

# 10 façons de repenser l'exploration des données

Nous avons tous nos petites habitudes auxquelles nous sommes attachées. Nous employons ainsi souvent les méthodes les plus éprouvées pour atteindre nos objectifs, mais celles-ci ne nous révèlent pas toujours les informations les plus intéressantes. Voici 10 approches innovantes de l'exploration des données, qui permettront à votre équipe de mettre au jour des tendances à la fois significatives et exploitables.

## 01 Revenir aux fondamentaux

Le temps est précieux, mais vous devez éviter de vous lancer dans une étude statistique avant d'en avoir établi précisément les bases. Pour permettre à votre équipe d'identifier les données requises, il convient de définir quatre éléments essentiels à cette démarche : le contexte, les besoins, la vision et le résultat.

**Contexte** : Votre analyse porte-t-elle sur des données microéconomiques ou macroéconomiques ? Vous intéressez-vous aux tendances générales ou aux cas particuliers ?

**Besoins** : À quels besoins spécifiques de l'entreprise souhaitez-vous répondre par l'extraction de ces données ? S'agit-il de besoins à court ou à long terme ?

**Vision** : Dans quel but souhaitez-vous répondre à ces besoins ? Est-il possible de prévoir le résultat par une solution logique ?

**Résultat** : Comment évaluer la réussite de l'opération et comment démontrer que les données ont permis d'arriver à ce résultat ?

## Adopter un cadre à l'ancienne

Le plus beau cadeau que vous puissiez faire à un analyste, c'est un cadre de travail bien construit. Dans notre univers numérique, nous n'avons pas toujours le réflexe de prendre un papier et un stylo pour poser nos idées. Or, cette démarche visuelle peut être déterminante dans le processus de réflexion. Encouragez votre équipe à définir de grandes lignes directrices au cours de cette phase ; les détails viendront ensuite. Réduisez les interférences, faites le tri dans vos données et représentez tout cela dans un schéma intuitif.

Vous n'avez pas de feuille sous la main ? Prenez une serviette en papier. Il paraît que certaines grandes idées sont nées de cette manière.

## 03 Faire participer tout le monde

Indépendamment de leurs compétences, les analystes passent le plus clair de leur temps, jusqu'à 80 %, à préparer manuellement les données (nettoyage, compilation et classement). Cela ne laisse que très peu de place aux analyses qui permettent de dégager des tendances révélatrices.

Offrez à votre équipe la possibilité de consacrer moins de temps à la préparation manuelle des données (et réduisez le nombre d'erreurs associées à ce processus) en adoptant des outils automatisés en libre-service. Grâce à ces outils, vos analystes travailleront plus efficacement, et tous les autres membres de l'organisation pourront effectuer différentes tâches d'analyse annexes. Du PDG au stagiaire, chacun de vos collaborateurs peut endosser la fonction informelle d'analyste-citoyen et participer ainsi à l'exploitation globale des données de votre organisation.

## Parler la langue de son interlocuteur

Si tout le monde parlait la même langue que les analystes, la présentation des informations en serait largement simplifiée. Mais nous savons bien que ce n'est pas toujours le cas.

Une fois que vous avez identifié des statistiques exploitables, vous n'avez fait que la moitié du travail. L'autre moitié consiste à présenter ces statistiques. Il ne suffit pas d'être le meilleur analyste, capable de faire preuve de la plus grande perspicacité face aux données les plus complexes. Il faut également savoir présenter son travail de façon claire, concise et pertinente, pour éviter que celui-ci finisse sur une étagère à prendre la poussière.

Si vous vous adressez à un PDG, par exemple, mettez l'accent sur les statistiques globales. En revanche, un responsable de la veille stratégique pourra s'intéresser à la méthodologie et aux techniques d'exploration des données. Vous devez donc penser à adapter la présentation de vos statistiques en fonction de votre interlocuteur, de manière à parler la même langue que lui. Cela jouera sur son adhésion ou non à votre argumentaire et déterminera l'impact que vos conclusions pourront avoir par la suite.

## 05 Vérifier la date d'expiration des données

Si l'y a bien une chose qui est certaine concernant les données, c'est qu'elles évoluent en permanence, tout comme les statistiques que l'on peut en extraire. Cependant, il nous arrive parfois d'oublier que ces statistiques ne sont pas éternelles. N'hésitez pas à rappeler aux membres de votre équipe que leurs données ont une durée de vie, car cela peut faire toute la différence. Ils interagissent avec celles-ci de façon radicalement différente, et les actions à effectuer découlant de ce travail seront elles aussi radicalement différentes. S'il peut être tentant de s'attacher à un ensemble de données que l'on connaît bien, dont la création a demandé de nombreux efforts ou sur lequel on a longuement travaillé, encouragez votre équipe à se débarrasser des données qui ne sont plus pertinentes.

## Confirmer les résultats par des tests

La corrélation n'est pas une relation de cause à effet, mais il est tout à fait humain de voir ce que l'on souhaite voir de temps à autre. En effet, les grands ensembles de données font souvent apparaître des corrélations apparemment évidentes, mais trompeuses. Lorsque l'équipe attend ou cherche un résultat précis, elle a tendance à donner une signification à ces corrélations pour étayer son hypothèse de départ. Afin d'éviter de telles erreurs, les tests et les expériences en conditions réelles comptent parmi les méthodes les plus efficaces. Ainsi, le test univarié est idéal pour identifier une conséquence directe et prévisible, mais il peut parfois s'avérer irréalisable (comprenez : il est long et coûteux). Alors, pourquoi ne pas essayer le test multivarié ? Dans la mesure où il est contrôlable, ce type d'encouragement expérimental est extrêmement précieux et permet d'obtenir des statistiques objectives donnant des indications fiables, même si celles-ci vont à l'encontre de l'hypothèse initiale.

## 07 Éviter la paralysie analytique : voir petit pour de grands résultats

Le volume de la base de données ne détermine pas la qualité des statistiques. En effet, les grandes bases de données peuvent entraîner une paralysie analytique effroyable. Comment trouver le bon équilibre avec un échantillon de taille à la fois suffisante et raisonnable ? Dans un article publié en 2013, Hal Varian, Chief Economist chez Google, expliquait que les équipes Google utilisent généralement des échantillons aléatoires représentant 1 % des données disponibles, voire moins, pour réaliser la plupart de leurs analyses. En effet, selon les experts, les grands échantillons peuvent avoir une incidence préjudiciable sur les résultats, et ce pour deux raisons :

- L'extraction et l'analyse de grands volumes de données demandent énormément de temps.
- Plus la base de données est volumineuse, plus la probabilité de justifier des hypothèses erronées est grande.

À éviter d'immobiliser vos équipes, en particulier lors des phases d'expérimentation, faites-leur étudier un petit échantillon, réellement aléatoire, pour voir ce que cela donne. Vous pourriez être surpris.

## Faire collaborer les données

Dans le cadre d'un projet, la facilité consiste à regrouper plusieurs ensembles de données distincts pour former "lacs de données". Nous recommandons par conséquent de nous appuyer sur des collaborateurs d'établir quelques liens entre ces données. Par exemple, les clients d'un magasin fournissent de nombreux éléments d'information. En règle générale, un détaillant peut identifier les articles qui sont achetés ensemble, les meilleures ventes selon les saisons et l'impact des bons de réduction sur des articles particuliers, entre autres. Cependant, pour appréhender un écosystème dans son intégralité, les équipes doivent s'appuyer sur de nombreux ensembles de données disparates pour y rechercher des corrélations pertinentes.

Quelle que soit votre activité, il est important d'identifier les tendances et d'anticiper leur évolution dans chaque ensemble de données, mais il est vital de savoir assembler toutes les pièces de ce puzzle pour avoir une véritable vue d'ensemble. À l'instar d'une équipe, les données sont d'autant plus efficaces et utiles qu'elles travaillent en bonne collaboration.

## 09 Adopter une méthode itérative, plutôt que linéaire

Chacun de vos projets peut être comparé à une course dont le tracé ne serait pas une ligne droite, mais une boucle, avec parfois un chronomètre en bruit de fond. Le colonel John Boyd de l'US Air Force a mis au point la célèbre boucle OODA, une méthode qu'il recommande à ses stratèges militaires, et qui consiste en quatre actions à répéter : Observer, s'Orienter, Décider et Agir. Tout l'intérêt de ce concept réside dans sa répétition. En appliquant cette méthode, une équipe peut tester les solutions proposées afin de voir si elles sont efficaces, puis les reprendre de manière différente pour les ajuster si nécessaire, plutôt que de procéder de façon linéaire. Cela permet de ne pas s'attarder trop longtemps sur l'examen d'une situation inédite ou sur la recherche d'une solution "idéale". Même si elle peut donner l'impression de tourner en rond, c'est une méthode efficace, quel que soit le projet.

## Donner un supplément de sensibilité humaine aux données

La technologie joue un rôle essentiel dans l'analyse des données, au même titre que l'intervention humaine. En effet, la technologie permet d'extraire et d'analyser d'immenses volumes de données, mais c'est à vous et à votre équipe qu'il appartient de donner au sens aux résultats ainsi obtenus et de définir la stratégie de votre entreprise. Voici quatre conseils qui vous permettront de garder un œil critique sur la situation :

- Concentrez-vous sur les tendances, pas sur des points de données isolés.
- Pensez à l'esthétique, pour faciliter la compréhension et la transmission du message.
- Cherchez des corrélations statistiques pertinentes, puis testez-les.
- Restez sceptique. Gardez à l'esprit que nous avons tendance à voir ce que nous souhaitons voir !

L'exploration des données prend une place toujours plus importante dans le processus décisionnel des entreprises. Toutes les équipes, même les plus efficaces, ont intérêt à revoir leur façon de traiter les projets en particulier, et les ensembles de données de manière générale. Nous espérons que vous et votre équipe pourrez mettre à profit ces dix conseils pour extraire ces bijoux enfouis dans vos données brutes et mettre en œuvre des stratégies payantes.