

Big Data : données sur les entreprises et marketing prédictif B2B

Par François BANCILHON
Sidetrade

Quoi de neuf sur les données ?

Le Big Data, c'est la conjonction de trois phénomènes : la disponibilité de données nouvelles et en grande quantité, la disponibilité de machines de plus en plus puissantes (la loi de Moore continue à faire des siennes) et la disponibilité de technologies de parallélisation qui permettent de traiter ces données massives grâce à ces machines puissantes.

Les données nouvelles sont essentiellement venues des réseaux sociaux, de l'Internet des objets, des données générées par le web et de l'open data. Les deux premières sources de données étant largement connues, nous détaillons uniquement les deux dernières.

Les données du web qui ont contribué au Big Data sont de deux types : les données du web elles-mêmes et celles liées au comportement des utilisateurs du web. La capacité à recueillir des données sur le web par *crawling* (aspiration de l'ensemble du contenu d'un site web sans essayer de le comprendre) ou par *scraping* (recueil du contenu d'un site web en connaissant la structure, connaissance utilisée pour extraire des informations structurées) permet à de nombreux opérateurs de construire des jeux de données riches et importants à partir du web. C'est cette activité qui fait que 52 % du trafic Internet soit généré par des robots plutôt que par des humains. La capacité à recueillir les données de comportement des utilisateurs du web a conduit à la mise en place, par les gestionnaires de sites web, de plateformes DMP (*data management platforms*) qui mémorisent les comportements des usagers du web (qui est venu sur mon site, quand et pour y voir quoi) et qui étendent cette information à des ensembles de sites.

L'open data (ouverture des données) est la mise à disposition des entreprises (pour réutilisation) et des citoyens (pour consultation et par besoin de transparence) des données dont se sert la puissance publique dans l'exercice de ses fonctions. Échappent bien entendu à cette ouverture tout ce qui relève du « secret défense » et les données personnelles. Le mouvement open data, lancé par l'administration Obama à la fin des années 2000, a trouvé un écho fort dans l'ensemble des pays démocratiques, à des degrés divers. Nous verrons plus bas sa traduction en France.

Pour traiter ces données massives, des technologies ont été développées, qui permettent à tous, de la plus petite *start-up* au plus grand groupe, de recueillir massivement des données, de les traiter et de construire applications et nouveaux usages pour les exploiter.

Le Big Data, basé sur ces données et la capacité à les exploiter, s'est donc développé dans de nombreux domaines : transport, médecine, ville intelligente, publicité, agriculture, marketing *Business-to-Consumer* (B2C), etc. *A priori*, aucun domaine ne devrait y échapper.

Cet article se concentre sur le sujet du marketing *Business-to-Business* (B2B).

Quoi de neuf pour les données sur les entreprises ?

Un grand nombre de données officielles sur les entreprises, qui n'étaient pas visibles, sont tout à coup devenues disponibles à tous grâce à l'open data. C'est le cas des données émises par les entreprises sur elles-mêmes et de celles produites à propos d'elles par d'autres (presse et médias Internet).

Les entreprises communiquent (et donc se décrivent) sur leurs sites web, leurs réseaux sociaux, leurs applications mobiles, leurs sites éventuels de commerce électronique et les sites de recrutement sur lesquels elles publient des offres de postes. Nous verrons plus loin le taux d'équipement des entreprises en chacun de ces outils. Par ailleurs, les médias Internet communiquent régulièrement sur les entreprises, ce qui est une deuxième source d'information.

Quoi de neuf en France sur les données sur les entreprises ?

Quelles données open data sont disponibles en France sur les entreprises ? Quelle communication ces mêmes entreprises font-elles sur le web ?

L'ouverture des données

Le mouvement d'ouverture des données, amorcé en France à partir de 2010, s'est fortement accéléré sous l'impulsion d'Etatlab à partir de 2013. De nombreux lois et décrets, sous les divers gouvernements, ont réaffirmé le principe de l'open data par défaut (qui n'est pas complètement appliqué, mais progresse régulièrement). Après l'État, les collectivités se sont emparées de l'idée. Ce mouvement s'est notamment traduit par de nombreuses ouvertures de données sur les entreprises. Le fichier SIRENE (Système national d'identification et du répertoire des entreprises et de leurs établissements), géré par l'Insee, a été ouvert au début de 2017. Le RNCS (Registre national des Commerces et des Sociétés), après un long combat d'arrière-garde d'Infogreffe pour en empêcher l'ouverture, a été ouvert début 2018. C'est finalement à l'INPI qu'il est revenu de mettre à disposition ces données, collectées par les greffes des tribunaux de commerce. L'INPI avait déjà en 2016 ouvert ses données (marques, brevets et dessins). Le Bodacc (Bulletin officiel des Annonces civiles et commerciales) avait lui aussi été rendu disponible à la même date par la Direction de l'Information légale et administrative (DILA). Le ministère de la Recherche et de l'Enseignement supérieur a contribué à l'édifice en publiant la liste des entités publiques de recherche et des projets coopératifs entre recherche et industrie, ce qui s'est traduit par la mise en place de l'application scanR qui offre l'accès à ces données. Le RNA (Registre national des Associations) est lui aussi ouvert depuis 2017.

Ainsi, en un intervalle de deux ans, c'est l'ensemble des données détenues par l'État sur les entreprises qui est devenu disponible en open data. Avant cette ouverture, la plupart de ces données n'étaient disponibles que de façon payante : pour l'ensemble de ces données, il fallait compter plusieurs centaines de milliers d'euros d'abonnement mensuel pour y accéder.

Le premier impact de cette ouverture a été l'accès immédiat des *start-ups* qui ont pu offrir des services innovants à partir de ces données. Le deuxième impact est l'amélioration de la qualité de ces données par le retour de ceux qui y ont accès.

Maturité numérique des entreprises

Les entreprises françaises participent au mouvement général d'accès au web. Elles disposent de sites web, elles sont présentes sur les réseaux sociaux, elles sont présentes sur Internet *via* des applications mobiles. Le tableau ci-après donne le taux d'équipement en sites web, sites e-commerce, Facebook, Twitter et LinkedIn. Pour les entreprises françaises, nous n'avons retenu que les « sociétés commerciales » (catégorie Insee).

Pays	Entreprises	Sites Web	ecommerce (% de SW)	twitter	FB	linkedIn
UK	4,260,628	23.89%	7.59%	8.29%	7.76%	3.66%
FR	2,454,396	25.20%	9.64%	4.06%	9.22%	1.80%
BE	1,700,340	6.69%	5.87%	0.73%	2.61%	0.50%
NL	2,760,265	30.64%	6.63%	5.34%	9.36%	2.68%

Le taux d'équipement croît assez vite dès qu'on considère des entreprises d'une certaine taille : si l'on se restreint aux entreprises d'un chiffre d'affaires d'au moins 1M€, le taux d'équipement en sites web est de 60 % pour la France.

Quel impact sur le marketing B2B ?

On peut diviser le marketing B2B en deux grandes catégories d'activités : la gestion de l'« entonnoir commercial », processus central du marketing opérationnel allant de l'identification du prospect à sa transformation en client, et les études de marché qui sont des études générales pour comprendre un ou plusieurs marchés et leur évolution.

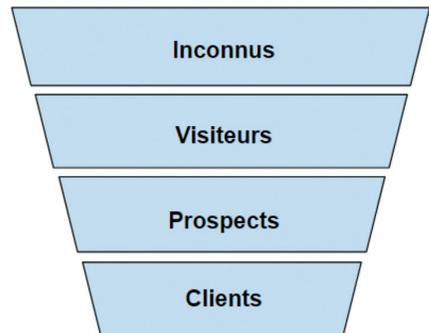
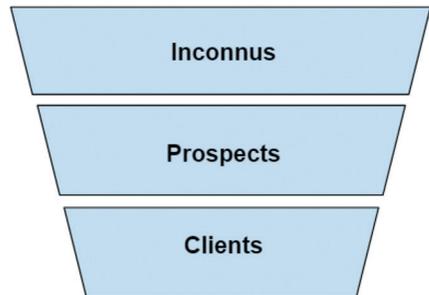
La gestion de l'entonnoir

La première des figures ci-dessous illustre l'entonnoir classique de la démarche marketing. Il s'agit d'identifier parmi l'ensemble des entreprises celles qui peuvent devenir clientes, et de les transformer en prospects avec lesquels on établit une relation commerciale, puis qu'on transforme en clients. Le passage du statut d'inconnu à celui de prospect se faisait jusqu'à maintenant par des méthodes traditionnelles : publicité, présence sur des salons, publipostage, télémarketing et plus récemment mailing de masse à partir de listes d'adresses acquises auprès de vendeurs de fichiers.

La seconde figure illustre le nouvel entonnoir marketing dans lequel une partie importante de l'activité se passe sur le site web, ce qui justifie l'apparition d'une nouvelle catégorie, celle des visiteurs de ce site.

Qui sont les entreprises ou les personnes à chaque étage de l'entonnoir ?

- Les inconnus : ce sont les 2,5 M d'entreprises que vous ne connaissez pas, qui sont pour certaines d'entre elles des clients potentiels de vos produits ou services ;
- Les visiteurs de votre site web sont de deux catégories :
 - non identifiés : ils sont venus sur votre site, attirés par votre *inbound marketing*, vous ne les connaissez pas, mais vous les voyez circuler sur vos pages et vous mémorisez leur parcours ;
 - identifiés : ils sont venus sur votre site, attirés par votre *inbound marketing*, ils se sont inscrits (pour avoir un livre blanc, une démonstration ou un autre contenu), donc vous les connaissez, vous les voyez circuler sur vos pages et vous mémorisez aussi leur parcours ;



- Les prospects : ils ont parlé avec vos commerciaux ou interagi avec votre système ;
- Les clients : ils utilisent vos produits ou services, vous voyez ce qu'ils en font ou ce qu'ils n'en font pas.

Le jeu de l'entonnoir consiste à faire descendre le plus possible d'entreprises ou de personnes dans l'entonnoir et à avoir un taux de transformation entre l'étage n et l'étage $n-1$ le plus élevé possible. Plus précisément :

- Les inconnus : il s'agit de trouver parmi eux ceux qui sont vos clients potentiels (donc ceux qui ont besoin de vos produits ou services) et de les faire venir sur votre site par du marketing *in-bound* (à partir de contenu ou de référencement) ou *out-bound* (par des mailings de masse) ;
- Les visiteurs non identifiés : vous pouvez en identifier certains (par *IP-tracking* qui consiste à rattacher l'adresse IP de votre visiteur à son entreprise), vous pouvez apprendre des informations supplémentaires sur eux (*via une data management platform* qui mémorise leurs trajets sur d'autres sites), vous pouvez aussi les convaincre de s'identifier pour bénéficier d'informations supplémentaires ;
- Les visiteurs identifiés : il s'agit de les transformer en prospects ;
- Les prospects : il s'agit de les transformer en clients ;
- Les clients : vous suivez leur comportement, vous voulez les convaincre d'acheter plus (*up-sell*), d'acheter d'autres produits de votre gamme (*cross-sell*) et de ne pas vous quitter (rétention).

Voyons ce que nous avons comme données sur les éléments à chaque étage de l'entonnoir :

- Les inconnus : l'ensemble des données sur les entreprises ;
- Les visiteurs non identifiés : leur comportement sur votre site, et éventuellement sur d'autres sites ;
- Les visiteurs identifiés : leur comportement sur votre site, et éventuellement sur d'autres sites, ainsi que les données sur leur entreprise ;
- Les prospects : leur comportement sur votre site, sur d'autres sites ainsi que les données sur leur entreprise et l'historique de votre relation commerciale avec eux ;
- Les clients : leur comportement sur votre site, et éventuellement sur d'autres sites ainsi que les données sur leurs entreprises et l'historique de votre relation avec eux et de leur usage de votre produit.

Voyons maintenant avec quelles technologies on peut utiliser ces données pour accomplir notre objectif de faire descendre vos interlocuteurs dans l'entonnoir :

- Les inconnus : outils de filtrage, de recherche et de *scoring* par *machine learning* ;
- Les visiteurs non identifiés : *IP tracking*, *scoring* par *machine learning*, *marketing automation* ;
- Les visiteurs identifiés : *scoring* par *machine learning*, *marketing automation* ;
- Les prospects : *scoring* par *machine learning* ;
- Les clients : *scoring* par *machine learning*.

En quoi consiste ce *machine learning* pratiqué à chaque étage de l'entonnoir ? À partir des données accumulées sur le comportement de l'entonnoir dans le passé, on construit un modèle prédictif plus ou moins sophistiqué qui nous permet d'identifier les candidats les plus aptes à être transformés.

Le tableau ci-dessous résume ce que nous venons de décrire.

Cibles	Objectif	Données	Technologies et outils
Inconnus	Identifier et engager	Données externes sur les entreprises	<i>Machine learning</i>
Visiteurs inconnus	Identifier et engager	Comportement sur le site	<i>IP Tracking et Machine learning Marketing automation</i>
Visiteurs connus	Transformer en prospects	Données externes sur les entreprises + comportement sur le site	<i>Machine learning Marketing automation</i>
Prospects	Transformer en clients	Données externes sur les entreprises + comportement sur le site + interaction commerciale	<i>Machine learning CRM</i>
Clients	Rétention, <i>up-sell, cross-sell</i>	Données externes sur les entreprises + comportement sur le site + interaction commerciale + utilisation du produit/service	<i>Machine learning CRM</i>

Les études d'ensemble

Sous cette catégorie, nous rangeons l'évaluation de la taille d'un marché et de son évolution, l'identification d'un écosystème et la segmentation d'un marché par type d'activité. Ce type d'étude permet par exemple d'identifier des entreprises innovantes ou de comprendre le basculement de la technologie du moteur à explosion du diesel vers l'essence. Il s'agit donc de qualifier ou de quantifier un phénomène. Pour faire ce type d'étude, on peut en passer commande à un cabinet de conseil. Le cabinet va identifier entre vingt et cent entreprises qu'il considère comme représentatives du secteur. Il va préparer un questionnaire, l'administrer à son panel, produire un rapport qualitatif et quantitatif, et faire des recommandations. Le rapport sera imprimé et distribué aux clients.

Ce qui caractérise l'approche traditionnelle de ce type d'étude est que le rapport est en papier (une version électronique en pdf sera certainement disponible mais cela ne change pas le problème), qu'il donne une photographie à un instant déterminé mais peut être dépassé six mois plus tard et qu'il se fonde sur un échantillon.

L'ensemble des données disponibles sur les entreprises en temps réel, la capacité à les maintenir à jour automatiquement dans le temps, la possibilité de développer des outils interactifs de visualisation, la capacité à maintenir des données sur l'ensemble des entreprises et non pas sur un échantillon, permettent désormais d'adopter une approche complètement différente, avec un outil :

- en grandeur nature : on ne prend pas un échantillon, mais l'ensemble réel des entreprises concernées ;
- en temps réel : on ne fait pas une photo, mais on suit l'évolution en temps réel des données ;
- interactif : ce n'est pas un rapport mais une application interactive qui permet de voir divers points de vue.

Ces tableaux de bord dynamiques, en temps réel, interactifs et exhaustifs, ne sont pas réellement concurrents des autres méthodes, plus qualitatives, ils tendent plutôt à les compléter.

Le Big Data est en train de changer la plupart, voire la totalité des activités économiques et sociales. S'il a pénétré un peu plus tardivement le domaine du marketing B2B, il est en train de bouleverser la façon d'aborder ce marketing. De nouvelles données et de nouvelles technologies fournissent des outils nouveaux qui offrent aux entreprises des résultats très fortement améliorés. Ils les aident à acquérir de nouveaux clients et à augmenter leurs revenus avec leurs clients existants. Ils leur fournissent une vision détaillée et en temps réel de leurs marchés.