Cómo crea Probative sus conjuntos de datos abiertos: informe de métodos

Versión 1.0. Publicado por Probative Growth, una agencia de marketing para firmas de abogados de EE. UU. CC BY 4.0. No es asesoría legal.

Qué cubre este informe

Este informe describe los métodos que comparten los conjuntos de datos abiertos que Probative Growth publica sobre el marketing de las firmas de abogados de EE. UU.: de dónde salen los datos, cómo se leen y se comprueban, cómo se versionan y corrigen, y cómo se citan. La página y el README de cada conjunto indican sus fuentes, columnas, cobertura y límites. La lista vigente, con cada versión y DOI, está en probativegrowth.com/es/datos/ y, en formato para máquinas, en probativegrowth.com/data/catalog.json.

De dónde salen los datos

Empezamos por el organismo que publica o hace cumplir el dato: agencias federales y estatales, tribunales, colegios de abogados y legislaturas, o la política que publica la propia plataforma. Cada registro nombra su fuente, la enlaza cuando existe un enlace e indica la fecha de nuestra última lectura.

El README de cada conjunto explica los códigos de estado que usa, para que el lector distinga un valor leído en una fuente primaria de uno que no lo es.

Cómo se leen

Cuando una fuente publica un archivo o una API oficial, un script lo lee según el calendario de la propia fuente, a un ritmo moderado y con un agente de usuario que nos identifica. Cuando una fuente no se puede leer con un script, una persona la lee en el navegador y la lectura queda registrada como una comprobación con fecha. No sorteamos inicios de sesión, CAPTCHA ni otros controles de acceso.

Cómo se comprueban

Las cifras de nuestras páginas se generan a partir de los archivos de datos o se registran contra ellos, y la compilación falla cuando una cifra ya no coincide con su fuente registrada.

Cualquier lector puede repetir la comprobación: copie un valor y el identificador de su registro, abra el enlace de la fuente de ese registro y compare el valor, la unidad y la fecha de lectura. Una edición más nueva de la fuente puede diferir de nuestro registro sin que ninguno esté mal; por eso importa la fecha de lectura.

Versiones, DOI y copias

Cada conjunto tiene números de versión y un DOI en Zenodo: uno por versión y un DOI de concepto que siempre apunta a la más reciente. Los cambios se publican en el registro de cambios de cada conjunto y en el canal de cambios. Los mismos archivos se publican en GitHub, Hugging Face y Kaggle, y una comprobación compara esas copias con el catálogo vigente.

Cómo se mantienen al día

Cada conjunto declara cada cuánto se vuelve a leer. Un conjunto, una página o una descarga que pasa su fecha de revisión hace fallar nuestra compilación hasta que se revisa de nuevo. El calendario de datos indica la próxima actualización de cada conjunto.

Correcciones

Una corrección es un valor que publicamos mal: una tabla mal leída, un cruce equivocado o una errata. Revisamos cada reporte con la fuente oficial. Si el dato está mal, lo corregimos, publicamos una nueva versión y dejamos constancia del valor anterior, el nuevo y la causa en probativegrowth.com/es/datos/correcciones/. Cuando una agencia revisa sus propias cifras, eso es una actualización y va en el registro de cambios.

Probative benchmarks

Las tablas identificadas como «Probative benchmarks, 2025–2026» son datos agregados y redondeados, con los grupos pequeños suprimidos. No nombran a ninguna firma.

Cómo leer los datos

Los conjuntos cuentan y describen; no califican a una firma, no dan asesoría legal ni dicen que un anuncio esté permitido. Eso le corresponde al abogado responsable, al colegio de abogados o a su asesor legal. Los resultados de búsqueda cambian según el lugar, así que una cifra leída en un resultado de búsqueda vale para la ubicación donde se leyó. Dos cifras parecidas pueden medir cosas distintas: compruebe primero el lugar, el periodo, la unidad y la definición.

Cómo citar y reutilizar

Los datos tienen la licencia CC BY 4.0: puede reutilizarlos, incluso con fines comerciales, si da crédito y un enlace. Forma estándar: Probative Growth, <título del conjunto>, <año>, https://doi.org/<DOI>. El kit de citas, en probativegrowth.com/es/datos/kit-de-citas/, tiene la línea exacta de cada conjunto.

Para citar este informe: Probative Growth, Cómo crea Probative sus conjuntos de datos abiertos: informe de métodos, 2026, https://probativegrowth.com/reports/metodos-de-datos-de-probative.pdf.