



3

ÉTAPES CONCRÈTES POUR MIEUX CONNAÎTRE LES CLIENTS AVEC LE BIG DATA

Maîtrisez le potentiel du Big Data pour
un business agile & performant

A PROPOS DES AUTEURS :



Gérard Danaguezian est un spécialiste des systèmes de collecte et d'analyse de données. Il a fondé et dirige depuis 1986 l'éditeur de logiciel de Marketing Intelligence Soft Concept. Il a piloté pour des grands comptes de nombreux projets de mise en place de systèmes de collecte de données, de reporting et de dashboarding. Il est diplômé de l'EM Lyon.



Alexia Lardon est la rédactrice en chef de Survey-Magazine, le magazine des nouvelles technologies de collecte et d'analyse de données. Elle a travaillé au département marketing des groupes L'Oréal, Société Générale ainsi qu'à la cellule digitale des Grands Magasins Printemps. Elle est diplômée de Kedge Business School et de l'Université Bocconi.

A PROPOS DE SOFT CONCEPT :

Soft Concept est l'un des principaux éditeurs de logiciels de Marketing Intelligence. Les outils Soft Concept aident des milliers de professionnels en France et dans le monde à collecter, analyser et visualiser tous types de données permettant de mieux connaître et satisfaire leurs clients.



Big Data

Passer des généralités au concret

Tout le monde parle de Big Data. Les livres blancs sur le sujet fleurissent. Les webinars, workshop et Mooc se multiplient. Des salons et congrès professionnels sont désormais consacrés à cette nouvelle discipline. Bref, le phénomène Big Data semble incontournable. Les entreprises qui ne s'y intéresseraient pas sont promises aux pires déconvenues alors que celles qui s'y lancent convoitent la fortune marketing et la prospérité commerciale.

Bien sûr, devant cette agitation et ce qui semble être le Graal du moment, tout responsable raisonnable se doit de réfléchir. Mais lorsqu'on regarde les choses de plus près, une question se pose inévitablement :

Comment mon entreprise peut tirer parti concrètement du Big Data ?

Pour y répondre, il convient de commencer par bien définir ce qu'on entend vraiment par Big Data. Essayez de le faire en quelques phrases sans tomber dans les sempiternels 3, 4, 5 ou n V qui sont davantage des caractéristiques du phénomène que le véritable socle d'une définition.

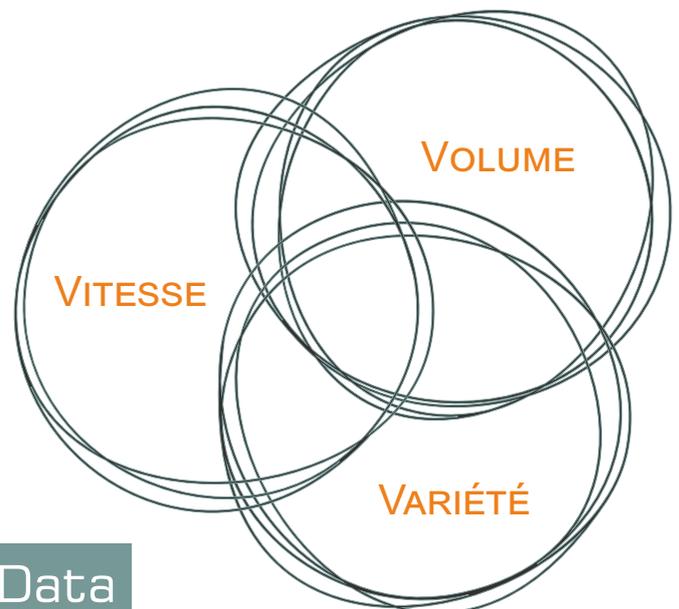
Certes le Big Data repose sur des Volumes faramineux de données que l'on illustre volontiers en évoquant les 7 milliards d'octets générés quotidiennement sur Twitter ou les 300 heures de vidéos prises en charge chaque minute par Youtube.

Certes la Vitesse phénoménale de circulation des informations ne cesse de croître, atteignant des niveaux parfois difficiles à appréhender.

Certes la Variété des données est désormais une constante avec laquelle les professionnels de l'informatique, habitués à des formats traditionnels et bien structurés, doivent pourtant composer.

Certains ajoutent aussi la Véracité, la Valeur, la Visibilité et d'autres V encore.

Tout cela étant dit et redit, on reste souvent sur sa faim. Tout le monde comprend bien qu'on dispose d'extraordinaires volumes de données, qui arrivent plus vite et qui se présentent dans des formats variés. Mais cela ne permet pas de s'imaginer concrètement ce dont il s'agit dans le contexte particulier de chaque entreprise et la manière concrète qui nous permettrait d'en bénéficier.



Les 3 V du Big Data

Oui, mais encore ?

En réalité, le problème principal qui empêche de bien appréhender l'exploitation que l'on pourrait faire du Big Data réside dans la nature même des données dont il est question. En effet, les entreprises ont toujours été habituées à manipuler des données qu'elles maîtrisaient à la fois en termes de localisation et de structure. Or les données évoquées lorsque l'on parle de Big Data ne correspondent que très minoritairement à ces délimitations.

Soyons clairs : chaque entreprise dispose aujourd'hui d'un volume important de données nouvelles qui lui parviennent quotidiennement. Il peut s'agir d'e-mails, d'appels de clients, de courriers de réclamations, de remontées d'informations des points de vente ou de contact, d'éléments issus des dispositifs de fidélité, de statistiques de fréquentation du site web, de résultats de campagnes de promotion multi-canal, etc. Tout cela vient s'ajouter aux données financières et commerciales, de chiffre d'affaires, de volumes des ventes, panier moyen, taux de transformation, fréquentation, parcours client, etc.

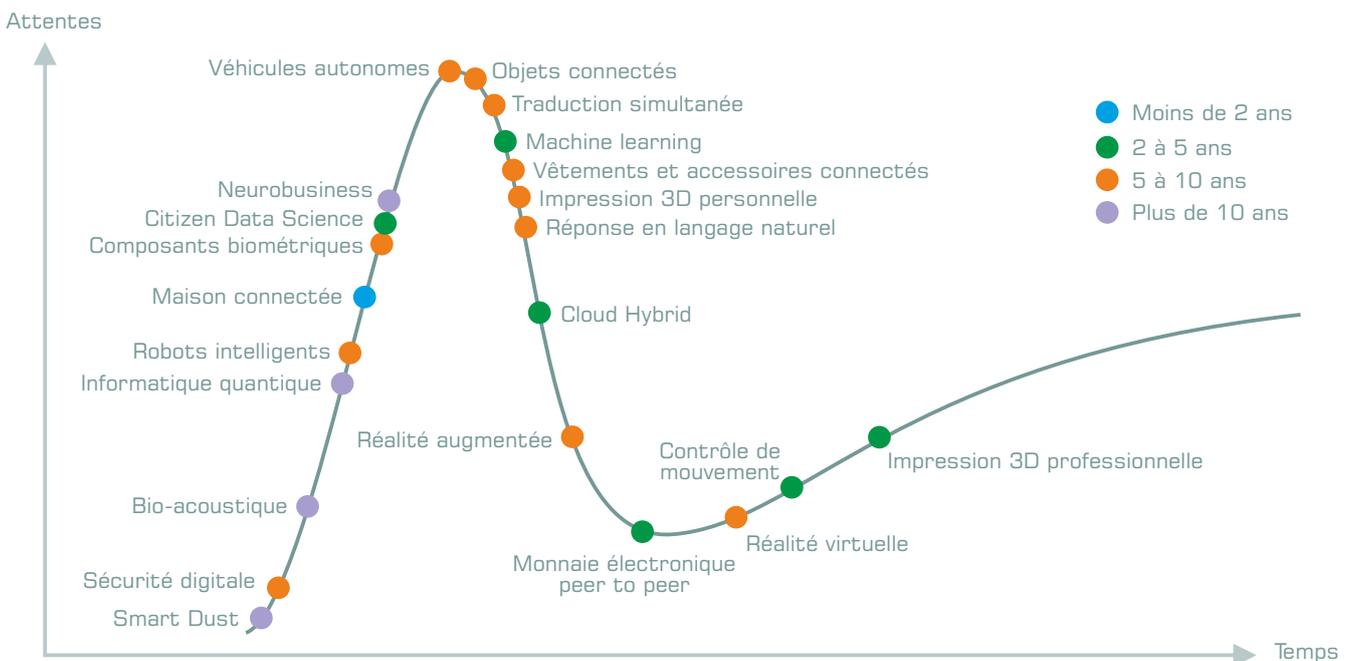
Une partie seulement de ces données est actuellement prise en compte et analysée avec les outils traditionnels de l'entreprise : CRM, ERP, SGBD relationnels, Outils BI, tableaux Excel, etc. Beaucoup de datas existantes restent inexploitées par manque de temps ou de ressources. C'est par exemple le cas pour beaucoup de données issues de systèmes transactionnels et de gestion de la relation clients.

Mais lorsqu'on parle de Big Data, on ne s'arrête pas seulement à ces données internes exploitées ou pas, même si leur volume a considérablement augmenté. On fait également référence à de nombreuses autres informations qui peuvent intéresser l'entreprise et qui sont disponibles ailleurs que dans l'entreprise. Il peut s'agir d'opinions émises sur les réseaux sociaux ou les sites d'avis consommateurs, de commentaires sur des blogs ou des forums, de contenus extraits de vidéos YouTube, d'actualités provenant de sites de presse, de sources statistiques externes, d'informations présentes sur des sites web institutionnels ou concurrents...

A première vue, ces données semblent aléatoires et parfois pas très intéressantes. En réalité, si on a vraiment les moyens de collecter les bonnes informations et de les exploiter de manière optimale, toutes ces données peuvent non seulement fournir une valeur ajoutée pour l'entreprise mais représenter également un véritable avantage concurrentiel et distinctif.

Courbe d'adoption des technologies

Source : Gartner Report 2015



Contrairement aux années précédentes, le Big Data n'apparaît plus sur la courbe. Cela doit alerter les entreprises n'ayant pas encore intégré cette technologie dans leur stratégie numérique.

Data Collection

Réunir les données

La clé du Big Data se trouve clairement dans la capacité à collecter, stocker et combiner des datas multi-sources et multi-formats, qui mixent des données de natures et de formats habituels avec des informations nouvelles, issues du web et d'autres sources externes et qui arrivent très majoritairement dans des formats textuels et non structurés. Avant d'envisager une quelconque exploitation, deux questions principales viennent à l'esprit.

Comment aller chercher des données dont on ne maîtrise pas la source ?

Pour répondre à cette question, des plateformes de veille et d'analyse des médias sociaux se sont développées sur le web qui promettent aux entreprises de scanner tous les médias en ligne à la recherche des mots clés qui pourraient les intéresser (marques, catégories de produits...). Les résultats s'expriment en tableaux de bord qui indiquent l'évolution de citation d'un produit ou d'une marque, assortie d'indicateurs de tonalité (positif/négatif) des commentaires.

L'approche est bigrement séduisante mais souffre malheureusement du triple syndrome de la « boîte noire », de la simplification réductrice et d'approximations parfois erronées. Qui plus est, on n'obtient là que des tableaux de bord et des indications qui restent dans un écosystème en mode SaaS, externe et propriétaire. Ces offres extrêmement performantes sur le plan commercial le sont donc bien moins du point de vue opérationnel et SI.

Quelles technologies de collecte et de stockage utiliser ?

La référence au Big Data s'accompagne généralement de l'évocation de différentes technologies nouvelles censées se substituer à celles en cours dans les entreprises. Hadoop, MapReduce et autres NoSQL inspirées des pratiques des géants du Web s'affichent comme les nouveaux outils et pratiques de stockage et de gestion des nouvelles sources de données.

A y regarder de près, la démarche qui consiste à opposer le monde du Big Data à celui des SI existants semble être une posture avant tout commerciale. En effet, ceux qui professent la mise au rebut des systèmes matériels et logiciels existants et leur remplacement par des nouveaux équipements et infrastructures destinés à prendre en charge les nouveaux formats de données, paraissent viser en réalité à vendre encore plus de hardware, de software et de prestations de consulting. Business as usual !

Le Big Data repose certes sur des approches nouvelles et des modes de prise en charge des données qui diffèrent de ceux que les professionnels de l'informatique ont apprises. Habituellement, on accole immédiatement les qualificatifs « Relationnelle » ou « Structurée » à la notion de Base de données. En effet, comment stocker des éléments si ce n'est de manière organisée et bien pensée. Pourtant, les nouvelles données en provenance du web, sont, par nature, textuelles et non structurées. Les volumes en jeu imposent également des modes d'organisation et de stockage non conventionnel. Imaginez par exemple que Google cherche à stocker tous les contenus qu'il génère sur un très gros serveur et dans une immense base de données relationnelle. C'est strictement impensable et c'est pour cela que les nouvelles structures de type Hadoop ont été mises en place.

Mais nous ne sommes pas tous Google. Nous n'avons pas forcément les mêmes besoins et les mêmes volumes. Ce n'est pas une raison pour ne pas s'intéresser aux nouvelles technologies de stockage réparti et non structuré. Mais ces technologies ne sont pas non plus des substituts purs et simples de nos systèmes habituels et il convient de les aborder en étant conscients qu'ils renferment souvent des coûts cachés et nécessitent des ressources particulières.

En se focalisant sur les choix technologiques, on ne fait en réalité que retarder la prise en compte du Big Data dans son activité. SQL ou NoSQL, that is not the question. Les vrais enjeux consistent plutôt à savoir si on est capable de se doter de la capacité d'aller rechercher ce dont on a besoin, facilement, rapidement et de manière automatisable. Loin de rendre obsolètes les données « classiques » de l'entreprise, les nouvelles sources doivent donc se concevoir comme un complément capable d'apporter une vision élargie et des performances supplémentaires.

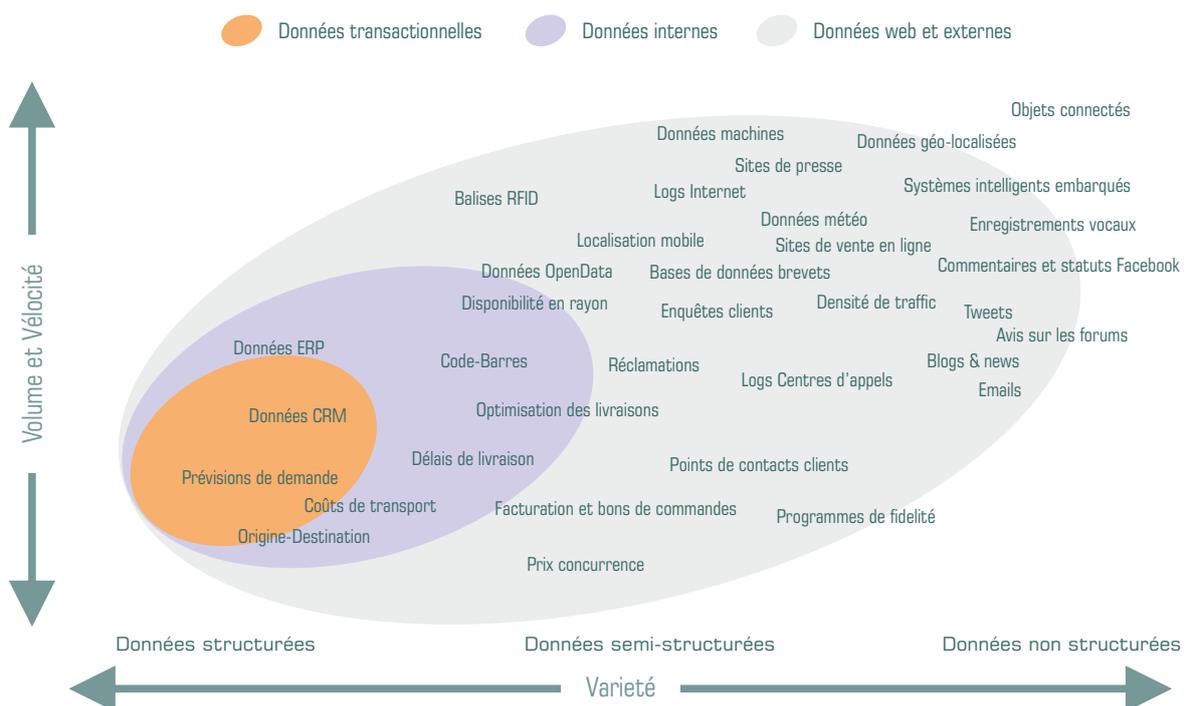
La première étape d'une opération de collecte de données consiste à identifier les sources utiles en fonction d'objectifs définis. Aujourd'hui la variété d'informations disponibles rend possible des réponses à une large panoplie de problématiques qui nécessitaient auparavant le recours à des études complexes et coûteuses. Ainsi, on peut chercher à mieux connaître ses clients, mieux comprendre leurs comportements et leurs usages, repérer des insatisfactions et des réclamations, évaluer son image de marque, mesurer l'impact de ses actions de communication, suivre ses concurrents, remonter des données pour mieux gérer une crise, repérer des tendances et des évolutions du marché, entre mille autres choses. Pour cela, il faut avoir la capacité technique de récupérer les données identifiées, dans un format exploitable. Ainsi, il est possible à présent de récupérer les tweets émis sur un sujet donné, d'identifier et de collecter les likes et les commentaires sur une page Facebook (et depuis peu, repérer les posts émis sur un thème défini en capturant les messages postés sur les pages concernées),

de charger les textes des messages d'un forum filtrés par période et par topic, de récupérer les évaluations qualitatives et numériques de ses produits et de ses concurrents sur les sites d'avis consommateurs. Toutes ces opérations ont comme intérêt supplémentaire de faire remonter des données concernant des segments nouveaux de clientèle, que l'entreprise peut ne pas avoir identifiés ou qu'elle n'a pas pour habitude d'adresser et de solliciter.

Mais le gisement de données pouvant intéresser l'entreprise va aujourd'hui bien au delà des seules conversations sociales. Il peut s'étendre aux données présentes sur des sites concurrents, aux nombreuses ressources OpenData désormais à la disposition de tous mais que peu de professionnels savent vraiment exploiter, aux informations disponibles sur les sites de presse, les sites boursiers, les bases de données de brevets, les sites institutionnels, les plateformes de vente en ligne...

Il faut donc être en mesure de capturer facilement tous les éléments composant le bouquet de sources utiles, de manière ponctuelle mais aussi en mettant en place des processus automatisés de capture périodique et de mise à jour des informations. Pour être utilisables toutes ces données doivent pouvoir être stockées dans des formats et des supports adaptés à une exploitation statistique et sémantique.

Répartition des données selon les 3V



Data Analytics

Analyser les données

Recueillir des données c'est bien. Savoir les traiter, c'est tout de même mieux.

Le problème avec les données issues du Big Data et des réseaux sociaux saute aux yeux : il s'agit en très grande partie de données textuelles que l'on n'a pas l'habitude d'analyser et d'exploiter. Certains estiment que ces données textuelles représentent 80% de l'ensemble des données disponibles. D'autres poussent jusqu'à 90% sans que l'on sache si ces estimations sont bien basées sur une évaluation scientifique (dont on a du mal à imaginer le protocole) ou issues de l'approche bien connue du doigt mouillé.

Quoi qu'il en soit, les faits sont là : on a bien affaire à des données particulièrement peu adaptées aux analyses statistiques habituelles.

Comment tirer le meilleur parti des données recueillies ?

Avant de s'intéresser au traitement des données textuelles, rappelons tout de même que celles-ci sont rarement isolées. En effet, toute donnée récupérée de quelque source que ce soit est généralement accompagnée de données complémentaires. C'est le cas par exemple d'un tweet qui a un émetteur, une date et heure d'émission, l'indication de la plate-forme source, le nom, le pseudo de l'auteur (@) et son nombre de followers... Ces métadatas permettent de caractériser les messages, de les classer et de les lister. Le contenu même des messages comprend des mots clés (hashtags) facilement repérables dans le cas de twitter puisqu'ils sont précédés du symbole #. Rien que ces quelques éléments laissent entrevoir certaines possibilités de traitement et de listage « classiques ».

Parfois, les éléments textuels sont accompagnés d'une note qui renseigne sur leur tonalité. C'est le cas des évaluations de produits sur les sites de consommateurs ou sur les sites marchands. L'utilisation de ces étoiles ou échelles de notation permet de catégoriser les avis selon leur note.

Cependant, le traitement véritable des données textuelles nécessite des méthodes complémentaires que l'on n'a pas l'habitude d'utiliser dans l'analyse quantitative et l'exploitation des données structurées. Il s'agit du Traitement Automatique de Langage (TAL) qui a fait d'énormes progrès. Les outils qui intègrent ce type de traitement permettent aujourd'hui de repérer facilement les sujets abordés dans les commentaires et de détecter automatiquement leur tonalité. Bien sûr, les méthodes automatiques n'ont pas une fiabilité absolue et présentent souvent des insuffisances dans la détection de nuances, la compréhension de l'ironie ou la décodification de structures de phrases complexes ou mal formulées. Les thuriféraires de l'analyse automatique des sentiments affirmeront toutefois que les algorithmes ont les avantages de leurs défauts puisqu'ils se trompent toujours de la même manière alors qu'un même texte présenté à plusieurs évaluateurs humains voire à une même personne à différents moments, peut se voir attribuer une tonalité différente.

Si le traitement des données textuelles prend en charge les unités inhabituelles que forment les mots, le résultat de l'analyse correspond finalement à une recodification en rubriques. Par exemple, un corpus de commentaires de clients peut se résumer par un nombre de citations positives ou négatives pour chacun des thèmes habituels : prix, délais, accueil, qualité de service, etc. Les résultats peuvent donc souvent être stockés dans des structures traditionnelles pour être réutilisés de manière classique (ex : graphe d'évolution des scores dans le temps pour chacun des critères identifiés).

Le traitement des données issues du Big Data ne doit pas être vu comme distinct de celui des datas décisionnelles habituellement manipulées par les entreprises, bien au contraire. En effet, les données du Big Data doivent venir compléter et enrichir les contenus habituels du système d'information de l'entreprise, en apportant des éléments nouveaux et non sollicités qui permettront d'élargir l'angle de vue du décideur. Ainsi les données issues des réseaux sociaux peuvent, par exemple, venir compléter utilement les enquêtes de satisfaction de l'entreprise ou les audits de points de vente. Tous ces éléments peuvent s'agrèger également avec les chiffres de vente et de performance, les données des programmes de fidélité, pour offrir un tableau complet et pertinent de la situation.

La confrontation d'éléments en provenance de sources diverses peut faire ressortir d'éventuelles corrélations ou effets induits. L'agrégation des différents types de données peut se faire sur la base d'éléments communs qui peuvent être des codes magasins, régions, clients, produits ou même de simples dates.

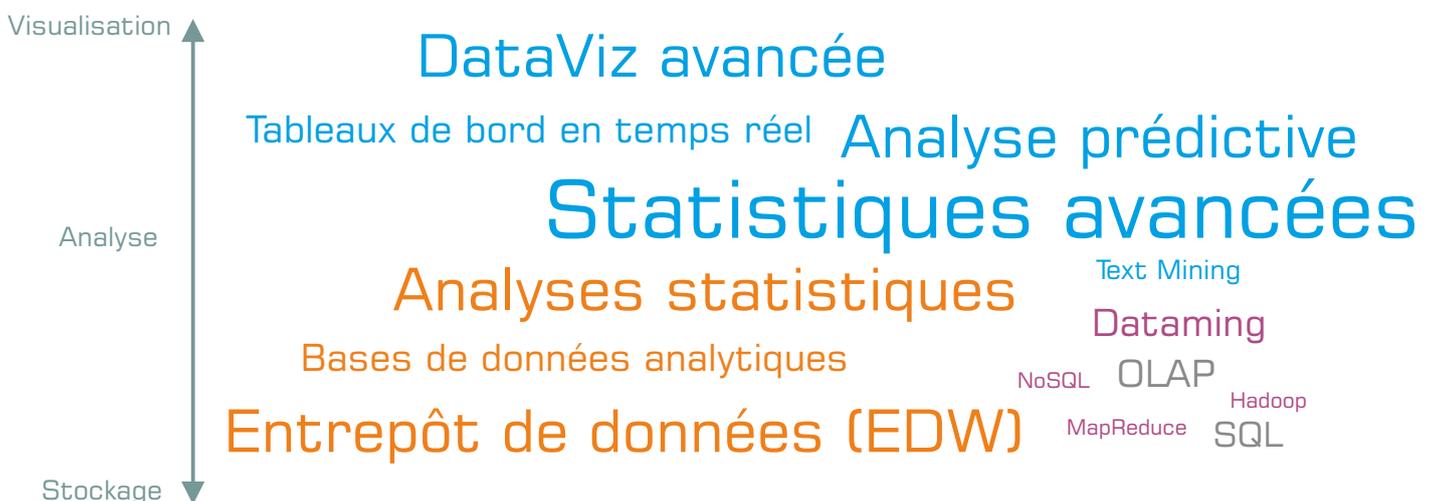
Les méthodes statistiques avancées permettent aujourd'hui de faire ressortir les relations intrinsèques entre les différentes variables. Les analyses factorielles, régressions, recherches de croisements significatifs, arbres de décision ou méthodes de détection automatique de l'importance des facteurs de satisfaction peuvent mettre en évidence des éléments opérationnels de grande utilité. Ces méthodes actuellement utilisées par les seuls spécialistes, mériteraient d'être mises à la disposition des équipes marketing et opérationnelles, dans des formes simplifiées et automatisées, permettant de repérer rapidement des éléments significatifs dans les masses de données analysées.

Concernant la partie textuelle, il est possible de recourir aux méthodes de traitement automatique du langage évoquées plus haut de manière plus ou moins poussée, selon son métier et le contenu des données textuelles collectées. Plutôt que d'être rebuté par la complexité de mise en oeuvre de processus sophistiqués d'analyse sémantique automatique ou du très en vogue concept de machine learning, on peut aborder le problème en mettant en place des systèmes plus simples d'identification dans les textes de mots-clés liés à sa profession et à ses produits. Le système automatique et simple ainsi mis en place peut servir à ventiler ces éléments qui seront ensuite validés par leurs utilisateurs. En effet, il ne faut pas perdre de vue que le traitement des commentaires a pour objectif d'effectuer des actions concrètes en conséquence (résolution d'insatisfactions, réponse à des réclamations, intervention auprès d'influenceurs...). Généralement, ce sont des humains qui s'en chargent, jusqu'à preuve du contraire, et ces humains peuvent déjà, si on leur facilite la tâche, agir rapidement en disposant des éléments à traiter, quitte à requalifier certains d'entre eux. Les buts opérationnels sont souvent aux antipodes des objectifs des chercheurs et ne relèvent pas d'une volonté de perfection ou d'omniscience. On a donc tout intérêt à penser chaque approche de manière pratique et efficace pour ne pas retarder l'exploitation des ressources Big Data.

Si l'entreprise dispose donc à la fois des capacités de traitement des données textuelles et d'outils permettant des analyses statistiques avancées, elle peut envisager une exploitation optimale de ses données Big Data combinées à ses données internes de performance ou de remontées clients.

Technologies les plus exigeantes en ressources

Potentiel de croissance : ● Très fort ● Fort ● Moyen ● Faible



DataViz

Visualiser et exploiter les résultats

La conjonction des technologies d'analyse rapide évoquées plus haut avec le flot ininterrompu de données multiples et variées peut provoquer une overdose informationnelle. C'est l'infobésité, mot taillé sur mesure pour illustrer « cette masse grasse formée par le bombardement d'informations qui étouffe nos processus intellectuels ». Extrapolé sur l'entreprise, l'excès de résultats peut bloquer les processus organisationnels et décisionnels.

L'objectif de la collecte et de l'analyse des données Big Data mais aussi de toutes les autres données opérationnelles utiles a pour objectif d'accélérer et de fluidifier la prise de décision et l'action dans un environnement de plus en plus complexe et mouvant. Pour ne pas que le torrent des analyses Big Data ne soit contre-productif, il faut s'interroger sur la meilleure manière d'organiser l'exploitation des résultats obtenus.

Comment transformer les résultats en actions ?

Evaluons d'abord les risques. Trop de résultats peuvent paralyser l'action surtout si leur prise en compte nécessite du temps de lecture et d'interprétation. C'est le cas lorsqu'on est en présence de tableaux et rapports interminables avec des éléments nombreux et détaillés.

De même, l'information peut ne pas arriver dans des temps raisonnables au bon interlocuteur capable de comprendre ses implications et de mener les actions qui s'imposent. Certaines observations et analyses remontées du terrain peuvent emprunter des cheminements complexes et arriver dans des délais qui ne permettent plus d'envisager des initiatives concrètes.

Un dernier écueil réside dans le caractère parfois trop général des résultats consolidés. Pour agir, chaque niveau a besoin de ses résultats précis, si possible mis en perspective par rapport à la situation d'ensemble et aux autres entités.

Pour permettre une véritable exploitation des données, il convient donc de s'intéresser à la forme des résultats, à leur mode d'accès et à leur caractère personnalisé. C'est sans doute ces exigences qui expliquent le développement des outils de DataViz destinés à simplifier des éléments complexes et à faciliter l'accès à l'information dans des formats agréables et adaptés.

Une image vaut mille mots. Un graphique peut valoir des dizaines de tableaux. Les différents types de mises en forme désormais accessibles permettent de traduire des résultats bruts en présentations éclairantes qui font ressortir les éléments essentiels. Dans le même temps, une représentation graphique adaptée permet d'avoir une vision globale et comparative des différents éléments, positionnés et/ou dimensionnés en fonction de leur importance respective.

La simplification de l'accès est également un enjeu majeur. Les systèmes de dashboards et de web-reporting dynamique permettent d'envisager des accès différenciés pour chaque utilisateur. On peut ainsi donner des résultats directement adaptés à chaque profil en mettant en avant les éléments sur lesquels la personne concernée est en responsabilité et peut agir.

Ces simplifications visuelles et en termes d'accès permettent non seulement d'identifier des données pertinentes de manière plus claire mais également d'accéder à des benchmarks rapides et informatifs. Une entité peut suivre ses résultats et voir d'un coup d'oeil sa situation comparée à celle d'autres entités ou de l'ensemble de l'entreprise. Des actions peuvent être déclenchées et suivies facilement par toutes les personnes concernées.

La visualisation graphique est la clé. Outre son efficacité, elle colle aux besoins d'immédiateté et de simplicité. Dans notre société de l'image, 90 % de l'information qui parvient à notre cerveau pour être scannée et assimilée est visuelle. La mise en adéquation des résultats avec nos prédispositions cérébrales ne peut être qu'efficace. Il est d'ailleurs possible de prolonger encore l'efficacité des tableaux de bord et rapports dynamiques de deux manières.

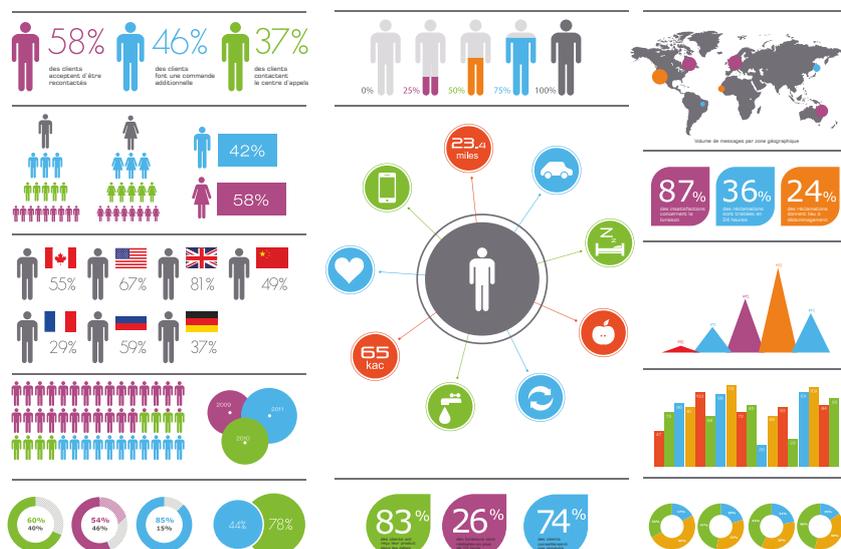
La première consiste à cultiver l'interactivité des résultats. Derrière chaque zone d'un graphique peuvent se cacher d'autres éléments graphiques, textuels ou tabulaires riches d'enseignements. En zoomant ainsi sur les éléments qui les intéressent, les utilisateurs peuvent explorer les données et afficher les informations sous-jacentes de manière plus facile et plus efficace qu'au travers d'interminables arborescences de menus. Ce mode d'utilisation correspond finalement aux interfaces graphiques habituelles de nos ordinateurs et tablettes qui permettent de naviguer en cliquant sur des objets graphiques plutôt qu'en tapant des commandes. En interagissant ainsi, les utilisateurs des tableaux de bord de résultats font donc appel à un mode de fonctionnement naturel et efficace, qui leur permet d'accéder plus rapidement aux résultats qui leurs sont utiles.

Le deuxième manière d'optimiser encore l'utilité de l'information graphique consiste à l'utiliser pour communiquer plus largement en interne et en externe. L'une des idées attachées au Big Data est l'influence. Les résultats graphiques peuvent être un vecteur de cette influence. Nous sommes habitués aujourd'hui à voir dans la presse de belles infographies, présentant de manière condensée et agréable de multiples données résumées en

métaphores visuelles. Certaines sont de véritables oeuvres artistiques et des coups de force intellectuels. La plupart arrive à résumer en peu de place des informations qu'il serait trop long et ardu de retranscrire sous d'autres formes. Sans viser un tel niveau de perfection, il est possible avec un bon outil d'analyse et de visualisation de produire de véritables infographies de qualité, qui pourront transmettre à des équipes internes ou à des tiers (journalistes, clients, visiteurs de la page facebook, internautes...) des résultats condensés et percutants. L'infographie a une viralité propre. On peut facilement la communiquer et la partager. Elle constitue donc un outil de communication de premier choix, qui permet de transmettre un message facile à comprendre et à relayer.

Notons que l'infographie est, encore plus que les dashboards ou web-reports, un moyen efficace de combiner des résultats sur des données de plusieurs provenances. Il est en effet aisé et naturel de représenter sur l'unique document que constitue une infographie des résultats issus de l'analyse des données internes de l'entreprise et de celles engrangées à l'extérieur.

Exemples de DataViz



EthnosData

Concret, ingénieux, puissant

Si ce qui se conçoit bien s'énonce clairement, l'inverse devrait également être vrai. C'est parce que nous connaissons bien le champ et les délimitations du Big Data que nous avons conçu l'outil idéal pour tirer facilement parti des nouvelles sources de données.

Nous pourrions vous parler de plate-forme, de système intégré, d'outil de pilotage, de solution complète. Nous allons plutôt vous parler d'un logiciel simple et efficace qui permet à ses utilisateurs de faire concrètement des tonnes de choses utiles dans le domaine du Big Data.

EthnosData couvre tout le spectre de la démarche évoquée dans ce livret : il sait collecter un nombre incroyable de données sur le web ou depuis les fichiers et bases de données de votre entreprise. Il propose plus d'une centaine de méthodes d'analyses statistiques avancées sur des données structurées ou textuelles. Il permet enfin de produire des tableaux de bord dynamiques, des reporting interactifs et des infographies d'une qualité exceptionnelle, qui peuvent être facilement mis en ligne et partagés.

Data Collection

EthnosData possède en natif des connecteurs permettant de récupérer des données de la plupart des sources web. Nous avons balayé plusieurs milliers de sources, de Twitter à Facebook en passant par les principaux forums, sites marchands, d'avis consommateurs et autres organes de presse ou sources OpenData, pour développer des connecteurs simples et efficaces.

Nous avons également prévu dans EthnosData un module d'apprentissage pour vous permettre de définir de manière assistée, vos propres connecteurs pour des sources nouvelles. EthnosData saura donc récupérer pour vous, ponctuellement ou à la fréquence de votre choix, toutes les données qui vous sont utiles. Il saura également se connecter à vos bases d'entreprise ou charger des données à partir d'Excel, Access, fichiers texte ou XML, etc.

Data Analytics

EthnosData est en mesure de traiter ensemble les données en provenance de toutes vos sources. Ainsi, les données de vos fichiers internes et les éléments récupérés sur le web social ou ailleurs peuvent être pris en compte et analysés dans un même ensemble pour donner lieu à des résultats croisés et consolidés. EthnosData peut leur appliquer plus d'une centaine de méthodes statistiques, des tris les plus simples aux techniques d'analyse automatique de satisfaction les plus avancées, en passant par le traitement de données textuelles et la détection de tonalité.

Tous les traitements d'EthnosData sont dynamiques et peuvent s'appliquer sur vos nouvelles données ou sur n'importe quelles parties de vos fichiers (point de vente, région, produit, types de clients...). La puissance du produit est à la portée de tous les utilisateurs et n'exige aucune compétence avancée.

DataViz

Quels types de graphiques produisez-vous habituellement en dehors de ceux d'Excel ? Avec EthnosData apprêtez-vous à générer tous vos graphiques usuels (histos, courbes, camemberts, radars...) mais également des dizaines d'autres formes efficaces et originales, adaptées à la représentation des relations entre les données et à l'intégration dans des rapports et infographies de qualité : matrices, mappings, nuages de mots, cartographies Google Maps, anneaux, abaques, compteurs, jauges, cibles, podiums, etc.

EthnosData intègre un module de création de rapports dynamiques. Les tableaux et graphiques peuvent être placés dans des pages avec des textes, images et objets graphiques. Ces rapports se rafraîchissent automatiquement à chaque mise à jour de données. L'utilisateur peut également produire des rapports partiels (par région, point de vente, période...) et générer ainsi dix, cent ou mille rapports en quelques clics (fonction d'éclatement ou « drill-down »). Ces rapports sont, bien entendu, exportables aux formats Powerpoint ou PDF. Ils peuvent également servir de base à une application de dashboard et de web-reporting (avec notre outil Web-Reports) pour donner un accès dynamique et en ligne aux différents résultats.

Enfin, EthnosData propose également un module spécifique de fabrication d'infographies, elles aussi dynamiques et qui peuvent être générées par critère et exportées en PDF ou en PNG. Ce module propose tout ce qui fait la spécificité du mode de communication infographique : capsules de titre et de bas de page, séparateurs, symboles et personnages...

Bref, EthnosData est un outil unique, beau, puissant et ultra-innovant, qui s'installe vite, fonctionne bien et s'intègre dans vos environnements techniques sans les bouleverser.

Jetez-y un oeil.

Fonctionnalités d'EthnosData

Data Analytics



Tabulations et croisements de données

- Tris à plat, croisés, multiples, NxN, combinés NxN sur tous types de variables (qualitatives, numériques, ordinales, date, heure...)
- Tableaux et croisements dynamiques multi-niveaux
- Pondérations simples ou multiples



Fiches et états récapitulatifs

- Fiches individuelles simples ou scorées
- Etats récapitulatifs multi-critères
- Rapports individuels avec phrases de diagnostics automatiques



Analyses comparatives

- Tableaux d'évolutions avec diagnostics
- Benchmarking, analyses top-flop, Net-Promoter-Score...



Tests et analyses statistiques avancées

- Comparaison de moyennes et de proportions, régressions, Anova...
- Qnalyse multi-dimensionnelles : ACP, AFC, ACM, typologies et classifications
- Matrices importance-satisfaction : Satimix, Tetraclasses



Analyse de textes et de questions ouvertes

- Classement thématique
- Lemmatisation
- Analyse de sentiment

Data Collection

Import de fichiers de données aux formats usuels

- Tableaux (Excel, Open Office...)
- Fichiers BD (Access, DBF...)
- Fichiers XML
- Fichiers délimités (CSV, TXT...)

Connexion à des bases de données

- SQL Server
- MySQL
- SQL Lite
- Hadoop

Récupération dynamique de données web et réseaux sociaux

- Aspiration de pages web
- Capture de forums
- Récupération de tweets et réseaux sociaux
- Extraction de fichiers Open Data

Prise en charge de données textuelles pour analyse de contenu

- Boîtes de messageries
- Documents Word ou RTF
- Fichiers PDF
- Entretiens qualitatifs, réponses ouvertes (Ethnos)

Chargement de fichiers d'enquêtes en provenance de différents supports

- Formulaires web (Net-Survey)
- Enquêtes mobiles (Mobi-Survey)
- Formulaires scannés (OMR Manager)
- Plateaux d'appels (Catiopée)



DataViz

Plus d'une centaine de représentations graphiques avancées de vos tableaux

- Graphiques usuels : histogrammes, camemberts, barres, courbes, aires, graphes radar...
- Graphiques spécifiques pour les représentations de données croisées et multi-niveaux : matrices, graphes miroir, abaques, anneaux, compteurs, jauges et graphes de niveaux...
- Graphes matriciels d'analyses de données et d'analyses importance-satisfaction
- Tag-clouds (nuages de mots), cartographie Google Maps ou cartes Flash, graphes dynamiques Google Charts...

Module de génération de rapports automatisable

- Création de rapports dynamiques, génération multiples par drill-down, exportables en PDF et PPT

Module de conception et de diffusion d'infographies

- Création d'infographies dynamiques avec capsules, séparateurs, bas de page, formats spécifiques, symboles...

Intégration avec les systèmes de Web-Reporting et de Dashboarding

- Connexion directe avec Web-Reports pour générer des applications d'accès contrôlé et en ligne aux résultats
- Intégration de filtres dynamiques multi-niveaux, d'affichage interactif des résultats...