleurs les interlocuteurs avec qui le candidat pourra être amené à travailler quotidiennement.

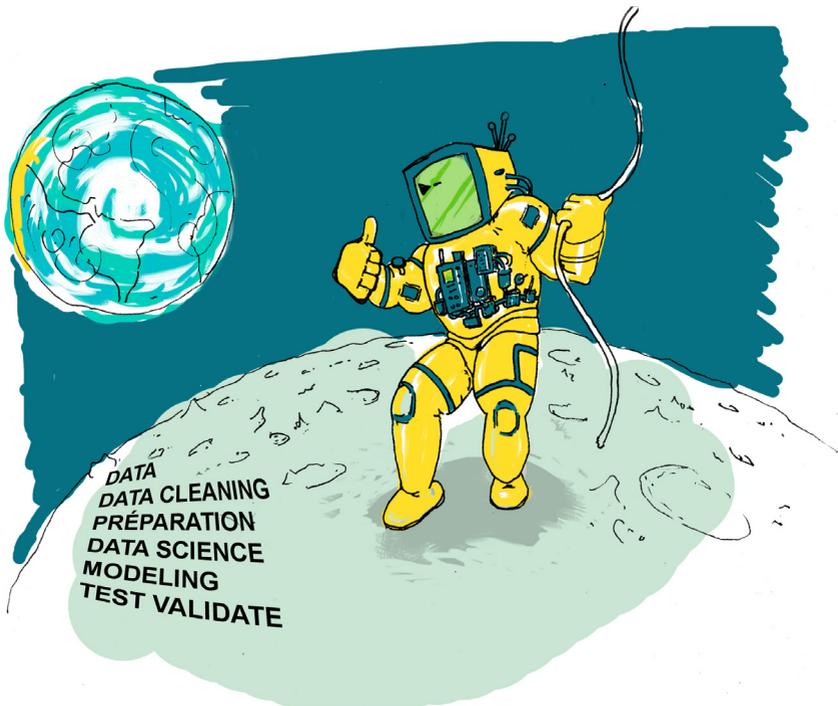
.....Olivier Kahn.

L'auto-formation en Data science

Un data scientist se devant de maîtriser de nombreux domaines différents, nous passerons en revue dans cet article les divers outils à sa disposition pour ce faire.

En premier lieu, plusieurs plateformes de cours en ligne existent, permettant de s'auto-former ou de simplement consolider ses connaissances sur les méthodes mathématiques appliquées à la data. Ces cours sont proposés gratuitement par plusieurs universités sous la forme de MOOCs (Massive Online Open Courses). Coursera et EdX sont deux plateformes connues qui centralisent de nombreux MOOCs, dont certains de qualité excellente. Un consensus a émergé autour des cours les plus pertinents, listés ci-dessous :

- « Machine learning » par Andrew Ng, Stanford University : probablement le cours le plus suivi au monde par la première génération de data scientists !
- « Algorithms : Design and Analysis » par Tim Roughgarden, Stanford University : ce cours en deux parties permet de poser les bases d'une bonne compréhension de la notion d'algorithme, ainsi qu'une bonne introduction à la notion de structure de données.



- Le parcours « data science » proposé par la Johns Hopkins University : il s'agit d'une formation introductive au métier de data scientist, centrée sur la manipulation de données
- « Mining massive data sets » par un groupe de professeurs de Stanford University : un cours plus poussé que le précédent sur la manipulation de données, avec une attention particulière portée sur les problématiques autour des très grandes bases de données.

D'autres cours plus pointus sur des thématiques particulières sont disponibles, dont l'intérêt dépendra des tâches concrètes que le data scientist sera amené à accomplir en entreprise.

Pour se perfectionner, le data scientist peut également utiliser l'ensemble des ressources mise à sa disposition sur internet ou sur des réseaux intranets. Par exemple, chez Quantmetry, un wiki a été mis en place pour permettre de trouver une réponse rapide aux questions fréquentes qu'un data scientist peut être amené à se poser. Cet outil peut également être utilisé pour l'aider à définir les points sur lesquels il doit travailler pour s'améliorer.

Ensuite, il est nécessaire pour un bon data scientist d'avoir une activité de veille sur les technologies utilisées en data science. Les outils incontournables pour faire une veille technologique sont les blogs d'entreprises actrices dans le Big Data et la data science, les publications dans des revues universitaires, ainsi que les réseaux sociaux : Twitter, LinkedIn, Google+. De nombreux outils permettent de créer des alertes par email pour se tenir informé des nouveaux articles contenant des mots clés pertinents. La présence aux conférences et aux Meetups spécialisés est aussi nécessaire. Les Meetups sont d'occurrence régulière et permettent de se tenir au courant des pratiques récentes tant en data science (recherche fondamentale, approches innovantes pour des domaines spécifiques, technologies Big Data et bonnes pratiques, etc.) que dans les domaines connexes (stratégies « data-driven » en entreprise, marketing digital, etc.). Les bases de connaissances sont également un bon moyen pour se tenir informé des technologies émergentes. Par exemple, le site Quora.com propose à ses utilisateurs de créer, d'éditer et d'organiser des topics sur un système de questions-réponses.

Un autre moyen de se former par la pratique et d'effectuer une veille technologique est de participer aux compétitions en data science proposées par Kaggle à l'international ou datascience.net en France. Il est possible de participer à ces compétitions en équipe, ce qui constitue un bon moyen de s'améliorer au contact des autres, de partager ses connaissances et d'apprendre le travail en équipe sur des projets de data science. Il s'agira en général de résoudre un problème concret qui se pose pour une entreprise à l'aide de données qu'elle aura fournies ; par exemple : créer un modèle de prédiction du nombre de visiteurs quotidiens des gares SNCF, ou des demandes de réservation sur un réseau de type Vélib. Sur ces plateformes, les données sont propres, bien organisées et le nombre de fichiers différents est limité. Ceci permet de se focaliser sur la partie « machine learning » et de progresser ainsi sur des aspects importants du métier de data scientist comme le « feature engineering » (création de variables explicatives pertinentes) ou le paramétrage des algorithmes d'apprentissage.

Cependant, les données réelles des entreprises sont rarement aussi bien organisées que celles des compétitions en ligne, et il n'existe pratiquement pas de formation de type Coursera sur le nettoyage et la structuration de données. Il s'agira donc pour cette partie d'une démarche d'auto-formation un peu plus improvisée et qui se fera principalement lors des premières tâches qu'aura à effectuer le

data scientist en entreprise. À noter que les méthodes les plus efficaces de nettoyage de données sont accessibles depuis le terminal avec des commandes comme « sed », « grep » et « awk » qui disposent d'un menu d'aide directement accessible depuis le terminal. Les langages de script comme python sont aussi très utiles pour ce genre de tâches, et lorsque l'on passe sur des plateformes Big Data, il est utile de connaître Pig, un langage de scripting sur Hadoop qui s'abstrait du paradigme MapReduce, permettant ainsi de se concentrer sur les tâches de nettoyage et de préparation des données.

Enfin, il ne faut pas négliger les aspects techniques du Big Data. Beaucoup de data scientists juniors se forment en mettant l'accent sur le machine learning et les algorithmes, mais ils négligent souvent de monter en compétence en programmation et dans la compréhension des technologies. A ce titre, il faut noter que les tutoriels des éditeurs de logiciels Big Data (Microsoft, IBM, HP, Cloudera, HortonWorks, MapR, etc.) fournissent d'excellents guides de prise en main et des versions d'évaluation très complètes de leur logiciels (avec des images virtuelles par exemple) pour une démarche de découverte non professionnelle.

En résumé, la data science nécessite une mise à contribution de nombreuses compétences, aussi bien en mathématiques qu'en informatique. Un data scientist doit avoir de solides bases dans ces deux domaines pour pouvoir mener à bien un projet en data science, mais il est important également qu'il soit capable de se former constamment. Les méthodes et outils évoqués dans cet article devraient l'aider dans ces tâches.

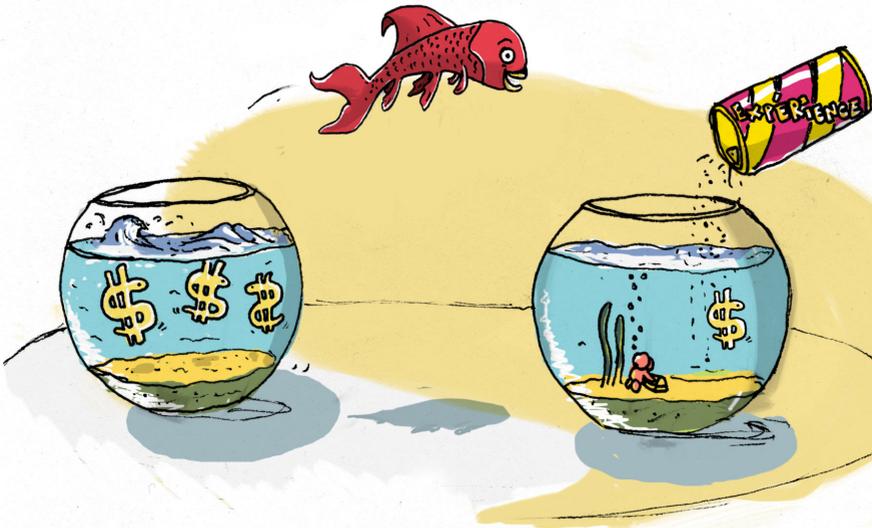
.....Issam Ibnouhsein & Rony Rozas.

Enjeux RH

Dans un contexte du marché de l'emploi largement en faveur aux data scientists, quels sont les leviers autres que le salaire disponibles aux entreprises pour minimiser le risque de départ de ses talents ?

Une fois les Data scientists embauchés et intégrés aux équipes, les entreprises doivent et devront à l'avenir faire face à un problème de taille : le turn-over. Il est important de gérer ce risque au plus tôt et d'avoir une stratégie qui vise la rétention de ces talents. Aujourd'hui, le marché de l'emploi est encore largement en faveur des Data Scientists, ils le savent et peuvent en jouer à n'importe quel moment.

L'EXPERIENCE : LE MEILLEUR MOYEN POUR UN DATA SCIENTIST DE NE PAS TOURNER EN ROND



La description dans ces articles du rôle clef joué par le data scientist dans leur organisation les amènent à considérer ce poste comme hautement stratégique pour leur développement. Elles sentent qu'elles devront mettre les moyens pour recruter les meilleurs profils et éviter à la fois de se priver d'un élément générateur de croissance et de voir cet élément aller à la concurrence.

Du côté des candidats, des vagues de profils débutants (0 à 3 ans d'expériences), issus de cursus généralistes orientés sur l'aspect analytique (Mathématique, Statistique et/ou Informatique décisionnelle) cherchent à trouver au plus vite le poste qui leur permettra d'obtenir l'expérience légitime pour se revendiquer data scientist expérimenté et ce, avant l'arrivée massive sur le marché de profils sortant des cursus spécialisés « data science ».

Cette dynamique entre les deux parties fait que les data scientists débutants ou confirmés sont conscients, à tort ou à raison, que cette situation idéale risque d'être éphémère. C'est donc pour eux, une course afin de pour trouver « Le job en or ». Ces profils sont sollicités quotidiennement via les réseaux sociaux professionnels et peuvent facilement céder à la tentation de changer d'entreprise s'ils estiment qu'une opportunité existe ailleurs.

Dans ce contexte très fantasmé, comment et avec quels leviers les entreprises peuvent développer auprès de ses data scientists la sensation qu'ils occupent déjà une position privilégiée ?

Voici quelques éléments de réponses autour de deux leviers, celui du challenge et du gain en compétence et celui de la reconnaissance du travail.

Levier challenge et gain en compétence:

La data science couvre de nombreux sujets différents mais complémentaires. Ceux qui embrassent ce choix de carrière sont par définition extrêmement curieux. Ils chercheront donc à maximiser leur gain en expérience en fonction de leur temps. L'entreprise doit donc être capable de leur proposer un environnement technique ainsi que des projets capables de satisfaire cette volonté de monter en compétence efficace. Dans la mesure du raisonnable et surtout pour des profils junior, ce levier «Challenge et gain en compétence» peut-être un levier plus puissant que le salaire.

De manière générale, une seule personne ne peut pas maîtriser l'ensemble de ces domaines de manière égale mais elle doit progresser dans plusieurs directions en parallèle. Elle cherchera à travers ses missions à privilégier sa progression, en particulier dans les domaines qui lui seront facilement transposables dans d'autres contextes. Pour la plupart des jeunes data scientists, l'ordre d'importance est souvent le suivant :

1. Mathématique / Statistique
2. Programmation / Manipulation des données
3. Communication / Data Visualisation
4. Expertise métier

Les missions et projets sur lesquels le data scientist sera amené à travailler devront donc lui

permettre d'apprendre sur chacun de ces axes.

On doit aussi lui laisser suffisamment de temps et de liberté pour expérimenter et s'approprier de nouvelles méthodes d'analyses ou de nouveaux outils. Ce temps libre est important et parfois encadré officiellement, par exemple sous la forme d'une journée hebdomadaire de veille technologique ou de réunion de partage entre data scientists.

Enfin, il est bon d'associer ceci à une certaine liberté technique. Pour expérimenter et progresser au mieux sans frustration, les moyens techniques doivent être disponibles de la façon la plus souple possible. Cela répond également à la problématique de la vitesse d'évolution des technologies Big Data qui s'accélère depuis plusieurs années. Un data scientist sera donc sensible à la qualité du hardware mis à sa disposition (ordinateur, serveur de calcul ou data center), mais aussi et surtout à sa liberté d'accéder aux technologies et aux données internes ou externes à l'entreprise de manière fluide.

Levier reconnaissance du travail :

Un autre levier puissant pour limiter le turnover de ses data scientists est celui de la reconnaissance du travail. Audelà du sujet de la rémunération qui a un impact direct sur cette sensation de reconnaissance voici d'autres éléments qui peuvent contrebalancer son importance :

- Partage, communication et reconnaissance des résultats : un data scientist doit pouvoir partager et démontrer que son travail apporte une réelle valeur. S'il y parvient, il faut que son discours trouve écho et soit relayé (blogs, 5 à 7, ...)+
- Les moments privilégiés avec leurs n+1, n+2... permet au data scientist de montrer ses capacités de vulgarisation d'un sujet complexe à ses supérieurs qui ne maîtrisent pas le contenu de leur travail tout en rappelant les valeurs qu'il a mis au service du projet, de l'équipe et de la société.
- La formation, participation à des séminaires, si possible, internationaux : ces événements qui sortent le data scientist de son quotidien pour qu'il échange permettent de matérialiser la volonté qu'a l'entreprise d'investir sur lui, autrement que par le salaire.
- L'entretien individuel de performances : De manière générale les entretiens individuels de performances sont des outils précieux pour faire le point avec un collaborateur (ils permettent d'établir un diagnostic sur l'affinité avec l'entreprise). Exploités intelligemment, ces moments clés peuvent permettre de dissiper les frustrations et de relancer la sensation d'appartenance à l'entreprise.

.....Matthieu Vautrot.

Références

1 - Compétences recherchées en Data Science.....

<http://pro.clubic.com/itbusiness/actualite693592datascientistmouton5pattescoeur-donnees.html>

<http://blog.udacity.com/2014/11/datasciencejobskills.html>

http://www.tias.edu/docs/defaultsource/Kennisartikelen/harvard_datascientistthe-sexiestjobofthe21s_tcentury_2012.pdf?sfvrsn=0

2 - Equipe Datascience.....

<http://www.forbes.com/sites/teradata/2014/11/19/5tipsforbuildingagreatdatascience-team/>

<http://www.lemondeinformatique.fr/actualites/lireanalysepredictiveledatascientistdoit-travaillerequipe59360.html>

<http://sas.developpez.com/actu/79082/UneetudeSASdefinitles10profilstypesdu-DataScientist/> <http://blog.jasoncrawford.org/twopizzateams> <http://fr.scribd.com/doc/113617905/ScalingAgileSpotify>

https://www.linkedin.com/pulse/7-new-big-data-roles-2015-martyn-jones?imm_mid=Ocb30b&cmp=em-da-a-na-na-newsltr_20150121

3 - L'encadrement d'une équipe.....

<http://sloanreview.mit.edu/article/getting-value-from-your-data-scientists/>

<http://www.ibmbigdatahub.com/blog/data-scientists-challenge-managing-stubbornly-autonomous-experts>

<https://hbr.org/2014/03/what-makes-big-data-projects-succeed/>

<http://fr.slideshare.net/joelhorwitz/agile-data-science-36258963>

5 - Marché du recrutement.....

<http://pro.01net.com/editorial/599331/fleurpellerindetaillesonplanbigdata/>

GRAPHISME ET ILLUSTRATIONS

AURÉLIEN GOMEZ

AURELIENGOMEZ@GMAIL.COM

NEOGRAFIX-CREATION.COM

IMPRESSION & BROCHAGE

SCRIPT LASER

29, BOULEVARD MALESHERBES

75008 PARIS

+

DESSINE MOI UN
DATA SCIENTIST

8 LEÇONS POUR LE RECRUTER

+

Le Big Data fait désormais partie du paysage économique français, européen et mondial et devrait représenter autour de 8% du PIB européen en 2020. Les entreprises ont pris conscience de la valeur des données pour améliorer leurs processus, augmenter leur business, mieux connaître leurs clients, anticiper les changements, fournir de meilleurs services et accomplir leur transformation digitale. Les données ouvertes et la recherche publique renforcent le phénomène en ouvrant les applications à des thématiques d'intérêt général.

Cette tendance de fond ouvre des perspectives importantes pour l'emploi. Les prévisions du Plan Big Data sont de l'ordre de 140000 emplois créés en France d'ici 2020.

Quels profils rechercher, quelles sont les offres de formation, comment constituer et encadrer une équipe de data scientists ? Vous trouverez ici les points de vue et les témoignages des experts de Quantmetry.



DATA JOB
2015

SAVE THE DATE

Jeudi 26 novembre
à L'espace Pierre Cardin

LE PLUS GRAND SALON DE RENCONTRE POUR
LES MÉTIERS DE LA DATA

Réservez votre stand dès maintenant

www.datajob.fr

contact@datajob.fr

Quantmetry
Data Science Consulting

www.quantmetry.com

REMORA
Talent Advisory

www.remora-talent.com