

iab españa

LIBRO BLANCO DE

DATA



1. Introducción _05

a. Definición de los distintos tipos de datos

1.1 Tipos de datos según su categoría:

- Datos estructurados
- Datos no estructurados

1.2 Tipos de datos con respecto a la clasificación para el Marketing Digital:

- Zero Party Data
- First Party Data
- Second Party Data
- Third Party Data

1.3 Tipos de datos con respecto a la categorización por su proceso de recolección:

- Datos determinísticos
- Datos declarativos
- Datos modelizados

1.4 Tipos de datos en cuanto a la categorización por la información que nos aportan:

- Datos sociodemográficos
- Datos comportamentales
- Datos intenders
- Datos Geolocalizados

b. ¿Quién es el propietario del data?

2. Tipos de actores _10

a. Agentes compradores de datos

- Anunciante
- Publishers
- Agencia de Medios
- Trading Desk
- DMP
- Otros

b. Agentes vendedores de datos

- Recolectores de datos
- Agregadores de datos
- Procesadores de datos



c. Data exchange (DXP)

3. El valor de los datos _13

a. Cuál es el valor del dato

b. Modelos de comercialización

c. Calidad de la data

4. Cross-Device: la solución multidispositivo _18

a. Quién lo utiliza

b. Modelos probabilísticos

5. Estandarizaciones de mercado _21

a. Organismos de la industria que regulan a los Data provider

• EDAA

• JICWEBS

• IAB Europe Transparency & Consent Framework

• ePrivacy Seal

6. Usos de la data _25

a. Customer journey: trazabilidad del usuario

b. En la activación de medios: recogida, segmentación y activación en media

c. Para la creatividad (no es solo retargeting)

d. Para la investigación

e. Modelización

f. Otros medios (off line)

g. Otros usos

7. Limitaciones del data _32

• La limpieza de los datos

• Migración de los datos

• La calidad de la data

• El coste y persistencia de la data

• Limitaciones en el grado de alcance de los datos



8. Buenas prácticas (transparencia, fraude) _33

- Definición de fraude
- Tipos de fraude en función de la tipología de data
- Provenientes del fraude publicitario digital:
- Requisitos de transparencia y ejemplo de buenas prácticas

9. Legal _37

10. Glosario _40

11. Agradecimientos _43



Los usuarios están utilizando cada vez más una gran variedad de canales y puntos de contacto tanto para comprar como para consumir contenido. Las interacciones y comportamientos a través de estos puntos de contacto (Webs de noticias e información, Redes Sociales, buscadores, Video Online, juegos, etc.) originan una gran cantidad de datos que inicialmente pueden parecer desconectados. Si se combina este hecho con el uso creciente de herramientas de adblockers, el ascenso de la publicidad programática y las nuevas formas de comunicación, el reto de crear publicidad relevante para el usuario se vuelve aún más complicado.

A medida que los consumidores navegan a través de estos canales y puntos de contacto, deben recibir mensajes publicitarios relevantes y de calidad y en este sentido los datos tienen un papel clave. Así, el aprovechamiento inteligente de los datos puede revelar y establecer conexiones entre estos puntos que permiten comprender y dirigirse a los clientes a medida que avanzan en el customer journey.

Los datos son un tema de plena actualidad y vital para todos los actores del ecosistema digital. Es por ello por lo que es importante entender los diferentes tipos de datos disponibles y cómo hacer un buen uso de estos. El objetivo de este libro blanco es ofrecer un marco de referencia, así como una orientación sobre los tipos de datos y su utilización en las estrategias de Marketing Digital.

a. Definición de los distintos tipos de datos

1.1 Tipos de datos según su categoría:

• Datos estructurados:

Hablamos de datos estructurados al referirnos a la información que encontramos en la mayoría de bases de datos. Son los archivos de tipo texto que podemos mostrar en filas y columnas con títulos. Estas bases de datos pueden ser ordenadas y procesadas de forma sencilla por todas las herramientas de tratamiento de datos.

• Datos no estructurados:

Por el contrario, los datos no estructurados son aquellos que no residen en las bases de datos relacionales, tales como documentos XML o datos almacenados en bases de datos NoSQL. Los más frecuentes son los generados a través de social media, ficheros de imágenes jpg, ficheros de audio mp3 o ficheros de video tipo flash.

1.2 Tipos de datos con respecto a la clasificación para el Marketing Digital:

• Zero party data:

Son datos declarativos de los consumidores recogidos fundamentalmente a través de encuestas.

• First Party Data:

Se conoce como First Party Data, a los datos recolectados a través de las posibles interacciones que un usuario puede realizar con cualquiera de los properties de una compañía. Estamos hablando de datos recogidos por el CRM, el call center, el paid media, y también recogen conductas, acciones o intereses demostrados a través de la navegación de su sitio web o datos creados a través de taxonomías realizadas a medida.

• Second Party Data:

Son datos que esencialmente provienen de acuerdos de compra y venta con diversos partners. Este modelo de datos permite a los anunciantes poder ampliar sus bases de datos (First Party Data) con información muy valiosa que de otra forma no podrían incorporar.

Además, dentro de esta categoría podemos encontrar tres casuísticas:

- Enriquecimiento de First Party: Se trata de datos de un segundo que dan nuevos matices acerca de nuestros datos. Por ejemplo, cruzar el código postal de nuestros clientes con el valor catastral ofrecido por el gobierno. En estos casos, la información adquirida puede ser gratuita o de pago.
- Intercambio: Dos compañías deciden llegar a un acuerdo recíproco de intercambio de cookies. En estos casos, las cookies recibidas, nunca habrán sido vistas y ofrecerán la posibilidad de ser usadas para prospección.
- Compra: Hay marcas que ponen a la venta sus cookies a otras marcas. En estos casos, las cookies recibidas, nunca habrán sido vistas y ofrecerán la posibilidad de ser usadas para prospección.

• Third Party Data:

El Third party data son datos procedentes de compañías que, recopilan datos y comportamientos de miles de webs, y las cuales agregan para venderlos con fines publicitarios. Las fuentes de estos datos, así como los tratamientos a los que han sido sometidos, nunca son compartidos. Estos proveedores ofrecen cookies con información como edad, género, categoría socio-profesional, interés, localización geográfica, etc.

1.3 Tipos de datos con respecto a la categorización por su proceso de recolección:

• Datos determinísticos:

Son los que se recolectan a través de fuentes 100% verificadas, como registros recogidos en distintas bases de datos como en el CRM de una compañía, y ofrecen un valor único del individuo.

• Datos declarativos:

El usuario ha facilitado información no contrastable, a través de cualquier tipo de fuente que nos permite asignarle un perfil de audiencia.

• Datos inferidos:

Se asigna un perfil de audiencia a un usuario en base a sus patrones de navegación.

• Datos modelizados:

Se generan realizando extrapolaciones de una muestra representativa de audiencia que hayan realizado una determinada acción:

- Lookalike: son segmentos caracterizados con diversas variables y donde buscamos audiencias similares.

1.4 Tipos de datos en cuanto a la categorización por la información que nos aportan:

• Datos sociodemográficos:

Edad, género, ocupación, clase social, etc...

• Datos comportamentales:

Behavioral Targeting consiste en construir perfiles de cada usuario, basándose en variables relacionadas con su comportamiento, como puede ser identificando las palabras clave relevantes de un contenido consumido.

• Datos intenders:

Audiencias que a través de alguna acción han mostrado interés en un bien o servicio determinado.

• Datos Geolocalizados:

Audiencias construidas a través de precisar el punto geográfico donde se encuentran.

b. ¿Quién es el propietario del data?

“El propietario del dato es el usuario, y esto le otorga una capacidad de decisión sobre los mismos”

Es necesario aclarar que el que genera la información es el usuario, y por tanto, es él el dueño de esa información. Hace tiempo existían por internet, webs o aplicaciones, que te permitían saber cuánto dinero podrías ganar si comercializabas tus datos de Redes Sociales. Eran simuladores que calculaban el nivel de influencia en tu red para poder extraer una estimación del ingreso que podrías llegar a generar.

El problema empieza con las zonas grises sobre la propiedad comienzan cuando consideramos el tercero que está viendo lo que hacemos, posiblemente porque lo estamos haciendo en su casa sitio web o en un espacio de su propiedad, ¿esto les convierte a ellos también propietarios de la información?

Anunciantes, Publishers, usuarios, etc., se muestran todos propietarios de algún tipo de dato y vamos a ver el porqué:

- **Caso Publisher:** el Publisher reclama los datos de los usuarios que se mueven en su entorno.
- **Caso Anunciante:** el anunciante Necesita todo el feedback posible para poder mejorar su producto.
- **Caso Usuario:** una página web sin usuarios que la visitan no vale nada, pero sólo con un usuario, tampoco vale... es útil cuando hay millones y millones. Aun así, todo lo que produce un usuario en cuanto a información es de su propiedad y sobre esto no hay discusión.

Dicho esto pasamos a analizar la situación: tenemos un Publisher que permite que la información se genere ya que ofrece un lugar donde generarla; tenemos un anunciante que hace campañas y permite que la información que se genere sea más rica; y finalmente tenemos el usuario que solo no hace nada, su información no vale si no es agregada.

Con este panorama es justo decir que todo cambiaría si fuera posible hacer análisis en un entorno determinista donde todo lo que conocemos no sea fruto de una inferencia partiendo de una muestra de usuarios sino que de ese usuario conocemos exactamente esa información.

La situación actual es de total inestabilidad en todos los frentes. Por un lado, las nuevas tecnologías de blockchain y hash-graph hacen soñar al usuario con ese día en el que podrá vender su propia información sin intermediarios. Por otro lado, los Publisher lanzan un mensaje muy importante a los defensores de esa teoría: es muy simple, si tuviéramos que pagar las tecnologías que usamos día a día en internet dejaríamos de pensar en ganar dinero por nuestra Data puesto que nos daríamos cuenta de que ya lo ganamos en forma de servicio.

El anunciante en todo esto se encuentra con la teoría de las estrategias Data-Driven y un mercado que a todos los efectos no es un mercado ni una industria del dato. Esto ha llevado a que cada anunciante quiera crear su propio ecosistema de extracción de datos para poder trabajar en un entorno que evalúa como el adecuado.

Unido al tema de la propiedad del dato va desarrollándose desde hace años la idea de reglamentar a nivel de la Unión Europea todo lo que se refiere a cesión de información a través de internet. Esto lo que provocará es una profesionalización de la industria del dato, dando el valor apropiado a todos los componentes de la cadena de generación de la información lo que permitirá efectivamente crear estrategias Data-Driven.

En definitiva , gracias a la regularización del mercado del data podremos entender de verdad cuál es el valor económico que aporta y sobre todo nos ayudará a unificar, estandarizar ciertos procesos que ayudarán a mejorar nuestras estrategias basadas en datos.



a. Agentes compradores de datos

En la publicidad tradicional es muy fácil identificar quienes son los factores implicados, al fin y al cabo siempre hay alguna empresa que ofrece un espacio que otra empresa está dispuesta a comprar. Con respecto a los datos, esta reducción también es válida y, por lo tanto, podemos distinguir entre empresas generadoras de datos y empresas compradoras de los mismos.

Los datos son información y la información es algo que puede comprarse, no solo por empresas sino por personas también. En el ecosistema publicitario actual, podemos analizar cada uno de los componentes y preguntarnos si comprarían o no datos y qué harían con ellos.

• Anunciante

El primer agente que encontramos es el anunciante. Este quizás sea el más claro, debido a su deseo de completar el customer journey y saber qué ofrecer y en qué momento, pues es lo que desde siempre les ha empujado a comprar datos, pero ahora ha cambiado la forma de uso.

Es el agente interesado en la difusión de la publicidad en los medios. Son los que determinan el presupuesto total que van a emplear en una campaña publicitaria, definen el público objetivo deseado y proporcionan los materiales creativos. En referencia a la compra de datos, pueden comprar datos para enriquecer sus CRMs o comprarlos para la activación directa si trabajan con una agencia de medios o un Trading Desk.

• Publishers

El segundo agente que encontramos es el Publisher, el titular de una página web. En publicidad online se utilizan también los términos afiliado, editor o partner. Pueden comprar datos para enriquecer su inventario con perfiles de audiencia o para enriquecer sus DMP.

Este no deja de ser una compañía que capta usuarios para vender usuarios, la calidad de su tráfico está directamente relacionada con la calidad de su estrategia de captación. En cuanto al dato, el Publisher es el nuevo agente comprador, en el pasado el Publisher había comprado tráfico pero nunca data.

Esta compra no tiene que ser por activación de campañas publicitarias sino que está más centrada en el concepto de personalización de contenido para sus usuarios.

2. Tipos de actores

• Agencia de medios

Es el agente comprador de espacios publicitarios para sus anunciantes finales con los que tienen acordados el contrato publicitario. Intermedian en la compra del tráfico para los anunciantes. Son los responsables de la consecución de los objetivos esperados por sus clientes finales, bien sea resultados de branding o de performance. Ellos disponen de los presupuestos de los anunciantes para ser invertidos en aquellos DSPs, ad exchanges, networks, etc. que consideren más oportunos para el éxito de la campaña.

• Trading Desk

Es el equipo técnico de personas que optimizan la compra programática de sus distintos anunciantes conectando con múltiples DSPs. Pueden formar parte de la agencia de medios o ser compañías independientes. Generalmente el trading desk está enfocado hacia la compra inteligente de audiencias mediante tecnologías de optimización en múltiples plataformas (ad exchanges, networks, DSPs, soportes) de manera tal, que comprará aquella impresión de manera más eficiente para los objetivos de sus clientes/anunciantes.

• DMP

Los DMPs (Data Management Platform) son plataformas tecnológicas en las que se recolectan, se integran y se gestionan grandes cantidades de datos estructurados y no estructurados. Tienen la capacidad de recopilar datos de audiencia procedentes de cualquier fuente.

• CDP

El CDP (Customer Data Platform) es un software paquetizado que crea una base de datos de clientes unificada y persistente que es accesible a otros sistemas. Básicamente, es un sistema prediseñado que centraliza los datos de los clientes de todas las fuentes y luego los pone a disposición de otros sistemas para campañas de marketing, servicio al cliente y todas las iniciativas de experiencia del cliente.

• Otros

Como ejemplo de otros agentes compradores de datos encontramos a todas aquellas empresas que ofrecen servicios de medición de audiencias.

b. Agentes vendedores de datos

• Recolectores de datos:

Los recolectores de datos son Publishers, Redes Sociales y otras empresas propietarias de datos que venden la data de sus usuarios tales como: datos de las interacciones con las campañas publicitarias que activan, datos de navegación de los usuarios por sus páginas, datos de registro, datos de ventas. (ej: Amazon, Telefónica, grupos editoriales, Facebook, Yahoo...)

• Agregadores de datos:

Los agregadores de datos son vendors que comercializan bases de datos de terceros tales como emailing, estudios de solvencia de clientes, datos de contratación de las empresas de telecomunicaciones, datos de las pasarelas de pagos, datos de pymes, etc. Actúan como intermediarios entre el propietario del dato y el comprador. Algunos ejemplos son empresas como zeotap, experian, Datmean, Antevenio...

• Procesadores de datos:

Los DMPs (Data Management Platform) son plataformas tecnológicas en las que se recolectan, se integran y se gestionan grandes cantidades de datos estructurados y no estructurados. Además de la capacidad de recopilar y agregar datos procedentes de cualquier fuente para la creación de segmentos de audiencia, tienen capacidades analíticas y conexiones con las distintas plataformas para la activación publicitaria. Como ejemplos encontramos a Blukai, Semasio, Weborama, Krux o Adobe DMP.

c. Data exchange (DXP)

Un data exchange es una plataforma que da acceso a múltiples fuentes de datos a través de las cuales los compradores pueden seleccionar el vendedor adecuado para los insights de campaña al precio que comprador y vendedor determinan. Funcionan como un Marketplace de datos. Algunos ejemplos son Exelate, Eyeota, Adsquare, Lotame...

3. El valor de los datos

a. Cuál es el valor del dato

Si pudiéramos dividir la publicidad en eras podemos decir que estamos en la era de la "digitalización digital", es decir, con la programática y con el uso de las tecnologías y de la información estamos en pleno proceso de optimización de nuestra metodología de hacer Marketing Digital.

Esto pasa porque en un primer momento de "digitalización" de la publicidad se han instaurado procesos muy manuales que incluso han creado puestos de trabajo para poder ser desarrollados. Toda esta ineficiencia de los inicios del sector digital ha puesto muy en duda ese medio, ya que su efectividad y su eficacia en algún caso eran cuestionables. Si nos paramos un momento a pensar y consideramos lo que algunos decían, la historia se repite, ahora mismo estamos en una era pre-Data o dicho de otro modo, en un momento en el que el Data ya existe en todos los planes de publicidad pero la planificación, los procesos, los costes hacen pensar que de momento es poco eficiente y en muchos casos poco eficaz

Otra cuestión que hay que tener en cuenta es que la evolución de una industria no depende solo de la tecnología sino también de las personas que la componen. En este sentido todavía es complicado encontrar perfiles que sepan aprovechar al 100% la tecnología y no hay suficiente experiencia en este sector como para encontrar perfiles senior. Esto es algo que lleva necesariamente tiempo y tendrán que pasar años para encontrar perfiles adecuados para la industria.

En definitiva, el valor del dato es al que compras una audiencia para su activación frente a comprar sólo espacio publicitario y es variable según su cualificación y su volumen. Además, no depende tanto del dato en sí, sino también del proceso y las herramientas que se usen tanto a nivel de recursos humanos como a nivel de tecnología.

b. Modelos de comercialización

En el ámbito publicitario, hace años que los datos se comercializan para la optimización de las diferentes campañas de Marketing Online, por ejemplo, compañías que comercializan bases de datos de direcciones de correo electrónico para campañas de email marketing.

Es con la llegada de la compra-venta programática donde se produce un auge importante de la comercialización de datos y se comienza a trabajar con lo que se conocen como datos de audiencia anonimizados. Ya no hablamos de datos

3. El valor de los datos

personales, como puede ser una dirección de correo electrónico, sino de datos comportamentales, geolocalizados o estadísticos que permiten relacionar a un usuario o grupo de usuarios con determinados comportamientos o intereses.

Gracias a la compra-venta programática se puede acceder a cantidades ingentes de inventario, pero su gran potencial son las capacidades de segmentación que ofrece, por este motivo son los DSPs (Demand Side Platform) y Ad exchanges las primeras plataformas que ponen a disposición distintas fuentes de datos para poder targetizar las campañas publicitarias.

Los traders pueden definir las estrategias de datos para segmentar sus campañas comprando segmentos de audiencias a través de estas plataformas. El modelo de transacción más utilizado en las plataformas es el del CPM (Coste Por Mil). Las empresas proveedoras de datos son las encargadas de transferir paquetes de audiencias en los DSPs o Ad exchanges para que puedan ser comprados.

Básicamente estas compañías, los data provider, cuelgan sus paquetes de audiencia en los escaparates de los DSPs o Ad exchanges para ponerlos a disposición del mercado. Cuando se detecta a un usuario que cumple con las segmentaciones marcadas en la campaña, se produce la puja o compra, el precio marcado de esta compra se divide en CPM fijado para el medio o publisher y otro CPM fijado para el proveedor de datos.

El crecimiento de este nuevo ecosistema publicitario hace que aparezcan nuevos players aportando su valor diferencial y podemos encontrarnos otros modelos de venta de bases de datos, como puede ser el pago de un fee que cubra toda la explotación de dicha bases de datos.

Con la llegada de las tecnologías DMPs (Data Management Platform), los anunciantes y Publishers pueden tener acceso a otras fuentes de datos y generar acuerdos de compra-venta de second party data. En estos casos, los modelos de negocio de compra-venta se definen de forma particular, pudiendo estar basados en CPM, en un coste por cookies transferidas o modelos de fees únicos o mensuales.

Además, estamos empezando a ver modelos de data colaborativa donde básicamente se produce un intercambio de audiencias entre compañías que no entran en conflicto sectorial pero pueden tener sinergias en sus targets.

c. Calidad del data

Cuando estamos usando o adquiriendo data, debemos tener en cuenta dos dimensiones: calidad y volumen.

Así, la calidad del data está reñida con los siguientes atributos:

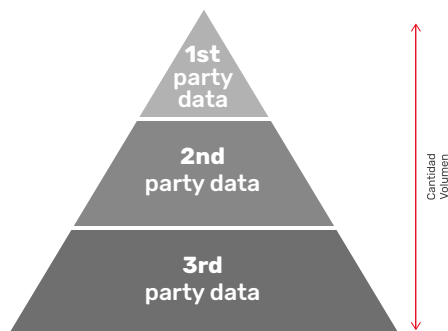
- **Transparencia:** conocer el origen y el modo en el que se ha procesado la información ayuda a determinar la fiabilidad del dato usado, así como el resultado esperado de su activación.
- **Persistencia:** al vincular la data a cookies, estas tienen un ciclo de vida limitado. Un usuario pueda borrarlas, si bien técnicamente se puede fijar duraciones extensas de meses. Dependiendo del tipo de información, una cookie con una vida larga puede no ser útil.
- **Obsolescencia:** el tipo de información asociada a una cookie puede ser muy efímera. Por ejemplo, cookies con información de intencionalidad de compra, su valor se reduce cuanto mayor sea el paso del tiempo, ya que el usuario ha podido adquirir el producto. En cambio, la utilidad de la información demográfica, no tiene porqué decaer con el paso del tiempo. Datos declarativos del género o edad de los usuarios, poseen un valor constante, ya que es un valor que no varía en cuestión de meses, y en caso de hacerlo, es probable que esas cookies ya hayan sido eliminadas.

Cuando contraponemos calidad y volumen respecto a los distintos tipos de data, obtenemos la siguiente relación:



3. El valor de los datos

Se dice siempre que el **First Party Data** tiene la mayor calidad. Esto es debido a que el origen del dato es conocido, son datos propios, cuyo proceso de adquisición es controlado por la propia empresa, así como a los procesamientos a los que han sido sometidos. Sin embargo, el volumen de data es siempre reducido, limitado al volumen de los activos digitales de la propia compañía.



El **Second Party Data** ofrece la posibilidad de ampliar el alcance de la audiencia de First Party Data y debido a cómo se ha accedido a él, ya sea través de un acuerdo entre compañías o un contrato económico, la calidad es elevada. Las constricciones de volumen vienen dadas por la disposición de alcanzar acuerdos o la capacidad económica para sostenerlos.

El **Third Party Data** ofrece los mejores volúmenes, si bien, la opacidad respecto al origen del data y al procesamiento que ha tenido, invita a preguntarse acerca de su calidad. No es extraño detectar incongruencias cuando se realiza un análisis sobre los segmentos de un data broker (vendedores de datos). Por ejemplo, empleando un DMP y usando los análisis de overlap, no es infrecuente ver que, la información demográfica ofrecida por un proveedor tiene solapamientos considerables entre sí, y que el segmento de hombres, contiene cookies clasificadas también como mujeres, y viceversa. A pesar de estas discrepancias, no se debe descartar el uso de Third Party Data, es aconsejable tener en cuenta que se trata de un negocio de volumen, en el cual se recurre a modelización como lookalikes o inferencias estadísticas para poder ofrecer el alcance necesario, y que aunque diluye la calidad y la efectividad, sigue guardando su utilidad cuando los costes de los medios son reducidos.

El modelo de negocio de los data brokers consiste en nutrirse de pequeñas websites, las cuales venden su data de manera anónima a través de ellos, no obstante algunas marcas comercializan sus audiencias a través de ellos usando su propia identidad comercial.

Muchos data brokers conscientes de la creciente mala fama respecto a la calidad de su data, han establecido ciertas medidas para mejorar la percepción de sus servicios como limitar la vida de cookies a un máximo de días, aumentar precios o

3. El valor de los datos

hacer públicos algunos de los sites de los cuales obtienen la información, y otros han optado la táctica comercial de autodenominarse cooperativas de Second Party Data generando cierta confusión con una verdad a medias. Es cierto que la data que ofrecen es proveniente de otros sites, sin embargo, en muchos casos ha sido agregada sin aplicar ninguna segmentación o se ha aplicado algún tratamiento estadístico que ha mermado su calidad.

El uso del Third Party Data está justificado desde la óptica de negocio en las situaciones en las cuales se carece de información de los usuarios y el uso de éste ayuda a realizar profilings para orientar campañas online u ofrecer audiencias de las que se carece información.

Respecto a la calidad del data cabe mencionar el importante rol de los publishers, los mayores suministradores de audiencias para las marcas. Es aconsejable que los soportes realicen algún tipo de procedimiento de validación de las audiencias que comercializan, especialmente en el caso de behavioural targeting o de audiencias inferidas a través de técnicas de Data Science. Dicha validación se puede realizar fácilmente mediante encuestas online a esos usuarios y es un potente argumentario de ventas.



4. Cross-device: la solución multidispositivo

La navegación online es cada día más una realidad multi-dispositivo: crece el número de personas con ordenador, tablet y smartphone, y algunos de ellos se tienen incluso por duplicado para el uso personal y el laboral.

Los usuarios recorren webs, apps y Redes Sociales saltando de un dispositivo a otro y continúan lo que estaban consultando desde un aparato en el siguiente. En este contexto, los sistemas cross-device analizan el comportamiento global de los usuarios, de dispositivo a dispositivo, para poder ofrecer publicidad optimizada a cada canal y mejorar así la eficacia de las campañas de marketing.

La combinación de los datos del usuario y de su comportamiento online en cada uno de los dispositivos, o en el conjunto de varios de ellos, permiten diseñar una publicidad más inteligente, centrada en las necesidades y preferencias del individuo, adaptada a la hora del día en la que se encuentra más receptivo y al dispositivo que esté utilizando. Es decir, gracias a la identificación de los usuarios y sus acciones a lo largo de todo el funnel que facilita el cross-device, se puede crear una experiencia publicitaria sincronizada y personalizada, traducible en un customer journey realista. Con ello, además, se evitan impactos innecesarios, con lo que se optimiza la rentabilidad de la inversión publicitaria y se consigue también una mejor precisión en la medición de las campañas de marketing.

Otra de las grandes ventajas del marketing cross-device, y la más importante en su próximo desarrollo, es la generación de perfiles de uso y de comportamiento que no solo permiten conocer a los usuarios ya existentes, sino también llegar a inferir cómo actuarán usuarios futuros a partir de sus experiencias anteriores.

a. Quién lo utiliza

Los sistemas cross-device resultan interesantes tanto para vendedores (publishers, ecommerce, marketplaces, etc.) como para compradores (trading desks, agencias de medios, empresa final, anunciantes...).

Para los compradores, además de poder personalizar el mensaje publicitario y contar con la posibilidad de limitar la exposición de anuncios, las soluciones cross-device permiten conocer el resultado exacto de cada canal, detectar audiencias duplicadas, evitar así una inversión innecesaria y mejorar la identificación de clientes realmente interesados en cada campaña.

4. Cross-device: la solución multidispositivo

Para los publishers, una solución cross-device aumenta la sofisticación en la creación del mensaje publicitario en función del dispositivo, lo que mejora de manera significativa la suite de productos de publicidad ofrecidos. El resultado es una nueva estrategia de monetización mediante la captación de ingresos de anunciantes premium que quieran maximizar el valor de su campaña. De esta manera, los vendedores ven crecer la fidelización de sus clientes gracias a este valor añadido, siempre y cuando, claro está, consigan incrementar el compromiso con el target definido.

b. Modelos probabilísticos

El éxito de los sistemas cross-device radica en un buen análisis de los datos que permita identificar usuarios únicos a través de sus diversas conexiones en múltiples dispositivos. Normalizar y explotar correctamente estos datos es el gran reto del cross-device y, en función de la naturaleza de dichos datos, se contemplan dos tipos de modelos: deterministas y probabilísticos.

Los modelos deterministas se basan en la información ofrecida voluntariamente por usuarios registrados: no solo el correo y los datos demográficos, sino sus hobbies y preferencias, las búsquedas que realizan o su actividad en las redes sociales. Es el caso de la navegación ligada a la autenticación de los usuarios mediante login o customer ID, habituales en ecommerce. Otra posibilidad es el modelo de Walled Gardens, en el que la información proviene de terceros como redes de afiliación, Google, Facebook e incluso las propias operadoras de telefonía. Los datos provenientes de redes sociales resultan especialmente interesantes, ya que permiten definir modelos de comportamiento muy útiles para construir clusters. La fiabilidad de los modelos deterministas es muy elevada, pero presentan grandes limitaciones a la hora de gestionar a los usuarios no registrados.

Los modelos probabilísticos van un paso más allá al incorporar matrices de datos y algoritmos capaces de crear patrones y establecer conexiones posibles. Es decir, a través de la información de la navegación, los datos de proximidad por GPS o incluso las direcciones IP, es posible inferir similitudes entre los usuarios no identificados y el comportamiento de los que sí han ofrecido información mediante registro.

4. Cross-device: la solución multidispositivo

Por ejemplo, un usuario registrado en un portal de búsqueda de empleo, por el que conocemos su edad, profesión y nivel de ingresos, accede a una página de venta de coches y centra su búsqueda en tres marcas concretas de vehículo. Los modelos probabilísticos analizan a los usuarios con un comportamiento de navegación similar a este e intentan establecer patrones de probabilidades con el fin de determinar si pertenecen a un perfil sociodemográfico parecido, a pesar de no contar con información directa en este sentido.

El modelo probabilístico más prometedor hasta la fecha es el device fingerprint, que se basa en los atributos de los ajustes del dispositivo y los navegadores, direcciones IP, conexiones de wifi y actividad online. La combinación de esta información con los datos de diversos DMPs, mediante técnicas de minería y Big Data, permite clusterizar e inferir comportamientos de usuarios teóricamente desconocidos. La precisión de los métodos probabilísticos dista aún mucho de la de los métodos deterministas, pero cuentan con la ventaja de ser muy escalables. Será, por tanto, una buena combinación de ambas opciones la que logre el equilibrio más adecuado para un buen resultado del marketing cross-device.

Aunque a día de hoy todavía existen importantes retos tecnológicos para optimizar el funcionamiento de los sistemas cross-device, sobre todo por la dificultad de identificar de manera inequívoca la navegación de un mismo usuario entre diversos dispositivos, no cabe duda de que se está trabajando fuertemente en esta línea y que constituirá una herramienta imprescindible y cada vez más potente en el futuro cercano.



a. Organismos de la industria que regulan a los Data providers

• EDAA (European Interactive Digital Advertising Alliance):

La Alianza Europea de Publicidad Digital Interactiva, de la que forma parte IAB Europe y gestiona el sistema de autorregulación de OBA (**Online Behavioural Advertising- OBA**). Gestionan el proceso de licencia del icono de Ad Choices (OBA) y también administran el programa de certificación según estas normas europeas. La iniciativa intersectorial de autorregulación fue desarrollada por los principales organismos europeos con el fin de introducir normas paneuropeas para mejorar la transparencia y el control de los usuarios de la publicidad comportamental online.

• JICWEBS (Joint Industry Committee for Web Standards):

Está compuesto por representantes de la ISBA (The Incorporated Society of British Advertisers), la IPA (Institute of Practitioners in Advertising), la IAB UK (Interactive Advertising UK) y el AOP (Association for Online Publishing). Fue creado por la industria de medios del Reino Unido e Irlanda para garantizar el desarrollo independiente de estándares y benchmarking de las mejores prácticas para el comercio publicitario online. JICWEBS ha desarrollado varias normas de información y principios de buenas Prácticas que abarcan el vídeo online, la seguridad de la marca, la viewability y la lucha contra el fraude publicitario.

• IAB Europe Transparency & Consent Framework

1.- El Marco permite a los usuarios tener plena autoridad sobre sus datos.

IAB Europe Transparency & Consent Framework (Marco de Transparencia y Consentimiento) tiene como principal objetivo ayudar a todas las partes de la cadena de publicidad digital a garantizar que cumplen con el Reglamento General de Protección de Datos de la UE y la Directiva de ePrivacy en el tratamiento de datos personales o el acceso y/o almacenamiento de información en el dispositivo de un usuario, tales como cookies, identificadores de publicidad, identificadores de dispositivos y otras tecnologías de seguimiento.

El Marco es particularmente relevante para los “first parties”, como los editores y otros proveedores de servicios online, que trabajan con “third parties” para servicios basados en datos. Utilizando el Marco, los first parties pueden permitir a los third parties procesar los datos de los usuarios en una de las bases legales de la regulación. El Marco estandariza la presentación a los usuarios de las solicitudes de tratamiento de datos de terceros que requieren un consentimiento “informado”

para el tratamiento de datos. El Marco permite la “señalización” de elección del usuario a lo largo de la cadena de suministro de publicidad. Es de código abierto, sin ánimo de lucro, con un gobierno de la industria basado en el consenso, liderado por IAB Europe, con un importante apoyo de las partes de la industria y de IAB Tech Lab, que proporciona la gestión técnica de las especificaciones de código abierto y el control de versiones.

2.- ¿Cómo funciona el marco?

La tecnología se basa en una API de JavaScript y permite visualizar la información de terceros seleccionados, el almacenamiento de las elecciones efectuadas por parte de los usuarios relacionadas con esos terceros y el paso de información sobre terceros aprobados.

La finalidad es que las partes en el ecosistema de publicidad online comprendan qué partes han sido aprobadas por los editores y han sido publicadas y aprobadas (o consentidas, según sea necesario) por sus usuarios.

Una pieza clave del Framework es un registro único conocido como Global Vendor List, que es una lista de terceros registrados y aprobados. En base a esto un editor puede, en relación con el procesamiento de datos personales en sus propiedades digitales, permitir el acceso.

Los editores pueden aprovechar el Marco de las siguientes maneras:

- Los editores pueden usar el registro para ver cuáles de sus socios han solicitado adherirse a las políticas del Marco y para qué fines utilizan sus datos personales para determinar con qué terceros eligen trabajar e incluir como proveedores en sus interfaces de usuario individuales.
- Los editores podrían proporcionar información a sus usuarios acerca de qué proveedor puede recopilar y usar legalmente los datos personales de sus usuarios en relación con la venta y medición del espacio publicitario.
- Los editores y los proveedores pueden comunicarse entre sí sobre qué proveedores pueden acceder o procesar los datos personales de los usuarios de los editores y para qué fines.
- Los editores y proveedores pueden capturar, divulgar y mantener un registro de auditoría de la elección del usuario acerca de esos proveedores y sus actividades de procesamiento de datos.

3.- Las especificaciones técnicas actualizadas Las especificaciones incluyen:

- Consent Management Provider Javascript API spec v1.1
- Cadena de consentimiento y especificación de formato de lista de proveedor global v1.1

Una especificación de borrador adicional que estará disponible para los comentarios del público:

- publisher-vendors.json spec v.1.0 (borrador)

Esta especificación de borrador adicional proporcionará un soporte más robusto para los proveedores que operan en diferentes bases legales, pistas de auditoría secundaria y controles más detallados para los editores sobre los propósitos aprobados para cada uno de sus proveedores aprobados.

Más información: <http://advertisingconsent.eu/>

- ePrivacy Seal

ePrivacyseal GmbH otorga el ePrivacyseal data protection seal tras una auditoría en profundidad de los productos mobile y online de una empresa. La certificación cubre los requisitos del Reglamento General de Protección de Datos (GDPR) para productos digitales. El catálogo de criterios para la certificación se adapta continuamente a la interpretación del GDPR y otras leyes de protección de datos.

Más información: <https://www.eprivacy.eu/en/privacy-seals/eprivacyseal/>



a. Customer journey: trazabilidad del usuario

El customer journey es el mapa que identifica las etapas por las que pasa un consumidor hasta adquirir un producto o cumplir un objetivo del anunciante independientemente del canal de comunicación utilizado. El uso de data permite controlar el número de impactos y la frecuencia de mensajes por cada usuario además de las interacciones con la campaña y la conversión final, es decir, el cumplimiento del objetivo marcado. Permite también establecer segmentos de usuarios según esos criterios y los distintos puntos del customer journey, con la intención de moverlo de un estado a otro. Es necesario tener en cuenta que el customer journey no incluye la actividad del consumidor en los walled garden como Facebook, Google, Amazon..., por la propia naturaleza de las plataformas.

b. En la activación de medios: recogida, segmentación y activación en media

En una primera fase para la recogida de datos, se pueden recolectar diferentes tipos de datos: datos sobre la navegación del usuario en una web en una app, sobre sus búsquedas online, datos sociodemográficos, de geo-posición, de interacciones con campañas publicitarias, propias o de terceros, de apertura de e-mails, de compra de productos o servicios, de gustos en redes sociales, etc.

Para que los análisis sean correctos, es fundamental la existencia de una capa de identidad que identifique al mismo usuario en las distintas bases de datos: un ejemplo sería el cross device, en donde estaríamos asociando al mismo individuo en dos dispositivos distintos. Pero del mismo modo, necesitamos saber que es el mismo individuo el que ha hecho click en los resultados de búsqueda en SEM y el que ha sido impactado en Display; en este caso a través de la unificación de Adservers podríamos obtenerlo, pero para ser capaces de ir más allá e incluso hablar de conexión con Walled Gardens, necesitamos el macheo que obtenemos con nuevos players como son los matching table providers, que identifican a un usuario en distintas tecnologías.

Tras la recogida pasaríamos a la fase de segmentación, donde los datos se procesan para construir segmentos basados en ese mayor conocimiento del usuario que permitirán la optimización tanto del mensaje como de otros criterios (frecuencia, canal, momento, entorno, predisposición, etc) en definitiva, extracción de insights a través del data para perfilar mejor a los consumidores.

Como última fase estaría la activación en medios que consiste en la activación de campañas utilizando dichas optimizaciones: desde el retargeting hasta la precisión en los targets y la planificación para incrementar el ROI. Este mayor conocimiento del usuario habrá dado lugar a una mayor eficiencia en las campañas.

c. Para la creatividad (no es solo retargeting)

La adaptación del mensaje al usuario en función de los intereses y el estado de éste. El uso de data y por tanto el mayor conocimiento del usuario puede ayudar a perfilar el mensaje, detectar el tipo de contenido y la profundidad en el mismo que requiere. Nos puede ayudar con el diseño del propio producto, el punto de venta o la imagen de marca, así como mejorar el engagement y facilitar la recomendación al acercarnos al usuario como algo relevante para él.

d. Para la investigación

Tradicionalmente la investigación se ha basado en la recogida de datos declarativos a través de estudios, focus group y encuestas para conocer y entender el consumo de medios de los consumidores, sus preferencias, intereses, etc. Esta aproximación permite considerar tanto medios offline como medios online. Un ejemplo de esto es el Estudio General de Medios o EGM que a través de encuestas telefónicas recoge datos sobre el consumo de radio, prensa, revistas, exterior, digital y televisión para facilitar una visión agregada del volumen o porcentaje de penetración de un medio en la población. Este tipo de enfoque también se ha venido aplicando a cualquier estudio realizado por un anunciante que se focalice en su audiencia o target particular.

En general, este tipo de estudios de investigación buscan obtener una visión del medio de comunicación y los resultados agregados de la penetración de cada uno de ellos.

Por otra parte, en el entorno digital, gracias a la traza que la actividad digital deja a nivel de cookie, el conocimiento de los consumidores puede extenderse a través de paneles digitales que permiten observar el comportamiento de las personas o panelistas en la actividad en apps, el consumo de medios digitales, impactos publicitarios digitales recibidos, etc. Este tipo de explotación del data permite compaginar el análisis del comportamiento de las personas junto con la ejecución de encuestas en base a las acciones que realizan, así como la declaración de datos sociodemográficos en el momento de adhesión al panel.

Este uso del data permite identificar pautas y patrones en la forma de relacionarse de las personas con los medios digitales en cuanto a intensidad (tiempo dedicado), momento del día y semana (tipo de actividad en función de cuándo se produce), tipo de soportes y plataformas que se consume (noticias, Redes Sociales, juegos, etc.). Adicionalmente, este tipo de investigaciones pueden contemplar distintos dispositivos de una misma persona (Tablet, Smartphone, PC, etc.).

Por otro lado, este tipo de investigaciones digitales tienen vocación en entender desde la perspectiva de las personas como es su consumo de medios. Es decir, si el consumo de medios de una persona fuera 100 minutos al mes, como se distribuye dicho consumo por todas las plataformas, en lugar de un enfoque orientado a cuántas de las personas objeto del estudio consumen un medio como ocurre en el caso del EGM.

Esta evolución en la investigación es fundamental para abordar varias de las grandes preguntas de hoy en día: ¿qué caracteriza a este consumidor?, ¿este consumidor responde a un perfil común a otros consumidores?, ¿este perfil explica un distinto interés en el consumo de medios y en determinadas preferencias?, etc. Y este enfoque es el que conduce a que cada vez más se hable de personalización en los contenidos y comunicación basada en las personas (persona planning), en lugar de comunicación basada en la penetración de cada soporte o medio (media planning).

Luego, con métodos más sencillos y descriptivos, o con métodos más avanzados como puede ser una segmentación no basada en filtros y reglas de negocio, la investigación es uno de los grandes consumidores de datos.

e. Modelización

Se habla de modelización cuando se aplican técnicas matemáticas y estadísticas para identificar el algoritmo o fórmula matemática que mejor explican casos como los siguientes:

- Qué características comunes diferencian a un grupo de personas. Por ejemplo, ¿qué atributos discriminan mejor a los consumidores que compran un producto de los que no?
- La ocurrencia de un evento. Por ejemplo, ¿por qué se han recibido todas estas visitas al site? ¿Por qué esta persona ha comprado este producto?

6. Usos del data

- La probabilidad de que ocurra algo. Por ejemplo, ¿cuántas visitas recibirá el site si invierto tanto en estos medios? ¿Cómo de probable es que esta persona compre ese producto?
- Cómo optimizar la inversión. Por ejemplo, ¿qué plan de medios maximiza las visitas o ventas?

Junto a las metodologías matemáticas, las tecnologías de big data han permitido que este tipo de técnicas puedan ser aplicadas con mayores volúmenes de datos y con una capacidad de computación mayor facilitando su uso a mayor escala.

Si bien hace décadas que se modeliza el impacto de la publicidad offline en los KPIs de negocio agregados (volumen de leads, de ventas, etc.), gracias a la digitalización es posible disponer de datos más granulares y continuos en el tiempo relativos al comportamiento de las personas o cookies. De esta forma, se está extendiendo el uso de modelos matemáticos para aplicaciones como los siguientes:

- **Atribución digital.** En lugar de aplicar reglas de negocio (last click, etc) sobre las conversiones digitales, los modelos matemáticos evalúan los customer journeys y los impactos digitales recibidos a lo largo del mismo, tanto de aquellos que usuarios que convierten como los que no, e identifican que soportes, placement, formatos, etc., son determinantes para que se produzca una conversión hasta facilitar una atribución y eficacia a cada medio y soporte.
- **Lookalike.** Se estima la similitud de ciertas personas respecto a otros consumidores que presentan un patrón conocido y de interés del anunciante. En caso de que esa similitud sobrepase un umbral y sea significativo, esas personas podrían ser objeto de campañas similares a las que los consumidores conocidos e interesados ya recibieron.
- **Propensión a realizar un determinado comportamiento.** Por ejemplo, se estima la probabilidad de que una cookie vaya a hacer clic o visitar una determinada web en base al patrón de comportamiento de otras cookies que sí lo han hecho previamente.
- **Optimización de los targets.** Es factible estimar que volumen de target es el más adecuado para conseguir un determinado CTR o CPA o, etc.

Para llevar a cabo todas estas capacidades, es fundamental una correcta recogida de datos y una validación y clasificación inicial que luego permita aplicar estos recursos de machine learning y big data.

f. Otros medios (off line)

El data se lleva utilizando muchos años en los canales tradicionales. Lógicamente, en la era digital actual, se asocia inevitablemente al mundo online dada la alta y constante interacción que tenemos con este entorno.

Anteriormente en los canales tradicionales, el data o la información sobre los clientes y consumidores se recogía de encuestas puerta por puerta o telefónicas, audiencias televisivas, paneles de consumo en el punto de venta o de la propia información generada gracias a la interacción de los clientes con las empresas. Esta información se utilizaba para lo mismo que se utiliza el data ahora, la única diferencia es que el volumen y el origen de la información disponible era notablemente reducido en comparación con la información y las fuentes de las que disponemos hoy. Pero la gran ventaja con la que contamos ahora, es que podemos combinar el data que proviene de los medios online y offline para aplicarlos, indistintamente a unos consumidores y a otros. Hoy en día un porcentaje muy elevado de clientes van a utilizar todos los medios para crear su decisión de compra. Por este motivo ambos tipos de medios tienen gran relevancia para todas las marcas.

• ¿Cómo se usa el data en los medios de comunicación masiva?

Como decíamos estos medios siguen siendo relevantes ya que se busca llegar a la mayor cantidad de personas posibles, generar mayor número de impresiones y hacer que un único mensaje llegue a la vez a todo ese público.

Las nuevas formas de consumir televisión nos permiten conocer mejor a las personas individuales que están delante de nuestra publicidad. Ahora se pueden recoger en el momento los datos sobre las películas, series o programas que ven y escogen los consumidores, lo que nos facilita servir la publicidad adecuada a ese público concreto. Un medio tradicionalmente masivo se transforma: ahora la publicidad es personalizada y además interactiva. Podemos obtener una respuesta en el momento sobre la campaña que estamos realizando.

En el caso de la radio, el uso del data no es un factor diferencial. Nos encontramos ante limitaciones técnicas que impiden que podamos segmentar los mensajes. La diferencia estriba ahora en las conexiones a la radio a través de los móviles, utilizando datos y wifi lo que permite la ubicación del terminal y facilitaría la personalización del mensaje, pero, por el momento no hay muchas empresas que ofrezcan soluciones en este sentido.

6. Usos del data

La prensa impresa no se está aprovechando del uso del data para la segmentación, excepto en el caso de publicaciones con un rango local muy marcado utilizando la segmentación geográfica sobre el área de influencia.

En el caso de la publicidad exterior, la cosa cambia. El uso del data aquí ha tomado un papel protagonista. Pasamos cada vez más tiempo fuera de casa y la Publicidad exterior ofrece una solución perfecta para contactar con nuestros clientes mientras estos se desplazan.

El uso del data nos permite conocer no solo el perfil de las personas que residen en una zona donde vamos a situar nuestra publicidad, sino también el de las personas que transitan delante del mismo. Tecnología GPS permite ubicar a los panelistas, dándonos información, incluso, de cuantas veces pasan por cada punto. Al conocer el lugar de residencia de dichos panelistas podemos incluir en la ecuación el factor actitudinal. Combinación perfecta para que el mensaje sea el adecuado para el perfil que lo verá y esté ubicado en el sitio por el que sabemos que nuestros clientes potenciales van a pasar.

Si a esta capacidad de segmentación le añadimos la posibilidad de interacción con el soporte exterior, lo que se llamamos Digital Out Of Home (DOOH), las posibilidades de servir el mensaje más personalizado posible, son infinitas.

El uso de la publicidad exterior en las inmediaciones de comercios permite generar tráfico a la tienda, el data nos ayuda a crear el mensaje que mejor va a llegar a las personas que pasan por delante de nuestra publicidad exterior.

En lo que respecta a eventos, las empresas organizadoras pueden facilitar datos de asistencia, para más tarde enriquecerlos con información sobre la forma de pago de las entradas y el lugar de residencia de los asistentes, por ejemplo. De esta manera conseguiremos un perfil muy claro del público que asiste a esos eventos y podemos proporcionar esa información a los anunciantes para que valoren las posibilidades de patrocinio en futuras ediciones.

El marketing directo es muy útil para poder personalizar mensajes, generar cercanía y afinidad con las personas, pero es fundamental la segmentación del público para asegurar la efectividad del mismo. Los principales medios utilizados para hacer campañas de marketing directo son el telemarketing, envíos postales o buzoneo selectivo y el marketing en el punto de venta.

En el caso de los dos primeros medios, el uso del data es muy claro: la segmentación. Gracias a herramientas y análisis que realizan algunas empresas podemos saber de una forma muy precisa las características de las personas ubicadas en una zona, hasta el punto de conocer sus gustos o preferencias, podemos llegar a saber si en un tramo de vía la gente que vive allí son familias sin hijos, qué les gusta leer, si les gusta viajar o si tienen mascotas. Esta información tan detallada permite realizar campañas de envío postal/buzoneo selectivo o telemarketing de una forma muy eficaz. Muy importante tener en cuenta que no estamos hablando solo de la captación sino de la retención o la rentabilización de clientes actuales mediante la venta cruzada; ya les conocemos, y nos será más fácil, con el uso del data, ofrecerles productos complementarios que puedan necesitar.

g. Otros usos

- Procesos de negocio: por ejemplo, con la optimización de los stocks basados en predicciones o la optimización de las rutas de reparto con datos de geoposicionamiento, predicciones meteorológicas en tiempo real etc.
- Optimización del rendimiento personal a través de los wearables.
- Aplicación del dato sobre la salud: ayudando en la prevención de enfermedades y en la predicción y control de epidemias.
- Smartcities: optimización de los recursos y de la organización de los servicios en las ciudades a través de los datos de uso, detección de necesidades, predicciones sobre la población, etc.



7. Limitaciones del data

Hoy en día, el registro de datos y la capacidad de almacenar grandes cantidades de ellos ya no supone un problema para las grandes compañías. En los mejores casos, existen departamentos internos destinados a estas acciones, o se externaliza a compañías especialistas en recoger todos los datos para su posterior utilización.

No obstante, a menudo esperamos que el big data nos aporte una respuesta directa y clara de cómo optimizar nuestras acciones o cómo mejorar inmediatamente las estrategias de negocio. Esta idea equivocada lleva a muchas empresas a no ser cuidadosas a la hora de elegir qué herramientas usar para el análisis o cómo interpretar los datos extrapolados, implementando estrategias incorrectas.

La recogida del data no es suficiente si hay una falta de análisis y grado de conocimiento para procesarla.

Gran parte del esfuerzo y tiempo debe estar dedicado al “data preprocessing” con el fin de entender y dar sentido a lo recogido. Es aquí donde la labor del perfil del data-scientist es fundamental, ya que es necesario un equipo cualificado capaz de saber analizar, comprender y evaluar los datos cuando hayan sido procesados.

Una vez entendido que el factor humano es clave en el proceso de recogida, análisis e integración de los datos, se mencionan a continuación una serie de limitaciones o retos del data.

• La limpieza de los datos:

Múltiples partners juegan un papel muy activo en el mundo del data. DSPs, DMPs, CRMs, entre otros, manejan diferentes tipos de data de diferentes fuentes y sistemas. En general, los datos vienen dados de forma no coherente e inconsistente: diferentes fuentes, múltiples formatos, fórmulas y campos que provocan incoherencias a la hora de analizarlos y buscar correlaciones óptimas entre ellos.

El mayor reto se centra en manipular los datos en el data preprocessing de tal forma que se integren únicamente aquellos que sean significativos y con valor además de ser capaces de definir un hilo común entre toda la información. En algunas ocasiones, esto significa no usar todo el data que se recoge. La calidad de nuestro output es proporcional al trabajo realizado en nuestro input.

7. Limitaciones del data

• Migración de los datos:

A lo largo del tiempo puede que sea necesario agregar, eliminar o modificar campos en el conjunto de los datos. Es importante vincular los datos antiguos y nuevos adquiridos con el fin de conseguir una integración completa entre ellos.

• La dificultad de escalabilidad del data:

Trabajar con big data conlleva a manipular múltiples fuentes. Es fundamental contar con una infraestructura que sea capaz de adaptarse e incrementar su escalabilidad a la vez que nuestros modelos/volumen de datos aumentan. El tamaño de data hace difícil su manipulación y transferencia a servicios web (clouds) o incluso el traspaso de una muestra a un ordenador convencional.

• La calidad del data:

Datos no verificados, expirados o la incorporación de valores predeterminados son algunos ejemplos de las limitaciones de la calidad del data. No debemos confiar en ningún conjunto de datos por defecto.

• El coste y persistencia del data:

Dificultad de tomar la decisión sobre qué hacer con los datos que no se están utilizando pero que tienen un alto coste de almacenamiento.

• Limitaciones en el grado de alcance de los datos:

Una vez que ya se ha constituido un modelo estadístico, nos encontramos con la dificultad de aumentar el alcance relevante y cualificado. Este apartado se centra en especial en 2nd y 3rd party data. La naturaleza de encontrar una tendencia relevante de datos hace que la proporción/cobertura se vea disminuida en comparación con el grupo de individuos identificados anteriormente. A la hora de implementar los insights generados a partir del data, hay que tener en cuenta que la posibilidad de conseguir acciones optimizadas con altas probabilidades de generar mayor respuesta, conlleva a tener un menor alcance.

8. Buenas prácticas (transparencia, fraude)

a. Definición de fraude

En primer lugar, ateniéndonos a la definición que nos proporciona la R.A.E. podemos comprobar que fraude tiene dos acepciones principales:

1. m. Acción contraria a la verdad y a la rectitud, que perjudica a la persona contra quien se comete.
2. m. Acto tendente a eludir una disposición legal en perjuicio del Estado o de terceros.

Antes de entrar en una definición de fraude publicitario hay que comentar que, desde un punto de vista objetivo y como ya hemos comentado, existen dos tipologías de fraude en función de quién o qué lo realiza:

- El primero tiene que ver con el fraude generado por la parte humana de la cadena. Y se produce en la propia venta o comercialización: con falta de información sobre lo que se compra, falta de transparencia e incluso la mentira al respecto de lo que se puede comprar y conseguir.
- El segundo provocado por los fraudes en la publicidad digital que influyen en que también haya fraude en los segmentos de data compuestos a partir de esa compra fraudulenta.

El fraude publicitario se podría definir como cualquier actividad que se realiza a propósito y que impide que los impactos publicitarios lleguen a personas reales. Es decir, atenta directamente contra la primera necesidad de la publicidad que es llegar a un público determinado para fomentar ciertos comportamientos en el mismo.

Como en toda acción de fraude o engaño la culpa puede recaer en varios actores de la cadena, desde los editores pasando por los partners tecnológicos, hasta ciertos tipos de tráfico que nada tienen que ver con estos dos players.

El panorama del fraude es cambiante y los implicados en su realización evolucionan a un ritmo frenético. Es por ello que es importante realizar una monitorización constante de las fuentes de tráfico en base a las que construimos segmentos de data.

8. Buenas prácticas (transparencia, fraude)

b. Tipos de fraude en función de la tipología de data:

• De First Party Data:

Si atendemos a la definición general de fraude, en su primera acepción, hay que hablar de un tipo de fraude que no suele ser catalogado como tal pero que es muy común en nuestro sector. Este fraude se basa en el aprovechamiento del desconocimiento existente en muchos anunciantes para vender proyectos que tienen un precio muy elevado y unas aplicaciones prácticas (reales) muy por debajo de las expectativas generadas por la sobreventa.

• De Third Party data:

Proveniente de la comercialización:

En este punto, nos referimos al fraude humano por la venta de expectativas que no son reales, como se ha comentado con anterioridad y que están relacionados con fraudes de soportes y no de un proveedor de datos.

c. Provenientes del fraude publicitario digital:

- Existe un tipo de fraude humano, dentro de este punto, generado por las impresiones servidas a personas fuera del target en las campañas digitales, y el desperdicio de presupuesto que ello conlleva. Esto se debe al uso de fuentes de datos poco fiables, la falta de transparencia en el origen del dato y la falta de precisión de los mismos, al abuso de sistemas de inferencia y modelaje, y a cookies de intencionalidad o interés con antigüedades mayores a 90 días.
- Bots. Programas de software que están preparados para realizar tareas automatizadas en un sitio web sin que haya intervención humana. En este punto hay que dejar claro que hay ciertos bots y crawlers que son necesarios para el buen funcionamiento del ecosistema, no deben ser considerados como fraude.
- Clicks fraudulentos. Se trata de acciones de click que se producen en anuncios que no son vistos por usuarios reales y que son generados también por bots.
- Impresiones fraudulentas. Al igual que en los clicks, se trata de impresiones que se solicitan a un Ad-exchange y que no van a ser visualizadas por usuarios reales.
- Pixel Stuffing (relleno de píxels). Se muestra un anuncio de 1x1 píxels en una zona de la página que ningún usuario ve pero que sí es contabilizado como una impresión.

8. Buenas prácticas (transparencia, fraude)

- Enmascaramiento de URL's. Se trata de una técnica que se basa en camuflar una URL suplantando a otra con el objetivo de conseguir que se publiquen anuncios que en principio deberían aparecer en la URL suplantada. El objetivo es generar más tráfico e ingresos a estos sites suplantadores.
- Ad stacking (Apilamiento de anuncios). Consiste en el apilamiento de anuncios dentro de un solo espacio publicitario. El fraude tiene lugar al cobrar al anunciante impresiones inexistentes que ningún usuario ve.

d. Requisitos de transparencia y ejemplo de buenas prácticas:

Los requisitos imprescindibles que cualquier anunciante o agencia tiene que tener en cuenta a la hora de comprar datos de un proveedor son:

Una de las condiciones centrales para que la programática dé buenos resultados es basar las campañas en data de alta calidad con la escala necesaria, sólo así se puede garantizar una segmentación de alta precisión. El problema de hoy en día es que la mayoría de proveedores de data no aportan transparencia, por lo que los anunciantes no tienen las garantías necesarias para asegurar resultados. Como proveedor de data, hay varias medidas a tomar en cuenta para solventar este problema, asegurar la calidad de la data y ganarse la confianza de los anunciantes.

El primer requisito es que el proveedor de data haya sido auditado y certificado conforme la nueva ley de protección de datos GDPR. Esta ley incluye temas como la obtención del consentimiento del usuario en el uso de datos, la anonimidad y encriptación de la información, la exclusión de los datos sensibles y el exhaustivo control de la seguridad de la misma.

En segundo lugar, es importante que los proveedores de data revelen las fuentes de los datos proporcionados, así como su naturaleza y su manera de recolección. La data que ha sido verificada por el DNI del usuario, por ejemplo, y que proviene de organizaciones de confianza como bancos, operadores de telecomunicaciones o sitios de e-commerce, siempre será más precisa que los datos inferidos o probabilísticos de otras fuentes, y sus métodos de recolección serán más transparentes y sencillos de monitorizar. Otro tipo de data con la que nos aseguramos calidad, es la de las grandes operadoras del país, al tratarse de data determinística, se consigue precisión y clara recolección por medio del consentimiento directo de los usuarios.

8. Buenas prácticas (transparencia, fraude)

Como tercer punto importante, se deben tener en cuenta los procesos internos que los proveedores de data llevan a cabo para garantizar la constante calidad de la data incluso cuando los volúmenes incrementan. Pruebas internas y externas de calidad de data, o la priorización y la actualización de la misma.

Finalmente, algo que también se recomienda es realizar auditorías externas sobre la precisión y veracidad de la data. En cada mercado existen verificadores oficiales que miden la precisión de audiencias en campañas, pero en algunos casos, estos se basan en paneles y por lo tanto pueden tratarse de metodologías insuficientes, principalmente en el entorno móvil. Es por esta razón que se recomienda optar por someter la base de datos a una auditoría externa con organismos oficiales, independientes y sin ánimo de lucro, que puedan llevar a cabo el proceso con mayores garantías, llevando a cabo metodologías fiables, asegurando un panel extenso y contactando a los usuarios directamente para mayor certeza y credibilidad de los resultados.

e. Reflexiones al respecto

No sirve de nada realizar una sobreventa de las capacidades y aplicaciones que tiene el data en la publicidad digital porque sólo conlleva la generación de falsas expectativas en el mercado. Por lo general, el manejo adecuado de expectativas es crucial para el buen funcionamiento de cualquier mercado.

El desconocimiento existente alrededor del mundo del data, como en todas las “novedades” que se van produciendo en el ecosistema publicitario digital, hace que se genere una desconfianza hacia todo lo que tenga que ver con ello y, por tanto, una burbuja que termine explotando y arrastrando así a muchos actores del mercado.

En definitiva, la honestidad es la base sobre la que se debe construir cualquier negocio si queremos generar la credibilidad suficiente para que se convierta en algo escalable y de uso generalizado. Sólo de esta manera generaremos la confianza necesaria por parte de los anunciantes.

Puede parecer una obviedad, pero para comprender el marco legal aplicable al data es necesario conocer los conceptos más básicos; y en este contexto, la unidad mínima y elemental es el dato, por lo que comenzaremos analizando su regulación, y lo haremos partiendo de una pregunta recurrente: ¿quién es el propietario de un dato?

En España, como en la mayoría de países de nuestro entorno, los contenidos se protegen a través de la normativa de propiedad intelectual, que se aplica a las “creaciones originales literarias, artísticas o científicas expresadas por cualquier medio o soporte”. Así, los autores de libros, piezas musicales, películas o programas de ordenador, pueden explotarlos económicamente y prohibir que sean reproducidos o utilizados sin su autorización. Sin embargo, y aquí viene la sorpresa, no ocurre lo mismo con los datos, porque no alcanzan el umbral de creatividad necesario para gozar de esta protección. Dicho de otra forma: no tienen propietario, porque son demasiado elementales; pero eso no quiere decir que podamos utilizar libremente todos aquellos datos con los que nos encontremos. De hecho, hay dos categorías de datos cuyo uso está muy acotado por la normativa, hasta el punto de exponer a severas consecuencias a quien los emplee de forma indebida: son los datos personales y los datos confidenciales.

Asentado lo anterior, aunque los datos, individualmente considerados, no tengan propietario, las bases de datos que los almacenen y organicen sí pueden estar protegidas: ocurrirá cuando su fabricante haya dedicado esfuerzos sustanciales a la obtención, verificación o presentación de la información. Esta protección permite a su titular impedir a terceros la extracción (“transferencia del contenido de una base de datos a otro soporte”) o reutilización (“puesta a disposición del público del contenido de la base de datos”) de la totalidad o de una parte sustancial de su contenido, por lo que esta protección constituye una herramienta idónea para proteger los datasets que las empresas utilicen a efectos publicitarios.

Entrando ya en los datos personales, para utilizarlos es preciso obtener el consentimiento de su titular, a menos que podamos basar nuestro tratamiento en otra de las condiciones que prevé el Reglamento General de Protección de Datos, como por ejemplo, el “interés legítimo prevalente”; y si optamos por el consentimiento, tendremos que tener en cuenta una amplia lista de variables:

- Debe ser libre, por lo que no puede vincularse al cumplimiento de un contrato o el uso de una tecnología, ni solicitarse cuando el interesado pueda percibir un riesgo en no otorgarlo;

- Debe ser específico, es decir, es preciso un consentimiento por cada finalidad para la que pretendamos tratar datos;
- Debe informado, por lo que el usuario debe poder conocer quién tratará sus datos, para qué finalidades, con qué extensión y cómo puede ejercitar los derechos que le reconoce el RGPD;
- Debe ser inequívoco, es decir, manifestado por el usuario o inferido de una “clara acción afirmativa”, que en todo caso ha de ser deliberada; y
- Debe ser revocable, en cualquier momento y de un modo tan simple como el empleado para solicitarlo.

La Agencia Española de Protección de Datos ha aclarado, además, que estos requisitos se aplican también al uso de cookies (y tecnologías similares, como el device fingerprinting) en los equipos de los usuarios, lo que unido a la ampliación del concepto de dato personal (que ahora abarca también a los “números de identificación” y a los “identificadores de sesión en forma de cookies”), obliga a realizar un esfuerzo de transparencia mucho mayor para obtener información de los usuarios de una web o app. ¿Cómo se concretará? La AEPD está trabajando en una actualización de la Guía sobre el uso de las cookies, pero en este momento solo podemos remitirnos a la última edición de la guía de cookies. La futura Guía de cookies abordará también la responsabilidad de los diferentes actores (publishers, ad servers, agencias...) a la hora de cumplir con la normativa, y en cuya redacción se han tenido en cuenta los planteamientos de IAB Spain, entre otras organizaciones representativas de la economía digital.

Dicho lo anterior, esta guía no parece la solución definitiva a la problemática de las cookies: en paralelo, la Unión Europea está actualizando su normativa sobre la privacidad y las comunicaciones electrónicas, por medio de un nuevo Reglamento de Privacidad Electrónica. De entrada, la Comisión anunció que trataría de racionalizar las reglas del juego, pues la norma actual, en sus propias palabras, “ha dado lugar a una sobrecarga de solicitudes de consentimiento a los usuarios de Internet”. A partir de ahí, las soluciones propuestas y su alcance difieren entre las distintas versiones del texto que circulan por Bruselas, pero sobre todas ellas ronda la misma pregunta: ¿por qué solicitar el consentimiento de los usuarios, siempre y en todo caso, para tratar su información a través de cookies? O, dicho de otra manera, ¿por qué el interés legítimo no es fundamento válido en este ámbito, y sí lo es en otros tratamientos de datos, conforme al RGPD?

De acuerdo con el RGPD, “el tratamiento de datos personales con fines de mercadotecnia directa puede considerarse realizado por interés legítimo”; pero que “pueda” considerarse que concurra esta base jurídica, no significa que “siempre” vaya a existir este interés. De hecho, las autoridades de protección de datos han aclarado, en múltiples ocasiones, que la aplicación de este fundamento requiere de una ponderación, de una “prueba de sopesamiento” entre los intereses del responsable (la empresa) y del interesado (el usuario) en la que se deben de tener en cuenta factores como la cantidad de datos tratados, el período de conservación, el riesgo que supone el tratamiento para los usuarios... Por ello, su aplicación debe realizarse con suma prudencia, y documentando adecuadamente la realización de esa prueba de ponderación. Por otro lado, habrá que tener en cuenta la versión definitiva del Reglamento de Privacidad electrónica y si este permite esta base legal para el tratamiento de cookies y device fingerprinting.



- **Ad exchanges:** Un ad exchange es un canal de ventas entre, por un lado soportes y redes publicitarias, y por otro anunciantes a los que se ofrece inventario agregado. Los ad exchanges proveen de una plataforma tecnológica que facilita la automatización de las subastas basadas en precio y las compras en tiempo real, lo que conocemos como RTB (Real Time Bidding).
- **Ad server:** Servidor de publicidad en inglés, es un software de gestión responsable de la planificación, gestión interna, control, determinación de inventario disponible y emisión de la publicidad contratada en un sitio web conforme a patrones comerciales determinados.
- **Adblocker:** Software en el navegador de un usuario que impide que se muestren los anuncios en pantalla.
- **Bases de datos NoSQL:** estructuras que nos permiten almacenar información en aquellas situaciones en las que las bases de datos relacionales generan ciertos problemas debido principalmente a problemas de escalabilidad y rendimiento de las bases de datos relacionales donde se dan cita miles de usuarios concurrentes y con millones de consultas diarias. Además, las bases de datos NoSQL son sistemas de almacenamiento de información que no cumplen con el esquema entidad-relación. Tampoco utilizan una estructura de datos en forma de tabla donde se van almacenando los datos sino que para el almacenamiento hacen uso de otros formatos como clave-valor, mapeo de columnas o grafos.
- **Behavioral Targeting:** es un método para dirigir las impresiones publicitarias digitales a una audiencia específica de consumidores basada en sus acciones anteriores, aquellas acciones que ocurren tanto online como offline.
- **Bots:** Programas de software que están preparados para realizar tareas automatizadas en un sitio web sin que haya intervención humana.
- **CDP:** El CDP (Customer Data Platform) es un software paquetizado que crea una base de datos de clientes unificada y persistente que es accesible a otros sistemas. Básicamente, es un sistema prediseñado que centraliza los datos de los clientes de todas las fuentes y luego los pone a disposición de otros sistemas para campañas de marketing, servicio al cliente y todas las iniciativas de experiencia del cliente.
- **CRM (Customer Relationship Management):** Amplio término que cubre conceptos usados por compañías para gestionar sus relaciones con clientes, incluyendo la colección, almacenamiento y análisis de la información de los mismos.

10. Glosario

- **Cross-device:** Forma de identificar el seguimiento de un usuario único a través de smartphones, tablets y ordenadores. Esto se realiza a través de métodos deterministas y probabilísticos, o una combinación de ambos.
- **Customer journey:** El customer journey es el mapa que identifica las etapas por las que pasa un consumidor hasta adquirir un producto o cumplir un objetivo del anunciante independientemente del canal de comunicación utilizado.
- **Data exchange (DXP):** Un data exchange es una plataforma que da acceso a múltiples fuentes de datos a través de las cuales los compradores pueden seleccionar el vendedor adecuado para los insights de campaña al precio que comprador y vendedor determinan. Funcionan como un Marketplace de datos.
- **Demand Side Platform (DSP):** Los DSP permiten la compra centralizada del inventario de distintas fuentes, tales como Ad Exchanges, redes de publicidad o plataformas de optimización publicitaria (SSPs). Introducen ventajas como la simplificación de flujos de trabajo, generación de informes integrados, optimización algorítmica de decisiones de compra y posibilidad de realizar subastas en tiempo real.
- **DMP:** Los DMPs (Data Management Platform) son plataformas tecnológicas en las que se recolectan, se integran y se gestionan grandes cantidades de datos estructurados y no estructurados. Tienen la capacidad de recopilar datos de audiencia procedentes de cualquier fuente.
- **First Party Data:** Se conoce como First Party Data, a los datos recolectados a través de las posibles interacciones que un usuario puede realizar con cualquiera de los properties de una compañía. Estamos hablando de datos recogidos por el CRM, el call center, el paid media, y también recogen conductas, acciones o intereses demostrados a través de la navegación de su sitio web o datos creados a través de taxonomías realizadas a medida.
- **Fraude publicitario:** se podría definir como cualquier actividad que se realiza a propósito y que impide que los impactos publicitarios lleguen a personas reales. Es decir, atenta directamente contra la primera necesidad de la publicidad que es llegar a un público determinado para fomentar ciertos comportamientos en el mismo.
- **Pixel Stuffing (relleno de píxeles).** Se muestra un anuncio de 1x1 pixels en una zona de la página que ningún usuario ve pero que sí es contabilizado como una impresión.

- **Publisher:** Titular de una página web. En publicidad online se utilizan también los términos afiliado, editor o partner.
- **Reglamento General de Protección de Datos (RGPD):** Las siglas RGPD atienden a el Reglamento General de Protección de Datos. En inglés “GDPR”, General Data Protection Regulation. El Reglamento entró en vigor el 25 de mayo de 2016, concediéndose un periodo de adaptación de 2 años siendo de aplicación a partir del 25 de mayo de 2018.
- **Retargeting:** El uso de un pixel tag u otro código para facilitar a un tercero el reconocimiento de usuarios concretos fuera del dominio donde esta actividad ha sido recogida.
- **Second Party Data:** Son datos que esencialmente provienen de acuerdos de compra y venta con diversos partners. Este modelo de datos permite a los anunciantes poder ampliar sus bases de datos (First Party Data) con información muy valiosa que de otra forma no podrían incorporar.
- **Third Party Data:** El Third party data son datos procedentes de compañías que, recopilan datos y comportamientos de miles de webs, y las cuales agregan para venderlos con fines publicitarios. Las fuentes de estos datos, así como los tratamientos a los que han sido sometidos, nunca son compartidos. Estos proveedores ofrecen cookies con información como edad, género, categoría socio-profesional, interés, localización geográfica, etc.
- **Trading Desk:** Es el equipo técnico de personas que optimizan la compra programática de sus distintos anunciantes conectando con múltiples DSPs. Pueden formar parte de la agencia de medios o ser compañías independientes. Generalmente el trading desk está enfocado hacia la compra inteligente de audiencias mediante tecnologías de optimización en múltiples plataformas (ad exchanges, networks, DSPs, soportes) de manera tal, que comprará aquella impresión de manera más eficiente para los objetivos de sus clientes/ anunciantes.
- **Zero party data:** Son datos declarativos de los consumidores recogidos fundamentalmente a través de encuestas.

11. Agradecimientos

Datmean

Salvatore Cospito

/ CTO & Co-Founder

Experian

Giovanna Benito

/ Digital Account Manager

illumina Technology

Maitane Torca

/ Head of International Programmatic

illumina Technology

Miren Ortigosa

/ Senior Strategy Manager Programmatic

Omnicom

Fernando Alonso

/ Head of Annalect Spain

Prisa

Óscar Lopez

/ Senior Digital Strategist & Data Management Platform

Programatic

Javier Jimenez

/ Socio Consejero

Rebold

José Ramón Mencías

/ Sales and Business Development Director

Schibsted

Andrea Martinez

/ Market Research & Insights Manager

Ymedia

Sonia Casado

/ Chief Data and Analytics Officer





www.iabspain.es
marketing@iabspain.es