



UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE INGENIERÍA

75.06/95.58 Organización de Datos

Segundo Cuatrimestre de 2018

Trabajo Práctico 1:

Web analytics de www.trocafone.com

Grupo 42:

Miranda, Pablo

Marinier, Sebastián

Farfán, Verónica

Basconcelo, Eliana

Link de Github: <https://github.com/vpfarfan/Organizacion-de-datos-20182>

1. Introducción	3
2. Preprocesamiento y descripción de Datos	5
2.1 Preprocesamiento de Datos	6
3. Análisis de Datos	6
3.1 Clientes	6
3.1.1 Total de Usuarios	6
3.1.2 Nuevos Clientes	7
3.2 Marcas	9
3.2.1 Modelos	11
3.3 Métricas del comportamiento de los usuarios	11
3.4 Tipos de evento por persona	13
3.5 Sesiones por usuario	14
3.5.1 Análisis por sesión	14
3.5.2 Ingreso de usuarios	14
3. Conclusiones	18

1. Introducción

El objetivo del siguiente informe es analizar los datos de un conjunto de eventos web de usuarios que visitaron el sitio web www.trocafone.com. Este sitio es la plataforma de ecommerce de Brasil de la compañía Trocafone, cuyo fin es comercializar smartphones seminuevos (refurbished) a través de tres canales: ecommerce, marketplace y tiendas físicas. Su sitio web indica que venden principalmente las marcas Apple y Samsung.

Los datos obtenidos se encuentran en un archivo .CSV en el cual se encuentra la siguiente información:

- **timestamp:** fecha y hora cuando ocurrió el evento
- **event:** Tipo de evento
- **person:** Identificador de cliente que realizó el evento.
- **url:** Url visitada por el usuario.
- **sku:** Identificador de producto relacionado al evento.
- **model:** Nombre descriptivo del producto incluyendo marca y modelo.
- **condition:** Condición de venta del producto
- **storage:** Cantidad de almacenamiento del producto.
- **color:** Color del producto
- **skus:** Identificadores de productos visualizados en el evento.
- **search_term:** Términos de búsqueda utilizados en el evento.
- **staticpage:** Identificador de página estática visitada
- **campaign_source:** Origen de campaña, si el tráfico se originó de una campaña de marketing
- **search_engine:** Motor de búsqueda desde donde se originó el evento, si aplica.
- **channel:** Tipo de canal desde donde se originó el evento.
- **new_vs_returning:** Indicador de si el evento fue generado por un usuario nuevo (New) o por un usuario que previamente había visitado el sitio (Returning) según el motor de analytics.
- **city:** Ciudad desde donde se originó el evento
- **region:** Región desde donde se originó el evento.
- **country:** País desde donde se originó el evento.
- **device_type:** Tipo de dispositivo desde donde se generó el evento.
- **screen_resolution:** Resolución de pantalla que se está utilizando en el dispositivo desde donde se generó el evento.
- **operating_system_version:** Versión de sistema operativo desde donde se originó el evento.
- **browser_version:** Versión del browser utilizado en el evento

El análisis a realizar se va a centrar en explorar y extraer información pertinente para la empresa en forma de insights que puedan aportar conocimiento de los usuarios y del comportamiento de los mismos cuando interactúan con el sitio web. Por ello, es necesario establecer que existen 11 eventos posibles para los usuarios:

- **“viewed product”:** El usuario visita una página de producto.
- **“brand listing”:** El usuario visita un listado específico de una marca viendo un conjunto de productos.

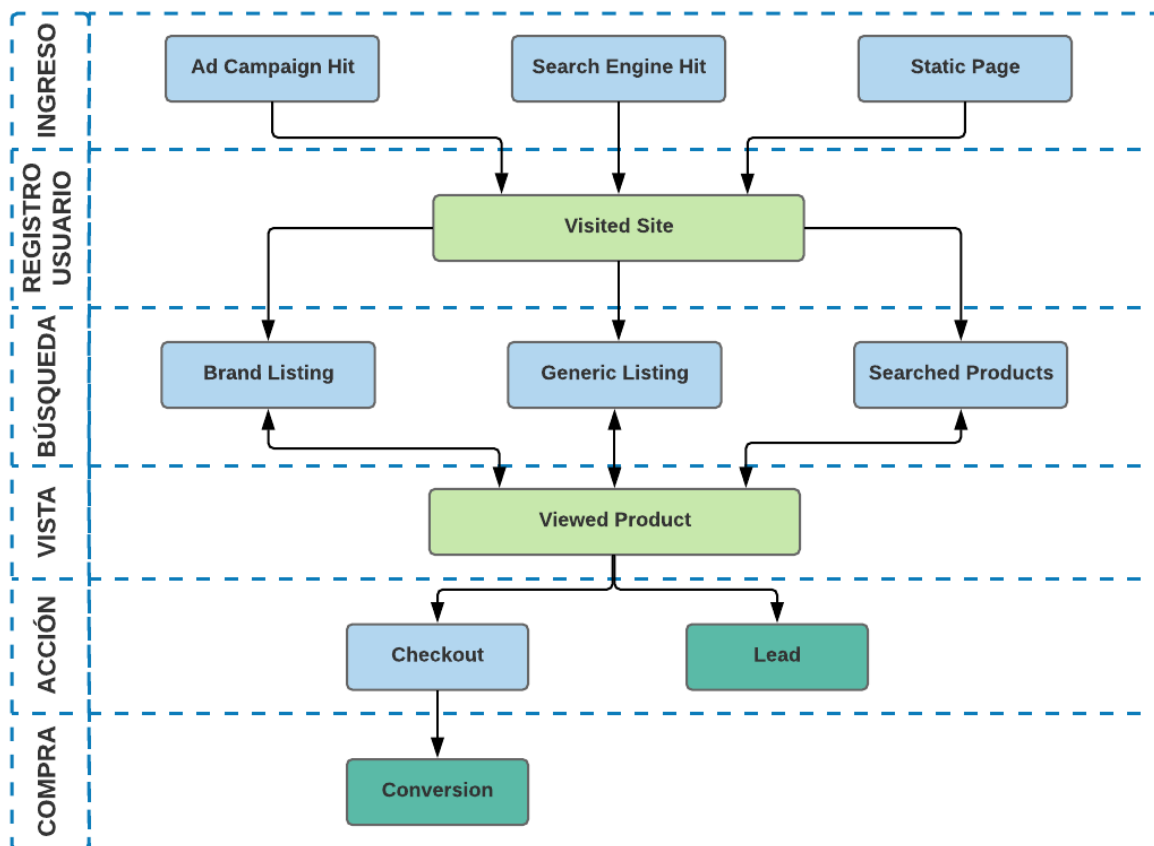
- **“visited site”**: El usuario ingresa al sitio a una determinada url.
- **“ad campaign hit”**: El usuario ingresa al sitio mediante una campana de marketing online.
- **“generic listing”**: El usuario visita la homepage.
- **“searched products”**: El usuario realiza una búsqueda de productos en la interfaz de búsqueda del site.
- **“search engine hit”**: El usuario ingresa al sitio mediante un motor de búsqueda web.
- **“checkout”**: El usuario ingresa al checkout de compra de un producto.
- **“staticpage”**: El usuario visita una página
- **“conversion”**: El usuario realiza una conversión, comprando un producto.
- **“lead”**: El usuario se registra para recibir una notificación de disponibilidad de stock, para un producto que no se encontraba disponible en ese momento.

2. Preprocesamiento y descripción de Datos

Establecida la información con la que contamos y el significado de cada columna y evento, procedemos al análisis general de estos datos mediante Python Pandas.

Para acotar la muestra relevante, hemos decidido considerar que utilizaremos únicamente los datos de usuarios cuyo país desde donde se originó el evento (country) sea Brasil. Esto se debe a que el volumen de transacciones originadas desde otros países es muy baja y porque Brasil es el mercado principal de la compañía. Este filtro nos garantiza también eliminar un número bajo de usuarios que solo tienen el evento checkout. Al no ser coherente con nuestra progresión de compra y no pertenecer a Brasil hemos decidido excluirlos.

El rango de fechas que se analiza en el set de datos va desde el 01 de enero de 2018 al 15 de junio de 2018. Como los eventos son interacciones con un sitio web, hemos encontrado cierta progresión de los mismos como muestra el gráfico a continuación:



2.1 Preprocesamiento de Datos

Para explorar los datos se agregaron columnas al set de datos relacionados con columnas existentes, a saber:

- day: día del evento - timestamp
- month: mes del evento - timestamp
- brand: marca del dispositivo - model
- session_id: identificador de sesiones para cada usuario
- last_timestamp: fecha del evento anterior de un mismo usuario
- new_session: primer evento de una sesión

Adicionalmente, se procesó el campo SKU. Una primer vista indica que para los eventos 'viewed product', 'checkout' y 'conversion' algunos datos contienen un .0 en su campo y hemos decidido eliminarlo para normalizar el contenido de este campo. También se eliminaron 2 eventos cuyo SKU era "undefined" porque no nos permite obtener información relevante para el análisis.

3. Análisis de Datos

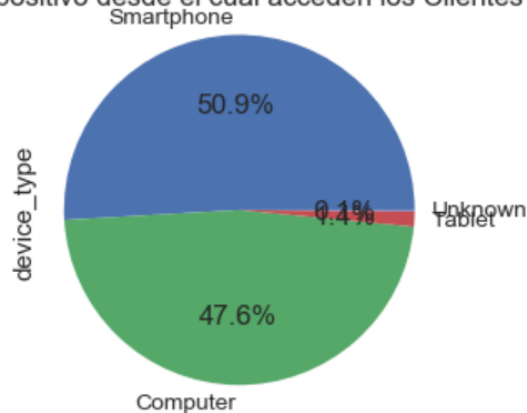
3.1 Clientes

3.1.1 Total de Usuarios

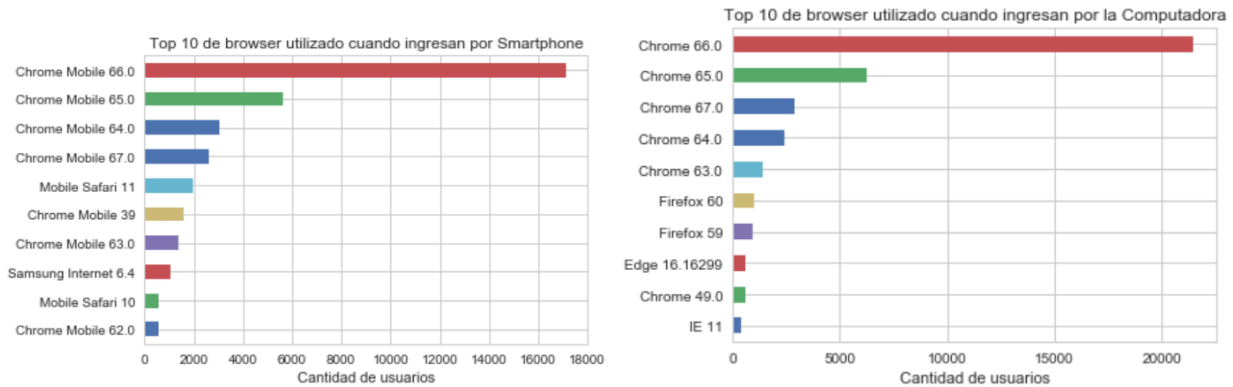
La cantidad total de usuarios únicos considerados clientes o potenciales adquirentes de smartphones que tuvieron actividad en el sitio web durante el período analizado es de 25.872. El dispositivo utilizado para acceder al sitio web es principalmente smartphone y en segundo lugar se encuentra la computadora.

```
Dispositivos desde el cual acceden los Clientes:
Smartphone    43481
Computer      40616
Tablet        1187
Unknown       114
Name: device_type, dtype: int64
```

Dispositivo desde el cual acceden los Clientes



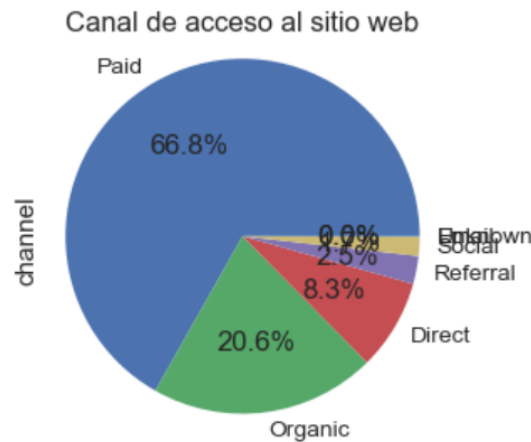
Aquellos que utilizan smartphones lo hacen a través de Chrome Mobile en sus diversas versiones y Mobile Safari aunque con un muy poco porcentaje. Y quienes utilizan computadoras lo hacen a través de Chrome en un 100% aunque no todos lo hacen con la misma versión. Este dato nos permite remarcar que, para asegurar una óptima experiencia de usuario, el sitio web debería estar optimizado para acceder desde Chrome web y mobile.



3.1.2 Nuevos Clientes

¿A través de qué canal llegan los clientes? Considerando los clientes que ingresan al sitio por primera vez en el período estudiado, la mayoría de los clientes ingresan mediante el canal paid.

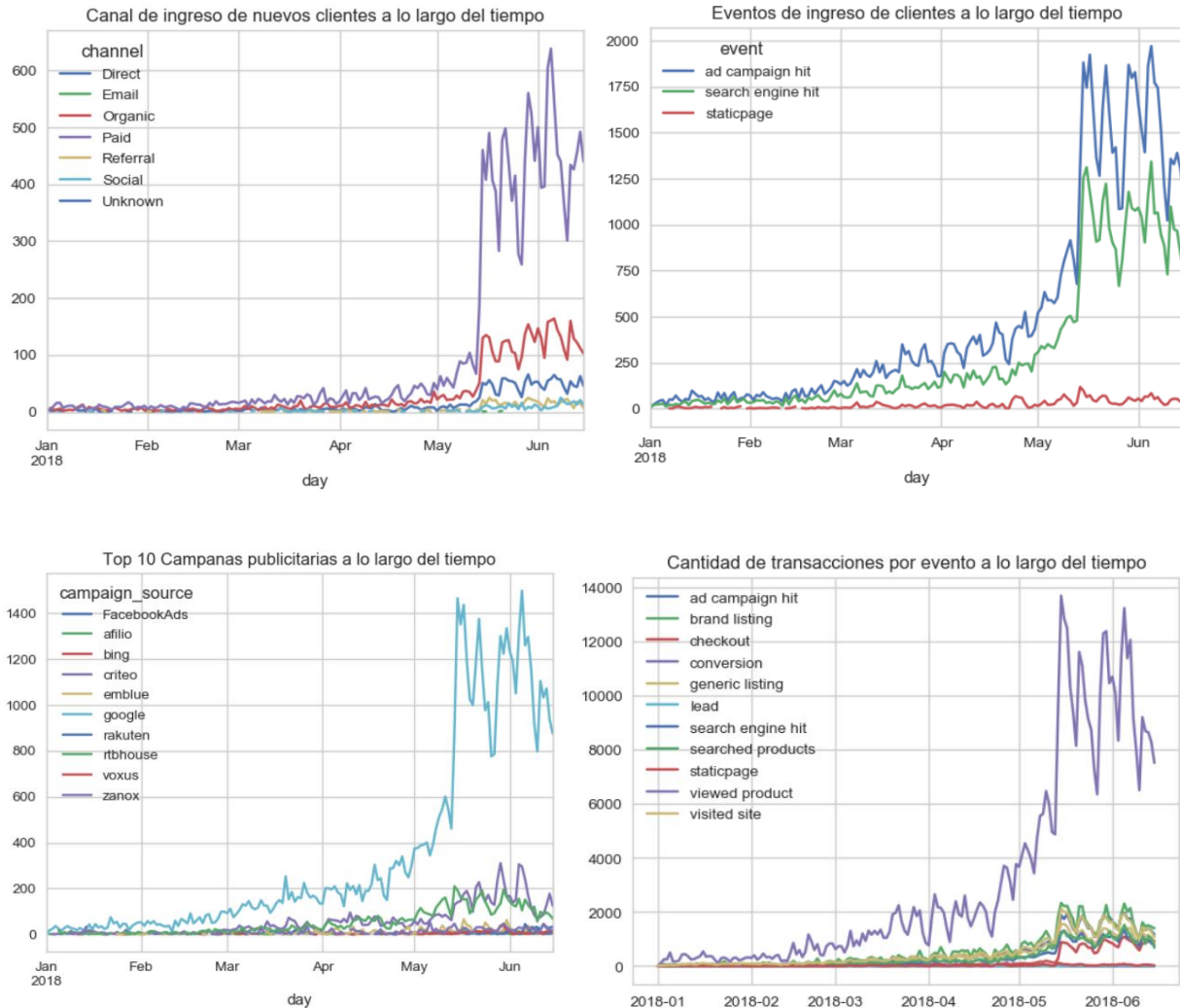
```
Nuevos clientes llegan principalmente por el canal Paid
Paid          17283
Organic       5331
Direct        2136
Referral       659
Social         451
Email          10
Unknown         1
Name: channel, dtype: int64
```



Para situar estos nuevos ingresos a lo largo del tiempo hemos decidido representarlo a través de un gráfico, esto da por resultado que hay un pico de nuevos clientes entre mayo y junio. ¿A qué se debe?



Si continuamos investigando este pico de ingreso entre mayo y junio y lo relacionamos con canales, descubrimos que llegaron a través del canal paid. Aunque también hubo un incremento del tráfico orgánico. ¿Puede deberse a una campaña en particular? A continuación veremos los gráficos de línea considerando distintos parámetros: canal, evento y campaña. Todos muestran similar comportamiento lo que nos permite concluir que, el aumento de clientes nuevos coincide con ad campaign hit (que es un canal de pago) y también con search engine hit (considerado un canal orgánico o de pago dependiendo si en el buscador se hace click en un link de campaña o no). La campaña llevada a cabo entre mayo y junio fue de Google, lo que nos indica un suceso en el ingreso de nuevos usuarios debido a la misma. Podemos afirmar que, Google al ser un buscador y un medio donde se puede hacer publicidad, ha provocado por una campaña (canal paid), el aumento de ingresos entre mayo y junio. Por supuesto, que una vez que un usuario ingresa al sitio, realiza el evento de mirar productos y este ha aumentado con el mismo comportamiento a medidados de mayo. Y crece y decrece acompañado de la efectividad del anuncio de Google.

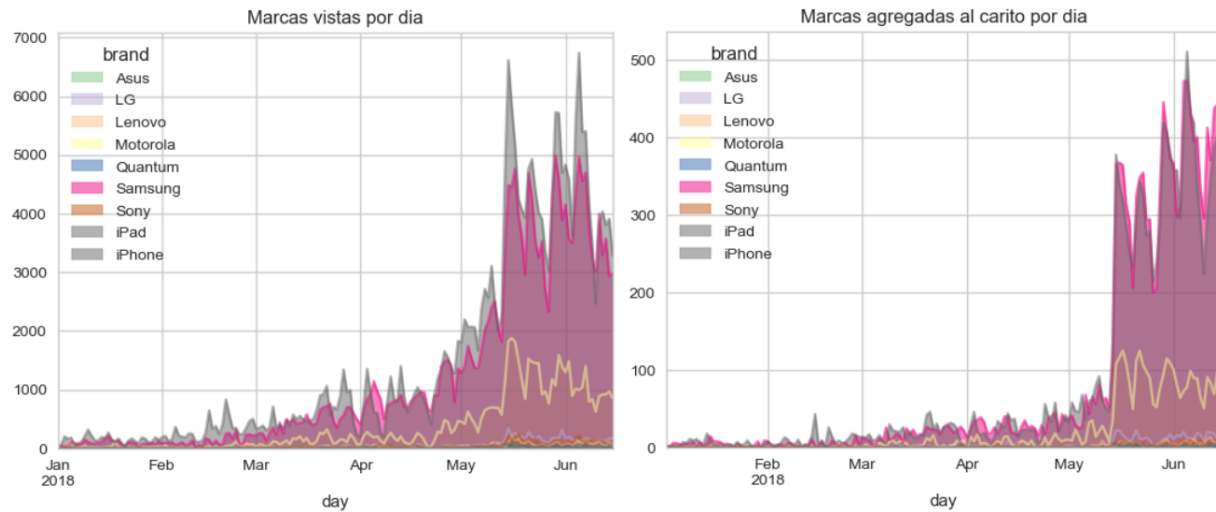


Nuestra recomendación es realizar Ads a través de los Google únicamente. Por otro lado, si desea hacer Ads por otros medios, recomendamos incrementar el targeting centrado en las marcas buscadas por ej.

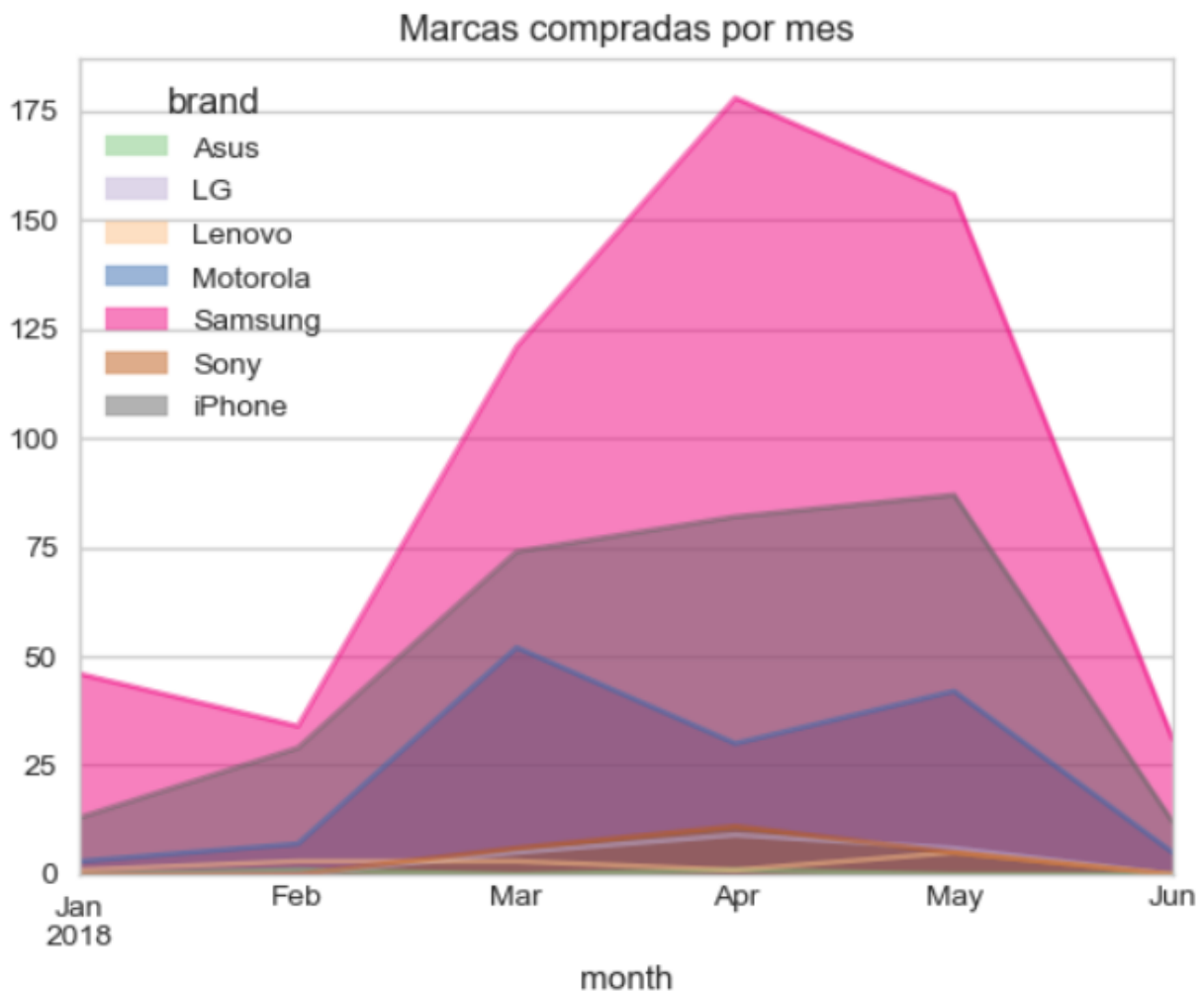
3.2 Marcas

Como hemos mencionado anteriormente, el sitio web www.trocafone.com indica que venden principalmente smartphones de las marcas Apple y Samsung. Para corroborarlo hemos estudiado la marca de dispositivo por evento: viewed product, checkout y conversion.

Las marcas más vistas por día (viewed product) y agregadas al carrito (checkout) son efectivamente iPhone (Apple) y Samsung. En tercer lugar se encuentra Motorola.



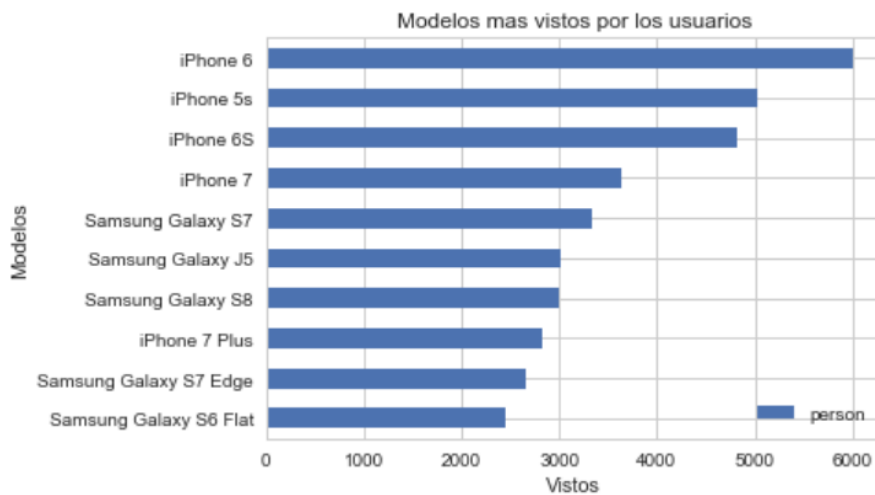
Cuando analizamos las marcas compradas (conversion) en total, obtenemos que Samsung es más comprada aunque iPhone sea la más vista.



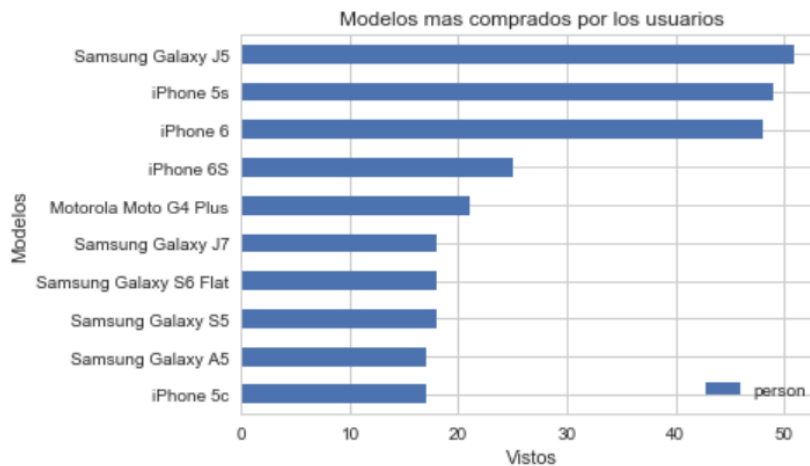
Curiosamente, cuando analizamos las marcas, también notamos un pico de vistas y agregados al carrito entre mayo y junio de dispositivos iPhone y Samsung pero no se observa la misma incidencia en las compras o conversiones realizadas. Si este pico se debe a la campaña de Google antes mencionada, ésta logró atraer nuevos clientes, aumentaron las visualizaciones y el checkout de productos, sin embargo el aumento en las compras fue momentáneo.

3.2.1 Modelos

Si bien ya establecimos que las marcas más vistas son Samsung y iPhone, a continuación veremos el comportamiento por modelos.



El modelo más comprado es: Samsung Galaxy J5, seguido de iPhone 5s, 6 y 6S. Pero en total, son más de la marca Samsung que iPhone.



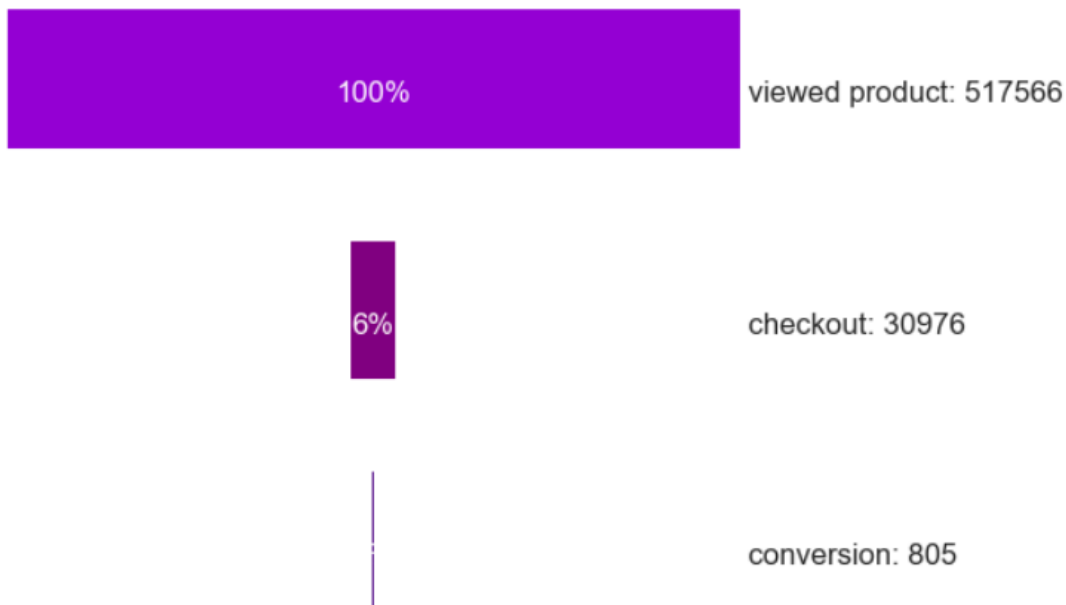
3.3 Métricas del comportamiento de los usuarios

¿Qué porcentaje de los productos “viewed product” son efectivamente colocados en el carrito de compras? ¿Qué porcentaje de los productos “viewed product” generan lead por falta de stock? ¿Qué

porcentaje de los productos “checkout” son efectivamente comprados (“conversion”)? ¿Cuánto representan estas compras sobre los productos vistos “viewed product”?

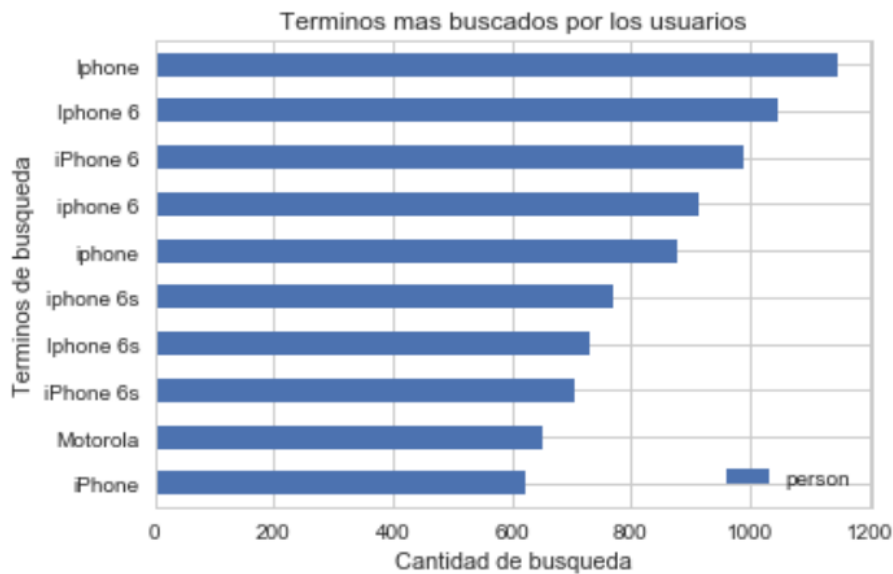
Responder a estos interrogantes le proporciona a Trocafone un insight del comportamiento de los usuarios y posibles insights acerca de la efectividad del sitio web para el recorrido del usuario en una compra.

La cantidad total de productos vistos es de 517.566, de los cuales 31.891 fueron colocados en el carrito de compras. Algunos de los productos colocados en el carrito de compras no pasaron por viewed product, significando que entraron directamente al link del carrito. De esto resulta que un 6% de productos vistos pasaron al carrito de compras (30.976). La cantidad de productos vendidos (conversion) es significativamente menor, tan solo 1.059, un 0.16% de los productos vistos y un 2.6% de los productos agregados al carrito.



Considerando esto, un usuario mira en promedio 20,57 veces el mismo producto antes de comprarlo, ofreciendo un insight del interés del usuario en un producto determinado. Se recomendaría a Trocafone realizar campañas publicitarias orientadas a usuarios que se muestren interesados en un producto (x cantidad de visitas del mismo producto) para lograr que se transforme de interés a conversión.

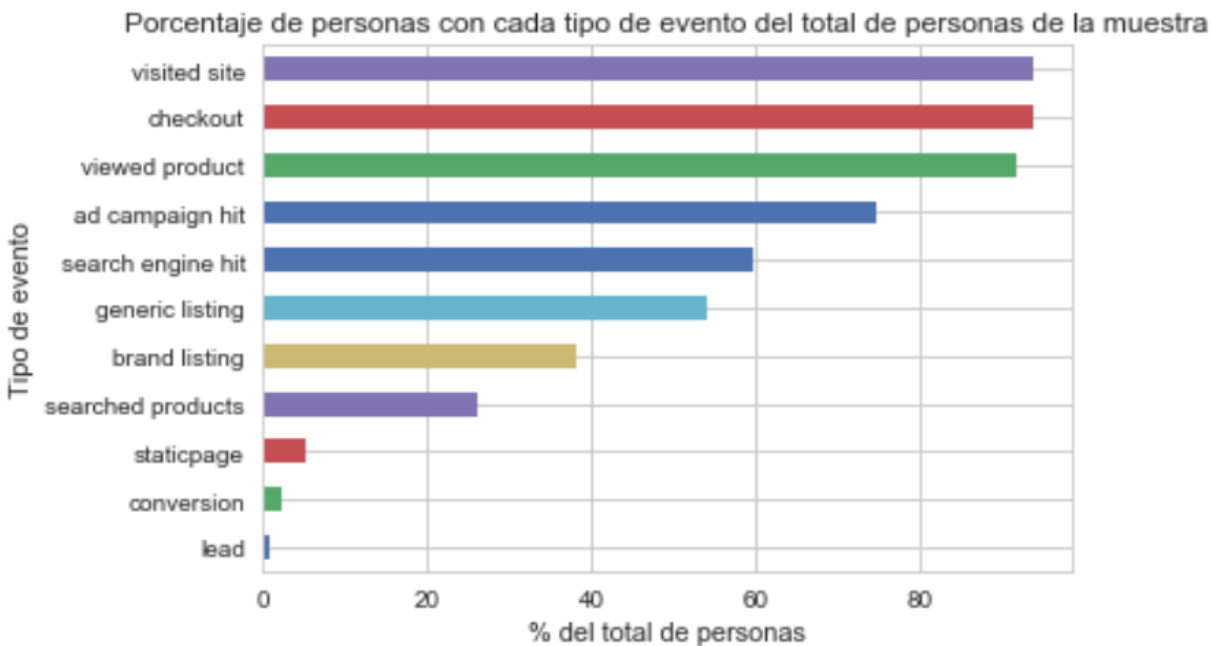
¿Qué buscan los usuarios cuando buscan? Las búsquedas van acompañadas de los productos que más se miran y con el 2°, 3° y 4° producto más comprado. Lógicamente los iPhone generan curiosidad en los usuarios pero por algún motivo, no todos los que buscan concretan una compra.



3.4 Tipos de evento por persona

Si analizamos la cantidad de personas con evento, al haber tomado sólo personas de Brasil, todas tienen al menos un evento visited site y un evento checkout.

Considerando lo anterior, ¿está la muestra tomada sólo para estos clientes? Si fuera así, recomendamos a la empresa Trocafone que tome todos los datos de las personas de su sitio ya que está perdiendo información de los usuarios que abandonan el sitio antes de hacer el checkout, de qué productos ven o qué patrones existen que evitan que lleguen al checkout.



3.5 Sesiones por usuario

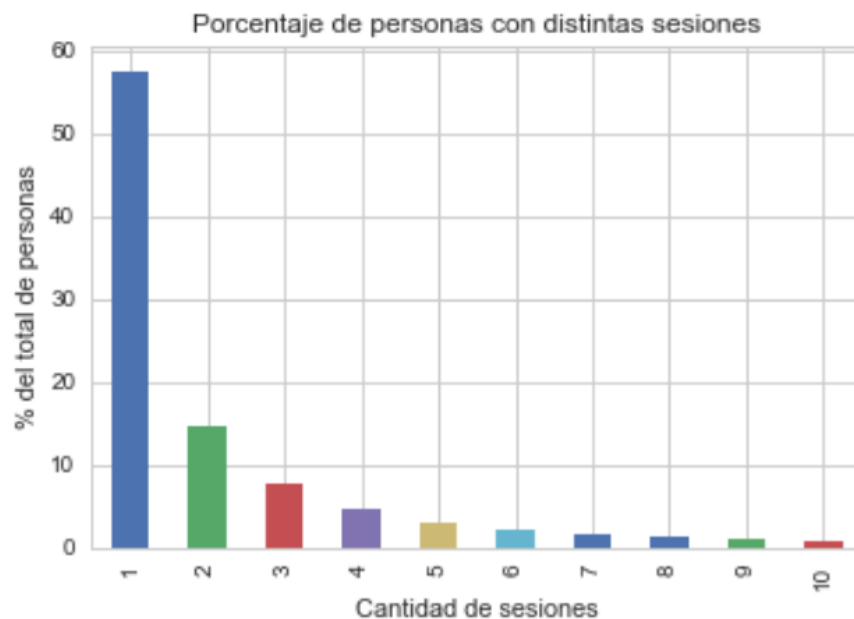
Para agrupar eventos por sesiones decidimos considerar que los eventos no estén separados entre sí por más de 30 minutos, a su vez cada sesión debería tener un solo evento visited site. De esta forma obtenemos:

- Cantidad total de sesiones: 85.995
- Cantidad de sesiones sin visited site: 1.566
- Cantidad de sesiones con más de un visited site: 65

Si bien implica que las sesiones no están delimitadas del todo bien, procedemos a aceptar esta separación ya que significan una porción muy pequeña del total de las sesiones.

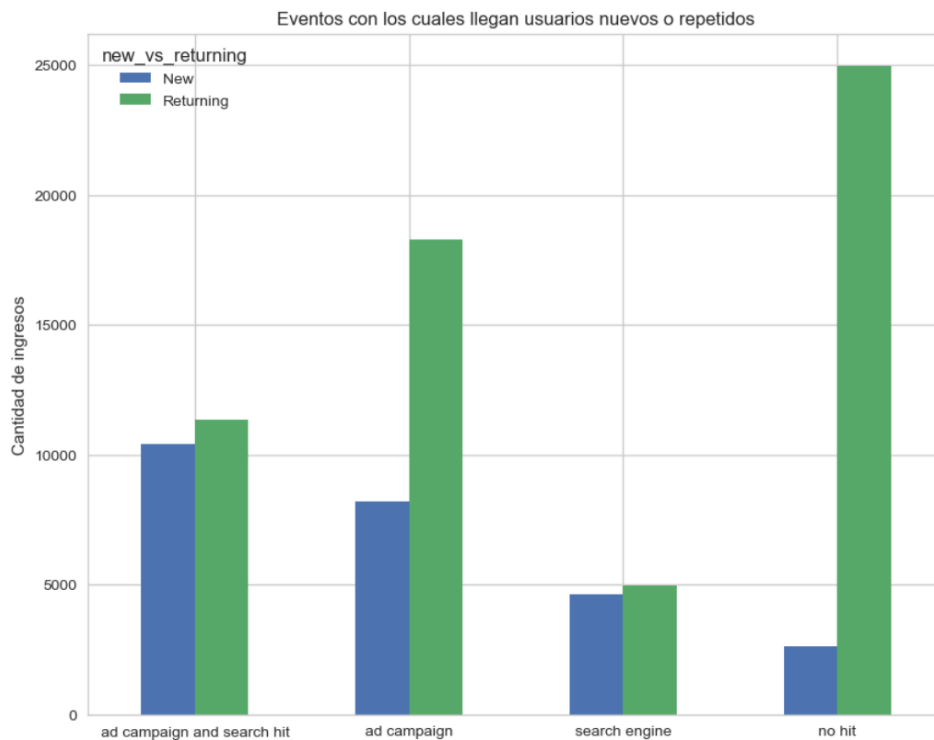
3.5.1 Análisis por sesión

Teniendo en cuenta el split propuesto, por usuario podemos ver que más del 50% tiene una sola sesión, y el 75% de las personas tienen entre 1 y 3 visitas. La cantidad de eventos por sesión sigue una distribución geométrica con una media de 11 eventos y una mediana de 7 eventos por sesión.

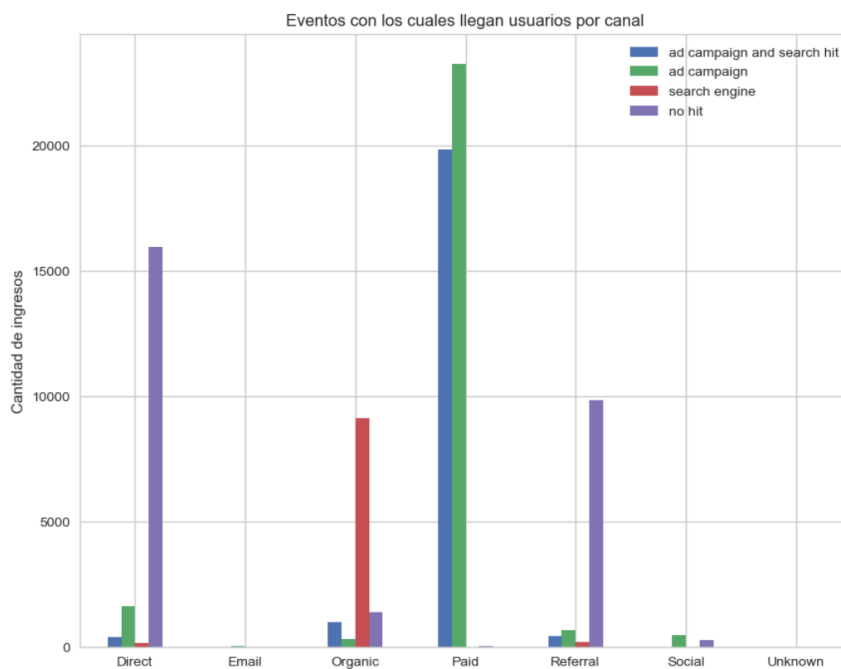


3.5.2 Ingreso de usuarios

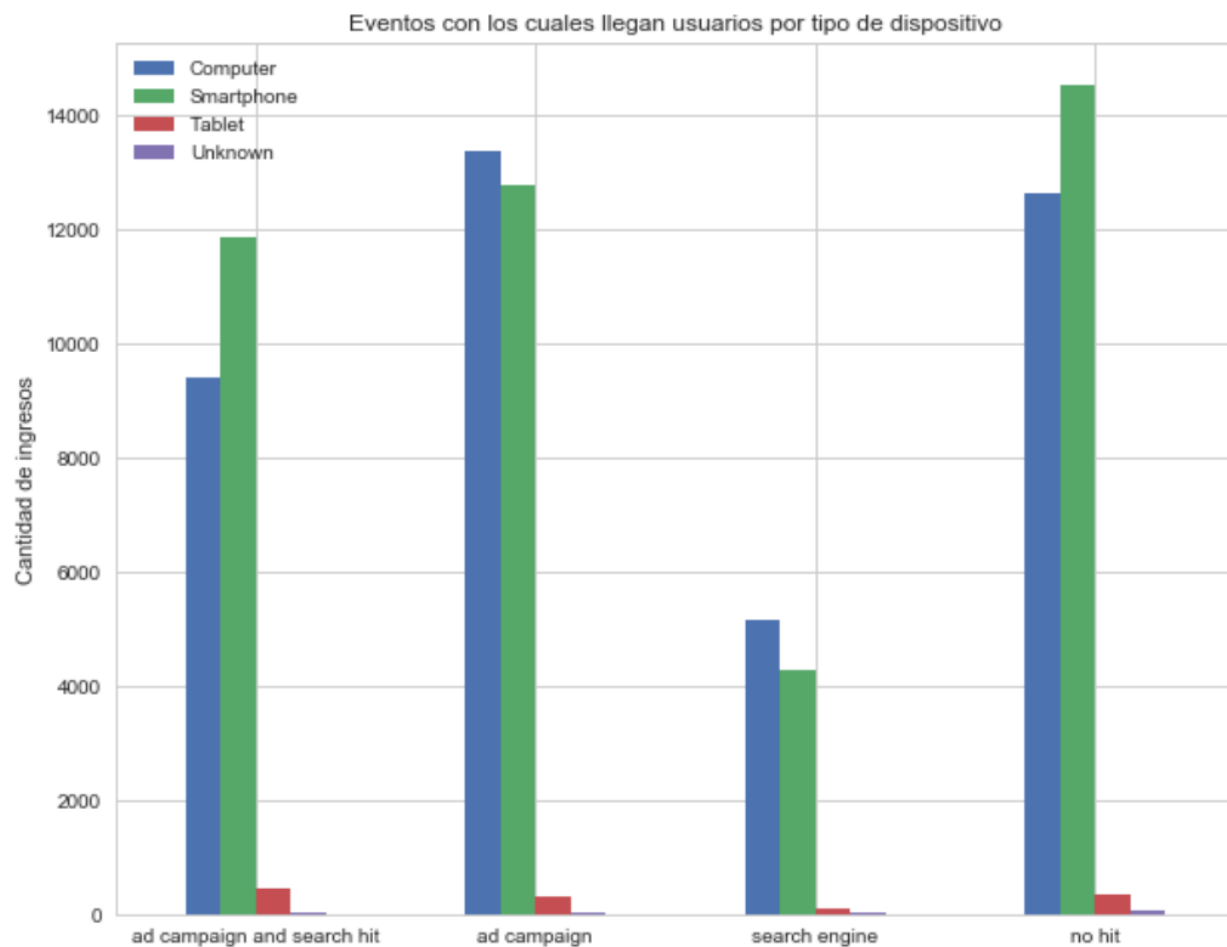
Si combinamos la distinción entre sesiones con el ingreso de nuevos usuarios podemos reforzar los descubrimientos anteriores. La mayoría de los usuarios nuevos llegan al sitio con dos eventos, ad campaign hit y search engine hit. Cuando se trata de usuarios que vuelven, lo más común es que no entren a través de ad campaign o search engine hit.



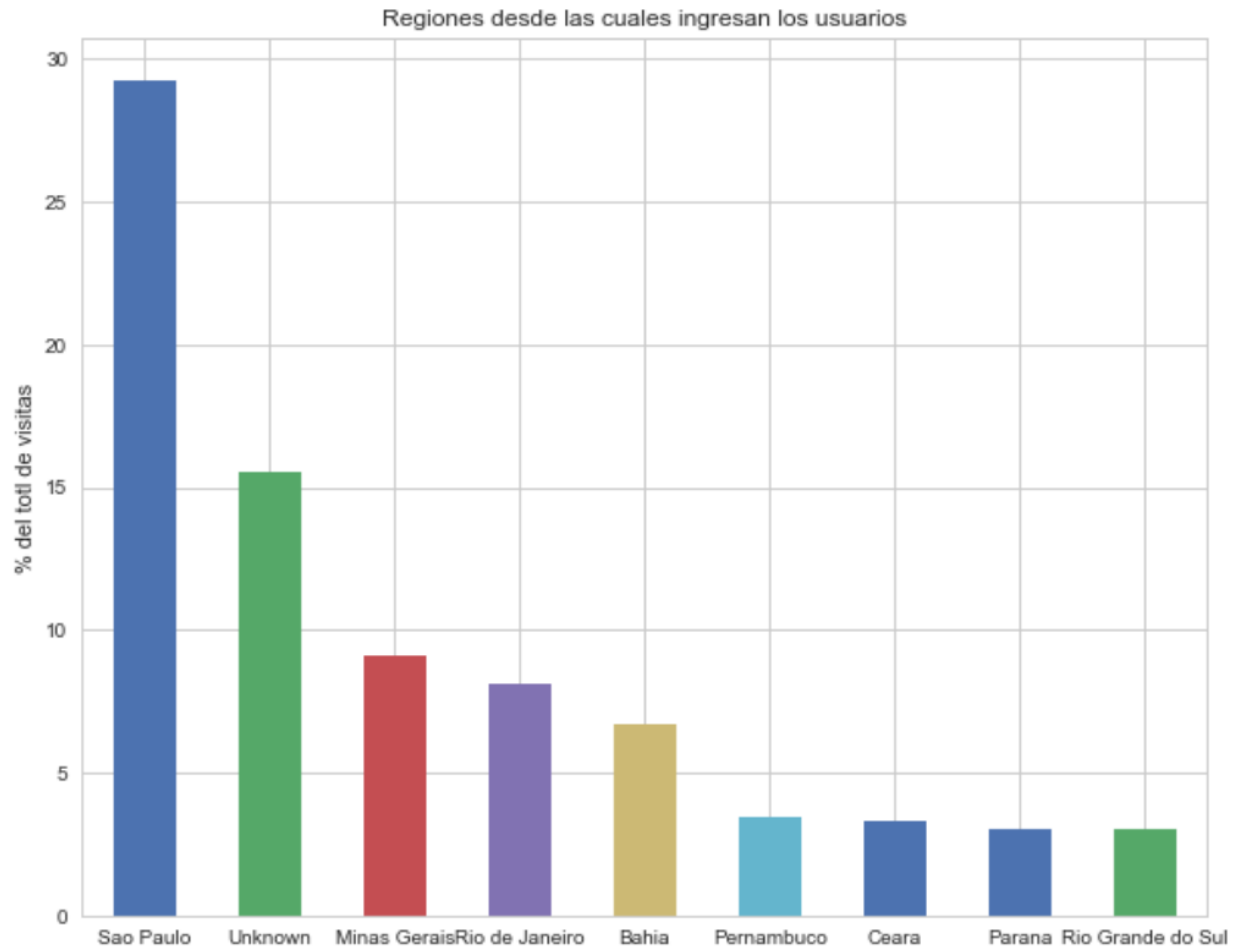
¿De qué canales vienen? Como establecimos anteriormente, la mayoría vienen de canales pagos cuando el evento con el que ingresan es ad campaign hit. La combinación de ambos (ad campaign y search engine hit) implican que el tráfico viene de una campaña en un buscador. Si ingresan sólo con search engine hit, el canal más frecuente es Organic. Cuando no llegan a través de los anteriores, los canales más frecuentes son Direct (es decir que ingresan con el url de la página) y Referral.



El medio más utilizado cuando hacemos el análisis por sesiones está distribuído entre Smartphones y Computadoras. Teniendo en cuenta el dispositivo por donde ingresan los usuarios, podemos deducir que las campañas publicitarias deben ser propuestas en computadora como en dispositivos móviles, y las campañas dentro de un buscador, tienen mayor efectividad en Smartphones.



Respecto de las regiones desde las cuales ingresan los usuarios, vemos que la mayoría ingresan desde Sao Paulo, mientras que hay una gran cantidad de visitas para las cuales no hay registro de la región.



3. Conclusiones

Luego del análisis realizado podemos decir que hemos obtenido interesantes insights que le servirán a Trocafone a entender a sus usuarios, a tomar medidas para mejorar y lograr el objetivo del negocio, es decir, concretar más ventas.

- Los usuarios de Trocafone, ingresan al sitio por Smartphones y Computadoras principalmente
- Tanto en sus Smartphones como en sus Computadoras, la mayoría elige Chrome como buscador

Con estos 2 insights Trocafone puede realizar acciones para optimizar su sitio web en vista Smartphone y Computadoras para que la experiencia de usuario sea mejor. Además, la optimización de interfaz debería ser a través de Chrome.

- El 66% del tráfico es pago y el 20% es orgánico
- La campaña publicitaria con más éxito en clicks, por ende visitas, fue realizada a través de Google
- Las campañas en buscadores (ej. Google) tienen mayor efectividad en Smartphones.
- Las campañas en otros sitios tienen igual efectividad en computadora como en dispositivos móviles

Esto significa que la empresa, debería realizar acciones para que su tráfico derive más a orgánico, como posicionarse en el buscador de Google. Pero, también implica que Trocafone debería centrar sus campañas de Ads en Google, ya que ha sido probada exitosa y hacerla disponible para individuos que entren por Smartphone o Computadoras.

- Los usuarios buscan en su mayoría dispositivos de la marca iPhone
- Apple y Samsung son las marcas más vistas y agregadas al carrito de compras (casi por igual)
- A la hora de comprar, los clientes eligen Samsung

Trocafone deberá analizar por qué los clientes desisten de comprar todos los iPhones que son buscados y/o colocados en el carrito de compras. ¿Será por el precio o por los medios de pago? Con este set de datos no lo podemos determinar, pero es un buen punto de partida para investigarlo.

- Cerca del 30% de usuarios que ingresan desde Brasil, lo hacen desde la región de San Pablo
- San Pablo, Minas Gerais, Río de Janeiro y Bahía concentran más del 50% de usuarios
- Lamentablemente hay muchos datos cuya región es Unknown

Para mejorar sus ventas, Trocafone debería hacer más campañas para ser reconocido en más regiones, sin perder de vista que tiene que mantener el volumen de usuarios de San Pablo y las regiones mayoritarias.

- Entre mayo y junio, hubo un gran salto en las visitas y la adición de productos al carrito de compras
- En el mismo período se observa una gran cantidad de hits de una campaña publicitaria en Google

- Este mismo salto no se vio reflejado en ventas, incluso las ventas de Samsung (la marca más comprada) decrecieron respecto del mes de abril. iPhone y Motorola tuvieron un leve crecimiento

Las conversiones no acompañaron bien a esta campaña de Google realizada a mediados de mayo. Pero, la campaña dejó un tráfico que acompaña al sitio hasta el día de corte de este estudio. Tal vez, la campaña logró hacer conocida la empresa e incrementar el número de futuros compradores pero no concretar ventas.

- 6% de productos vistos pasan al carrito de compras
- 0.16% de los productos vistos son comprados
- 2.6% de los productos agregados al carrito son comprados

Si bien todas las métricas son significativas para la estrategia de Trocafone, la que más relación directa tiene con los ingresos son estas 3. ¿Por qué los usuarios buscan tanto antes de decidirse agregarlo al carrito? ¿Por qué una vez agregado al carrito deciden no concretar la compra? En esto pueden influir muchos factores, tal vez la descripción de los productos o el abanico sea muy amplio y los usuarios no puedan decidirse por uno fácilmente. Considerando las características del negocio donde venden Smartphones refurbished (con calificación Excelente, Muy Bueno, Bueno) esto puede ser verdadero y la decisión al potencial comprador le lleva mucho tiempo y vistas de comparación. Pero el porqué de una vez agregado al carrito no se concreta la compra, tiene que ver con otros motivos que pueden ser envío, método de pago, o que prefieren ver el móvil en una tienda física, que exceden a los datos de este set para el análisis.

Con todos estos insights y conclusiones, Trocafone podría establecer estrategias, las que recomendamos u otras, para lograr un mayor tráfico de usuarios en su sitio, mejorar el targeting de Ads, optimización de su sitio e interfaz, mejora en la descripción de productos para disminuir las múltiples vistas de los mismos; todo esto en pos de incrementar la cantidad de productos en el checkout y por ende la cantidad de productos comprados.