

ITW mentor : Nicolas Bancel

Tu as travaillé dans le data engineering. Peux-tu décrire la différence entre le métier de data analyst et celui de data engineering ?

Le métier de data ingénieur consiste à amener les données externes relatives à une entreprise, par exemple ses ventes d'un produit depuis un site de e-commerce d'une autre entreprise, vers sa base de données interne. Il faut donc créer des « pipelines » pour acheminer ces données, puis assurer leur maintenance. Il faut également s'assurer que ces données soient exploitables. On parle de « cleaner » les données.

Le data analyst lui, manipule les données internes, en utilisant les langages SQL et Python, pour, par exemple, faire des recommandations business.

Tu as travaillé dans l'aéronautique. Peux-tu citer quelques missions du data analyst dans ce type de secteur ?

À l'époque, j'étais assistant de recherche dans un laboratoire. Nous travaillions sur une étude pour Fedex qui cherchait à évaluer le potentiel de faire voler des avions en formation, en se projetant sur un maximum de 5 avions ensemble. Les avions volant ensemble, cela diminue les forces de frottements et donc la consommation en kérosène. Il fallait donc étudier la faisabilité du projet. On évaluait les cost savings associés au fait de faire voler des avions ensemble. Nous avons eu par exemple à étudier le trafic aérien, tous les points de départ et d'arrivée de chaque avion, pour voir à quel endroit ils pourraient se regrouper et sur quelle distance ils pourraient voler en formation avant de se disperser vers leur point d'arrivée respectif. Et c'est à partir du calcul de ces trajets qu'on pouvait quantifier, connaissant la distance parcourue ensemble - en formation - la réduction de consommation de carburant, et évaluer les réductions de coût.

Dans quelle société travailles-tu aujourd'hui et quel est ton rôle ?

Je suis freelance, ce qui offre beaucoup de flexibilité dans mon travail. Avant, j'ai fait beaucoup de data analyse et de data engineering. Aujourd'hui, j'ai envie de faire du machine learning, qui consiste notamment à créer des modèles de prédiction, et la flexibilité du freelance me permet d'avoir du temps pour y parvenir.

En ce moment, je travaille pour un fonds d'investissement. Des données d'entreprises y sont actuellement éparpillées sur chaque ordinateur des personnels. Cela peut être par exemple des données issues de Linked In en format Excel. Voici la problématique de cette mission : comment réunir toutes ces données, les réorganiser, les « cleaner » puis enfin, les rendre accessibles à tous les membres de l'équipe?

Il faut donc recueillir les données, les acheminer vers une base de données via des scripts rédigés spécifiquement. Puis on stocke toutes ces données sur un serveur externe, et on met en place une infrastructure qui permet à des personnes extérieures d'accéder à cette donnée.

Peux-tu décrire le quotidien d'un data analyst ?

Il y a principalement trois types de mission :

- **La recherche d'opportunités** : tu travailles en autonomie et analyse les données internes. Tu dois trouver des opportunités business et on attend de toi d'être source de proposition.
- **L'investigation** : Il s'est passé quelque chose d'anormal (typiquement la mission numéro 1 du projet 2). Pourquoi ? Comment ?
- **L'accompagnement de projets** : tu pars d'une information émanant de l'analyse de données. Ça peut être par exemple une tendance d'achat de deux produits qui ne sont pas initialement liés. Tu transmits alors l'information qu'il faut changer quelque chose dans le site Web, par exemple créer un système de recommandation du deuxième produit lors de l'achat du premier. Tu étudies l'impact sur le taux de conversion. Tu es le lead technique sur ce projet.

Quelle a été une de tes journées les plus stressantes ?

Il y a du stress quand tes supérieurs hiérarchiques ne maîtrisent pas SQL et te demandent de trouver les données dans des délais très courts, voire trop courts. Tu « détiens la vérité », dans le sens où ce sont les data qui ont toujours raison. Il y a donc une dépendance : les conclusions sur le succès ou l'échec d'un projet viennent du data analyst. Et souvent, la hiérarchie veut avoir des réponses dans des délais très courts parce qu'ils ont besoin de communiquer sur ces résultats.

Par exemple, j'ai travaillé dans le domaine de customer support analytics. On travaillait sur un projet nécessitant d'employer 1500 agents externes qui travaillaient sur des tickets. On a réalisé que, tel que le modèle de pricing était initialement défini, on perdait beaucoup d'argent dans l'emploi de ces agents. J'ai donc eu à aménager toutes les données de coût relative à ces agents pour déterminer s'il était plus avantageux de les payer à l'heure ou encore au ticket résolu... Et ainsi proposer un nouveau modèle de pricing. Il fallait trouver rapidement une solution de sorte à ne plus perdre d'argent.

Que fais-tu pour maintenir tes compétences à jour ? Combien de temps y accordes-tu ?

Comme je l'expliquais, je suis freelance, ce qui offre beaucoup de flexibilité et je souhaite m'orienter vers le machine learning ou me renforcer sur les problématiques d'architecture de données. J'ai donc en ce moment tout mon temps pour apprendre et me spécialiser dans ces domaines.

C'est un luxe que je peux me permettre en tant que freelance car en entreprise c'est bien sûr différent. Il arrive parfois que maîtriser un nouveau champ de connaissance te permettrait de gérer un projet de manière optimale même si tu es déjà en mesure de le terminer dans le temps imparti. Cependant, cette option nécessiterait d'inclure une période de formation à un nouvel outil et rallongerait donc le délai. Il faut alors négocier avec son chef et c'est souvent un débat classique : est-ce que je livre un projet plus rapidement avec des outils que je maîtrise, ou est-ce que je profite pour explorer d'autres technologies où il y a peut-être un impact un peu négatif sur le court terme, mais très bénéfique - pour tout le monde - dans le long terme.

Quel serait ton emploi idéal dans l'analyse de données ?

Je peux passer beaucoup de temps sur les données d'une entreprise qui me plaît. Je pense qu'il faut aimer les données, le sujet des données que tu manipules. Si tu adores la musique, c'est cool de bosser sur des données musicales, dans une boîte comme Spotify ou encore Soundcloud. Et c'est top aussi si tu n'es pas cantonné à un rôle très spécifique sur la data et de pouvoir jongler entre différents aspects du travail de Data Analyst. C'est justement ce que je fais en freelance.