

Guide du gain de temps à l'usage du data scientist

Votre fonction de data scientist - ou expert en métadonnées - se situe à l'interface de la science, de l'ingénierie, de la business intelligence et des mathématiques. Elle vous amène à jongler avec bon nombre de composantes, depuis le codage jusqu'à la visualisation des données en passant par le nettoyage des datasets et la gestion des responsabilités ponctuelles quotidiennes. Sauf strict suivi effectif de votre temps, vous seriez surpris de découvrir comment ces tâches mangent le temps. Dans une récente enquête menée par HP auprès de 350 data scientists de par le monde, 48 % déclarent consacrer plus de temps à organiser leurs données qu'à les analyser réellement.¹

Votre temps est précieux, d'où l'importance cruciale des gains d'efficacité dans ce métier. Qui aurait cru que l'excellence dans votre fonction peut être autant tributaire de la gestion du temps que de la maîtrise des subtilités de SQL ? Voici donc six conseils pour optimiser votre workflow et utiliser votre temps au mieux.



Table des matières

Gagner du temps grâce à une communication proactive	3
Gagner du temps par l'appréhension anticipée de vos données	4
Gagner du temps en dotant votre espace de travail des accessoires utiles	5
Gagner du temps en dégagant vos schémas de productivité	6
Gagner du temps grâce à des outils et configurations appropriés	7
Gagner du temps grâce à l'automatisation du workflow	8
L'avantage des stations de travail HP	8



1

Gagner du temps grâce à une communication proactive

Pour gagner du temps, il est primordial de créer des points de contact, en termes de communication, de manière à prendre les bonnes décisions tout au long du projet. 40 % des data scientists interrogés indiquent qu'ils commencent souvent à travailler sur les données avant même de bien comprendre les objectifs opérationnels.¹ Souvent, cette carence de communication conduit les responsables à caresser des attentes irréalistes vis-à-vis des résultats du projet.

Ce même problème peut emprunter d'étonnants détours. En fin de compte, les parties concernées et les data scientists présentent tous un profil plutôt technique - si ce n'est qu'ils emploient simplement des vocabulaires et des styles de communication différents et en viennent ainsi parfois à parler sans se comprendre. Distinction critique : Les parties prenantes ont tendance à penser en termes de résultats binaires, alors que la data science - science des données - se pare de nuances d'incertitude. Il faut impérativement

accorder votre approche orientée données à celle de vos parties prenantes.

“Les parties prenantes ont tendance à penser en termes de résultats binaires, alors que la data science se pare de nuances d'incertitude.”

Quelque technique que soit le thème, il importe d'instaurer une communication accessible pour les personnes qui sont amenées à le mettre en œuvre. Ken Jee, ambassadeur des stations de travail HP² et head of data science auprès de Scouts Consulting Group, se charge d'analyses dans le domaine sportif. “Beaucoup d'athlètes sont réticents à appliquer une solution qu'ils ne peuvent comprendre. Le fait de commencer par des modèles linéaires plus simples, même si les résultats risquent de ne pas être aussi bons, peut être un moyen très efficace de mettre en lumière la motivation intrinsèque d'une décision spécifique.” En d'autres termes, votre projet peut être mis en œuvre plus rapidement si tous les participants en comprennent la pertinence.

Gagner du temps par l'appréhension anticipée de vos données



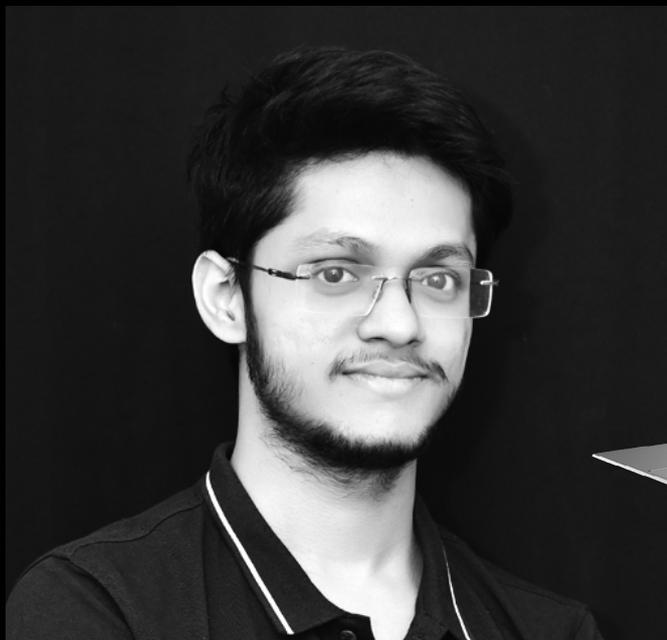
Louise Ferbach

En plus d'être précieux, votre temps se trouve souvent partagé entre des projets à long et à court termes. S'il est primordial d'utiliser votre temps au mieux, l'optimisation du temps au début d'un projet peut s'avérer particulièrement payante par la suite. Il est une erreur fréquente, et source de perte de temps en aval, qui consiste à lancer la phase de modélisation trop précocement, avant l'appréhension effective de vos données. À l'entame d'un nouveau projet, sans doute êtes-vous impatient de commencer à modéliser - c'est après tout la dimension passionnante de la mission - mais les data scientists expérimentés se gardent bien de céder à cette tentation.

Louise Ferbach, ambassadeur des stations de travail HP² et data scientist en actuariat en France, est d'avis que le temps le plus rentable consacré à un projet est peut-être le temps consacré à ce stade - celui de la recherche de corrélations et, comme le dit Ferbach, de "l'appréhension de ses données." Le temps qui y est consacré, fût-ce un ou deux jours, vous permet de dégager des schémas qui vous aideront à alimenter votre modèle. Au final, il y va d'un gain de temps considérable.

De la même manière, la documentation représente une part significative de tout projet de data science. La collecte précoce de la documentation utile offre un moyen déterminant d'améliorer l'efficacité et de gagner du temps. En d'autres termes, veillez à documenter correctement votre propre code au fur et à mesure. Un code mal documenté est une mauvaise habitude commune aux software engineers et aux data scientists. Donc, ne croyez jamais que vous vous souviendrez, ne fût-ce qu'une semaine ou, a fortiori, un mois plus tard, de ce que vous aviez l'intention de faire. Le temps que vous consacrez à la documentation vous évitera d'avoir à déchiffrer votre code à chaque fois que vous y revenez.

Gagner du temps en dotant votre espace de travail des accessoires utiles



Paras Varshney



Lors du choix du matériel, votre priorité ira sans doute au processeur et au processeur graphique. Il faut prévoir d'emblée une puissance de calcul conséquente pour répondre aux besoins de vos tâches quotidiennes et gagner du temps.

Souvent, les choses les plus simples peuvent faire une grande différence au quotidien. En général, le choix des accessoires relève d'une décision personnelle motivée par le confort, la commodité et les préférences - il n'existe pas d'accessoire panacée qui permette à chacun de gagner du temps et d'accroître son efficacité.

En matière de souris, par exemple, chacun a ses préférences. Paras Varshney, ambassadeur des stations de travail HP² et data scientist auprès de LogicAI en Pologne, n'utilise que rarement le pavé tactile de son PC portable : la souris est en effet l'un de ses gadgets préférés. Selon ses dires, il travaille beaucoup plus rapidement avec une souris.

Jee, pour sa part, ne jure que par sa trackball. Souvent amené à travailler en avion, il se réjouit de la dimension statique de la trackball, qui lui permet de travailler plus vite sur la petite tablette.

La priorité de Varshney va à l'écran : "J'adore utiliser un écran partagé. Depuis que je possède l'écran incurvé HP Z38c, je l'utilise pour garder plusieurs fenêtres ouvertes en même temps. Ainsi, nul besoin de changer de fenêtre ou de passer d'un onglet à l'autre - ils sont tous ouverts en même temps devant moi." De par sa nature même, votre mission implique d'opérer en multitâche dans de nombreuses fenêtres ouvertes, notamment des applications, des onglets de navigateur et des tableaux de bord.

En d'autres termes, l'écran du PC portable ne peut que vous ralentir quand vous ouvrez et fermez ou gérez votre espace de travail. De ce fait, le choix des accessoires utiles - de la souris aux écrans incurvés - peut vous aider à optimiser votre workflow et à gagner du temps.

Gagner du temps en dégageant vos schémas de productivité

La reconnaissance de liaisons fait partie de votre ADN de data scientist. De la même manière que vous passez vos journées à optimiser des modèles, il est possible d'optimiser vos routines quotidiennes en vue d'une meilleure productivité.



Qishen Ha

Chaque personne étant différente, un même flux de travail ne convient pas à tout le monde. À une époque où nombreux sont ceux qui choisissent de travailler à domicile, Qishen Ha, ambassadeur des stations de travail HP² et machine learning engineer auprès de LINE, reconnaît qu'un environnement de travail solitaire ne lui convient pas ; il est trop facilement distrait. Il préfère au contraire l'environnement du bureau, entouré de personnes plongées dans leur travail. "Je pense que cette formule favorise un travail plus intense", dit-il.



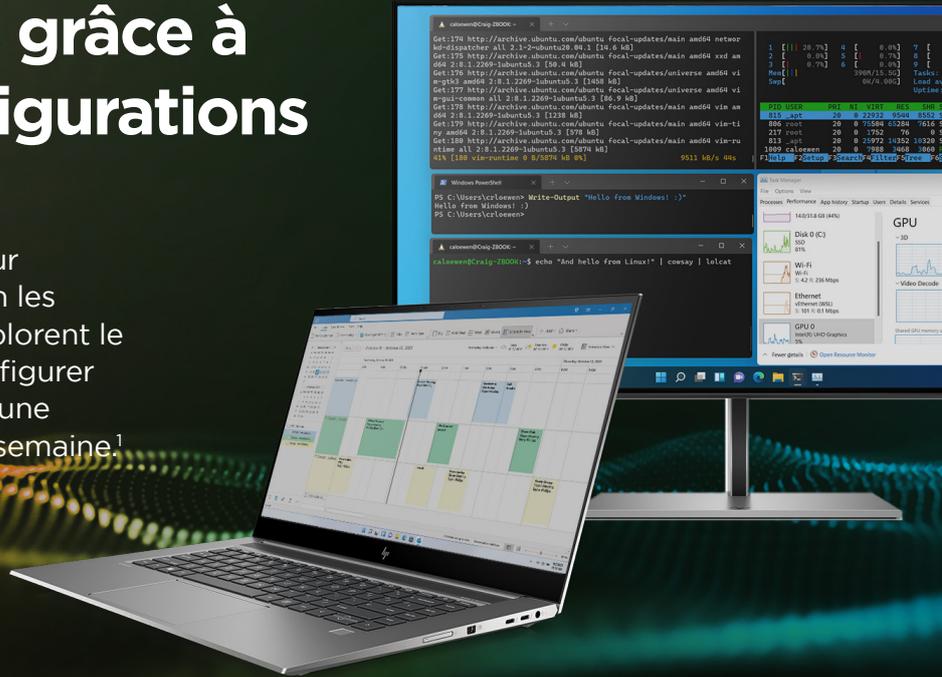
Ken Jee

Télétravailleur à plein temps, Jee a pour sa part longuement réfléchi à son workflow et réserve des plages de temps à la gestion des e-mails pendant la journée. Il déplore la perte d'efficacité induite par la dimension multitâche de sa fonction et l'envoi de courriels ; il souligne à ce propos que chaque consultation d'un e-mail pendant son travail lui fait perdre de précieuses minutes à tenter de revenir à sa tâche principale. À une époque où l'on s'attend à ce que tout le monde réagisse aux messages e-mail ou Slack pratiquement en temps réel, il importe de contrecarrer cette tendance en ménageant des plages de temps pour les e-mails de manière à optimiser l'utilisation du temps.

5

Gagner du temps grâce à des outils et configurations appropriés

La configuration d'un nouvel ordinateur constitue toujours un défi et, quand on les interroge, 42 % des data scientists déplorent le fait qu'ils passent trop de temps à configurer leur environnement de données, avec une moyenne de cinq heures perdues par semaine.¹



Pour améliorer son efficacité, il peut être intéressant d'adopter le Windows Subsystem for Linux® - WSL 2³ qui permet de faire tourner virtuellement des outils, utilitaires et applications Linux directement dans Windows sans recourir à une configuration dualboot ou à une machine virtuelle. Jee, par exemple, affirme ne plus avoir besoin d'accéder à distance à son poste de travail Linux quand il travaille sur son PC Windows - une amélioration qui réduit les frictions et accélère son workflow.

Pareillement, les piles logicielles (software stacks) préconfigurées sont apparues comme une véritable révélation pour les data scientists qui ont eu le plaisir d'en utiliser une.

“Avant d’y être confronté, je ne savais pas qu’une telle chose existait, et ce constat a entraîné un véritable chambardement”, explique Ferbach. “Le PC Data Science qui vous est envoyé par HP est équipé de ce qu’ils appellent la Data Science Software Stack.” Cette pile logicielle se compose essentiellement d’une suite complète d’applications et environnements - tous préchargés avec mises à jour automatiques, ce qui permet de prévenir les inévitables incompatibilités logicielles et les temps de dépannage qui pèsent sur les configurations de routine.

Gagner du temps grâce à l'automatisation du workflow

En fonction du stade où vous en êtes dans votre carrière en data science, sans doute recourez-vous à l'une ou l'autre forme d'optimisation de votre workflow. Il importe toutefois de rester attentif à d'autres formes, plus sophistiquées, d'optimisation de votre temps à mesure que vos compétences et votre expérience se développent. "Des compétences simples, comme le partage d'écran, sont accessibles aux novices de la data science", explique Varshney, se référant aux modalités d'utilisation optimale d'un grand écran.

Un outil plus sophistiqué pour votre boîte à outils ? Automatisation. Chaque data scientist quelque peu expérimenté connaît les avantages d'une automatisation de son workflow. Après tout, il ne faut pas longtemps pour se rendre compte que certaines tâches exigent un traitement manuel conséquent, voire souvent répétitif, au point que l'automatisation de ces tâches induit des gains de temps significatifs. Quelques commandes suffisent pour autoriser l'exécution autonome de tout un processus.

L'avantage des stations de travail HP

Comme en témoignent les conseils ci-dessus, nombreuses sont les habitudes à adopter qui contribuent à accroître l'efficacité de votre travail de data scientist. Quelques simples petits gestes peuvent contribuer à multiplier les gains de temps. Ceci étant dit, dans le cas des data scientists c'est la puissance du poste de travail qui est et restera toujours la principale source de gain de temps. Voilà pourquoi HP innove constamment pour offrir aux data scientists les stations de travail, écrans et outils hautes performances dont ils ont besoin pour une gestion homogène et transparente de leurs tâches. Découvrez [les outils et bibliothèques inclus dans la pile logicielle préconfigurée](#), proposée par HP, qui sont susceptibles de générer des gains de temps, ou découvrez [WSL 2](#) tel qu'il est proposé sur certaines stations de travail data science de HP.



¹ Etude HP : Comprendre les data scientists, novembre 2021.

² Disclaimer (Avis de non-responsabilité) : Ken Jee, Louise Ferbach, Qishen Ha et Paras Varshney participent au programme HP Workstations Data Science Ambassador et à ce titre, reçoivent des produits et équipements.

³ WSL 2 tourne sous Windows 10 ou supérieur, processeur Intel Core i5 ou supérieur et est disponible sur certaines stations de travail Z. Configuration minimale : Windows 10 version 21H2 et supérieure (Build 19044 et supérieur) ou Windows 11.