

Big Data : comment Trainline a construit sa gare de triage de données

Comment mettre en valeur, sur un support numérique, toute l'offre de voyages en train d'Europe ? Et, demain, comment afficher pour chaque visiteur l'offre la plus pertinente ? C'est à ces défis que s'attaque le laboratoire de Data Science que Trainline – l'ex-Captain Train – a monté voici environ deux ans. « *L'objectif était d'extraire les données intéressantes, de mener des analyses sur ces données et, éventuellement, d'en extraire des prévisions*, explique Cédric Raud, le responsable des développements au sein de l'équipe marketing. *Dans la réalité, avec une petite structure – à l'époque un data scientist et deux développeurs -, ces tâches demandaient beaucoup de temps, notamment l'extraction* ». Conséquence : le petit labo consacre l'essentiel de ses ressources à mener des extractions ponctuelles non automatisées et à intégrer ces données dans des analyses menées sous Excel. « *On pouvait par exemple passer beaucoup de temps sur la seule définition du ticket de train* », s'amuse Cédric Raud.

Déjà rendue difficile par le système transactionnel de Captain Train – qui privilégie la rapidité, à la simplicité d'accès aux données -, le travail du labo de Data Science est encore complexifié par les multiples outils d'analyse utilisés au marketing pour suivre les actions (SEO, bannières, réseaux sociaux, affiliations...) et par le fait que toutes ces solutions ont leur propre définition des notions qu'elles manipulent. « *Chaque analyse impliquait d'aller chercher la donnée dans l'outil adéquat, reprend Cédric Raud. D'où notre volonté d'automatiser avec un outillage adapté.* » Après avoir étudié les solutions de reporting et les pipelines de données, Trainline se tourne finalement vers la plate-forme de Data Science de la start-up française Dataiku. « *Cette technologie présentait l'intérêt de s'intégrer à notre infrastructure et était extensible à des solutions tierces, via des connecteurs Python. Elle permet aussi la collaboration entre les développeurs et les data scientists.* » Un point d'autant plus intéressant que, depuis son rachat par le leader des plates-formes de réservation de trains outre Manche – valorisant la start-up française environ 200 millions d'euros -, l'équipe data s'est étoffée, avec au total 8 personnes (dont 3 data scientists).

Bannières de pub 'data-driven'

Et, surtout, de premières applications sont passées en production. A commencer par un suivi affiné des performances des campagnes. « *Des connecteurs vers Google Analytics ou Adwords fournissent des données qui, agrégées avec la source principale de données venant du système cœur, permettent de générer des rapports mis à jour automatiquement* », précise Cédric Raud. A partir de ce même croisement de données, les équipes du marketing ont aussi accès à un outil de dataviz leur permettant d'explorer les données.

L'existence de cette base de données consolidée (dans PostgreSQL) permet aujourd'hui au voyageur de déployer plus rapidement de nouveaux scénarios d'usage. Comme l'affichage de statistiques (tarifs, tendance des prix...) sur ce que Trainline appelle ses pages SEO. Celles-ci concentrent des informations sur un trajet donné et permettent d'amener du trafic depuis les moteurs de recherche. Ou encore comme la création dynamique de bannières de publicité en fonction des préférences des internautes à un instant t. « *La seconde version de cette application intègre*

désormais le paramètre de la marge que génère telle ou telle vente », explique le responsable des développements. Selon Cédric Raud, ce scénario a généré des gains significatifs, de l'ordre de 10 %, qui a conduit à sa pérennisation. « La construction de cette application nous a, à la fois, délesté de la création des bannières – celles-ci sont générées à partir de templates – et permis de réaliser davantage d'itérations. Car, en plus du contenu dynamique, la maquette aussi peut changer, amenant de multiples tests comparatifs d'efficacité par AB Testing. »

Personnalisation du site Web

L'agence de voyage online réfléchit aujourd'hui à la personnalisation du site en fonction du profil du visiteur. Une voie sur laquelle s'engage également Voyages-SNCF, principal concurrent de Trainline en France. Ce dernier en a même fait une de ses priorités pour 2017. « Cette personnalisation aurait du sens, particulièrement à l'échelle européenne où les comportements sont très différents d'un marché à l'autre, détaille Cédric Raud. Elle pourrait démarrer par la géolocalisation et par la définition du profil des visiteurs, les voyageurs d'affaires n'ayant pas les mêmes attentes que les personnes réservant des billets de train pour leurs vacances. »

A lire aussi :

[Florian Douetteau, Dataiku : « Le GDPR va remodeler les applications Big Data »](#)

[Pour Air France, le Big Data est un atout maître dans la relation client](#)

[Big Data : il est temps de grandir... ou de mourir](#)

Crédit photo : zigazou76 via [VisualHunt](#) / [CC BY](#)