



Trabajo Práctico N° 1

Análisis Exploratorio

Teoría de Organización de Datos

Nombre grupo: *The beauty and the data*

Nº grupo: 38

Todo el trabajo realizado puede encontrarse en el siguiente repositorio de github:
<https://github.com/sebalogue/tp1-datos.git>

Participantes	Nº padrón	Mail
LOIS, Lucas Edgardo	98923	lucaslois95@gmail.com
LOGUERCIO, Sebastian Ismael	100517	seba21log@gmail.com
MARIANI, Santiago Tomás	100516	santiagomariani2@gmail.com
MARIJUAN, Magalí	100070	maguimar001@gmail.com

Índice

1. Introducción	2
2. Análisis Preliminar	2
2.1. Eventos	2
2.2. Viewed product	4
2.2.1. Análisis individual de las características	4
2.2.2. Análisis temporal	5
2.2.3. Análisis cruzado	8
2.3. Ad campaign Hit	10
2.3.1. Análisis individual de las características	10
2.3.2. Análisis temporal	12
2.4. Checkout	13
2.4.1. Análisis individual de las características	13
2.4.2. Análisis temporal	15
2.5. Lead	15
2.5.1. Análisis Individual de las características	15
2.6. Visited site	16
2.6.1. Análisis individual de las características	16
2.7. Generic listing	22
2.7.1. Análisis individual de las características	22
2.8. Searched product	22
2.8.1. Análisis individual de las características	23
2.9. Search engine hit	24
2.9.1. Análisis entre eventos	24
2.9.2. Análisis temporal	24
2.10. Static page	24
2.11. Conversion	24
2.11.1. Análisis individual entre las características	25
3. Análisis entre eventos	26
3.1. Visitas por campaña publicitaria VS conversiones a lo largo del tiempo	26
3.2. Relación entre número de compras y visitas para un mismo usuario	26
3.3. Análisis sobre las campañas de Marketing	27
3.3.1. Análisis general	27
3.3.2. Modelos más comprados por usuarios que cliquearon una determinada campaña	29
3.4. Análisis de dispositivos por el que se conectan los usuarios	32
3.4.1. Análisis de sistemas operativos	33
3.4.2. Análisis de resoluciones	36
3.5. Productos más populares por ciudad de Brasil	36
3.6. Conversiones en base a checkouts	37
3.7. Conversiones en base a visitas de productos en la página principal	37
3.8. Analisis de checkouts por persona	38
4. Conclusiones finales	38

1. Introducción

Este trabajo práctico consiste en realizar un análisis sobre un conjunto de eventos de Web Analytics de usuarios que visitaron www.trocafone.com, su plataforma de ecommerce en Brasil. Trocafone es un side to side Marketplace para la compra y venta de dispositivos electrónicos que se encuentra actualmente operando en Brasil y Argentina.

Para realizar dicho trabajo utilizaremos el lenguaje de programación *Python*. Para el análisis de datos usaremos la librería *Pandas* y para la realización de gráficos utilizaremos las librerías *Matplotlib* y *Seaborn*. Por otro lado, trabajaremos usando el sistema de control de versiones *GIT*.

El objetivo del trabajo práctico es analizar los distintos eventos e interacciones que pueden realizar los usuarios en la plataforma y a partir de estos establecer diversas conclusiones. Haremos énfasis en encontrar patrones de comportamiento y entender cuáles son las preferencias de los usuarios a la hora de comprar y vender productos teniendo en cuenta diversos parámetros, como la región, momento del día, semana del año, palabras clave, etc.

Utilizaremos la siguiente abstracción para realizar el trabajo práctico: Se realizará una pregunta general, y a partir del uso de las herramientas mencionadas anteriormente se buscará dar una respuesta a partir del comportamiento de los datos.

2. Análisis Preliminar

En un primer análisis intentaremos entender cómo está compuesto el dataset. Es decir, cuáles son los eventos que lo componen, cómo interactúan en conjunto y cuáles son las características propias de cada evento.

2.1. Eventos

El dataset contiene los siguientes eventos:

- 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 campaña de marketing online.
- Generic listing: El usuario visita la homepage.
- Serched products: El usuario realiza una búsqueda de productos en la interfaz de búsqueda del sitio.
- 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.

La cantidad total de ocurrencias es: 1011288. Las cuales están distribuidas de la siguiente manera.

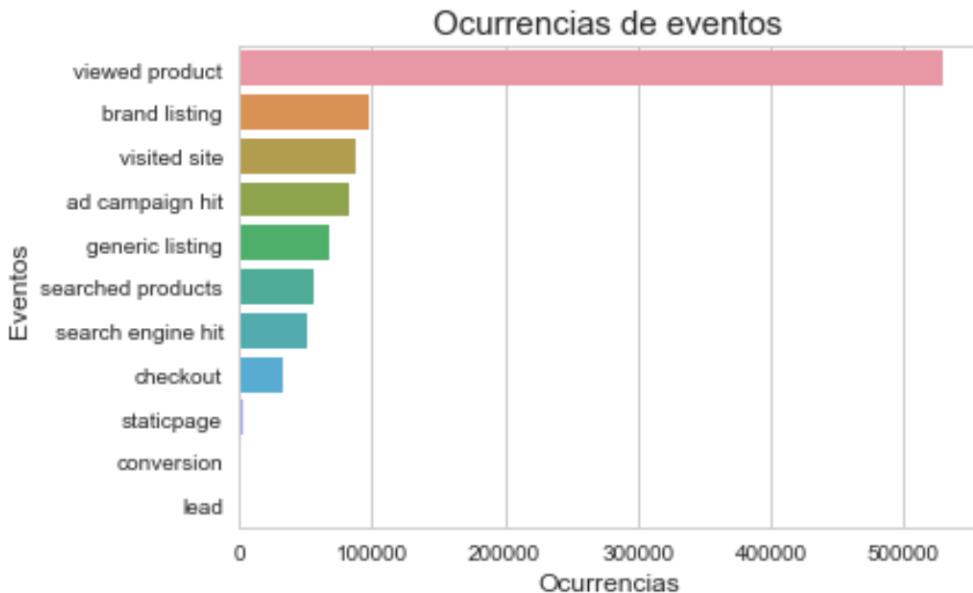


Figura 1: Cantidad de apariciones de cada evento.

Es importante mencionar que no todos los campos del dataset participan en todos los eventos. Cada evento utiliza una determinada cantidad de columnas del dataset. Por lo tanto, mediante un análisis de elementos nulos llegamos a las siguientes conclusiones:

- Todos los eventos tienen timestamps que indican exactamente en qué momento se realizó dicho evento. Además de poseer información sobre quien lo realizó.
- Para el evento **viewed product** sus campos obligatorios son: timestamp , sku , model, condition, storage, color.
- Para el evento **brand listing** su campo obligatorio es: skus.
- Para el evento **visited site** sus campos obligatorios son: channel, new vs returning, city, region, country, device type , screen resolution, operating system version y browser version.
- Para el evento **ad campaign hit** sus campos obligatorios son: url y campaign source.
- Para el evento **generic listing** su campo obligatorio es: skus.
- Para el evento **searched product** sus campos obligatorios son: skus y search term.
- Para el evento **searched engine** su campo obligatorio es: search engine
- Para el evento **checked out** sus campos obligatorios son: sku, color, storage, model y condition
- Para el evento **static page** sus campo obligatorio es: static page
- Para el evento **conversion** sus campos obligatorios son: sku, model, color , condition y storage
- Para el evento **lead** su campo obligatorio es: model

Analizaremos a partir de los timestamps cuál es el momento del día en que se registra mayor actividad.

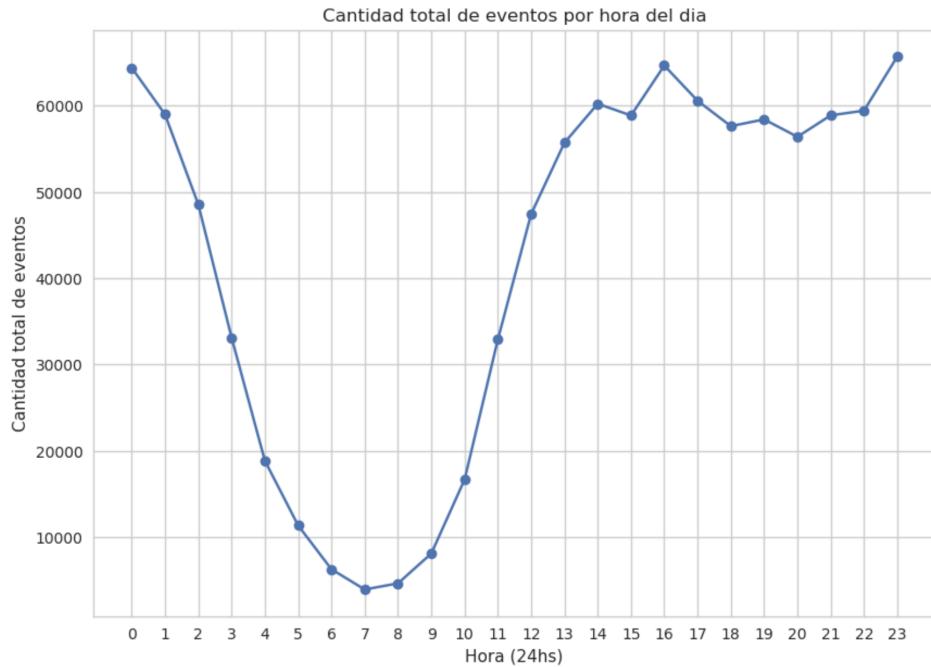


Figura 2: Distribución de los eventos a lo largo del día.

Podemos observar que la mayoría de los eventos se efectúan entre las 14 hs y 2 hs.

2.2. Viewed product

2.2.1. Análisis individual de las características

La primera columna que decidimos analizar de éste evento es model. Podemos observar cuáles son los modelos más visitados en el sitio:

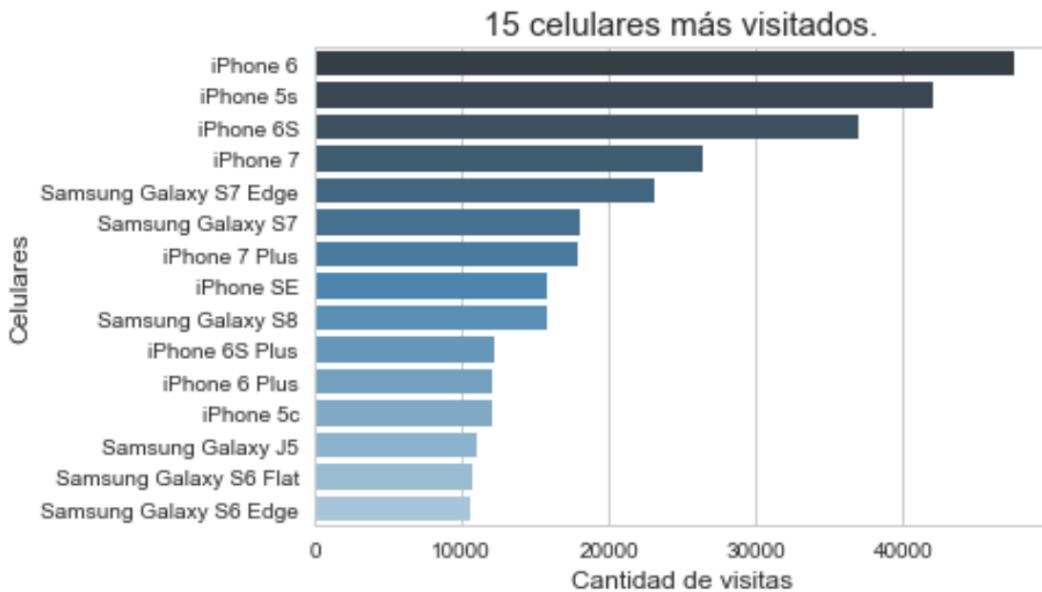


Figura 3: Cantidad de visitas por celular.

Sobre las características de esos modelos primero analizamos el color de los productos vistos y concluimos que el 50 % de las visitas se concentraron en los siguientes colores:

- el 23 % de las visitas es hacia productos negros.
- el 20 % de las visitas es hacia productos dorados.
- el 11 % de las visitas es hacia productos gris espacial.
- el 10 % de las visitas es hacia productos blancos.
- el 9 % de las visitas es hacia productos plateados.
- el 6 % de las visitas es hacia productos rosas.

Despreciamos el resto de los colores ya que el porcentaje de visitas es poco significativo.

Luego analizamos las visitas hacia el almacenamiento de los productos y concluimos que:

- El 33 % de las visitas es hacia productos con almacenamiento de 16 GB.
- El 32 % de las visitas es hacia productos con almacenamiento de 32 GB.
- El 17 % de las visitas es hacia productos con almacenamiento de 64 GB.
- El 6 % de las visitas es hacia productos con almacenamiento de 8 GB.

Por último analizamos las condiciones que más visita la gente y concluimos que:

- El 42 % de las visitas es hacia productos que son de calidad buena.
- El 27 % de las visitas es hacia productos que son de calidad excelente .
- El 27 % de las visitas es hacia productos que son de calidad muy buena.
- El 2 % de las visitas es hacia productos que no tienen identificador de huella digital.
- El 0,02 % de las visitas es hacia productos nuevos.

2.2.2. Análisis temporal

Estudiaremos en primer lugar cuáles son los días en los que se registraron la mayor cantidad de visitas.

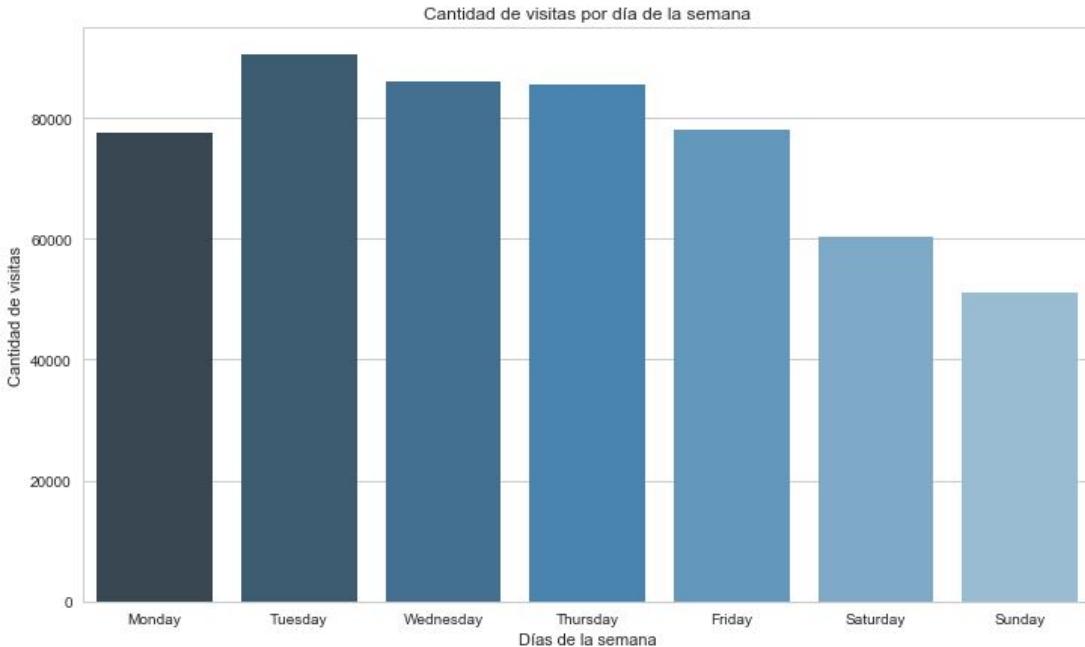


Figura 4: Cantidad de visitas según día del año.

Lo más destacado que se puede ver en este gráfico es que los días sábado y domingo poseen una baja cantidad de visitas al sitio, comparado con los otros días de la semana.

Luego, realizaremos un análisis sobre cómo es el progreso de las visitas a lo largo del semestre.



Figura 5: Cantidad de visitas según día del año.

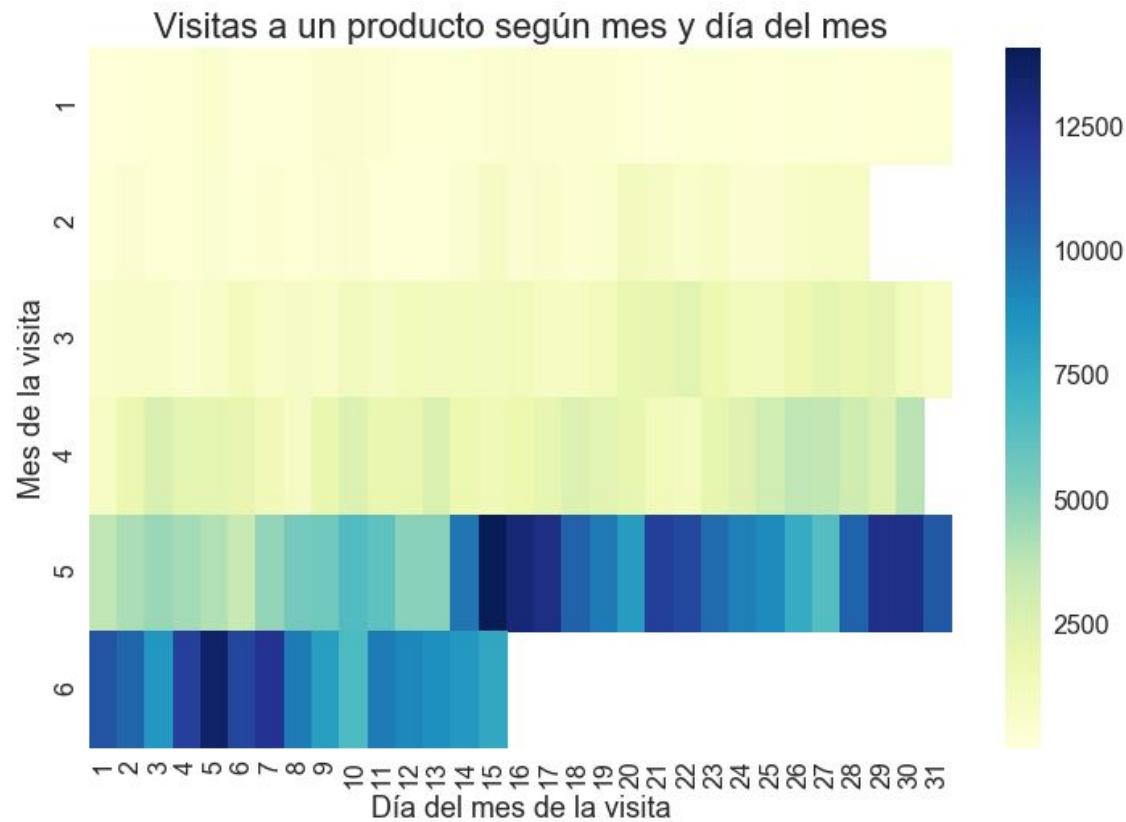


Figura 6: Cantidad de visitas por cada día de cada mes.

Podemos notar que hay un fuerte incremento de actividad a partir del 15 de mayo.

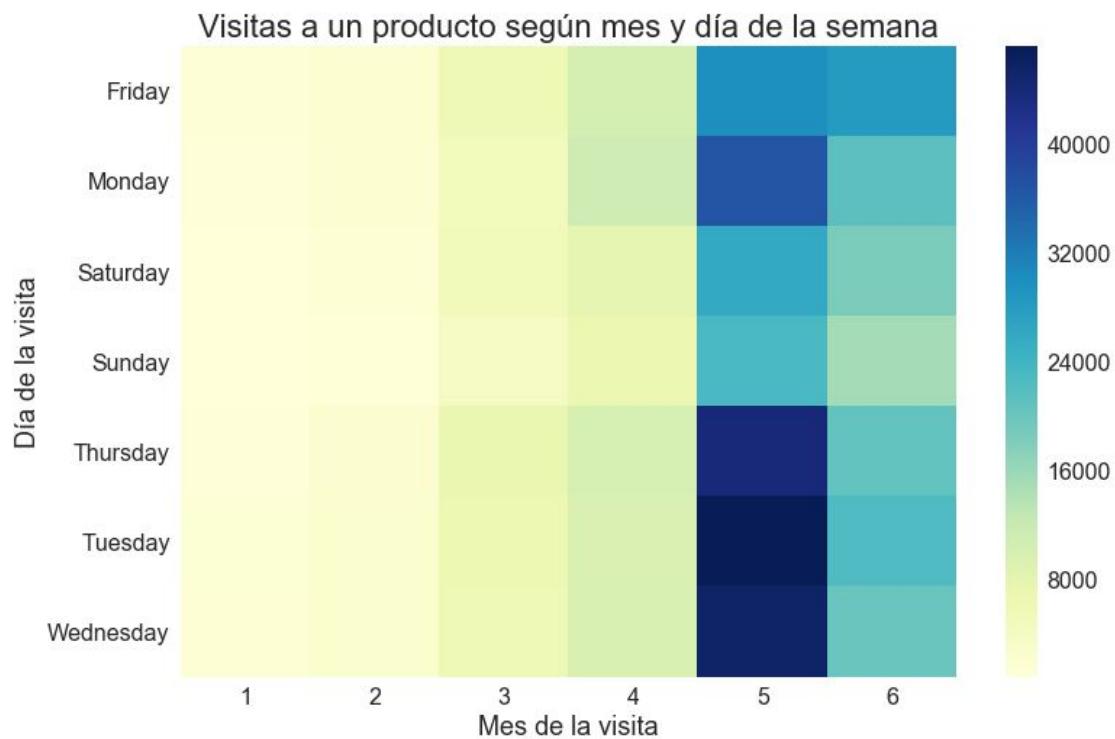


Figura 7: Cantidad de visitas por cada día de semana de cada mes.

Nuevamente hay un fuerte incremento a partir del mes de mayo. Sin embargo, no parece haber una predominancia cuando hablamos de los días de la semana. Lo que podemos concluir es que los domingos hay menos actividad.

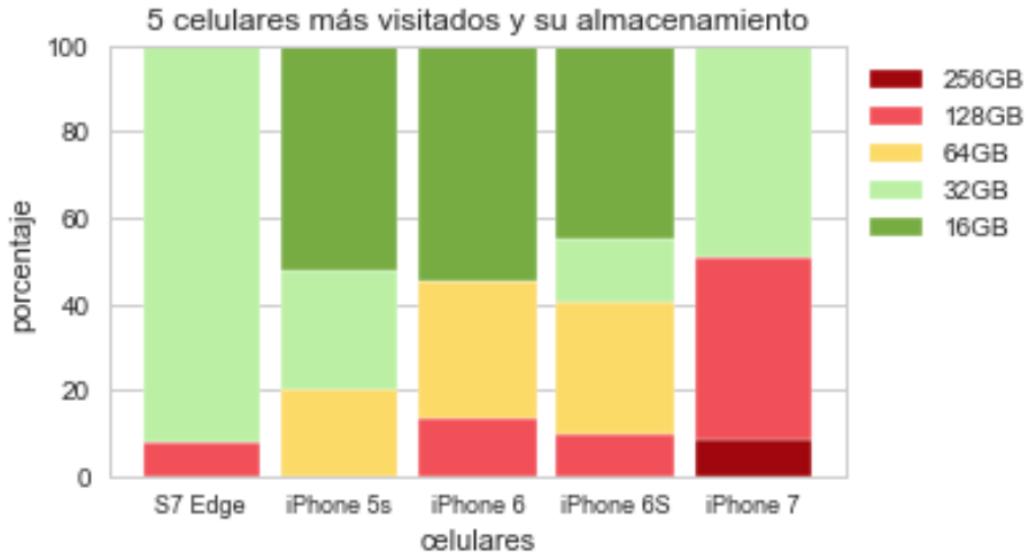
Es importante destacar en el análisis temporal que el mes de junio no está completo. Sólo tenemos datos de su primera quincena.

2.2.3. Análisis cruzado

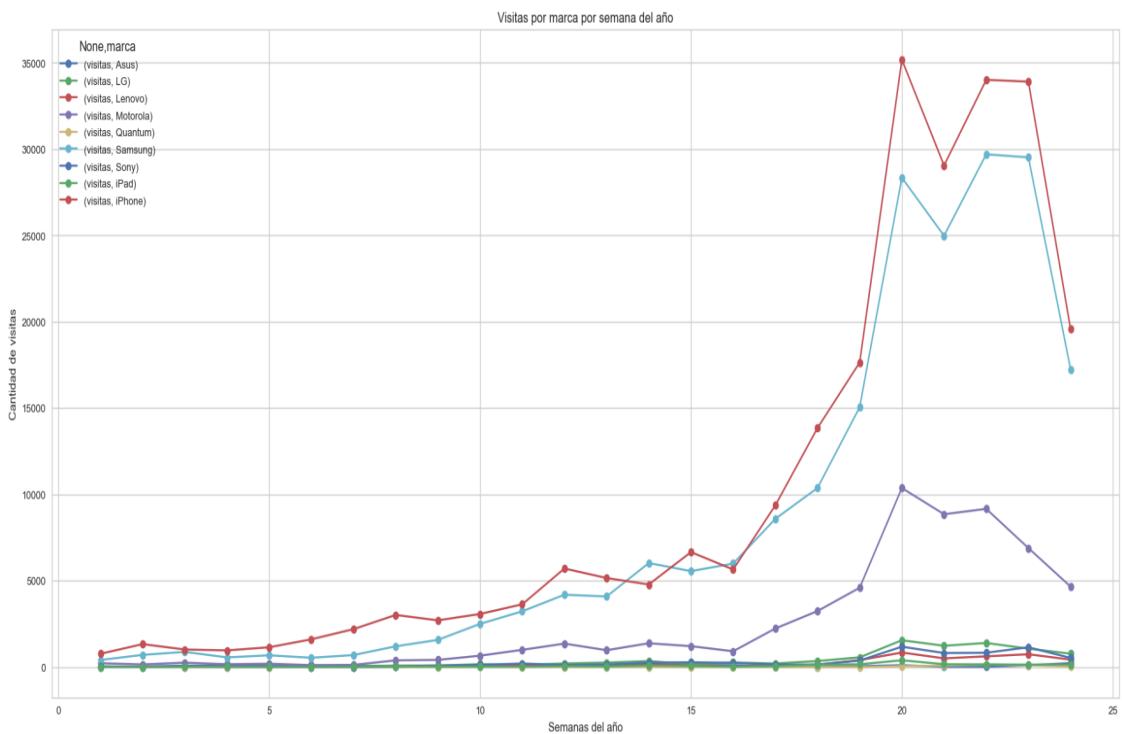
Por último relacionaremos todas las características analizadas anteriormente entre sí. Primero entrelazaremos la información entre los cinco Iphones más vistos y su color (Analizaremos los iphones ya que su esquema de colores es uniforme).

**Figura 8:** Top 5 de celulares según su color.

Por último, entrelazaremos información entre los cinco modelos más visitados y su almacenamiento.

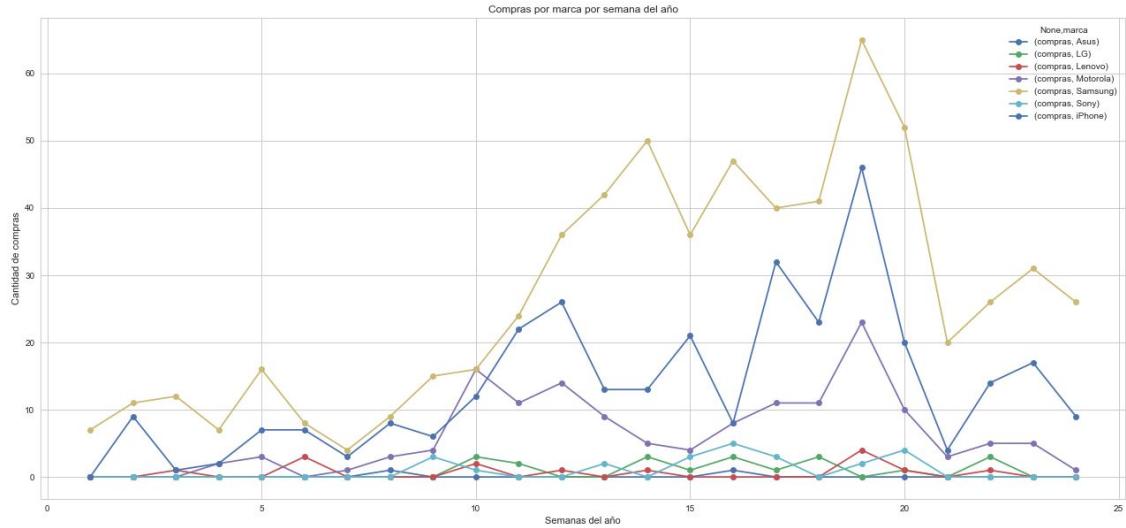
**Figura 9:** Top 5 de celulares según su almacenamiento.

Analizamos la cantidad de vistas de cada marca en cada semana del año.

**Figura 10:** Progreso de visitas de cada marca por semana del año.

Concluimos que a partir de la semana 20 se produjo una explosión de visitas a los productos de marca Apple y marca Samsung.

Por otra parte, analizaremos cuál es el progreso de compra de cada marca a lo largo del tiempo.

**Figura 11:** Progreso de ventas de cada marca por semana del año.

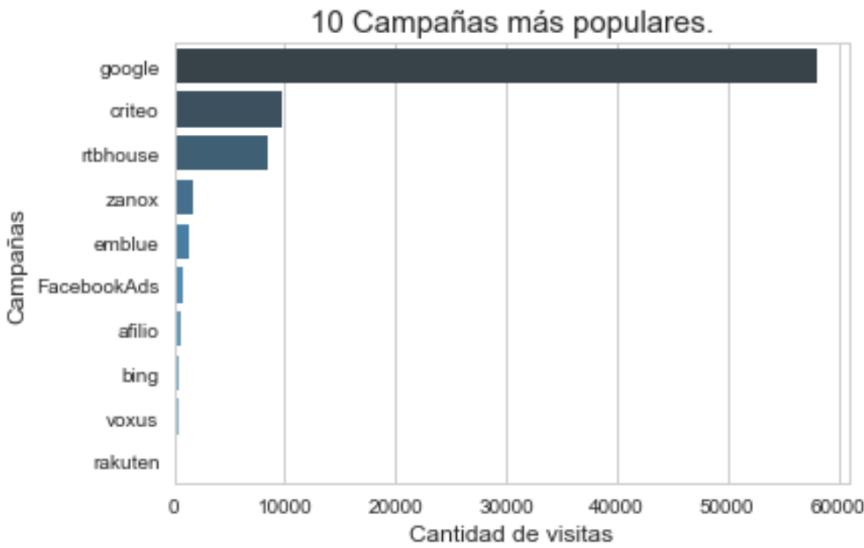
Podemos notar que si bien iPhone lidera las visitas del sitio a lo largo de las semanas, la mayor cantidad de compras las poseen los productos de Samsung.

2.3. Ad campaign Hit

Indica cuando un usuario se conecta a partir de una campaña publicitaria.

2.3.1. Análisis individual de las características

Comenzamos analizando este evento fijandonos cuál es la fuente que produce la mayor cantidad de visitas.

**Figura 12:** Las fuentes de campañas publicitarias que generaron la mayor cantidad de visitas.

Ad campaign hit se divide en tres tipos de accesos: página principal, ventas y compras. Obtuvimos que:

- El 65 % de los accesos es a compras.

- El 34 % de los accesos es a la página principal.
- El 0,26 % de los accesos es a ventas.

El acceso según ventas puede clasificarse según la marca. Obtuvimos lo siguiente:

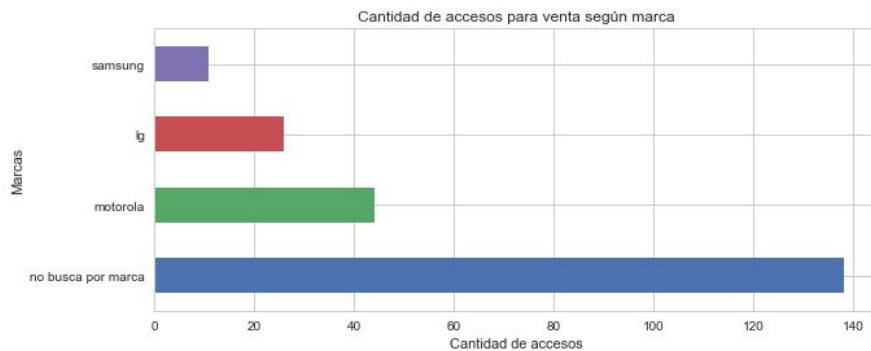


Figura 13: Cantidad de accesos para venta según la marca.

Este acceso tambien puede clasificarse según el modelo. Obtuvimos lo siguiente:

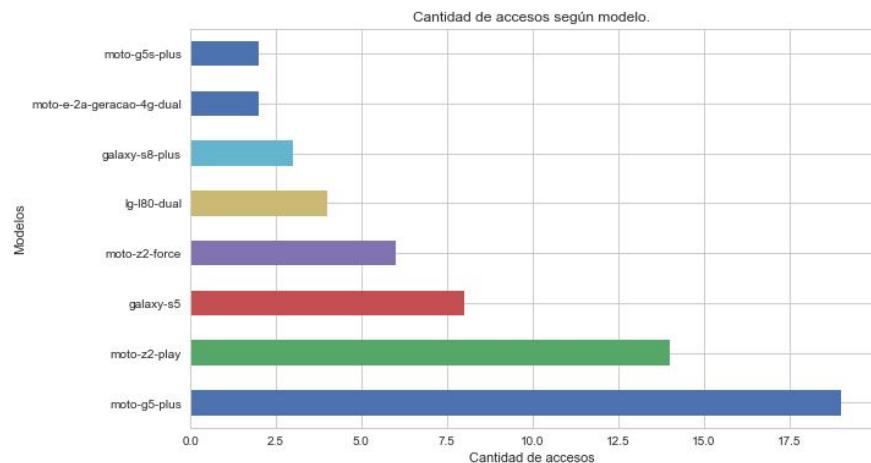


Figura 14: Cantidad de accesos para venta según el modelo.

El acceso según compras puede clasificarse según la marca. Obtuvimos lo siguiente:

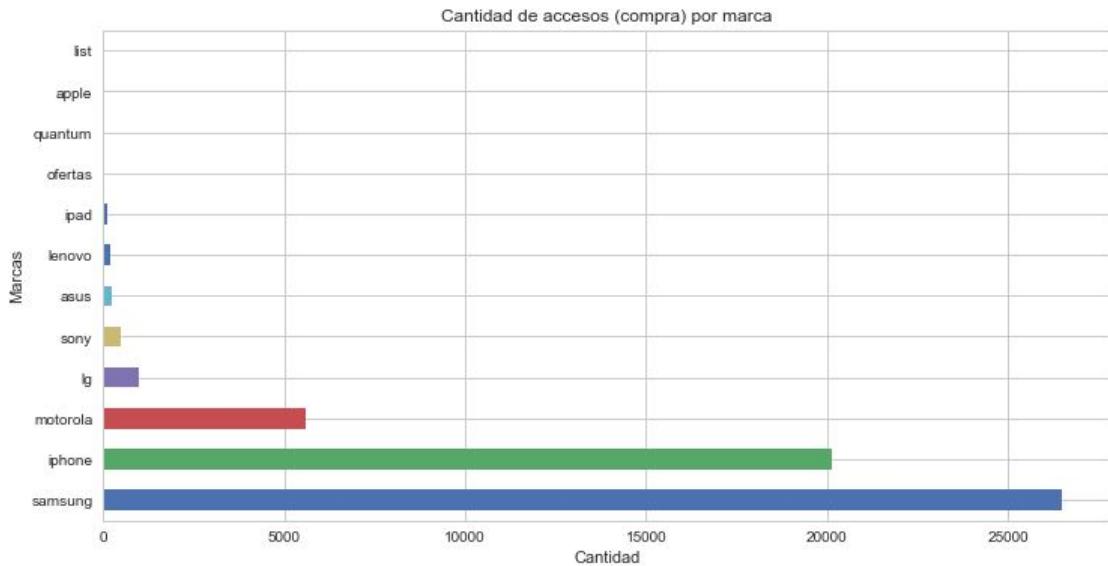


Figura 15: Cantidad de accesos para compra según la marca

El acceso de compras por oferta es poco significativo, por eso no lo excluimos del gráfico.
Hacer un análisis de las compras según modelo es muy complicado y no creemos que valga la pena.
Esto se debe a que un mismo modelo puede tener urls distintos.

2.3.2. Análisis temporal

Analizamos las cinco campañas publicitarias más clickeadas a lo largo de los meses. Lo obtenido es:



Figura 16: Cantidad de clicks según campaña por mes.

Analizamos así tambien el top5 de campañas publicitarias según la cantidad de clicks en cada día del año. Obtuvimos lo siguiente:

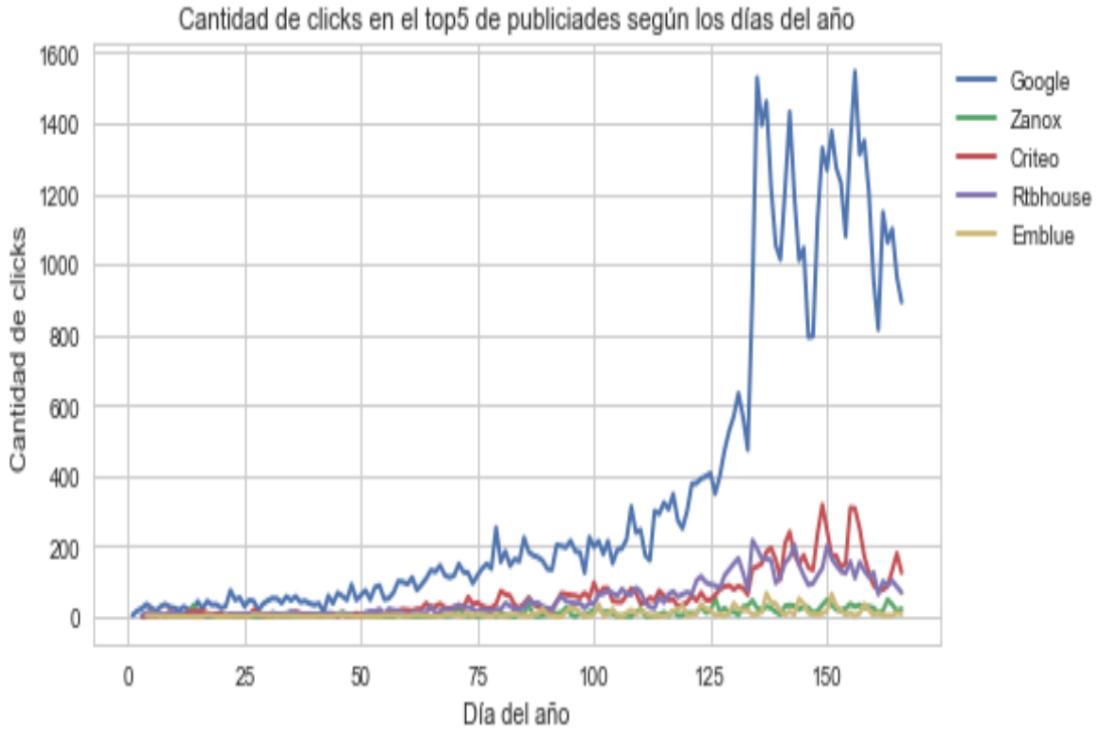


Figura 17: Cantidad de clicks según campaña por día del año.

Claramente se puede ver en los gráficos que existe una gran alza de clicks en las publicidades en los días del mes de mayo; además de que incrementan gradualmente desde enero. Esto coincide con la subida de visitas al sitio visto en el mismo mes.

2.4. Checkout

2.4.1. Análisis individual de las características

Empezamos analizando los 10 modelos que más llegaron al checkout. Ellos son:

- iPhone 6 con el 10 % de los checkouts.
- iPhone 5s con el 8 % de los checkouts.
- iPhone 6S con el 7 % de los checkouts.
- Samsung Galaxy J5 con el 6 % de los checkouts.
- Samsung Galaxy S7 con el 4 % de los checkouts.
- iPhone 7 con el 4 % de los checkouts.
- Samsung Galaxy S8 con el 3 % de los checkouts.
- iPhone 7 Plus con el 3 % de los checkouts.
- Samsung Galaxy J7 Prime con el 3 % de los checkouts.
- Samsung Galaxy S6 Flatcon el 3 % de los checkouts.

Analizamos la cantidad de checkouts seg\xfcren el almacenamiento. Lo obtenido es:

- El 37 % de los checkouts fue de dispositivos de 16GB
- El 29 % de los checkouts fue de dispositivos de 32GB
- El 16 % de los checkouts fue de dispositivos de 64GB
- El 11 % de los checkouts fue de dispositivos de 8GB
- El 5 % de los checkouts fue de dispositivos de 128GB
- El 1 % de los checkouts fue de dispositivos de 4GB
- El 1 % de los checkouts fue de dispositivos de 256GB
- El 0,1 % de los checkouts fue de dispositivos de 512MB

Analizamos la cantidad de checkouts seg\xfcren el color. Lo obtenido es:

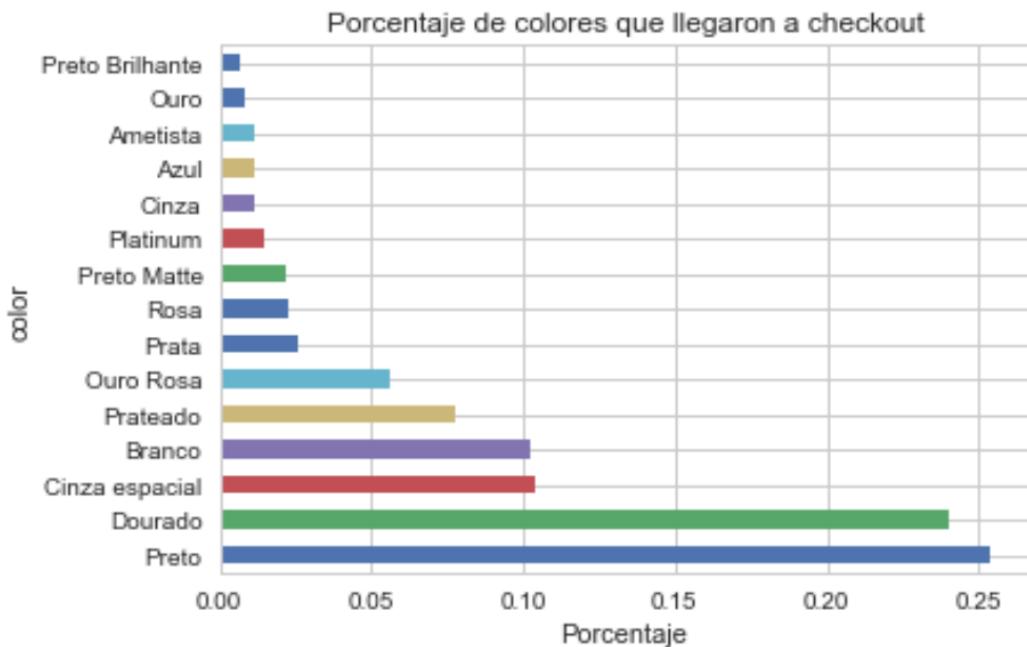


Figura 18: Porcentaje de colores que llegaron al checkout.

Por \xfaltimo analizamos la condici\xf3n de los productos que llegaron al checkout. Lo obtenido fue:

- El 45 % de los productos son de condici\xf3n buena.
- El 27 % de los productos son de condici\xf3n excelente.
- El 24 % de los productos son de condici\xf3n muy buena.
- El 3 % de los productos son de condici\xf3n buena con touch id.
- No es representativo la cantidad de productos nuevos.

2.4.2. Análisis temporal

Analizamos la cantidad de checkouts según los días del año. Obtuvimos lo siguiente:



Figura 19: Cantidad de checkouts según día del año.

Nuevamente se puede observar que esto coincide con el incremento de las visitas y los clicks en campañas publicitarias en los últimos dos meses.

2.5. Lead

2.5.1. Análisis Individual de las caracterísiticas

Analizamos cuáles son los modelos no disponibles con más pedidos de notificación de stock. El resultado es:

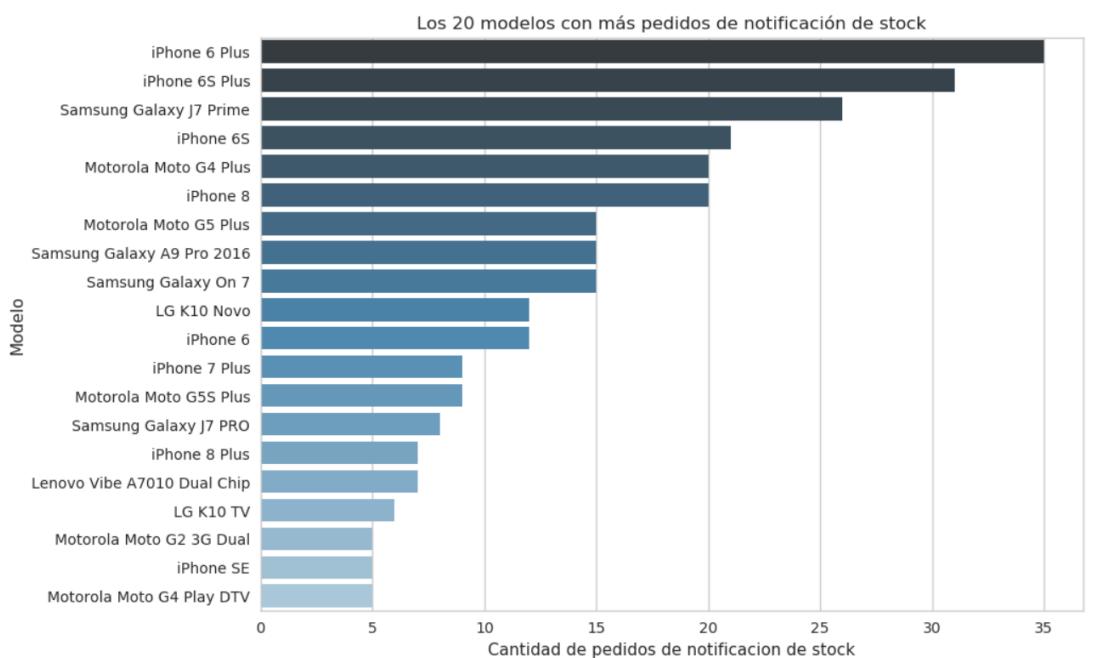


Figura 20: Los 20 modelos con más pedidos de notificación de stock.

En el gráfico se puede ver una gran predominancia de celulares de Apple, Samsung y, en menor cantidad, Motorola.

Al mirar la cantidad de compras y de Leads de cada usuario, obtenemos lo siguiente:

Existe una relación del 8% entre las veces que una persona agrega un determinado producto a Lead y este es comprado por la misma persona (ya sea por causa del Lead o por causa externa, no podemos determinarlo).

Sabemos que como mucho el 8% de las personas que agregaron un celular al Lead terminaron comprando este producto luego de haber recibido la correspondiente notificación. A pesar de que el porcentaje de conversiones en base al lead es bajo, es un factor importante a tener en cuenta ya que puede indicar cuánto interés tienen los usuarios por comprar un producto cuando este no se encuentra en stock. O puede indicar que tan útil o eficiente es este tipo de evento.

2.6. Visited site

Visitas recibidas en la página principal.

2.6.1. Análisis individual de las características

Comenzaremos por analizar cuáles son los países desde donde se efectúan las visitas, los más importantes son:

- Brasil con el 96% de las visitas.
- Unknown con el 3% de las visitas.
- United States con el 1% de las visitas

Como Unknown representa un porcentaje tan pequeño decidimos omitirlo, de igual manera con el resto de los países.

Como Brasil es el país predominante nos focalizaremos en analizar su comportamiento.

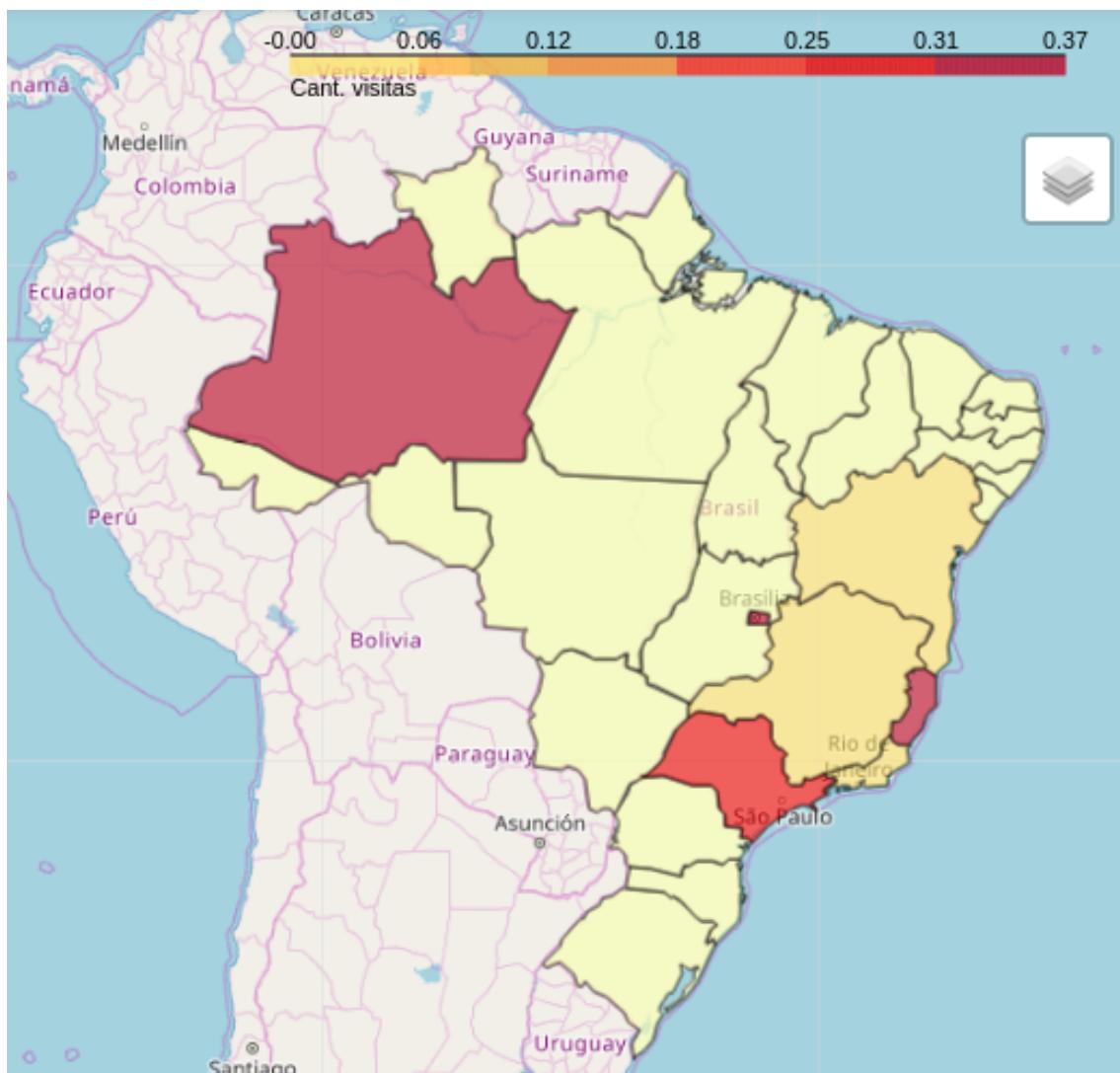


Figura 21: Cantidad de vistas según estado de Brasil

Podemos ver que la mayor cantidad de visitas se realizan desde Sao Paulo, Minas Gerais y Rio de Janeiro .

Analizaremos ahora las ciudades desde las que se produjeron la mayor cantidad de visitas.

Visitas	
Ciudad	
São Paulo	11711
Rio de Janeiro	3538
Belo Horizonte	2568
Salvador	2314
Brasília	1530
Osasco	1148
Curitiba	1132
Fortaleza	1084
Recife	913
Goiânia	853

Como podemos observar, São Paulo es la ciudad que posee la mayor cantidad de visitas. Visualizaremos la relación entre las visitas de las ciudades a partir del siguiente gráfico.

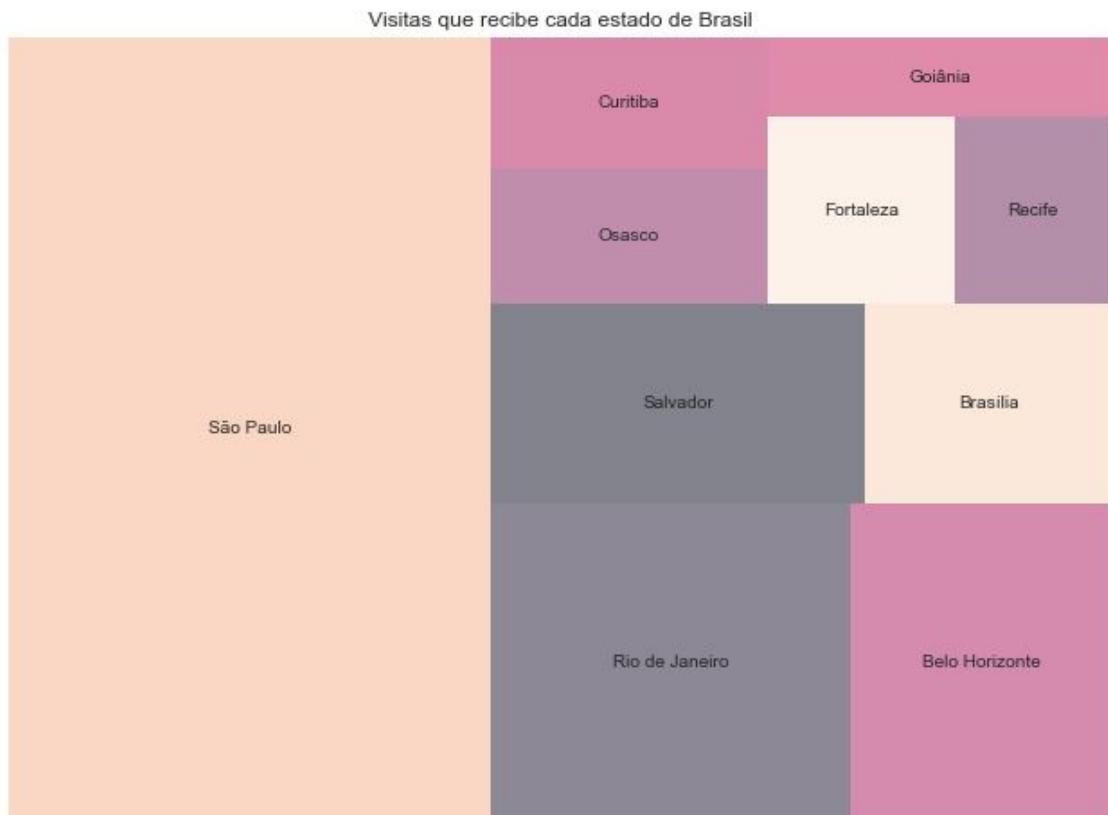


Figura 22: Cantidad de vistas según ciudad de Brasil

Por otra parte, procederemos por analizar los canales por los que se realizan las visitas. Los canales pueden ser:

- Display: Interacciones con un medio de *display* o *cpm*. También incluye las interacciones de Google Ads con la red de distribución de anuncios configuradas como *content*.
- Paid Search: Tráfico desde la red de motores de búsqueda, con un medio de *cpc* o *ppc*.
- Other: Sesiones etiquetadas con un medio de *cpc*, *ppc*, *cpm*, *cpv*, *cpa*, *cpp* o *affiliate* (excluida la publicidad en buscadores).
- Organic Search: Tráfico de búsqueda gratuita en cualquier motor de búsqueda.
- Social Network: Tráfico de cualquiera de las aproximadamente 400 redes sociales (que no están etiquetadas como anuncios).
- Referral: Tráfico de sitios web que no son redes sociales.
- Email: Sesiones que están etiquetadas con el medio de *email*.
- Direct: Sesiones en las que el usuario ha escrito la URL del sitio web en el navegador o ha llegado al sitio a través de un marcador.

Analizamos la cantidad de visitas por canal. El resultado es:

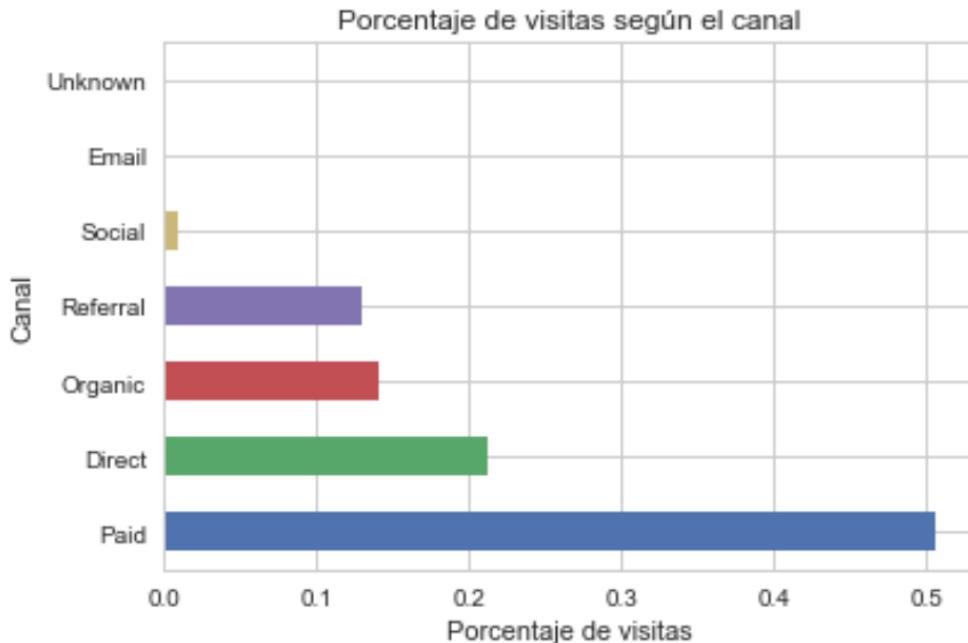


Figura 23: Porcentaje de visitas según el canal

Podemos ver que el canal predominante es paid seguido por direct.

De este evento podemos obtener información acerca de cuales son los usuarios nuevos y cuales volvieron. La información obtenida es:



Figura 24: Porcentaje de usuarios nuevos respecto de usuarios que volvieron.

Podemos ver que el 69 % de la actividad registrada en el sitio proviene de retornos y el 31 % de los usuarios nuevos.

Para analizar el retorno de los usuarios a la página más detalladamente, pudimos obtener la cantidad de usuarios que hubo por número de regresos al sitio. Es decir, mirando el gráfico: existieron y cantidad de usuarios que regresaron x cantidad de veces a la página.

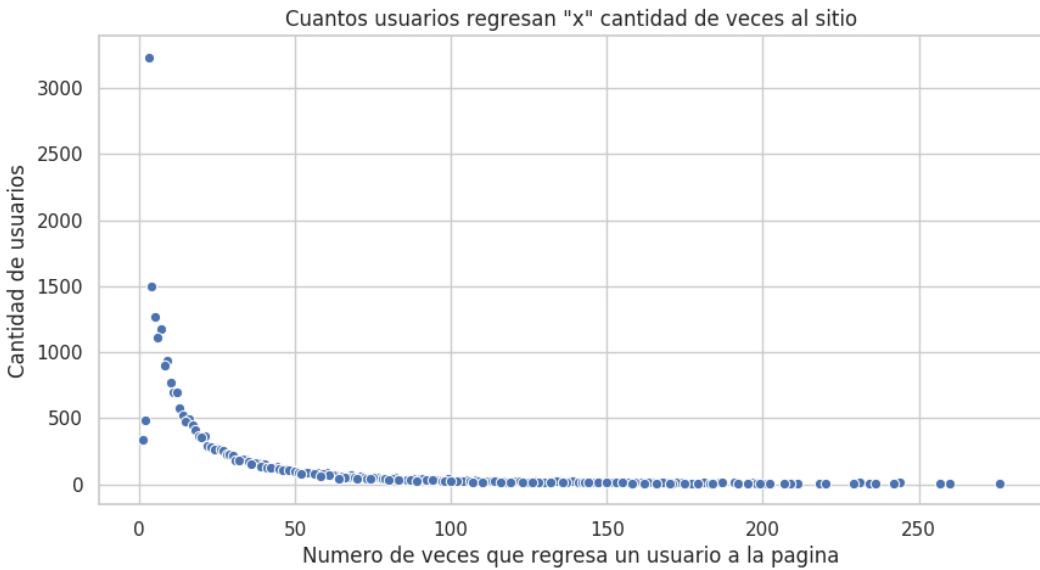


Figura 25: Cantidad de usuarios por número de regresos a la página

Se puede ver que gran parte de los usuarios están más a la izquierda del gráfico. Es decir, el número de regresos a la página es menor. La cantidad de usuarios disminuye a medida que se mira más adelante en el eje horizontal (a más cantidad de regresos al sitio).

Es interesante ver que existe un pico que se destaca en el número 3 (usuarios que regresan 3 veces) y ver que hay más usuarios que regresan entre 3 y 14 veces que los que regresan 1 o 2 veces. Aquí se muestra un gráfico de barras mostrando lo último descrito.

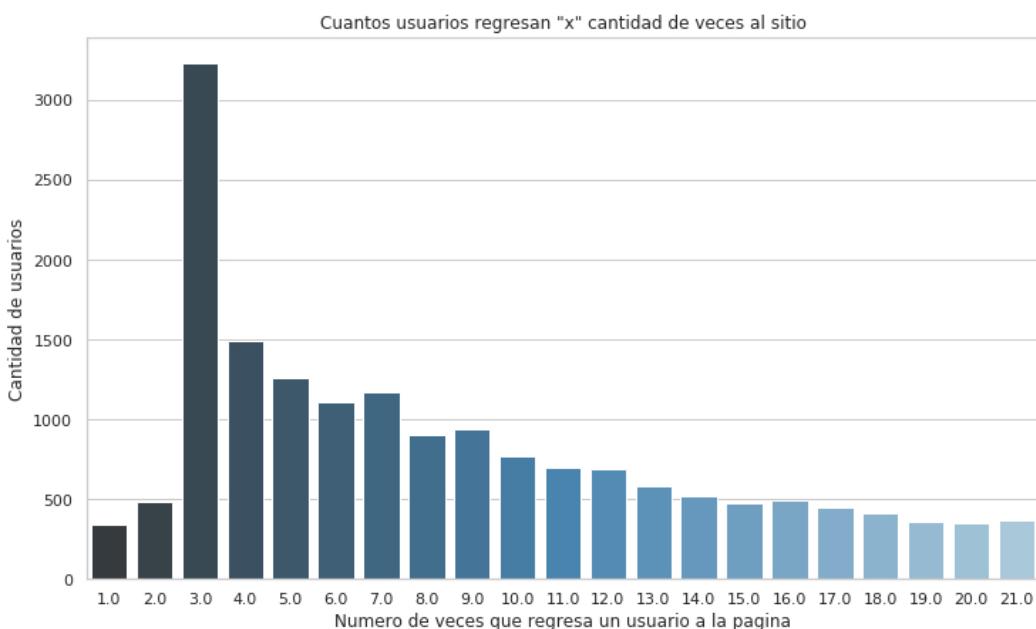


Figura 26: Cantidad de usuarios por número de regresos a la página

Tratamiento de datos nulos:

Para análisis anterior, decidimos tratar los datos nulos de la siguiente manera: Seleccionamos a todos los usuarios que alguna vez tuvieron un registro de New en la columna newvsreturning (cuando entraron a la pagina por primera vez). Luego, a todos los datos nulos en esa columna de los usuarios seleccionados los consideramos como Returning. Es decir, como un regreso a la página.

Por otra parte, analizando los dispositivos desde el cual acceden los usuarios, obtuvimos que:

- El 50 % accede desde su teléfono
- El 47 % accede desde su computadora
- El 2 % no se sabe.
- El 1 % accede desde su tablet

Procederemos por analizar las visitas desde los 15 sistemas operativos más populares.

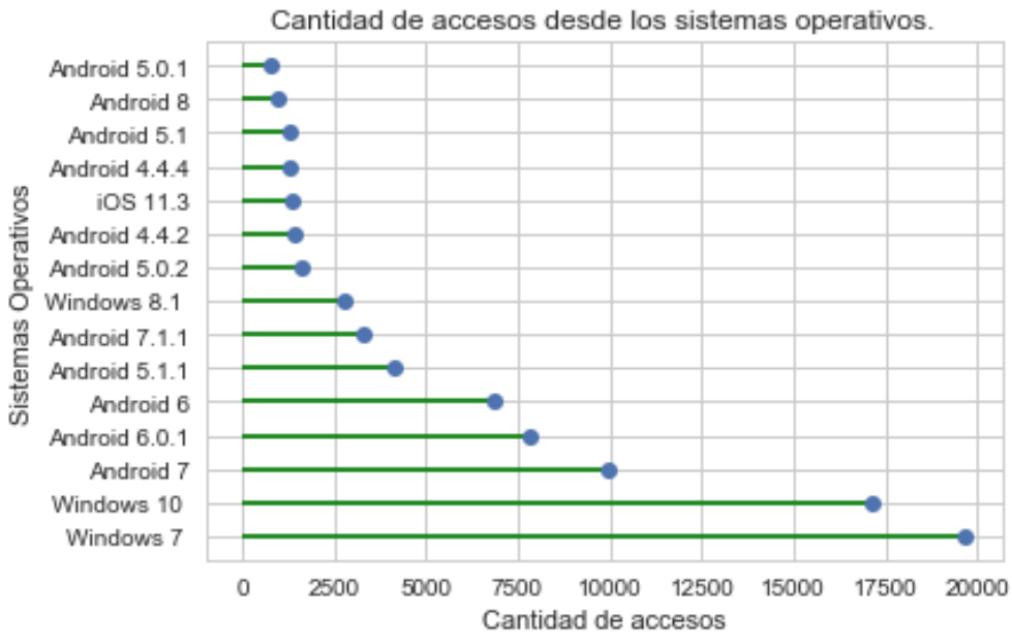


Figura 27: Cantidad de accesos por sistema operativo. Se muestran los 15 sistemas operativos más populares.

2.7. Generic listing

Visitas a la página principal

2.7.1. Análisis individual de las características

Analizaremos en primer lugar cuáles son los productos que reciben mayor cantidad de visitas en la página principal.

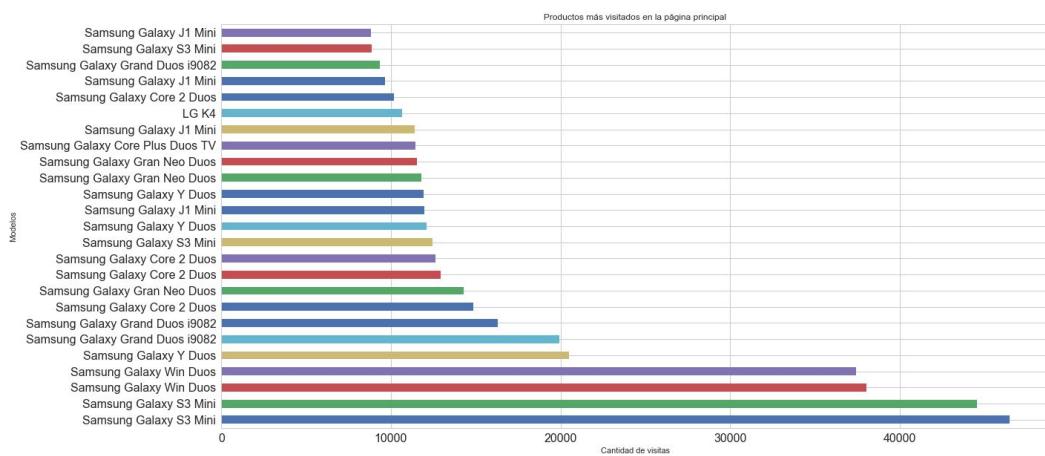


Figura 28: Cantidad de usuarios por número de regresos a la página

Es evidente una clara superioridad por parte de los productos de Samsung en visitas a la página principal.

2.8. Searched product

Búsqueda de productos

2.8.1. Análisis individual de las características

Decidimos analizar qué es lo que más buscan los usuarios por palabras clave. El resultado fue el siguiente:



Figura 29: Las palabras que más buscan los usuarios.

Las palabras que aparecen más grande son aquellas que más se buscan.
Por otro lado, analizaremos puntualmente cuáles son los productos más buscados

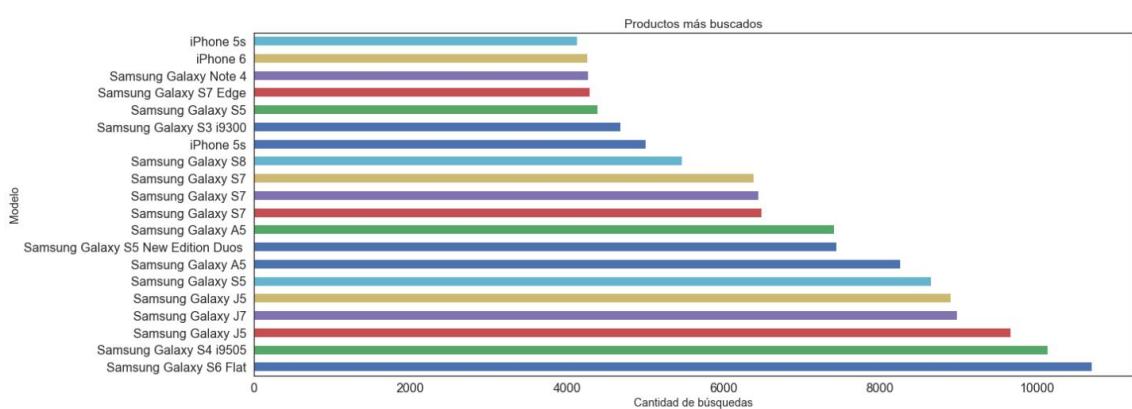


Figura 30: Los productos que más buscan los usuarios.

Al igual que ocurre con las visitas hacia la página principal, podemos notar una clara superioridad por parte de Samsung a la hora de buscar modelos.

2.9. Search engine hit

Ingresos al sitio a través de motores de búsqueda

2.9.1. Análisis entre eventos

Analizamos la cantidad de apariciones de los buscadores. El resultado fue que el 99 % de las búsquedas se realizan desde Google, entonces, el impacto de los demás buscadores se los puede considerar como despreciable.

2.9.2. Análisis temporal

Analizamos los ingresos a través de un motor de búsqueda por día del año. El resultado fue:

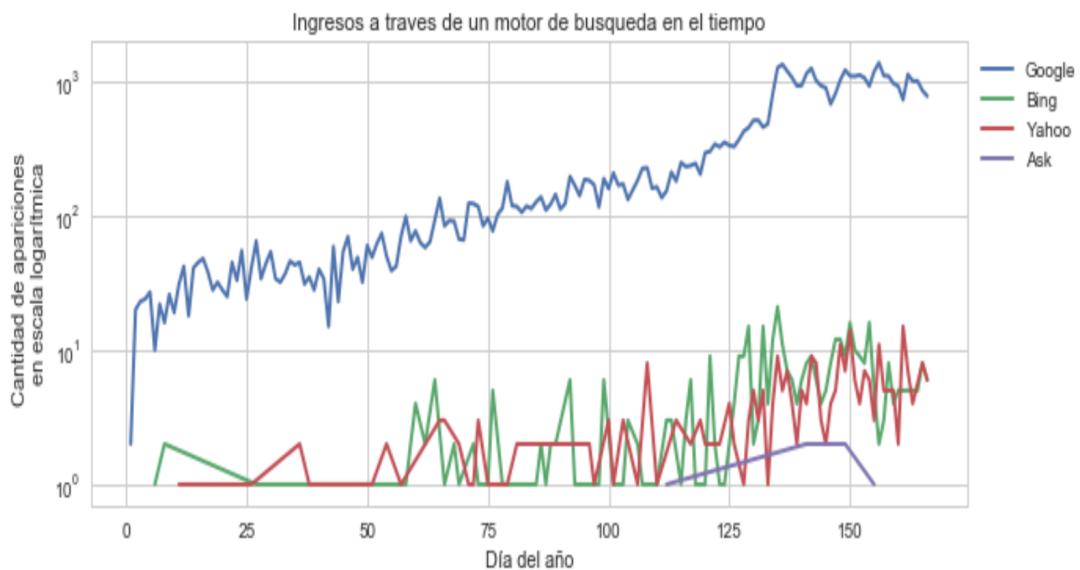


Figura 31: Ingresos a través de los motores de búsqueda a través del tiempo.

Es importante mencionar que elegimos realizar el gráfico en escala logarítmica ya que la diferencia de visitas entre los buscadores es demasiado grande y no es posible apreciar la comparación a escala.

Podemos observar que Google es el único buscador que posee un crecimiento estable a lo largo del tiempo, mientras que el comportamiento del resto de los buscadores es irregular.

2.10. Static page

Luego de analizar este evento nos dimos cuenta que no aporta información relevante en ninguno de sus campos. Las páginas estáticas poseen un volumen de visitas muy pequeño y no pudo encontrarse correlación en sus datos.

2.11. Conversion

Productos comprados por el usuario

2.11.1. Análisis individual entre las características

Analizamos los productos más comprados. Los 10 más comprados son:

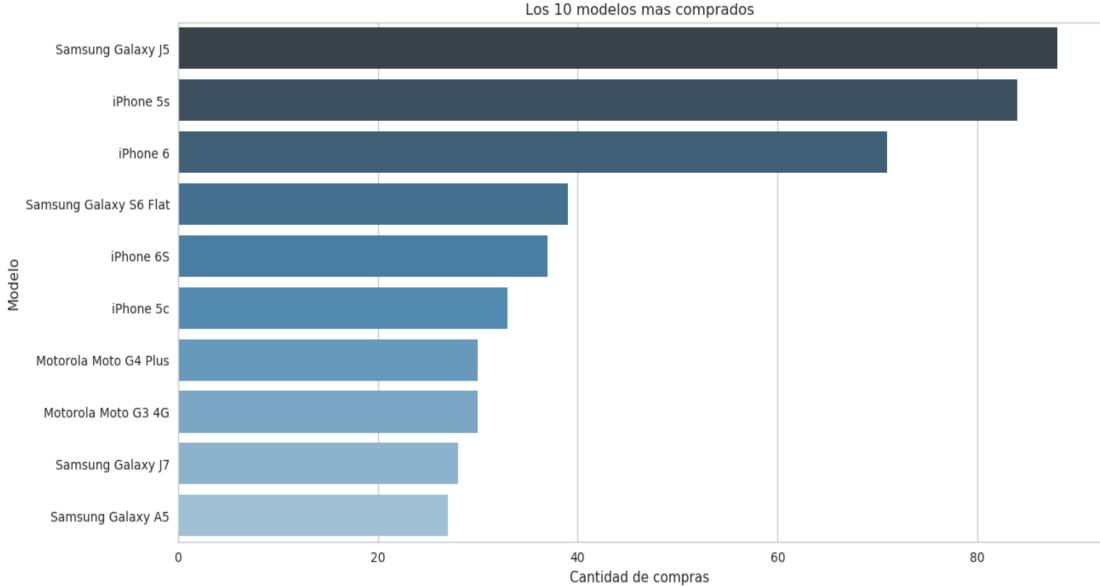


Figura 32: Cantidad de compras según modelo. Se muestran los 10 modelos más populares.

Los tres más comprados que se pueden ver son: Samsung Galaxy J5, Iphone 5S y Iphone 6. Esos son los que más se destacan por encima del resto. Luego no hay brechas muy grandes de compras entre distintos productos. Algo que tambien es posible destacar es que, de los diez más comprados, cuatro de ellos pertenecen a la familia de los Samnsung Galaxy, cuatro son de la marca Apple, y dos pertenecen a la familia de los Motorola Moto G

Analizaremos la condición de los productos comprados:

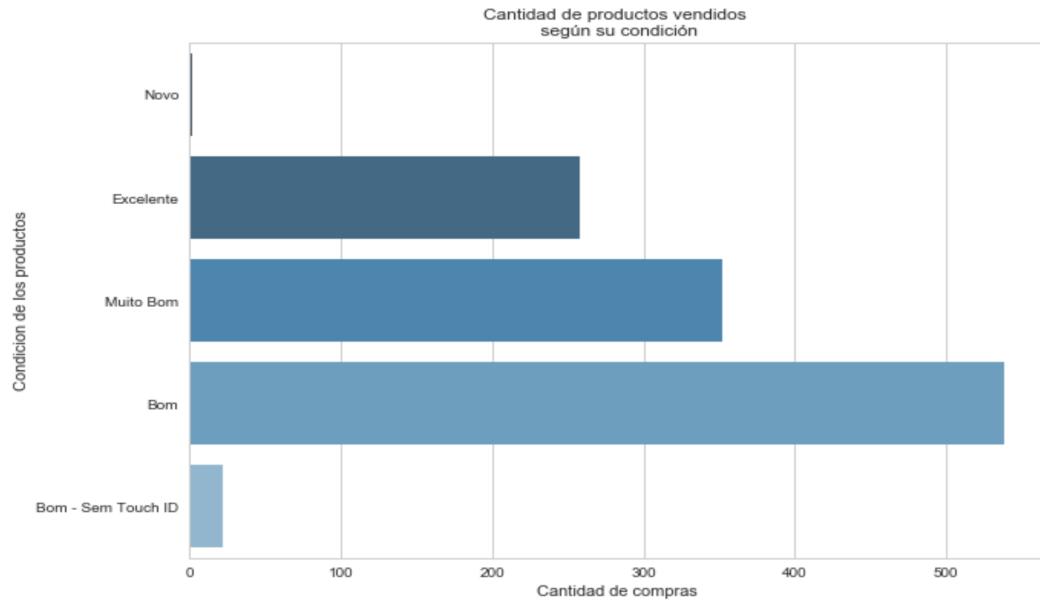


Figura 33: Cantidad de productos comprados según su condición.

Aquí se puede observar que los que más compras tienen son los que tienen, relativamente, peor condición (condición: Bien). Esto puede llevar a pensar que gran parte de los usuarios al hacer

una compra no priorizan la condición del producto. Aunque cabe destacar que se debería tener en cuenta cuantos productos de cada condición posee la pagina.

3. Análisis entre eventos

3.1. Visitas por campaña publicitaria VS conversiones a lo largo del tiempo

En esta sección analizaremos la relación entre eventos. Se buscará una relación entre los *clicks* de las campañas publicitarias con las visitas y las compras de los productos. Lo obtenido fue lo siguiente:

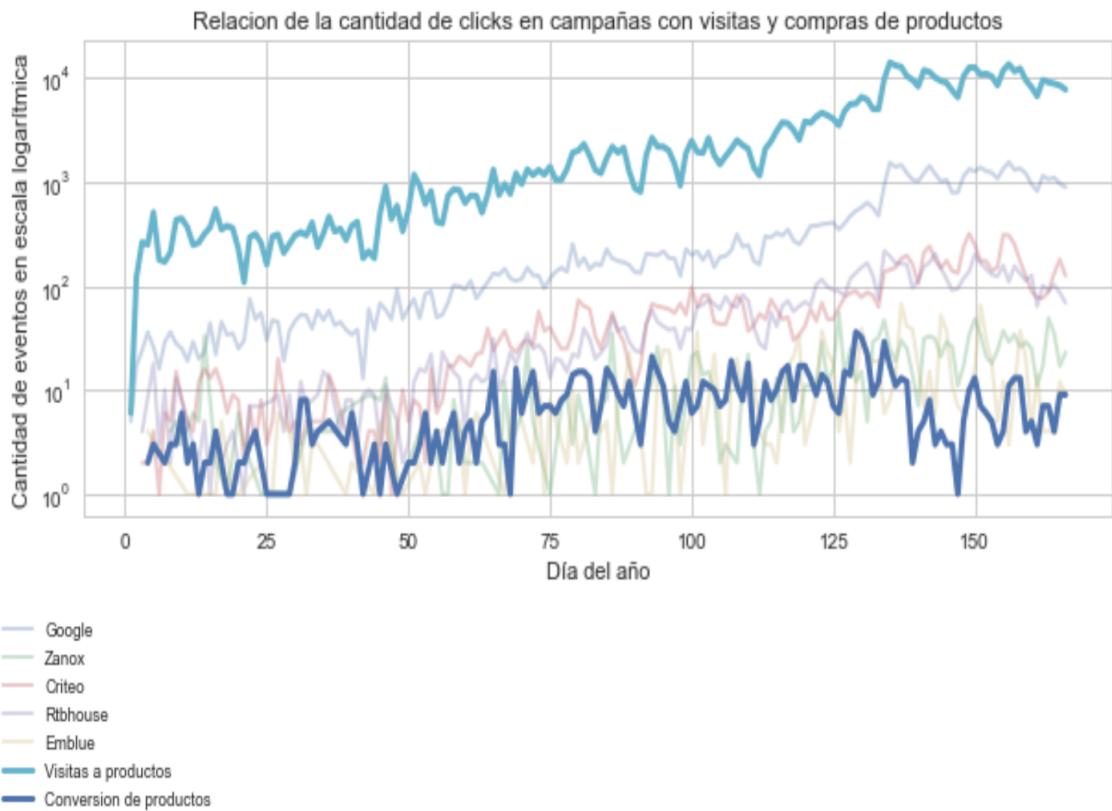


Figura 34: Relación entre los clicks en las campañas publicitarias con las visitas y compras de los productos .

Podemos ver que a medida que aumentan los clicks en las campañas publicitarias, también lo hacen las visitas a productos, lo cual es lógico. Algo quizás no tan esperado es que aunque aumentan las visitas y los clicks en las publicidades, las conversiones no parecen aumentar. En la primer mitad podemos ver que era más baja la cantidad de conversiones, pero luego distinguimos que estas se mantienen aproximadamente constantes a pesar del incremento en las visitas.

3.2. Relación entre número de compras y visitas para un mismo usuario

Un posible análisis para esta sección es tomar cada usuario que entró alguna vez al sitio y ver cuantas compras y cuantas visitas a productos posee.

Para eso realizamos un scatter plot en donde cada punto en el plano es un cliente que posee una cantidad total de visitas al sitio (eje *y*) y una cantidad total de compras (eje *x*).

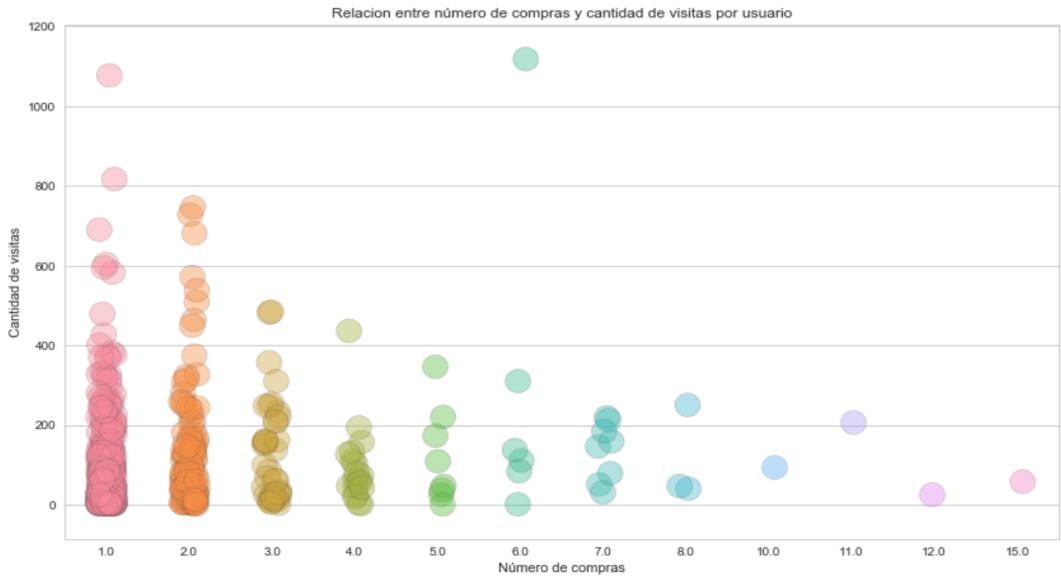


Figura 35: Relación entre el número de compras y cantidad de visitas por usuarios.

En el gráfico claramente se puede observar que la mayoría de los usuarios hace 1 o 2 compras como mucho y visitan la pagina una cantidad relativamente baja de veces. Tambien es posible observar que, en general, los compradores que hacen más visitas son usuarios que hacen pocas compras. A mayor cantidad de compras, se ven menos cantidad de usuarios y, ademas, son usuarios que visitan pocas veces el sitio en comparación a los otros compradores.

3.3. Análisis sobre las campañas de Marketing

3.3.1. Análisis general

Un punto importante que podemos analizar es : ¿Qué campañas de marketing se las puede considerar más *eficientes*?

Para intentar responder a esta pregunta, vimos cuantos usuarios entran por cada campaña y de éstos filtramos a los usuarios que compraron uno o más productos alguna vez. Así obtuvimos a la cantidad de compradores que ingresan al sitio por campaña.

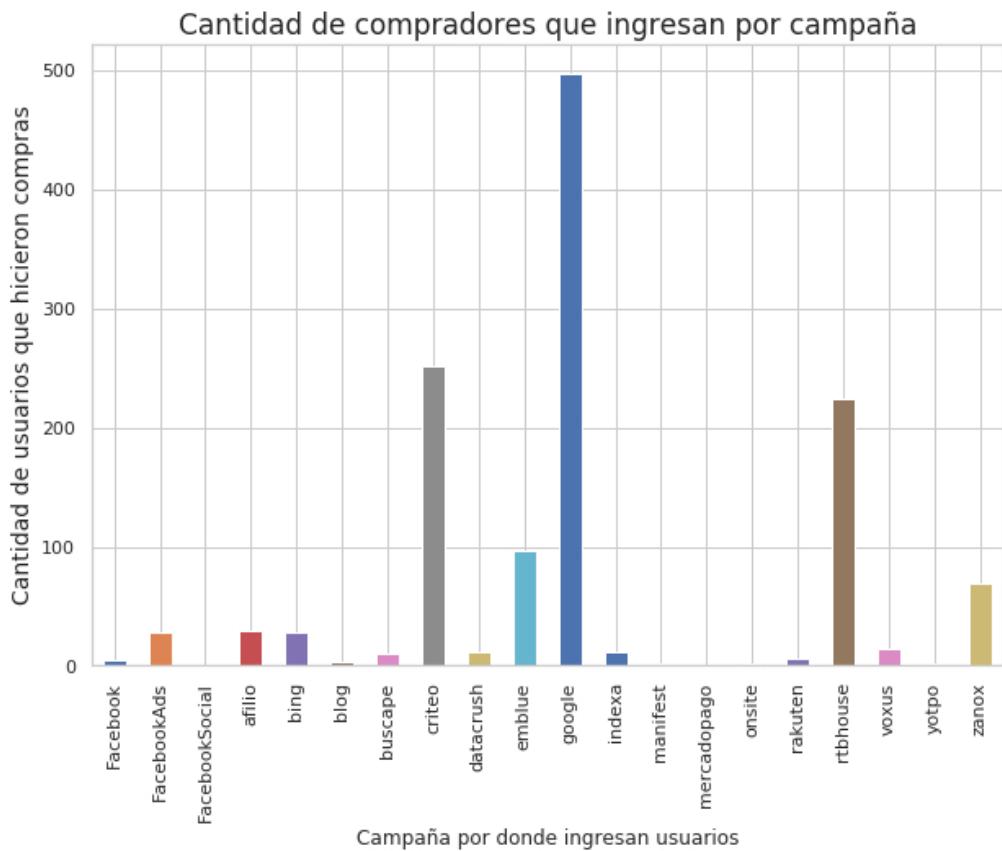


Figura: 36 Cantidad de compradores que ingresan por x campaña de marketing

Aquí se puede ver qué cantidad de usuarios que ingresaron alguna vez al sitio por una campaña de marketing hicieron una (o más) compras. Esto no quiere decir que estos mismos usuarios entraron al sitio mediante la campaña y realizaron una compra inmediatamente. Sólo se tomaron en cuenta los que hicieron alguna compra, sin restricciones temporales. Además, un usuario puede haber entrado al sitio por más de una campaña a lo largo del año.

Como se ve en el gráfico, la gran mayoría ingresa por la campaña de Google. Los dos siguientes más destacados son Criteo y Rtbhouse.

Para hacer una mejor comparación, obtuvimos la cantidad total de usuarios que ingresan por cada campaña y, con los datos del gráfico anterior, pudimos calcular qué porcentaje del total de éstos que ingresa por cada campaña son compradores.

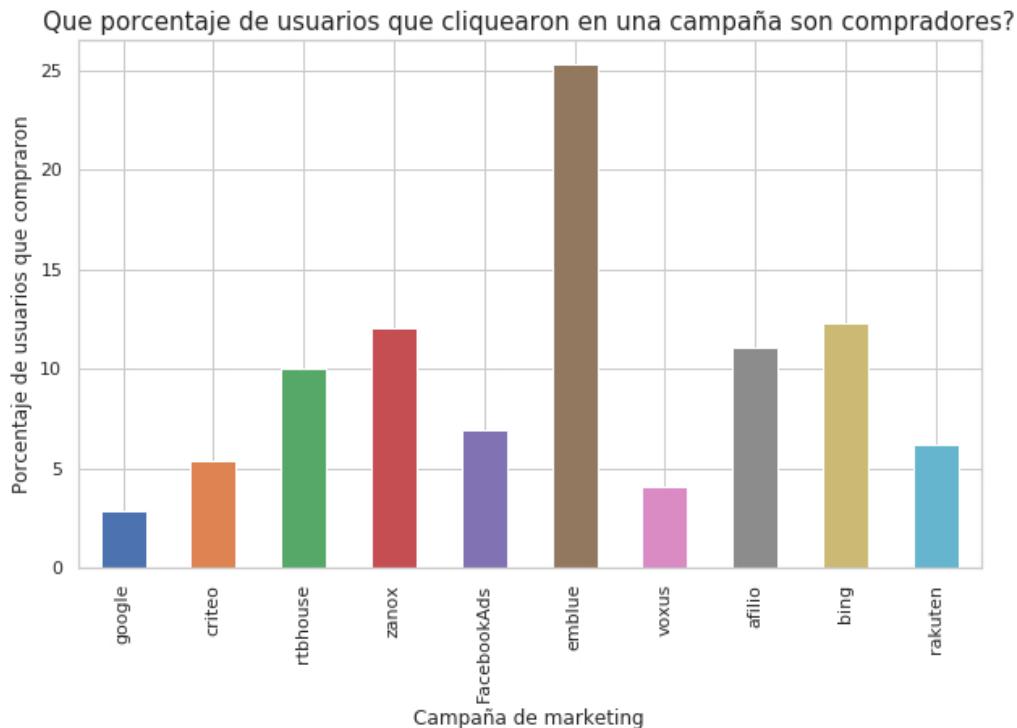


Figura 37: Porcentaje de usuarios que ingresan por cada campaña de marketing que son compradores

Los datos que parecen ser más relevantes son los siguientes:

- Campaña *emblue* con un 25.26 porciento de compradores (97 de 348)
- Campaña *bing* con un 12.28 porciento de compradores (28 de 228)
- Campaña *zanox* con un 12.0 porciento de compradores (69 de 572)
- Campaña *rtbhouse* con un 9.99 porciento de compradores (224 de 2241)
- Campaña *google* con un 2.86 porciento de compradores (497 de 17372)

Cabe destacar que, a pesar de que la campaña de marketing de *Google* en este gráfico es la de menos porcentaje de compradores, es la campaña en la cual se registró la mayor cantidad de compradores (497) y mayor tráfico de usuarios (17372 usuarios distintos).

Algo a resaltar es que la campaña de marketing de *emblue* es la que más se destaca del resto en el último gráfico. Es la sexta campaña con más usuarios que cliquearon (348) y la que mejor porcentaje de compradores tiene (25.26 %). Es decir: 97 de 348), siendo la cuarta con mas cantidad de compradores.

3.3.2. Modelos más comprados por usuarios que cliquearon una determinada campaña

La pregunta que nos motivó en el análisis de esta sección fue la siguiente: ¿Se comportan de forma distinta los usuarios en cuanto a las compras de los diferentes modelos dependiendo de la campaña por la cual ingresan?

Para eso, vimos cuantos usuarios entran por cada campaña. De estos filtramos a los usuarios que compraron uno o más productos alguna vez y seleccionamos esos productos, obteniendo la cantidad de usuarios que compraron cada producto.

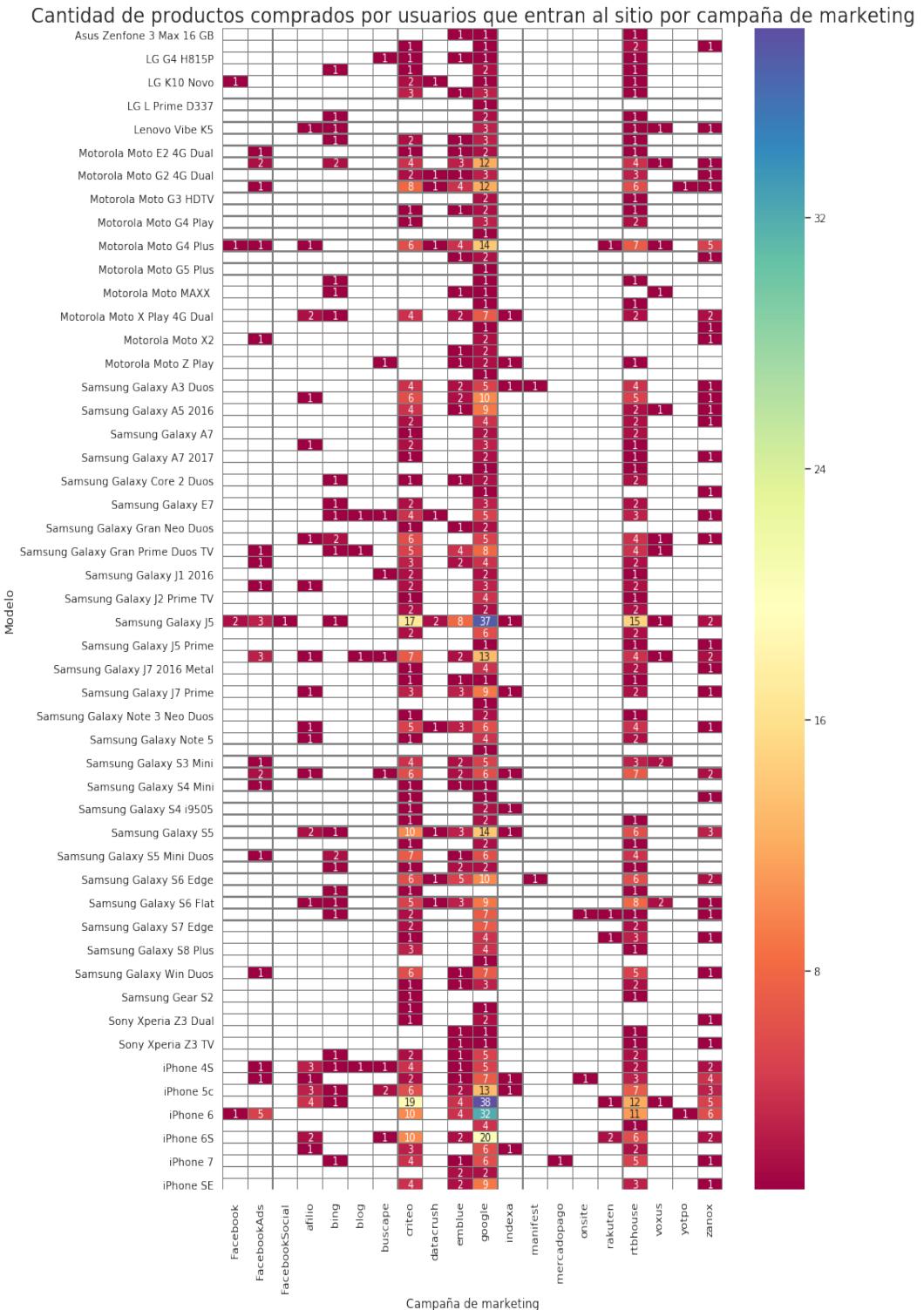


Figura 38: Cantidad de modelos comprados por x campaña de marketing que son compradores.

Hacemos una restricción y miramos las tres campañas de marketing más relevantes:

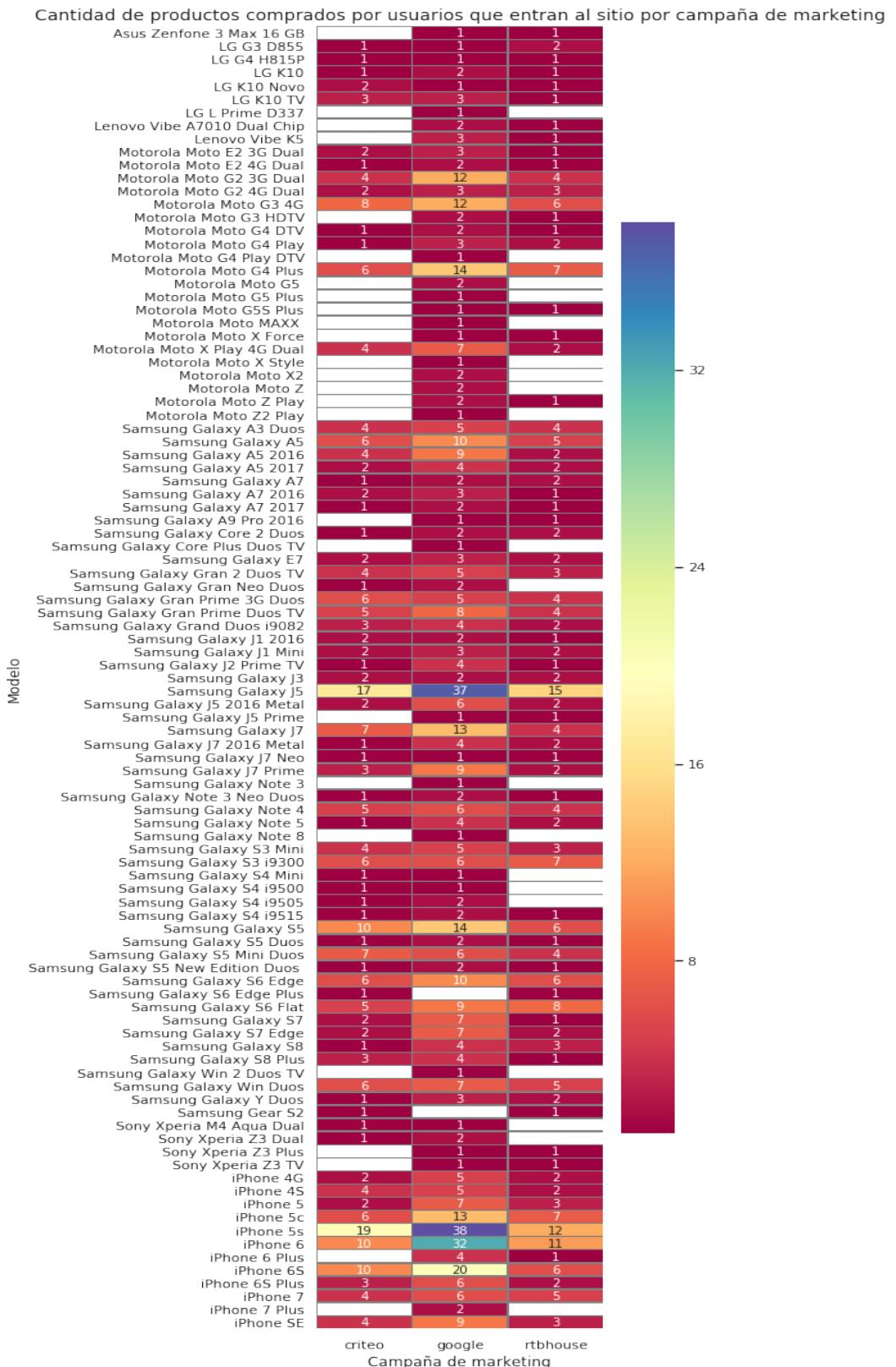


Figura 39: Cantidad de modelos comprados por x campaña de marketing que son compradores

Podemos ver solo los productos que estan en todas las filas:

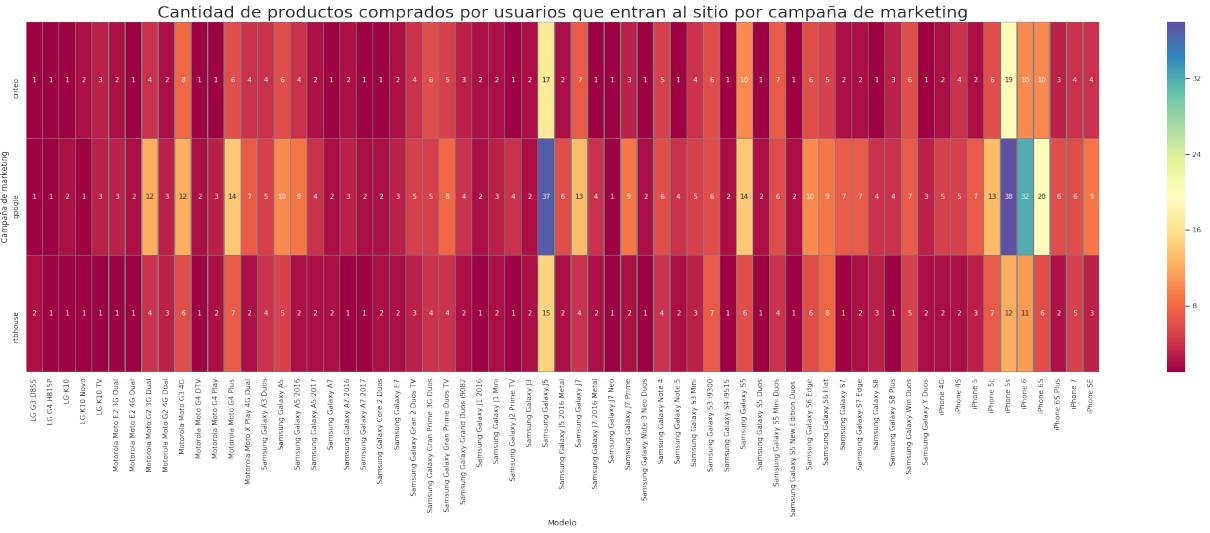


Figura 40: Cantidad de modelos comprados por x campaña de marketing que son compradores

Observando estos graficos pudimos concluir que no es posible destacar un comportamiento de usuarios haciendo una division por campaña de marketing. Más allá de la diferencia de cantidades de compradores entre campañas, la tendencia a comprar determinados productos, en general, es la misma. Por esto no se puede decir que los usuarios que entran al sitio por una campaña específica tienen distinto comportamiento a los usuarios que entran mediante otras campañas (respecto a las compras de los diferentes productos).

Esto puede relacionarse con el grafico de tiempo que muestra las compras y las entradas por distintas campañas.

3.4. Análisis de dispositivos por el que se conectan los usuarios

El 98% de los usuarios que han ingresado al sitio lo han hecho con un mismo tipo de dispositivo. Siendo que el 2% ha ingresado al sitio con varios tipos de dispositivo (Computadora, Smartphone, tablet, etc). Despreciando este 2%, podemos decir que los usuarios que ingresan al sitio lo hacen utilizando siempre el mismo tipo de dispositivo.

Partiendo de esta premisa, podemos realizar los siguientes análisis.

3.4.1. Análisis de sistemas operativos

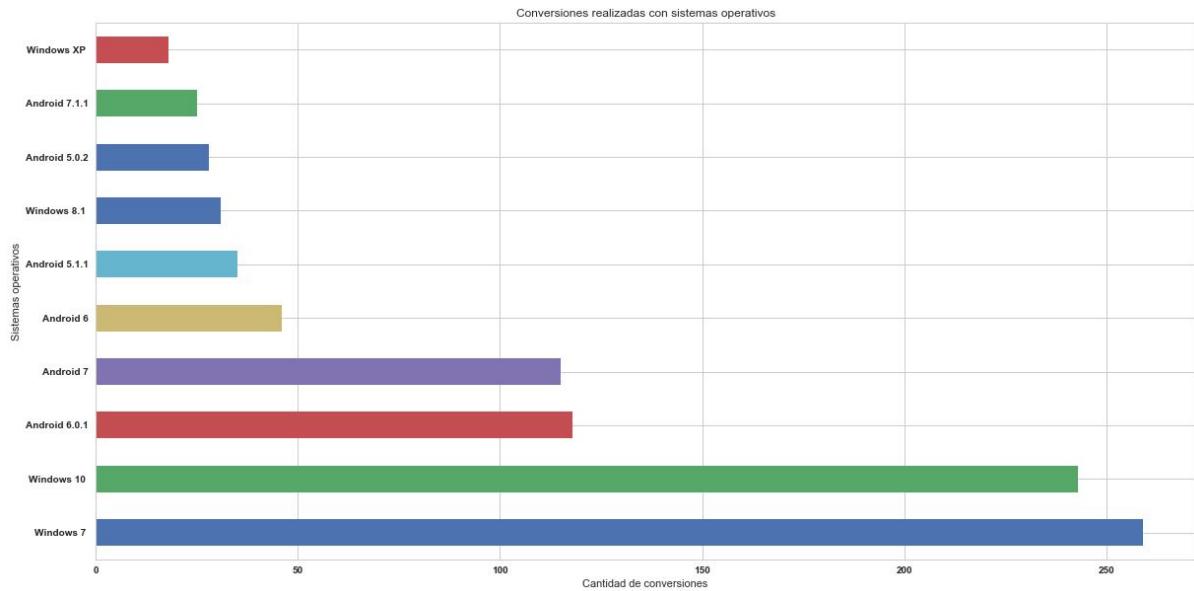


Figura 41: Cantidad de modelos comprados por x campaña de marketing que son compradores

Podemos observar que la mayoría de las conversiones se realizan desde sistemas operativos Windows 7 y Windows 10 (Computadoras).

Observación Si bien el sistema operativo más usado para visualizar productos es Windows, la mayor cantidad de visitas en total se producen a través de Smartphones, se supone que esto se dará a causa de la gran cantidad de sistemas operativos que hay en mobile, lo que provoca que la categoría Smartphones se disperse.

Intentaremos encontrar un patrón en base al sistema operativo por el cual acceden los usuarios y los productos más visitados y convertidos.

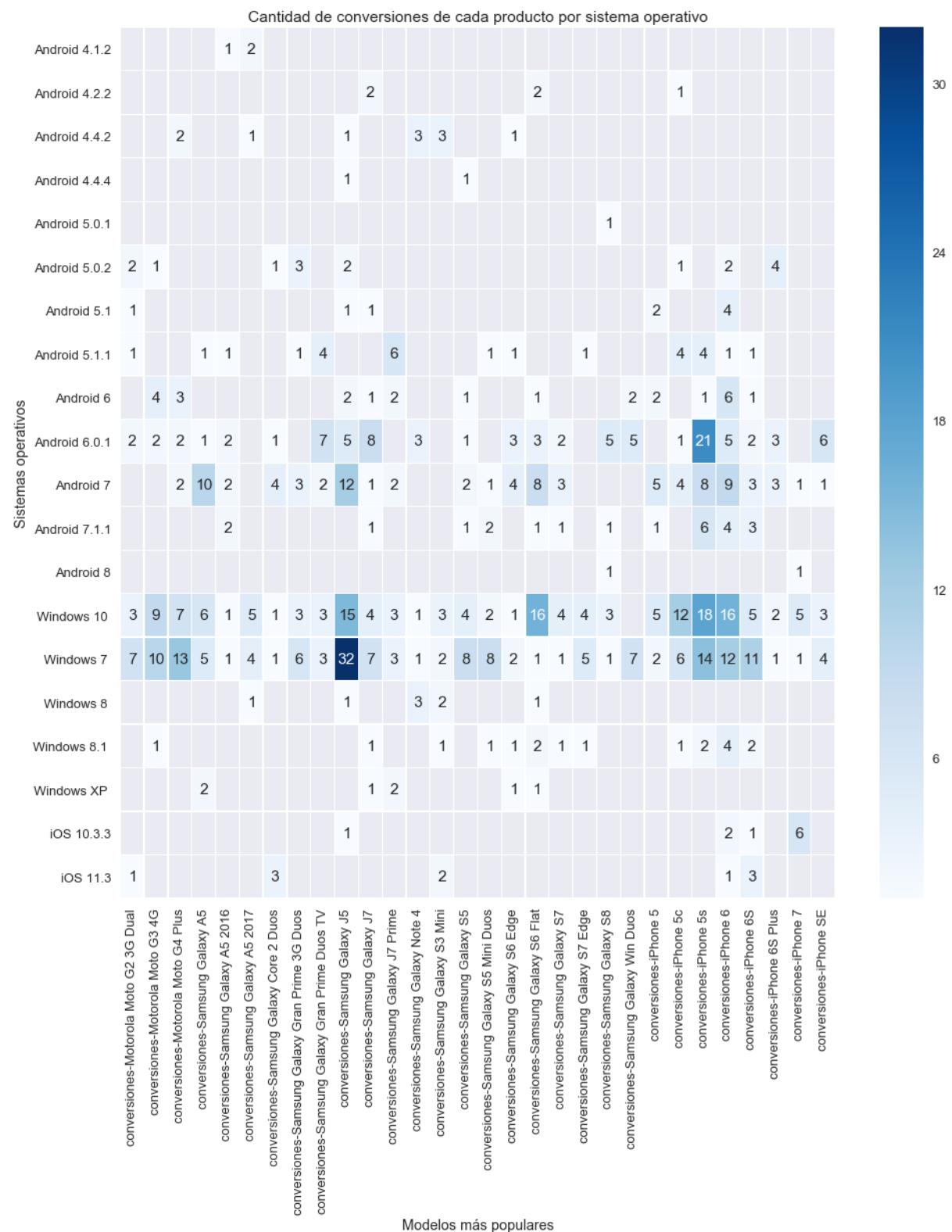
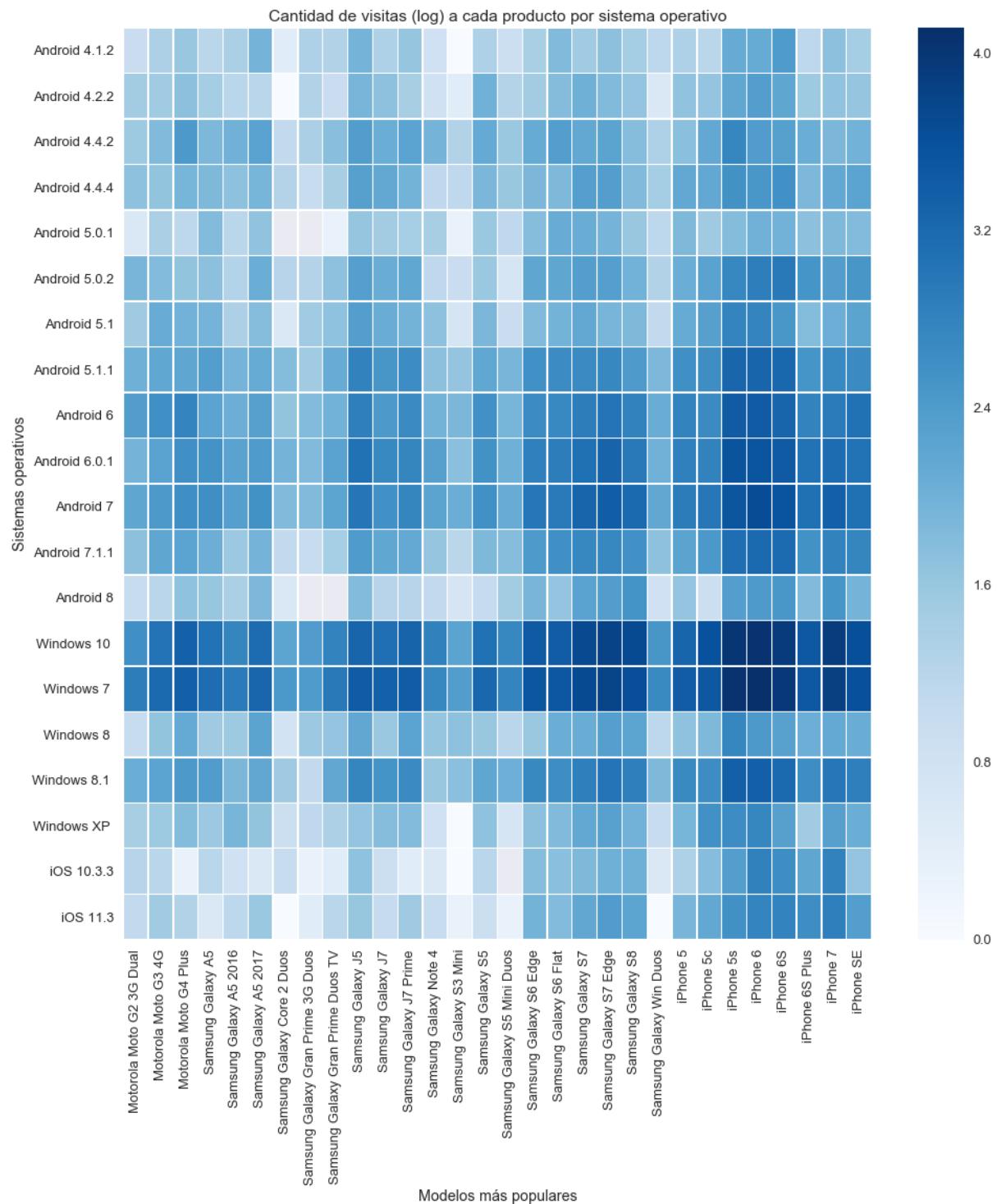


Figura 43: Cantidad de conversiones de cada producto por sistema operativo

En este caso no establecer ningún comportamiento ni patrón de compra. Se observa la clara predominancia de Windows 7 y Windows 10 para la compra de productos.

**Figura 44:** Cantidad de visitas (log) de cada producto por sistema operativo

Aquí podemos ver un curioso patrón donde los usuarios de iPhone (que tienen sistemas operativos iOS) tienen una clara preferencia por los dispositivos de la misma marca.

3.4.2. Análisis de resoluciones

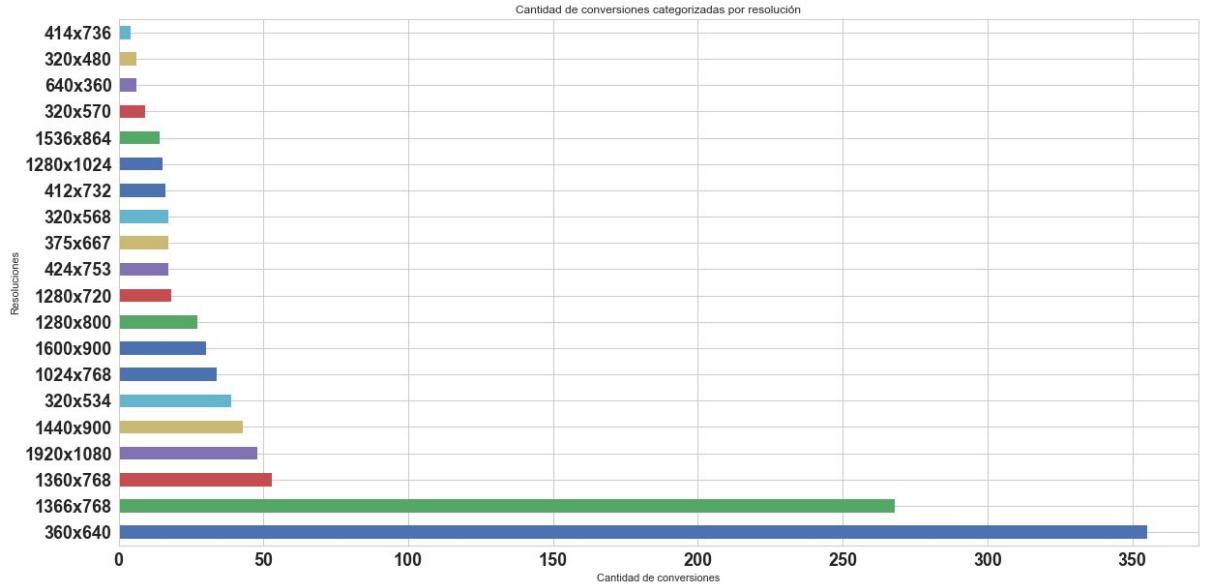


Figura 45: Cantidad de conversiones por resolución

Podemos notar que la mayor cantidad de conversiones se realizan desde dispositivos con resolución de computadora. Sin embargo hay un dato que genera ruido.

3.5. Productos más populares por ciudad de Brasil

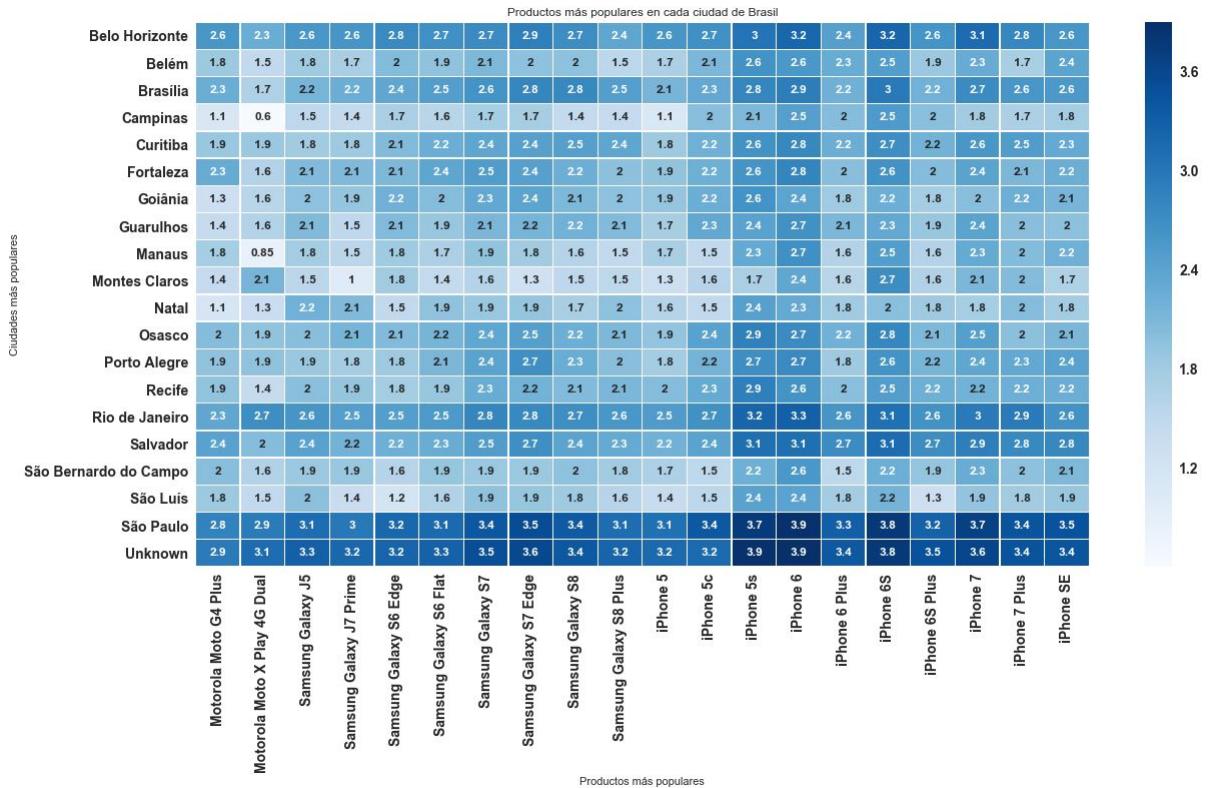
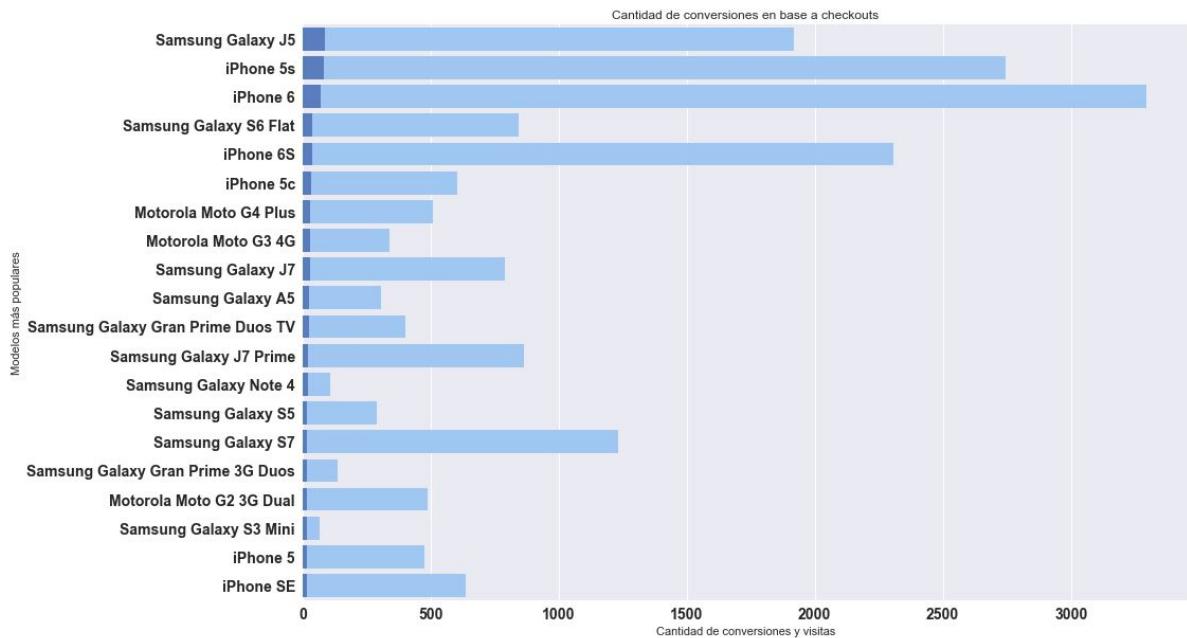


Figura 46: Cantidad de modelos comprados por x campaña de marketing que son compradores

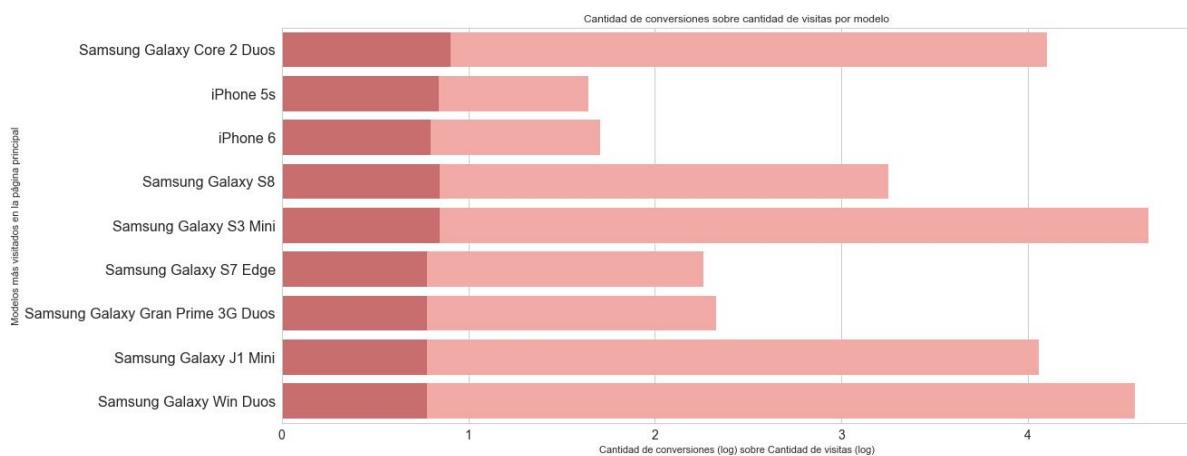
3.6. Conversiones en base a checkouts

**Figura 47:** Cantidad de conversiones en base a checkouts

El color más claro representa los checkouts y el color más oscuro representa las conversiones.

Podemos observar que si bien el Samsung Galaxy J5 no es el producto que más veces fue agregado a Checkout, sí fue el que más cantidad de conversiones recibió.

3.7. Conversiones en base a visitas de productos en la página principal

**Figura 48:** Conversiones en base a visitas de productos en la página principal

El color más claro representa las visitas y el más oscuro representa las conversiones.

En este gráfico podemos observar que si bien el iPhone 5s y iPhone 6 poseen una muy pequeña cantidad de visitas en la página principal, son uno de los productos con la mayor cantidad de conversiones.

3.8. Análisis de checkouts por persona

¿Cuántas personas ingresan a la etapa del checkout?

Analizando el set de datos para poder responder esta pregunta nos encontramos con que todos las personas que entraban a la página, identificadas por el campo person, terminaban haciendo al menos una vez un checkout. Pasaba que a veces los usuarios entraban a la página y no hacían checkout, pero en algún momento del periodo de tiempo con el que trabajamos estos volvían y realizaban un checkout. Este comportamiento nos pareció muy extraño, pero puede deberse a que el set de datos con el que trabajamos es una versión reducida del real, con lo cual puede ser que se haya tomado como criterio para reducir el dataset, el que los usuarios hicieran al menos una vez un checkout. Para comprobar esto lo que hicimos fue analizar para cada persona cuáles eran los eventos en los que participaban y contar las ocurrencias de cada uno, luego nos quedábamos con el evento checkout y analizamos si algún usuario tenía 0 ocurrencias de dicho evento.

4. Conclusiones finales

Las conclusiones que pudimos sacar y que nos parecieron las más relevantes e interesantes son las siguientes:

1. Es evidente que hay un fuerte incremento de actividad en el sitio a partir de mediados de mayo. Si bien no es posible concluir cuál es la causa de este incremento sospechamos que pueda deberse a un fuerte empuje en las campañas de publicidad lo que desembocó en el incremento de tráfico. Sin embargo es posible que el incremento de visitas por campaña publicitaria sea consecuencia de otra causa desconocida.
2. Hay una clara diferencia entre las visitas y conversiones de productos de Apple y Samsung con respecto a los demás. Son los productos por los que las personas están más interesadas. Quizás sería conveniente enfocar las campañas publicitarias a este tipo de productos, o publicitar y dar a conocer más sobre otros tipos de marcas.
3. Si bien Google es la fuente que produce la mayor cantidad de visitas, si miramos los porcentajes de compradores sobre cantidad de usuarios, emblue es la fuente que tiene un mejor ratio de conversión por visita. Puede ser buena idea darle mayor empuje a esta fuente. Sin embargo, se debe tener en cuenta que Google, al producir más cantidad de visitas, es de donde provienen la mayor cantidad de compradores.
4. El porcentaje de conversiones a partir de los Leads es muy pobre, quizás convendría alentar el uso de esta funcionalidad.
5. Hemos notado que la fuente por la cual ingresan los usuarios a Trocafone a partir de campañas publicitarias no impacta significativamente en los tipos de productos que convierten. No se generan patrones ni se categoriza el comportamiento. La mayoría de las personas siempre acude a los productos más populares y de las mismas marcas sin importar el sitio por el cual ingresan.
6. Se registró que un gran porcentaje de los usuarios regresan una poca cantidad de veces a la página. Esto parece tener sentido teniendo en cuenta que la idea es que los usuarios entren a comprar un producto. Sin embargo no se vió una gran cantidad de compras si lo compara con la cantidad total de visitas/checkouts.