ESTRATEGIA | Empresas

Más no siempre es mejor

En base de datos hay de todo, pero desordenado y confuso

Las empresas acumulan datos de todo tipo, pues cada sector genera y acopia los que se relacionan con sus tareas. Al intentar ensamblar tanta información, cuando varias bases de datos son combinadas, se generan registros inconsistentes, erróneos, innecesarios y/o repetidos y, como consecuencia, se hace imprescindible realizar un reordenamiento de estos.

Por Elisabeth Rassore

Para juegos de datos existentes, la solución a estos problemas es la de intentar subsanar los datos de algún modo. El proceso de encontrar registros incompletos, incorrectos, faltantes, inexactos, irrelevantes, etc. para luego corregirlos, modificarlos o eliminarlos se conoce como limpieza de datos o *data cleansing.*

El proceso de *data cleansing*

En una primera etapa los datos deben ser auditados a fin de descubrir anomalías y contradicciones entre ellos. Generalmente se utilizan métodos estadísticos y los errores que pueden hallarse, entre otros, son:

- errores de sintaxis
- registros duplicados
- diferencia de formato entre registros similares
- omisiones
- información incorrecta
- datos innecesarios
- registros mal cargados

Estos errores deben ser corregidos. Para lo cual es requisito considerar las causas de todas las anomalías encontradas a fin de proceder con los cambios adecuadamente. A título de ejemplo:

- ante registros duplicados se procede con la deduplicación de los mismos;
- si hay registros similares con distinto formato se realiza una normalización de los datos; como ser la separación de nombre y apellido, conversión a mayúsculas o minúsculas, eliminación de caracteres extraños /, #, etc.
- en el caso de omisiones relevantes para el negocio, como ser teléfonos de contacto, variables sociodemográficas, coordenadas geográficas, códigos postales, sexo o edad.

Elisabeth Rassore. *El proceso de* data cleansing.

se procede con el enriquecimiento de dicha información.

Finalmente, resta inspeccionar los resultados para verificar las correcciones hechas. Esta etapa es tan importante como todo el trabajo previo pues puede suceder que las modificaciones planteadas sean incorrectas o insuficientes y, consecuentemente, hay que volver a encarar este proceso.

De ser posible, tanto en tiempo como en costo, validar y actualizar la información susceptible de sufrir modificaciones (direcciones postales, *e-mails,* teléfonos, etc.).

Tanto la actualización de los datos preexistentes como la captura de nuevos, puede hacerse a través de campañas u otra técnica de marketing que sirva para recoger información fidedigna.

El recelo a entregar los datos a un tercero

Si la entidad no cuenta con un departamento idóneo en la minería de sus datos necesitará recurrir a especialistas en este tema. Sin embargo, ante el hecho de entregar a un consultor externo los datos propios de la empresa, aparece el lógico recelo a proporcionar tanta información. Sin embargo, es cardinal proveerse de profesionales competentes, que reúnan tanto habilidades analíticas como el entendimiento del negocio.

Es fundamental la facultad del profesional para decidir qué herramientas de análisis usar y cómo interpretar los resultados obtenidos sobre la base del negocio. Lo primero evita malgastar tiempo y dinero en estudios incongruentes. En cuanto a lo segundo, un error en la comprensión puede llevar a conclusiones que le hagan tomar decisiones incorrectas.

Opciones a tener en cuenta:

- Los datos críticos, pueden entregarse codificados. Así, el profesional solo va a ver códigos pero no sabrá qué significa cada uno. La desventaja es que este proceder puede reducir las conclusiones interpretativas de quien esté haciendo el análisis.
- Existen contratos de confidencialidad que el profesional externo puede firmar. Este tipo de contrato es fundamental para salvaguardar la confidencialidad tanto de los datos como de la información de la empresa y para que no se ponga en riesgo los secretos de su éxito u operatoria en general.
- Asimismo, no es imperativo entregarle al profesional toda la información disponible; solo basta con darle la que quiera que analice.

Una vez que se logra tener una base de datos confiable, es muy poco lo que no se pueda hacer con ella, pues se vuelve un recurso excelente tanto para realizar acciones de marketing como para descubrir patrones y tendencias ocultas en los datos. **M**