أبعاد جودة البيانات

تكون قراراتنا بجودة البيانات التي تقوم عليها فقط. وجودة البيانات تعني مدى ملاءمتها للغرض الذي نستخدمها فيه. وتصف DAMA، جمعية إدارة البيانات التي تقف وراء دليل DAMA-DMBOK، جودة البيانات من خلال أبعاد، وهي جوانب قابلة للقياس. وأوصى فريق عمل في DAMA UK عام 2013 بستة أبعاد أساسية: الدقة، والصلاحية، والاكتمال، والتفرد، والاتساق، والحداثة. ويمكن أن نسير في أربعة أجزاء: الدقة والصلاحية، والاكتمال والتفرد، والاتساق والحداثة، والقياس والتحسين.

1. الدقة والصلاحية

تسأل الدقة هل تصف البيانات الشيء الحقيقي وصفاً صحيحاً. فرقم هاتف الزبون دقيق فقط إذا كان فعلاً رقم ذلك الزبون. وتسأل الصلاحية هل تتبع البيانات قواعد الشكل والقيم المسموحة، مثل تاريخ بصيغة صحيحة أو بلد من قائمة مسموحة. وقد تكون البيانات صالحة وغير دقيقة، لأن رقم هاتف مكتوباً بصيغة سليمة قد يعود لشخص آخر.

مثال: يطلب نموذج تاريخ ميلاد. وبتطبيق الدقة والصلاحية، يفشل تاريخ مثل الحادي والثلاثين من فبراير في فحص الصلاحية، بينما يجتازه تاريخ حقيقي كُتب بسنة خاطئة لكنه غير دقيق.

في العمل: يفحص رامي قائمة الزبائن. وبتطبيق الدقة والصلاحية، يجد عناوين بريد إلكتروني بصيغة معطوبة، وهي غير صالحة، ورقم هاتف يعود لزبون سابق، وهو غير دقيق. فيرسل الحالتين للتصحيح.

2. الاكتمال والتفرد

يسأل الاكتمال هل البيانات التي نحتاج إليها موجودة. فالسجل الذي ليس فيه عنوان توصيل ناقص لغرض التوصيل. ويسأل التفرد هل يظهر كل شيء حقيقي مرة واحدة فقط، من دون تكرار. ويؤدي التكرار إلى العد المزدوج والمراسلات المتكررة والحيرة في معرفة أي سجل هو الصحيح.

مثال: في قائمة ضيوف حفلة اسمان فارغان، ويظهر ضيف مرتين. وبتطبيق الاكتمال والتفرد، يملأ المضيف الفراغات ويحذف التكرار، فيصح عدد المقاعد.

في العمل: يجد رامي أن أربعمئة زبون بلا رمز بريدي، وأن بعضهم يظهر مرتين. وبتطبيق الاكتمال والتفرد، يطلب الرموز البريدية الناقصة ويدمج المكررات. فتنخفض كلفة المراسلات، لأن كل زبون يتلقى رسالة واحدة.

حقوق الطبع والنشر محمية © - أكاديمية أساسيات (شركة أساسيات للخدمات الاستشارية)

3. الاتساق والحداثة

يسأل الاتساق هل تتفق الحقيقة نفسها في أماكن مختلفة. فإذا قال نظام الفواتير إن الزبون يسكن مدينة وقال نظام المتجر إنه يسكن مدينة أخرى، فالبيانات غير متسقة. وتسأل الحداثة هل البيانات حديثة بما يكفي لاستخدامها. وقد يصلح جرد مخزون السنة الماضية لتقرير سنوي، لكنه لا يصلح لقرار طلب اليوم.

مثال: يتغير عنوان شخص في تطبيق ولا يتغير في آخر. وبتطبيق الاتساق والحداثة، يختلف التطبيقان، ويكون الأقدم منهما غير محدث.

في العمل: يقارن رامي نظام الفواتير بنظام المتجر. وبتطبيق الاتساق والحداثة، يجد زبائن لهم عنوانان مختلفان، ونظاماً يحدّث مرة واحدة أسبوعياً فقط. فيطلب تحديثاً يومياً.

4. القياس والتحسين

يمكن فحص كل بعد باختبارات بسيطة. فمثلاً نعد نسبة السجلات التي تنقصها قيمة، ونسبة المكررات، أو عمر آخر تحديث. ثم نضع مستهدفاً لكل بعد، بحسب أهمية البيانات. وعندما تظهر مشكلة، من الأفضل إصلاح سببها عند المصدر، مثل حقل في نموذج، بدل تنظيف الخطأ نفسه مراراً. ويبقى وجود مالك واضح لكل مجموعة بيانات الفحوص مستمرة.

مثال: تعد مدرسة كل فصل عدد سجلات الطلاب التي ينقصها رقم هاتف ولي الأمر. وبتطبيق القياس والتحسين، تجعل المدرسة هذا الحقل إلزامياً في النموذج، فتنخفض النسبة.

في العمل: يقيس رامي نسبة الزبائن بلا رمز بريدي: اثنا عشر في المئة. وبتطبيق القياس والتحسين، يجعل حقل الرمز البريدي إلزامياً في النموذج الإلكتروني. وفي الشهر التالي، تنخفض النسبة.

الخاتمة

وخلاصة القول، جودة البيانات تعني ملاءمتها للاستخدام، ويمكن وصفها من خلال أبعاد. فالدقة والصلاحية تسألان هل البيانات صحيحة ومسموحة، والاكتمال والتفرد يسألان هل هي كاملة وغير مكررة، والاتساق والحداثة يسألان هل تتفق وهل هي حديثة. والأبعاد الستة التي أوصى بها فريق DAMA UK نقطة بداية شائعة. وبقياسها وإصلاح الأسباب عند المصدر، نجعل القرارات أكثر جدارة بالثقة.

المصادر

- فريق عمل DAMA UK (أكتوبر 2013): The Six Primary Dimensions for Data Quality Assessment (ورقة بيضاء).
- DAMA International: دليل DAMA-DMBOK لمجموعة معارف إدارة البيانات (الطبعة الثانية، 2017) – https://www.dama.org/cpages/body-of-knowledge
- مركز جودة البيانات في الحكومة البريطانية: Meet the data quality dimensions – https://www.gov.uk/government/news/meet-the-data-quality-dimensions
- DAMA وDAMA-DMBOK علامتان تجاريتان لمنظمة DAMA International.

حقوق الطبع والنشر محمية © - أكاديمية أساسيات (شركة أساسيات للخدمات الاستشارية)