

Jean-François (suite)

Conjoint/Famille : Ma copine et moi formons une petite famille.

Animaux domestiques : Aucun pour le moment, mais un jour j’aurai un perroquet.

Sports et passe-temps : Hockey, hockey à roulettes, salsa et natation.

Le week-end idéal : Un week-end qui durerait plus de deux jours :) – dans les Cantons de l’Est, à mon chalet près du lac.

Mets préférés : C’est triste d’avoir à choisir. Mais puisqu’il le faut, je dirais des sushis avec un Chardonnay coulant.

Si je pouvais faire le métier que je veux (à part être conseiller SAS), je serais... archéologue et je ferais des fouilles pour récupérer des objets plutôt que des données. :)

Lorsque je ne participe pas à un projet de consultation SAS, j’aime... faire du patin à roues alignées, aller au cinéma, danser la salsa et assister à des dégustations de vin.

Ce que j’ai appris en effectuant du travail de consultation pour SAS et qui, selon moi, profiterait à d’autres utilisateurs SAS...

Débuter un premier projet d’exploration de données (*data mining*) constitue toujours un défi emballant. Que vous soyez membre de l’équipe d’informations commerciales du client, statisticien, analyste d’affaires ou directeur du service d’analyse, à titre de futur explorateur de données, vous avez intérêt à vous assurer que vous disposez de toutes les ressources nécessaires.

En exploration de données, le succès repose essentiellement sur l’intégration. Les trois éléments primordiaux sont : l’objectif d’affaires, les outils et les données.

Trop souvent dans les projets d’exploration de données, on se concentre sur les variables explicatives disponibles. Le projet est approuvé, puis la course s’engage pour recueillir toutes les données, les rassembler et les intégrer au jeu de données magique.

Malheureusement, on sous-estime souvent le temps nécessaire à ce processus. Le dénombrement est également un processus fastidieux, et il nécessite l’intervention du groupe de TI pour la préparation des données et la distribution de ces dernières aux unités d’affaires qui les utiliseront.

Les ressources qui interviennent à chaque étape du processus sont si nombreuses qu’il convient d’entreprendre le projet avec un objectif clair, sinon on risque de déraiser rapidement.

Avant le début du projet et le développement des données, il faut répondre à un certain nombre de questions.

Par exemple, si une entreprise décide de se pencher sur le roulement de personnel, comment définira-t-on le roulement, l'attrition et la conservation du personnel ? Quelles sont les dimensions qui qualifient ce phénomène ? Le roulement volontaire ? Le roulement involontaire ? La durée de l'association ? Devrait-on prendre en compte le roulement de cette année seulement puisqu'il est à la hausse ? Hum... Est-ce que cela a une incidence sur la façon dont nous créerons notre jeu de données ? Bien sûr !

La préparation des données est une tâche monumentale pour toute entreprise. D'ordinaire, l'information requise est dispersée dans différents systèmes et ne se situe pas au niveau d'agrégation nécessaire à l'exploration de données. Étant donné tout le travail qu'exigent ces transformations, il importe de définir correctement la question d'affaires à cerner et d'en dégager une cible.

L'objectif d'affaires doit être établi dès le départ et expliqué à tous les intervenants du projet. On débute normalement un projet par une réunion de lancement avec chacun des groupes impliqués (TI, marketing de bases de données, crédit). Le but de cette réunion est de veiller à ce que tous les intervenants aient une vision claire du processus global en question et d'assurer leur participation au projet.

Les objectifs initiaux de l'explorateur de données consistent à cerner le problème d'affaires que l'entreprise tente de résoudre, puis à comprendre quelle information est disponible et sous quelle forme.

L'information est normalement disponible au niveau transactionnel, ce qui ne convient pas à l'exploration de données. Pour être utiles, les données transactionnelles doivent être colligées au niveau du client, de façon à avoir du sens dans le contexte de l'entreprise et de la question d'affaires.

Une fois que l'explorateur connaît l'objectif d'affaires, il devrait documenter l'évolution de l'événement ou de la cible. L'analyse de l'évolution et de son importance au fil des mois facilite l'établissement du calendrier d'exécution.

Si les données historiques démontrent qu'il s'agit d'un événement rare au sein de la population, l'explorateur pourrait sélectionner des clients qui ont vécu l'événement au cours de différents mois. Il existe divers scénarios pour établir des cibles. L'un des dangers courants consiste à utiliser une variable en tant que cible, car celle-ci est disponible dans les systèmes; autrement dit, elle a reçu le sceau d'approbation des services financiers ou de tout autre groupe de l'entreprise. Si votre cible est articulée sur le mode de fonctionnement de l'entreprise, elle vous indiquera comment vous traitez les clients et comment votre entreprise fonctionne. Si c'est la question d'affaires qui vous intéresse, alors vous êtes sur la bonne voie; mais n'êtes-vous pas plutôt intéressé par le comportement des clients et la façon dont vous pouvez influencer ce comportement ?

Grâce à la bonne compréhension de l'objectif d'affaires, l'explorateur de données est bien placé pour formuler une recommandation de ce qui, dans le domaine, est

communément appelé la « cible », c'est-à-dire le point d'équilibre de l'exercice d'exploration de données.

Si la cible ne reflète pas l'objectif d'affaires, vous aurez beau explorer les données de fond en comble, même si vous faites des trouvailles, elles ne seront d'aucune utilité pour votre entreprise.