

Deloitte Review

No. 21 | Julio 2017



Previsiblemente inexacto♦

La prevalencia y los peligros de
los datos grandes malos

Por John Lucker, Susan K. Hogan, y
Trevor Bischoff
Ilustración por Jon Krause

Deloitte.

Acerca de Deloitte

Deloitte se refiere a uno o más de Deloitte Touche Tohmatsu Limited, una compañía privada del Reino Unido limitada por garantía ("DTTL"), su red de firmas miembro, y sus entidades relacionadas. DTTL y cada una de sus firmas miembros son entidades legalmente separadas e independientes. DTTL (también referida como "Deloitte Global") no presta servicios a clientes. En los Estados Unidos, Deloitte se refiere a una o más de las firmas de los Estados Unidos miembros de DTTL, sus entidades relacionadas que operan usando el nombre "Deloitte" en los Estados Unidos y sus respectivas afiliadas. Ciertos servicios pueden no estar disponibles para atestar clientes según las reglas y regulaciones de la contaduría pública. Para conocer más acerca de nuestra red global de firmas miembro, por favor vea www.deloitte.com/about.

Deloitte presta servicios de auditoría, impuestos, consultoría, y asesoría financiera a clientes públicos y privados que se extienden a través de múltiples industrias. Con una red globalmente conectada de firmas miembros en más de 150 países y territorios, Deloitte ofrece capacidades de clase mundial y servicio de alta calidad para los clientes, entregándoles los conocimientos que ellos necesitan para abordar sus desafíos de negocio más complejos. Los más de 200,000 profesionales de Deloitte están comprometidos en volverse el estándar de excelencia.

Esta comunicación solo contiene información general, y nadie de Deloitte Touche Tohmatsu Limited, sus firmas miembros, o sus entidades relacionadas (colectivamente, la "Red de Deloitte"), por medio de esta comunicación, está prestando asesoría o servicios profesionales. Ninguna entidad de la red de Deloitte será responsable por cualquier pérdida tenida por cualquier persona que confíe en esta comunicación.

Copyright © 2017 Deloitte Development LLC. Reservados todos los derechos.

♦ Documento original: "*Predictably inaccurate: The prevalence and perils of bad big data*", Deloitte Review, Issue 21 – July 31, 2017. Written by John Lucker, Susan K. Hogan, and Trevor Bischoff - Illustration by Jon Krause by Jon Krause.

<https://dupress.deloitte.com/dup-us-en/deloitte-review/issue-21/analytics-bad-data-quality.html>.

Traducción realizada por Samuel A. Mantilla, asesor de investigación contable de Deloitte & Touche Ltda., Colombia, con la revisión técnica de César Cheng, Socio Director General de Deloitte & Touche Ltda., Colombia.



Previsiblemente inexacto

La prevalencia y los peligros de los datos grandes malos

Por John Lucker, Susan K. Hogan, y Trevor Bischoff
Ilustración por Jon Krause

"No somos mucho más inteligentes de lo que solíamos ser, a pesar de que tenemos mucha más información – y eso significa que la verdadera habilidad ahora es aprender cómo escoger la información útil proveniente de todo este ruido."

- *Nate Silver*¹

¿NUESTRO AMORÍO CON LOS GRANDES DATOS NOS ESTÁ LLEVANDO POR EL MAL CAMINO?

LA SOCIEDAD y los negocios se han enamorado de los grandes datos. Pero no podemos obtener suficiente: a más recaudamos, más deseamos. Algunas compañías acaparan datos, inseguras de su valor o no teniendo claridad de si o cuándo les serán útiles, pero en todo momento, renuentes a eliminarlos o no capturarlos por miedo a perder el potencial valor futuro. Alimentando este apetito está el crecimiento total en el volumen, velocidad, y variedad de los datos.

Por encima de todo, muchos líderes de negocio ven potencial alto en una cuarta V: *valor*. Dada nuestra capacidad para tener acceso y (potencialmente) entender cada movimiento que hagan nuestros clientes actuales y potenciales, vinculada al acceso

a sus datos demográficos, biográficos, y sicográficos, parece lógico que debemos ser capaces de formar con ellos una relación más íntima, significativa. Cada punto de datos debe mover al negocio al menos un paso más cerca del cliente.

Sin embargo, a pesar de toda la miga de pan digital, resulta que los vendedores pueden saber menos acerca de los consumidores individuales de lo que piensan. Los números no mienten - ¿o lo hacen? ¿Qué pasa si muchos de estos datos son menos exactos de lo que esperamos sean?

Los peligros que van desde vergüenzas menores hasta completa alienación del cliente pueden esperar a las empresas que de manera creciente dependen de los grandes datos para guiar las decisiones de negocio y seguir estrategias de mercadeo de micro-segmentación y micro-orientación. De manera específica, el exceso de confianza puesta en la exactitud de los datos tanto originales como comprados puede llevar a un falso sentido de seguridad que puede comprometer esos esfuerzos en tal extensión que menoscabe la estrategia general.

Este artículo explora las consecuencias potencialmente adversas de nuestro actual amorío con los grandes datos. La evidencia proveniente de nuestra investigación primaria anterior² y actual, respaldada por investigación secundaria, destaca la prevalencia potencial y los tipos de datos inexactos provenientes de corredores de datos basados en los Estados Unidos, así como también los factores que pueden estar causando esos errores. La buena noticia es que existen estrategias y barandas de seguridad para ayudar a los negocios a mejorar la exactitud de sus conjuntos de datos, así como también para disminuir los riesgos asociados con el exceso de confianza puesto en los grandes datos en general.

DATOS PERSONALES QUE SON TANTO INCOMPLETOS COMO INEXACTOS

“Es aterrador cómo los datos recaudados acerca de usted pueden estar equivocados – especialmente si las personas toman decisiones importantes con base en esta información incorrecta. Esto se vuelve aún más aterrador en la medida en que más y más decisiones se vuelven basadas-en-información.”

- Encuestado

PARA calibrar de mejor manera el grado y los tipos de inexactitudes de los grandes datos y la disposición del consumidor para ayudar a corregir cualesquiera inexactitudes, realizamos una encuesta para probar qué tan exactos es probable que sean los datos de los corredores-de-datos comerciales – datos a partir de los cuales muchas firmas se basan para mercadeo, investigación y desarrollo, administración de producto, y numerosas otras actividades. (Para detalles vea el recuadro “Metodología de la encuesta”). Algunos de los hallazgos clave son:³

- Más de dos tercios de quienes respondieron la encuesta señalaron que los datos proporcionados por terceros acerca de ellos solo eran entre 0 y 50 por ciento correctos en su conjunto. Un tercio de quienes respondieron percibió que la información era entre 0 a 25 por ciento correcta.
- Si los individuos nacieron en los Estados Unidos tendieron a determinar si eran capaces de localizar sus datos dentro del portal del corredor de datos. De quienes no nacieron en los Estados Unidos, el 33 por ciento no podría localizar sus datos; inversamente, de quienes nacieron en los Estados Unidos, solo el 5 por ciento tenía información engañosa. Además, ninguno de quienes respondió nació fuera de los Estados Unidos y residió en el país por menos de tres años podría localizar sus datos.

- El tipo de datos sobre individuos que estaba más disponible era la información demográfica; el menos disponible eran los datos del hogar. Sin embargo, incluso si la información demográfica estaba disponible, no era completamente exacta y a menudo estaba incompleta, con el 59 por ciento de quienes respondieron juzgando que sus datos demográficos solo eran entre 0 y 50 por ciento correctos. Incluso tipos de datos aparentemente disponibles de manera fácil (tales como fecha de nacimiento, estado civil, y número de adultos en la vivienda) tenían variaciones amplias en exactitud.
- Casi el 44 por ciento de quienes respondieron dijo que la información acerca de sus vehículos era 0 por ciento correcta, mientras que el 75 por ciento dijo que los datos del vehículo eran entre 0 y 50 por ciento correctos. En contraste con los datos del vehículo, los datos de la vivienda fueron considerados más exactos, con solo el 41 por ciento de quienes respondieron juzgando que sus datos eran entre 0 y 50 por ciento exactos.
- Solo el 42 por ciento de los participantes dijo que su actividad registrada de compras en línea era correcta. De manera similar, menos de un cuarto de los participantes sintió que la información sobre sus gastos en línea y fuera de línea y los datos sobre sus categorías de compra eran más de 50 por ciento correctos.
- Si bien la mitad de quienes respondieron era consciente de que este tipo de información acerca de ellos existía entre los proveedores de datos, la mitad restante estuvo sorprendida o era completamente inconsciente de la escala y amplitud de los datos que estaban siendo reunidos.

METODOLOGÍA DE LA ENCUESTA

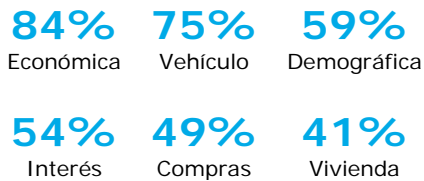
Nuestra encuesta les pidió a 107 profesionales de Deloitte US que privada y anónimamente revisaran los datos disponibles por un corredor de datos del consumidor líder, un corredor con un portal públicamente disponible, basado-en-la-red, que presenta a los usuarios una variedad de datos personales y de la vivienda. Quienes respondieron, todos entre 22 y 67 años de edad, completaron la encuesta de 87 preguntas, de respuesta rápida, entre enero 12 y marzo 31, 2017.

Quienes respondieron vieron sus perfiles de datos de terceros junto con una serie de variables específicas (tales como género, estado civil, y afiliación política), agrupados en seis categorías (económica, vehículo, demográfica, interés, compra, y vivienda). Para calcular el “porcentaje correcto” para cada variable individual, tomamos el número de participantes que señaló que el punto de datos del tercero para esa variable era correcto, y lo dividimos por el número total de participantes para quienes datos de terceros estaban disponibles para esa variable. Para determinar los puntos de vista de quienes respondieron sobre la exactitud de los datos por cada *categoría*, les pedimos que señalaran si sentían que la categoría de datos era 0 por ciento, 25 por ciento, 50 por ciento, 75 por ciento, o 100 por ciento exacta.

Figura 1. Exactitud reportada, de datos del consumidor de terceros, por quienes nos respondieron

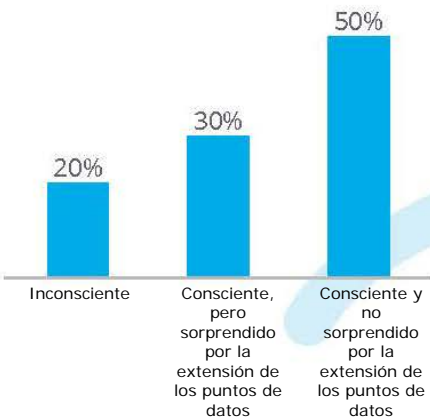
Inexactitud de la categoría de datos

Porcentaje de participantes que juzgó que sus datos en cada categoría eran solo entre 0 y 50% correctos.



General: **71%**

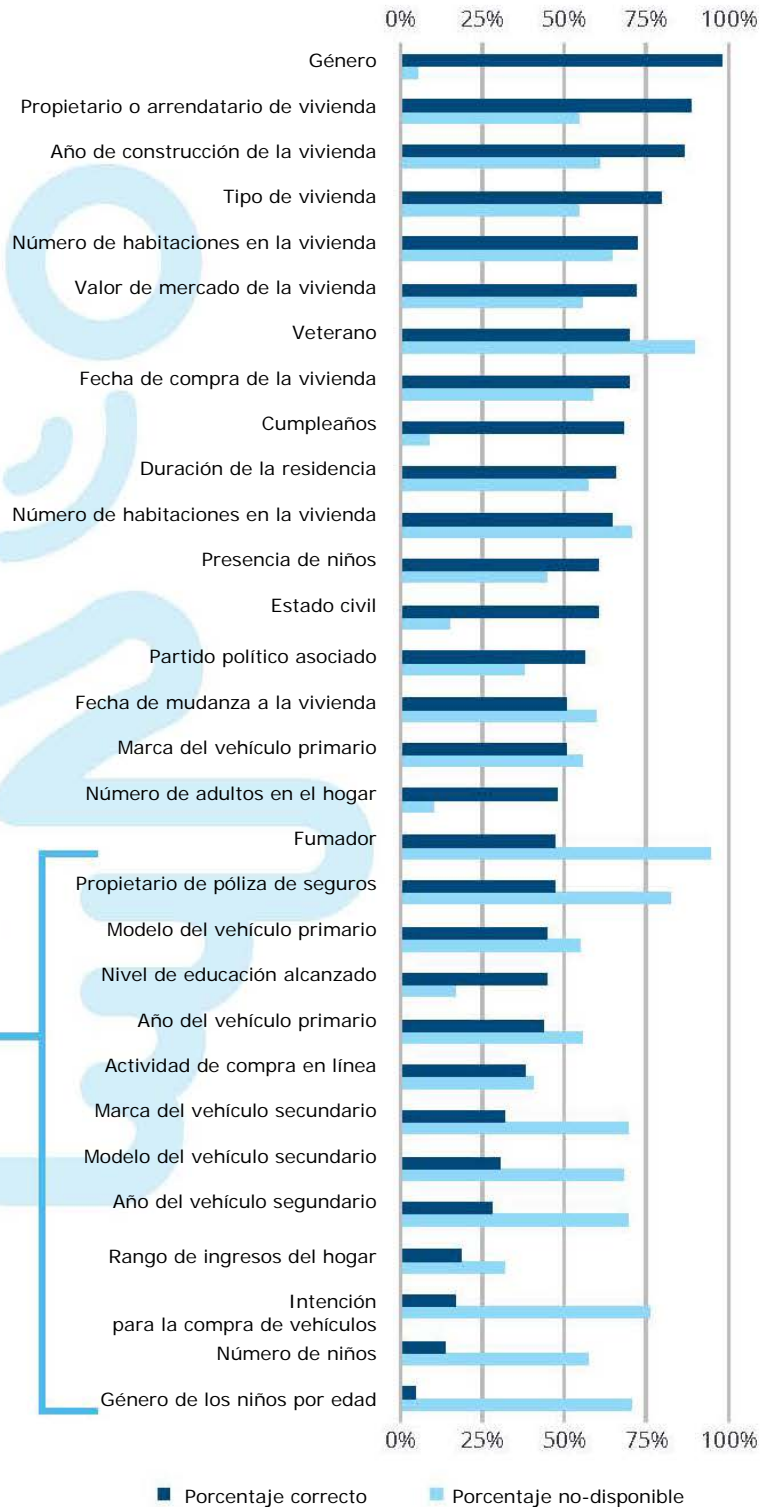
Conciencia del recaudo de datos del consumidor



Los datos para casi la mitad de las variables examinadas eran solo 50 por ciento o menos probable que sean exactos – equivalente a la exactitud obtenida mediante lanzar una moneda.

Para cada variable encuestada, el **porcentaje correcto** fue definido como la proporción de participantes señalando que la respectiva variable era correcta para con el número total de participantes para quienes la respectiva variable estaba disponible. El **porcentaje no-disponible** fue definido como la proporción de participantes señalando que la respectiva variable no estaba disponible para con la vista del número total de participantes que proporcionó una respuesta.

Variables encuestadas



La Figura 1 esboza otras inexactitudes u omisiones relacionadas con fecha de nacimiento, nivel de educación, número de niños, afiliación política, e ingresos del hogar. Claramente, todos esos tipos de datos son potencialmente importantes para los vendedores dado que ellos tienen como objetivo diferentes segmentos del consumidor.

¿Podemos contar con que los individuos corrijan sus propios datos?

“Si bien no me sorprendió la extensión de los datos recaudados, fue interesante verla. Me sorprendió qué tan pocos datos había acerca de mí (Yo soy un ávido comprador en línea), y qué tan incompleta es la imagen del ‘ciber yo.’ Pero no me quejo de ello.”

- Encuestado

A quienes respondieron la encuesta se les proporcionó la oportunidad para elaborar sobre por qué pensaban sus datos podían estar equivocados o incompletos. Muy comúnmente, la información disponible estaba desactualizada – especialmente los datos del vehículo. Muchos otros vieron los datos como que caracterizaban sus familias o a otros miembros del hogar (esposas o niños) más que a sí mismos. El sentimiento más mencionado entre quienes respondieron fue sorpresa – no por la cantidad de los datos correctos disponibles, sino porque la información era demasiado limitada, de pobre calidad, e inconsistente. En esencia, para muchos de quienes respondieron, los datos parecieron, como acertadamente lo dijo un encuestado, “rancios.”

“Había mucha información que no existía acerca de mí. Y de los datos que existían, muchos parecían inconsistentes con otros datos.”

- Encuestado

De manera interesante, incluso luego de haberseles dado la oportunidad para editar sus datos vía el portal en línea del corredor de datos, pocos de quienes respondieron escogieron hacerlo. Si bien aproximadamente dos tercios de quienes respondieron reportó que al menos la mitad de su información era inexacta, solo el 37 por ciento optó por editar sus datos.

La mejor razón más común para la decisión para editar (dada por el 31 por ciento de quienes respondieron que escogieron editar) fue mejorar la exactitud de la información. La segunda respuesta más común fue la decisión de editar solo lo que parecía relevante (proporcionada por el 17 por ciento de quienes respondieron que escogieron editar). Otro 11 por ciento de quienes respondieron que optó por editar citó privacidad y nerviosismo acerca de que sus datos estuvieran “por ahí.” Otros que respondieron señalaron el deseo de reducir o evitar mensajes dirigidos y correos políticos – así como también la esperanza de mejorar la calificación de su crédito (si bien, presumiblemente desconocido por ellos, este tipo de datos de mercadeo no tiene vinculación directa con cómo se derivan las calificaciones del crédito). Las categorías más comúnmente editadas fueron datos demográficos y datos del partido político.

¿Por qué muchos de quienes respondieron eligieron no editar sus datos? Muy a menudo, las personas citaron preocupaciones de privacidad. Otras razones incluyeron

que no percibieron valor en editar y ambigüedad en relación con cómo terceros pueden usar los datos. La Tabla 1 ofrece una vista de conjunto de las razones más comunes para la decisión de editar o no.

“Soy escéptico y cauteloso acerca de qué se podría hacer con estos datos. Asumiendo la mejor de las intenciones y la integridad de las personas que pueden consumir estos datos, no puedo imaginar un escenario que también podría ser en mi mejor interés o en el de mi familia. Yo actualmente prefiero que exista públicamente menos información personal acerca de mí. Por lo tanto, los datos oscuros, inexactos, o no-confiables son lo que considero es la mejor cosa que sigue.”
 - Encuestado

LOS PELIGROS DE CONFIAR EN DATOS MALOS

LOS hallazgos de nuestra encuesta sugieren que los datos que los corredores venden no solo tienen serios problemas de exactitud, sino que pueden ser menos actuales o completos que los compradores de datos esperan o necesitan. Dado que la mayoría de los principales corredores de datos de mercadeo de los Estados Unidos aloja el portal públicamente disponible usado para nuestra encuesta, esos hallazgos pueden ser considerados una representación creíble de todos los datos de mercadeo de los Estados Unidos disponibles a partir de numerosos corredores de datos. Los impactos de los datos inexactos o incompletos son muchos, variando desde *oportunidades perdidas* hasta *solo falta*.

Tabla 1. Razones comunes que orientan las decisiones para editar o no editar datos

¿Por qué usted editó sus datos?	¿Por qué usted no editó sus datos?
<ul style="list-style-type: none"> • Para hacer los datos más exactos/menores • Corregidos solo cuando percibí que era valioso /valía la pena el esfuerzo • Privacidad/nerviosismo de que estos datos estén por ahí • Para reducir/evitar publicidades/ofertas específicas 	<ul style="list-style-type: none"> • Privacidad • No valor percibido / no valen la pena el tiempo y la energía • No interesado / no me importa qué datos tengan sobre mí • Cautela / no-claridad sobre cómo la información será usada • Carencia de tiempo para editar • No es mi trabajo hacer su trabajo / corregir errores • Contra mercadeo dirigido

Fuente: Análisis de Deloitte

Oportunidad perdida 1: Subestimar el valor del cliente y no capitalizar el poder del hábito

“Ojalá gastara solo eso. Mis datos de compra parecen significativamente subestimados de lo que sé gasto en las categorías señaladas.”
- Encuestado

Entender el comportamiento del gasto y el poder de los clientes actuales y potenciales es muy importante para las firmas. Muchos vendedores extrapolan esta información con base en tres categorías clave: ingresos corrientes, valor neto modelado, y comportamiento anterior de compra. Los consumidores son creaturas de hábito – nuestro comportamiento pasado de gasto es uno de los mejores indicadores para que los vendedores determinen no solo qué tanto gastaremos en el futuro, sino qué tipos de elementos es probable que compremos. Esto puede guiar las predicciones sobre qué tantos ingresos ordinarios una compañía puede esperar ver en el próximo año, así como también cualesquiera esfuerzos cruzados de venta o de aumento de venta.⁴ Dada la importancia que esta información tiene para los vendedores, y el número increíble de migajas de pan digital que los consumidores dejan atrás, estuvimos sorprendidos con encontrar tal nivel alto de inexactitud. Más a menudo que no, quienes respondieron señalaron que los datos sobre los ingresos del hogar proporcionados por el corredor eran incorrectos, con los datos de compra a menudo subestimados, sugiriendo que los vendedores que confían en esta información para guiar sus esfuerzos dirigidos pueden estar dejando en la mesa potenciales ingresos ordinarios.

Oportunidad perdida 2: Lealtad e ingresos ordinarios del cliente disminuidos

“[Los datos] señalan que tengo una propiedad que actualmente es de mis padres, y al mismo tiempo, fallaron en listar la propiedad que actualmente tengo.”
- Encuestado

Otra área de inexactitud importante fue la residencia y la propiedad del vehículo, lo cual fue bastante sorprendente dado los registros públicos fácilmente disponibles para cada uno de ellos. Tal y como se señaló previamente, los datos de la vivienda eran más exactos que los datos del vehículo, pero todavía considerablemente inexactos en general. Quienes respondieron sugirieron que los datos en esas dos categorías a menudo estaban desactualizados – potencialmente entre cinco a diez años.

Uno de los períodos de gasto más alto en la vida de un individuo es cuando se cambia de residencia. No solo esos movimientos son costosos – los hogares incurren en importantes gastos auxiliares, incluso cuando los movimientos son locales. Cuando se mueve de una geografía a otra con un clima diferente, el consumidor a menudo comienza desde cero en numerosas categorías de producto (armario nuevo, muebles para el hogar, equipo al aire libre, y similares). Un vendedor no esperaría perder este momento de transición, en el cual los consumidores gastan más dinero de lo que normalmente sería como adquirir nuevos comportamientos – incluyendo rutinas y lealtades de compra. Con una descripción oportuna y relativamente exacta de los cambios de residencia de un consumidor, el vendedor podría perder influir en una compra momentánea, compras posteriores adicionales, y, potencialmente, construir lealtad del cliente en el largo plazo.

Figura 2. ¿Qué piensan las personas acerca de sus propios perfiles de grandes datos? Una muestra de comentarios de quienes nos respondieron.



Deloitte University Press | dupress.deloitte.com

Corroborando nuestros hallazgos, el estudio sobre la calidad de los datos de terceros encontró que el 92 por ciento de las instituciones financieras confían en información defectuosa para entender mejor a sus miembros, una tasa probablemente atribuible a errores y defectos humanos en la manera como fueron combinadas las múltiples fuentes de datos. El 80 por ciento de las cooperativas de ahorro y crédito consideran que las ineficiencias han afectado su línea de resultados, causando un promedio del 13 por ciento en los ingresos. Adicionalmente, el 70 por ciento de las instituciones

financieras culpan a la pobre calidad de los datos por los problemas continuos con sus esfuerzos de lealtad.⁵

Falla 1: Mover demasiado rápido la relación con el cliente

“Me molesta que ya no haya nada privado. De todos modos, rara vez uso los anuncios para las decisiones de compra, y me gustaría dejar de recibirlos por completo.”

Encuestado

No debo continuar sin decir que los mensajes micro-destinados están llenos de trampas – independiente de la exactitud de los datos en los cuales se basen. Tome, por ejemplo, el padre que conoció acerca del embarazo de su hija mediante ofertas de minoristas que llegan en el correo luego que el minorista detectó comportamiento de compra correlacionado con el embarazo.⁶ Si bien la evidencia sugiere que los consumidores se están volviendo más perceptivos ante el mercadeo personalizado, los vendedores todavía necesitan ser reflexivos y pisar suave en esta área.⁷ Esta palabra de alarma es consistente con la investigación reciente que identifica similitudes entre el desarrollo de relaciones interpersonales y las relaciones de negocios y con el cliente,⁸ así como también las teorías existentes en relación con el desarrollo de relación saludable. Particularmente, la auto-revelación de información personal está destinada a seguir un curso recíproco y progresivo, con inicial compartir mutuo de información personal a nivel superficial que con el tiempo evoluciona a un nivel más íntimo de intercambio.⁹ Demasiada, muy pronto de parte de cualquier parte puede volverse invasiva y horripilante – y generar disrupción en la relación que se ha desarrollado hasta ahora. Esto significa que demostrar temprano un conocimiento aproximado de su cliente puede ser más benéfico que demostrar un conocimiento íntimo o preciso. La investigación reciente ha corroborado esta idea, sugiriendo que la publicidad semi-adaptada o personalizada puede llevar a un 5 por ciento de incremento en la intención de compra. Sin embargo, la publicidad que es demasiado específica, que parezca cero en un individuo en oposición al perfil demográfico de un grupo, puede ser percibida como invasiva y muy poco cercana a ser cómoda. Esta última situación puede conducir a un 5 por ciento de *disminución* en la intención de compra.¹⁰

Falla 2: Entregar mensaje micro-dirigido equivocado o inapropiado

“¡Algunas de las fallas realmente fueron malas, como mi partido político y mi interés por el tabaco!”

Encuestado

Probablemente peor que estar demasiado cerca es estar equivocado. Cuando un vendedor intenta hacer una conexión personal mediante mensaje que usa información equivocada o inapropiada, los efectos pueden variar desde humorísticos – tal como un veinteañero recibe invitaciones de membrecía a AARP¹¹ - muy triste. El último fue el caso con una oferta de descuento recientemente enviada por correo que, si bien se envió a una persona viva, incluyó una referencia (inexacta) a no solo un familiar recientemente muerto sino la manera como esta persona murió – inmersa en la dirección de correo del destinatario. La firma que había enviado la oferta, que no consideró que podría enviar este correo hasta tanto recibir la prueba física, reclamó que este error fue el resultado de una lista de correo alquilada a un tercero proveedor.¹² Si bien los casos reportados como los de este último ejemplo son raros, basar en información equivocada o inapropiada un mensaje personalizado, y subsiguientemente enviar el mensaje equivocado micro-dirigido a clientes, puede no solo disminuir el efecto de los esfuerzos de mercadeo, sino hacer más daño que bien. Esta reacción adversa a menudo es referida como efecto bumerang: causar que el cliente se mueva de una actitud neutral, no-existente, o positiva frente a la compañía, hacia una negativa.¹³

Falla 3: Valorar de manera inexacta el riesgo

Las instituciones de atención en salud tanto privadas como públicas a menudo confían en modelos de grandes datos para entender las necesidades futuras de sus pacientes y la potencial esperanza de vida. Tales modelos de riesgo, sin embargo, van más allá de administrar la línea de resultados del asegurador mediante ayudar a identificar los clientes de riesgo-alto.¹⁴ Los datos inexactos pueden inducir valoraciones inexactas tales como la determinación de riesgos financieros,¹⁵ expectativas de vida,¹⁶ y necesidades de atención médica, que pueden llevar a pagos inapropiados de seguros.¹⁷ En el peor de los casos, si los grupos de salud pública que usan esos modelos de riesgo para guiar las decisiones estratégicas alrededor de las iniciativas globales de salud pública fallan en señalar, ello puede contribuir a muertes. Esas muertes podrían ser debidas a la identificación equivocada de poblaciones vulnerables o en-riesgo, lo cual podría ser evitado si para ellas estuvieran disponibles los tratamientos correctos.¹⁸

Falla 4: Predecir resultados inexactos

Si bien la mayoría de nosotros ha aprendido a recortar pronósticos del tiempo un poco flojos, nos fijamos en muchas bolas de cristal “científicas” y “estadísticamente importantes”: modelos usados para predecir los resultados de nuestras elecciones,¹⁹ juegos de fútbol, y carreras de caballo. Sin embargo, los modelos llevan a determinar que las precauciones a tomar a menudo estén fuera de lugar. Por ejemplo, en el año 2013, un motor de búsqueda basado en el modelo de seguimiento de la gripa pronosticó un incremento en las visitas al médico relacionadas-con-influenza que fue más del doble que lo que predijeron los Centers for Disease Control and

Prevention (CDC).²⁰ Si bien los CDC basaron sus predicciones en varios reportes de vigilancia de laboratorio recaudados a través de los Estados Unidos, el culpable detrás de la herramienta de seguimiento de los medios de comunicación sociales arrojó resultados salvajemente diferentes que algunos investigadores han denominado “auto-confianza excesiva de los grandes datos”: la equivocación de asumir que los grandes datos pueden ser sustituto para, más que complemento de, los métodos tradicionales de recaudo y análisis de datos.²¹

¿CÓMO LOS GRANDES DATOS SE HICIERON TAN MALOS?

DESAFORTUNADAMENTE, los hallazgos de nuestra investigación primaria no son únicos, sino, más aún, un vistazo al estado general de los amoríos: los grandes datos a menudo son inexactos,²² y las compañías que confían en grandes datos inexactos pueden sufrir consecuencias importantes. Dado que nosotros revisamos solo los campos disponibles para nosotros, es importante observar que las inexactitudes casi ciertamente se extienden más allá de los campos y atributos resaltados en este artículo, especialmente los campos menos comunes o más esotéricos, tales como si un individuo es un veterano.

Entonces, ¿cómo termina esta información fuera del mercado? Hay muchas causas posibles, tales como error humano, errores de recolección o de modelación, e incluso comportamiento malicioso. Para empeorar las cosas, un conjunto de datos a menudo es víctima de más de un tipo de error. Algunos ejemplos de cómo pueden surgir errores:

Los grandes datos a menudo son inexactos, y las compañías que confían en grandes datos inexactos pueden sufrir consecuencias importantes

- Información desactualizada o incompleta puede persistir debido al costo y/o esfuerzo de obtener información actualizada.
- Una organización que usa múltiples fuentes de datos puede de manera incorrecta entrelazar conjuntos de datos y/o ser inconsciente de relaciones causales entre puntos de datos y la carencia de mecanismos adecuados de gobierno de datos para identificar esas inconsistencias.
- Una organización puede caer presa de errores de recolección de datos:
 - Usando poblaciones de muestra sesgadas (sujetas a sesgos de muestreo basados en conveniencia, auto-selección, y/u opciones de optar por no, por ejemplo)²³
 - Hacer preguntas líderes o evaluativas que incrementen la probabilidad de efectos de demanda (por ejemplo, proporcionarles a quienes responden lo que ellos consideren sea la respuesta “deseada” o socialmente aceptable versus su verdadera opinión, sentimiento, creencia, o comportamiento).
 - Recolectar datos en escenarios sub-óptimos que también puedan llevar a efectos de demanda (por ejemplo, encuestas de salida, encuestas públicas, o cualquier mecanismo o entorno en el cual quienes respondan no sientan que sus respuestas serán verdaderamente anónimas).
 - Confiar en datos auto-reportados versus comportamientos (actuales) observados.²⁴
- Los errores del análisis de datos pueden llevar a inexactitudes debidas a:
 - Inferencias incorrectas acerca de los intereses de los consumidores (por ejemplo, inferir que la compra de una revista de ala delta sugiere un estilo de vida riesgoso cuando el verdadero motivo del comprador es un interés por una fotografía).²⁵
 - Modelos incorrectos (por ejemplo, supuestos incorrectos, aproximaciones, o presumir una relación causal que no existe).
- Partes maliciosas pueden corromper datos (por ejemplo, actividad criminal cibernética que altera datos y documentos).²⁶

Entender las causas de esos errores es un primer paso para evitarlos y rectificarlos. La siguiente sección explora los siguientes pasos que las compañías pueden tomar en el camino para utilizar los grandes datos de la manera correcta.

EL LIBRO DE JUEGO DE LOS GRANDES DATOS: PRESCRIPCIONES PARA EL ÉXITO

HAY reconocimiento creciente de que buena parte de los grandes datos está construida a partir de información inexacta, llevando a acciones incorrectas, sub-óptimas, o desventajosas. Están en camino algunos esfuerzos iniciales para poner en funcionamiento regulaciones alrededor del gobierno y la administración de grandes datos.²⁷ Agencias regulatorias, tales como la Federal Trade Commission y la National Association of Insurance Commissioners, están comenzando a considerar más vigilancia a los corredores de datos así como también a cómo son usados los modelos que utilizan sus datos. Sin embargo, las firmas expertas ya comprometidas en grandes datos no deben esperar que las agencias actúen, especialmente dada la incertidumbre alrededor de qué tan efectivas o restrictivas podrán ser cualesquiera regulaciones eventuales. Con base en nuestra experiencia y observaciones del mercado, aquí hay algunas guías, consejos, y remedios a considerar para ayudare a evitar a que usted mismo se dispare en el pie cuando utilice grandes datos.

Incremente la probabilidad de que más de sus grandes datos serán exactos

“Si fueran más inteligentes, podrían hacer referencia cruzada de los datos del hogar con los datos de los ingresos familiares para encontrar discrepancias importantes.”

- Encuestado

Pida y espere más de los corredores de grandes datos. Quizás nuestras expectativas para los grandes datos sean demasiado altas – pero es posible que estemos pidiendo muy poco de los corredores de datos, especialmente dados los resultados del estudio que aquí describimos. El rol de los corredores de datos ha evolucionado con el tiempo. Tradicionalmente, las firmas buscaron que los proveedores de datos les proporcionaran listas y etiquetas de correo de clientes prospectivos y, quizás, para administrar las listas de correo y hacerle seguimiento al comportamiento de compra de los clientes. Sin embargo, la información que los corredores proporcionan ahora juega un rol mucho más integral en nuestras estrategias, interacciones digitales, y modelos analíticos. En consecuencia, debemos estar pidiendo más *accountability*, transparencia, y diálogo continuo con esas organizaciones. (Vea el recuadro, “¿Qué les pide a sus corredores de datos?”)

Conozca las fuentes de los datos. Si bien usted ciertamente desea entender de dónde provienen sus propios datos, conocer la fuente y el linaje es parte importante de la información que usted obtiene a través de los corredores de datos. Sin embargo, nuestra investigación sugiere que los corredores de datos caen en un espectro cuando se trata de revelar sus fuentes. No todos los corredores generan orgánicamente los datos que venden; más aún, muchos licencian información unos con otros, dado que diferentes corredores atienden diversos casos de uso de datos y diversos nichos de negocio.

QUÉ LES PIDE A SUS CORREDORES DE DATOS

Demande transparencia en relación con:

Fuente(s) de los datos: el linaje de los campos y valores de los datos, oportunidad del mantenimiento, procesos de actualización.

Recaudo de datos, validación, y métodos de corrección.

Cualesquiera relaciones e interdependencias – por ejemplo, interrelación entre fuentes de datos e inputs del modelo.

Inputs y supuestos del modelo.

Asegure comunicaciones continuas con las fuentes de los datos en orden a mantenerse al tanto de cualesquiera:

Inexactitudes encontradas en los conjuntos de datos existentes.

Cambios a modelos y/o supuestos y la racionalidad para tales cambios, así como también la transparencia para la lógica del modelo y los meta-datos.

Cambios a las categorías y la racionalidad para tales cambios.

Verifique lo apropiado de la manera como usted está usando sus datos:

Explíquelo al corredor cómo usted está usando los datos, y verifique que su información es apropiada y suficientemente exacta para su contexto.

Considere especificar estándares de exactitud y desempeño en sus contratos con los corredores de datos.

Dé pasos para verificar que los corredores de los cuales usted se abastece tienen control adecuado sobre la exactitud de sus datos, incluyendo control sobre y transparencia en relación con sus fuentes de datos. Entienda el proceso de vigilancia que tienen en funcionamiento con esas fuentes para hacerle seguimiento a los cambios, medir la exactitud, y asegurar consistencia. Desarrolle y mantenga procesos para ser notificado de inexactitudes en los datos, y entienda qué tan a menudo la información es validada o actualizada. Considere la importancia de una diferencia de cinco años de edad: personas de 20 años están comprando productos diferentes que quienes tienen 25 años, al igual que

quienes tienen 25 están en una etapa diferente en la vida que quienes tienen 30 años.

Explore usted mismo los datos. Antes que usted use cualesquiera grandes datos (especialmente los obtenidos externamente) para guiar sus decisiones y estrategias de mercadeo, haga usted mismo un análisis exploratorio de los datos. Si es posible, pruebe una muestra por inexactitudes o inconsistencias contra los campos de datos que usted tenga o pueda validar. Por sí mismo, considere excavar en los datos y hacer verificaciones de validez, análisis exploratorio, y minería de datos contra información individual y de la industria. ¿Lo que usted está viendo tiene sentido? Por ejemplo, uno de los

autores de este artículo fue etiquetado como que tiene una conexión telefónica de Internet desactualizada más que la conexión actual de banda ancha.

Alternativamente, contrate un experto para que mire esos datos. También, tenga en cuenta que la información generada internamente a menudo confía en una combinación de fuentes – las cuales podrían ser externas o desactualizadas – y que también está propensa a error humano, de manera que allí también se tienen que realizar las mismas pruebas de verificación. Una estructura apropiada de gobierno de los datos puede llevar un camino largo para ayudar a asegurar que su información es exacta, oportuna, y valiosa.

Considere que los grandes datos con una herramienta más de la caja de herramientas, no el reemplazo de la caja de herramientas.

Mantenga bajo control las expectativas por los grandes datos. A menudo es el caso de que los grandes datos pueden ser direccionalmente correctos, pero aun así inexactos a nivel individual. La buena noticia para las firmas y para los vendedores es que las analíticas de grandes datos que se construyen en tal información “semi-exacta” pueden proporcionar poder predictivo en general. Sin embargo, es un error esperar que las micro-predicciones individuales conlleven el mismo nivel de exactitud.²⁸

Use y extraiga juiciosamente conclusiones derivadas de los grandes datos. Los grandes datos son una enorme herramienta para los vendedores, pero deben ser pensadas como una herramienta en la toma de decisiones y en la caja de herramientas de mercadeo, no como un reemplazo para las herramientas ya existentes. En consecuencia, no confíe demasiado en un número limitado de puntos de datos, especialmente si la exactitud es un peligro potencial. Si usted decide enviar cualquier micro-mensaje, considere limitar sus geografías y alcance para evitar algunos de los peligros que discutimos antes.

Adicionalmente, solicitar la retroalimentación del cliente sobre los datos no solo mejora el prospecto de datos más exactos – incrementa la transparencia dentro de la relación. Sin embargo, tal y como lo sugieren nuestros hallazgos, usted no puede contar con sus clientes para llenar las brechas adecuada y exactamente.

Complete los grandes datos con otras herramientas para la toma de decisiones. Si bien los grandes datos son y permanecerán siendo una herramienta poderosa para las firmas y para los vendedores cuando son usados apropiadamente, ya hemos explorado los peligros del exceso de confianza puesta en ellos – lo cual también podría resultar en que los vendedores pierdan la fe en su propia experiencia e intuición para ayudar a guiar las decisiones.²⁹ Por consiguiente, los ejecutivos deben complementar las decisiones derivadas de los grandes datos con sus propios conocimientos basados en la experiencia y otros métodos y fuentes de investigación (tal como investigación cualitativa con una muestra pequeña). Independiente de la calidad de los datos, una buena regla es no confiar demasiado en los datos y externalizar muchas decisiones.³⁰

Continuamente conéctese con los clientes

Sea ágil y sensible. Continuamente valore las fuentes de los datos y lo apropiado de las metodologías, los modelos, y los supuestos; frecuentemente revise y valore que las preguntas y las categorías se ajusten con la demografía y las categorías objetivo cambiantes. También, mida qué tan exitosos han sido los esfuerzos de mercadeo desde la incorporación de los conocimientos derivados de los grandes datos. Más allá de medidas cuantitativas u objetivas, cree oportunidades de retroalimentación dentro de su micro-orientación. Luego de recaudar la retroalimentación, gaste tiempo revisando, incorporando, y ajustando sus estrategias con base en esta retroalimentación. Cuando sea apropiado, responda directamente a quienes proporcionen retroalimentación –

la investigación reciente sugiere que esto puede no solo incrementar la probabilidad de retroalimentación individual, sino también hacer que el cliente se sienta más valorado y fomente el diálogo continuo.³¹

Recompense a los clientes por corregir sus datos. Si

bien nuestro estudio sugiere que los consumidores es improbable que corrijan la información proporcionada por una fuente de grandes datos, vale la pena explorar su disposición a tomar acción correctiva para sus propios datos si la solicitud viene de una firma con la cual tiene una relación – y para la cual pueden ver más valor directo derivado de tal acción. Adicionalmente, en un esfuerzo para darles las gracias a los clientes no solo por su patrocinio sino por actualizar la información personal, las

firmas pueden ofrecer incentivos por sus esfuerzos correctivos. Los beneficios podrían ser muchos: datos actualizados del cliente; una línea de comunicación directa, activa; y, en últimas, una conexión más profunda con los clientes.

Independiente de nuestro actual amor ciego por los grandes datos, tenemos que recordar que los datos nunca deben ocupar un lugar central a expensas del cliente. Las firmas que entienden las limitaciones (y las ventajas) de los grandes datos pueden adicionarlos a su arsenal de mercadeo y analítico, ayudando a fomentar y preservar la relación con el cliente y la confianza de que trabajan duro para desarrollarla y mantenerla.

John Lucker es un asesor principal en Deloitte & Touche LLP, líder del mercado de Global Advanced Analytics & Modeling, y líder en los Estados Unidos de Deloitte Analytics.

Susan K. Hogan es un gerente de inteligencia de mercados en Deloitte Services LP.

Trevor Bischoff es consultor senior en el sector de servicios financieros en Deloitte & Touche LLP.

Los autores desean agradecer a Ashley Daily, Adam Hirsch, y Michael Greene, quienes sirvieron como inspiración e incentivaron nuestro entusiasmo por esta investigación. También agradecemos a Negin Rood, Junko Kaji, Aditi Rao, y Kevin Weier por sus contribuciones.

NOTAS FINALES


- ¹ NPR, "'Signal' and 'noise': prediction as art and science," October 10, 2012, <https://n.pr/UPXRS4>.
- ² John Lucker, Ashley Daily, Adam Hirsch, and Michael Greene, "Predictably Inaccurate: Big data brokers," *LinkedIn Pulse*, November 18, 2014, www.linkedin.com/pulse/20141118145642-24928192-predictably-inaccurate-big-data-brokers.
- ³ Todos los porcentajes relacionados con quienes respondieron fueron calculados con base en el número de quienes respondieron de quienes datos de terceros actualmente estaban disponibles en las categorías de interés; los cálculos excluyeron a quienes respondieron para quienes no estaban disponibles datos de terceros.
- ⁴ Lucker et al., "Predictably inaccurate."
- ⁵ Thomas Schutz, "Want better analysis? Consider the data," *Credit Union Journal*, May 16, 2014, www.cujournal.com/opinion/want-better-analysis-consider-the-data.
- ⁶ Morgan Hochheiser, "The truth behind data collection and analysis," *John Marshall Journal of Information Technology and Privacy Law* 33, no. 1 (2015): pp. 32–55, <http://repository.jmls.edu/jitpl/vol32/iss1/3>.
- ⁷ Victoria Petrock, "Are consumers warming to personalized marketing services?," eMarketer brief, July 26, 2016, www.emarketer.com/Brief/Consumers-Warming-Personalized-Marketing-Services/5500941.
- ⁸ Susan K. Hogan, Rod Sides, and Stacy Kemp, "Today's relationship dance: What can digital dating teach us about long-term customer loyalty?," *Deloitte Review* 20, January 23, 2017, <https://dupress.deloitte.com/dup-us-en/deloitte-review/issue-20/behavioral-insights-building-long-term-customer-loyalty.html>.
- ⁹ Irwin Altman and Dalmas A. Taylor, *Social Penetration: The Development of Interpersonal Relationships* (New York: Holt, Rinehart and Winston, 1973).
- ¹⁰ Ithaca College, "Online creep: Targeted ads may have opposite effect of marketers' intent," *Science Daily*, April 8, 2015, www.sciencedaily.com/releases/2015/04/150408171201.htm.
- ¹¹ Joshua Lederman, Twitter post, September 29, 2016, 1:48 p.m., <https://twitter.com/joshledermanap/status/781596504351907840>.
- ¹² Natasha Singer, "Oops! Health insurer exposes member data," *New York Times*, November 10, 2014, <https://nyti.ms/2qycdOU>.
- ¹³ Sharon S. Brehm and Jack Williams Brehm, *Psychological Reactance: A Theory of Freedom and Control* (New York: Academic Press, 1981).
- ¹⁴ Leslie Scism, "Life insurers draw on data, not blood," January 12, 2017, *Wall Street Journal*, www.wsj.com/articles/the-latest-gamble-in-life-insurance-sell-it-online-1484217026.
- ¹⁵ Ankur Aggarwal et al., "Model risk—daring to open up the black box," *British Actuarial Journal* 21(2), December 22, 2015, http://journals.cambridge.org/abstract_S1357321715000276.
- ¹⁶ Scism, "Life insurers draw on data, not blood."
- ¹⁷ Rachel S. Karas, "Stakeholders urge CMS to factor Rx drugs in risk assessment pay, question other CMS ideas," *InsideHealthPolicy's Daily Brief*, April 28, 2016.
- ¹⁸ Jane Freemantle et al., "Indigenous mortality (revealed): The invisible illuminated," *American Journal of Public Health* 105(4), April 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC4358192/.
- ¹⁹ Jim Rutenberg, "A 'Dewey defeats Truman' lesson for the digital age," *New York Times*, November 9, 2016, <https://nyti.ms/2JL43lb>.
- ²⁰ Vasileios Lamos, Andrew C. Miller, Steve Crossan, and Christian Stefansen, "Advances in nowcasting influenza-like illness rates using search query logs," *Scientific Reports* 5 (2015), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4522652/>.
- ²¹ Larry Greenemeier, "Why big data isn't necessarily better data," *Scientific American*, March 13, 2014.
- ²² Daniel A. McFarland and H. Richard McFarland, "Big data and the danger of being precisely inaccurate," *Big Data & Society*, July–December 2015, pp. 1–4; <http://journals.sagepub.com/doi/full/10.1177/2053951715602495>.
- ²³ StopDataMining.me, "Opt out list," www.stopdatamining.me/opt-out-list/, accessed May 2, 2017.
- ²⁴ McFarland and McFarland, "Big data and the danger of being precisely inaccurate."
- ²⁵ Scism, "Life insurers draw on data, not blood."
- ²⁶ Mark Ward, "How fake data could lead to failed crops and other woes," *BBC*, March 21, 2017, www.bbc.com/news/business-38254362.
- ²⁷ Carten Cordell, "Transparency advocates pitch the next big thing in big data," *Federal Times*, March 23, 2017, www.federaltimes.com/articles/with-data-act-reports-looming-transparency-advocates-pitch-the-next-big-thing.
- ²⁸ Lucker et al., "Predictably inaccurate."

²⁹ Ulrich Hoffrage, "Overconfidence," in Rüdiger F. Pohl, editor, *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory* (New York: Psychology Press, 2004); Tore Håkansson and Tim Carroll, "Is there a dark side of big data—point, counterpoint," *Journal of Organization Design* 5(5), July 12, 2016, <http://link.springer.com/article/10.1186/s41469-016-0007-5>.

³⁰ Nigel Harvey, "Confidence in judgment," *Trends in Cognitive Sciences* 1(2), May 1997, pp. 78–82, [www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613\(97\)01014-0](http://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(97)01014-0).

³¹ Susan K. Hogan and Timothy Murphy, *Loving the one you're with*, Deloitte University Press, June 17, 2016, <https://dupress.deloitte.com/dup-us-en/focus/behavioral-economics/how-behavioral-factors-influence-customer-rewards-incentives.html>.

Deloitte.
University Press

 Siga @DU_Press #DeloitteReview

Para recibir actualizaciones por correo electrónico
suscribase en dupress.deloitte.com

