

Trabajo Práctico N°1

Segundo cuatrimestre
2018.

75.06 Organización de datos

Grupo 29:

X

Integrantes:

1)Augusto E. Hoszowski

2) Diego Durante

3)Manuel Barros

1 - Introducción	2
2- Conocimiento de dominio	3
2a - Productos	3
2b - Estados de los productos	6
2c - Estructura de la Página	6
2d - Cuestiones Geográficas	8
3 - Análisis preliminar del set de datos	9
3a - Detalle del datasets	9
3b - Columna 'Events'	11
3ba - Evento 'visited site'	12
3bb - ¿El evento 'visited site' indica un inicio de sesión?.	13
3bc - ¿Cómo se relaciona los inicio de sesión y las conversiones?	14
3bd - ¿Los usuarios que ingresaron directamente tuvieron actividad anterior en la página o es el único registro que existe de ellos?.	16
3c - Análisis Geográfico	17
3d - Análisis Temporal	17

1 - Introducción

En el siguiente informe se presentan los resultados del análisis realizado sobre los datos brindado, amablemente, por *Trocafone* de su plataforma ecommerce de Brasil. Los datos se encuentran en un archivo CSV. La información que contiene la misma es información sobre Eventos realizados por usuarios identificados. Esta información brinda principalmente:

- 1) Información geográfica y temporal del evento.
- 2) Firma o usuario que genera el evento y tipo de plataforma desde donde se realizan las solicitudes.
- 3) Tipo de evento realizado.
- 4) Características de los productos a los que se acceden.

Más adelante se realizará una apertura de estos puntos mediante el análisis. Para realizar el mismo se ha considerado que tipo de servicio brinda *Trocafone* para poder comenzar a preguntarse qué información, que extraigamos de los datos, puede llegar a tener algún valor.

2- Conocimiento de dominio

Trocafone se encarga de unir básicamente a la persona que tiene un celular usado y el cliente que quiere comprar un celular, dando una garantía sobre el estado real del dispositivo. Realiza Trade-In, se encarga de la toda la logística tanto la recolección como la entrega, del refurbishing, y la venta por distintos canales de ventas. En este caso nos toca analizar el punto de venta ecommerce de Brasil, que se obtienen de la interacción de su pagina (<https://www.trocafone.com/>).

2a - Productos

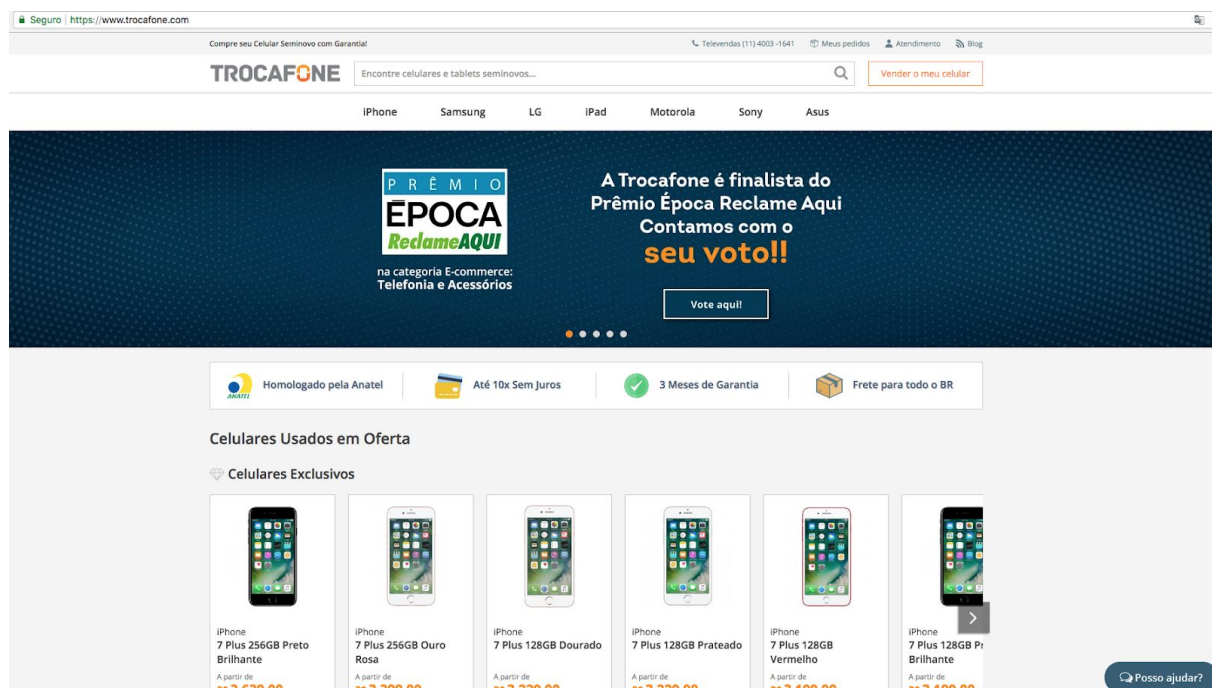


Imagen 1.1: Página principal de Trocafone Brasil.

La división por línea de equipos que tienen son:

- iPhone
- Samsung
- LG
- iPad
- Motorola
- Sony

-
- Asus

Y realizando un búsqueda en la página obtenemos los siguientes elementos (Conocimiento_dominio.ipynb):

Resultado Web Scrapping:

```
-----iPhone-----
iPhone 5
iPhone 5c
iPhone 5s
iPhone 6
iPhone 6 Plus
iPhone 6S
iPhone 6S Plus
iPhone SE
iPhone 7
iPhone 7 Plus
-----Samsung-----
Galaxy S4 i9515
Galaxy S5
Galaxy S5 Mini Duos
Galaxy S5 Duos
Galaxy S Duos 2
Galaxy Note 3
Galaxy S4 i9500
Galaxy S4 i9505
Galaxy S4 Mini Duos
Galaxy A7
Galaxy A5
Galaxy E7
Galaxy A3 Duos
Galaxy S6 Edge
Galaxy Note Edge
Galaxy J5
Galaxy J7
Galaxy Note 5
Galaxy S5 New Edition Duos
Galaxy Note 3 Neo Duos
Galaxy S7
Galaxy S7 Edge
Galaxy J7 Prime
Galaxy J5 Prime
Galaxy A3 2016
Galaxy A5 2016
Galaxy A7 2016
```

Gear Fit 2 Grande
Galaxy J5 2016 Metal
Galaxy A5 2017
Galaxy E5 4G Duos
Galaxy S3 Slim Duos
Galaxy S3 Duos
Galaxy S8
Galaxy S8 Plus
Galaxy Note 8

-----LG-----

G3 Beat D724
G4 Beat H736
G4 H818P
G4 H815P
G4 Stylus HDTV H540T
G5 SE
G6

-----iPad-----

iPad Air 2 Wi-Fi
iPad Air Wi-Fi + 4G

-----Motorola-----

Moto MAXX
Moto X2
Moto G3 4G
Moto X Style
Moto X Force
Moto Z Play
Moto Z
Moto Z Power Edition
Moto Z2 Play

-----Sony-----

Xperia Z3 TV
Xperia Z2
Xperia Z3 Compact
Xperia Z3 Plus
Xperia Z3 Dual
Xperia Z5
Xperia M4 Aqua Dual

-----Asus-----

Zenfone 5

2b - Estados de los productos

	Novo	Excelente	Muito Bom	Bom
Software	100% Funcional	100% Funcional	100% Funcional	100% Funcional
Tela	Qualidade de fábrica	Arranhões pequenos, quase imperceptíveis	Arranhões leves	Sinais de uso mais nítidos, como arranhões.
Carcaça	Qualidade de fábrica	Arranhões pequenos, quase imperceptíveis	Arranhões ou amassados leves	Sinais de uso mais nítidos, como arranhões ou pequenos amassados.
Condição da bateria	100%	Acima de 85%	Acima de 85%	Acima de 85%
Controle de Qualidade	✓	✓	✓	✓
Limpeza antibactericida	✓	✓	✓	✓
Desbloqueios	✓	✓	✓	✓
Reset de dados	✓	✓	✓	✓

¹ Cuadro de estados de la página de Trocafone Brasil.

2c - Estructura de la Página

La página cuenta con una entrada, que es la estructura básica, o el index (<https://trocafone.com>). Luego se puede acceder a compra o ventas a través de la estructura: “/comprar/” o “/vender/”, respectivamente. Dentro de cada uno se puede acceder en primera instancia la línea de producto ejemplo: si queremos acceder a Samsung sería: “/comprar/samsung/”, caso de “Iphone” sería: “/comprar/iphone”, etcétera. El mismo caso se aplica para la venta de productos.

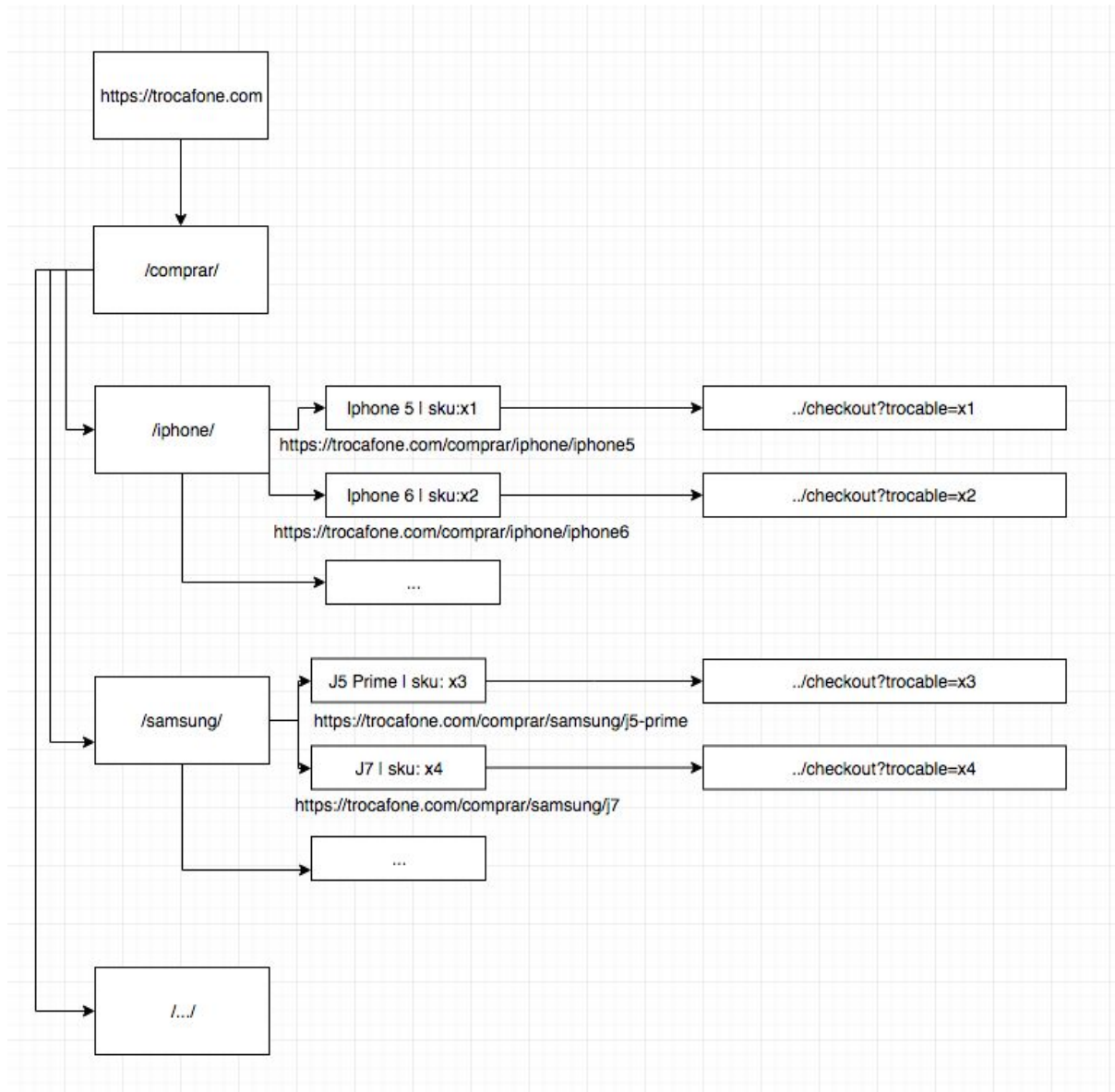
Por otro lado podemos encontrar el “checkout”, en la estructura de la página está en: “/comprar/checkout”. Cuando un producto es seleccionado y pasa al Checkout pasa con un número SKU que ya contiene las características del producto, es decir el estado del mismo.

El sitio dispone de un buscado, deriva al sector “/comprar/list?q=” que arroja como resultados productos que dirigen directamente a los productos.

El resto son páginas estáticas y de contacto, que contiene información de trocafone.

¹ Se copia el cuadro de la página de Brasil ya que difiere del de Argentina.

A continuación vemos la estructura de la página para las compras:



En la imagen podemos ver la composición de las Urls.

2d - Cuestiones Geográficas

Si bien el negocio de Trocafone se realiza a través de plataformas tiene presencia física también. Las oficinas principales están en Buenos Aires Argentina, pero dispone de 9 sucursales ubicadas en Brasil llamados "Quiosques Trocafone". Es un dato que se quiere tener en cuenta para ver si hay una incidencia mayor de conversiones o ventas en estas zonas.

3 - Análisis preliminar del set de datos

3a - Detalle del datasets

Cantidad de registros: 1011288

Cantidad de Columnas: 23

Columns	Qty Rows not null	Type
timestamp	1011288	object
event	1011288	object
person	1011288	object
url	82756	object
sku	563838	object
model	564284	object
condition	563836	object
storage	563836	object
color	563836	object
skus	221699	object
search_term	48967	object
staticpage	3598	object
campaign_source	82796	object
search_engine	50957	object
channel	87378	object
new_vs_returning	87378	object
city	87378	object
region	87378	object
country	87378	object
device_type	87378	object
screen_resolutio	87378	object
operating_system_version	87378	object
browser_version	87378	object

Tabla 3.a.1

Lo primero que se ve en el Dataset es que las tres columnas que tienen siempre contienen un valor no nulo son:

- **timestamp**
- **event**
- **person**

Teniendo en cuenta que cada registro es un Evento la relación de igualdad numérica entre distintas columnas nos indica que son campos relacionados, o que obtiene información cuando se genera un evento o acción particular.

En las otras columnas se puede notar otra relación:

- *channel* 87378
- *new_vs_returning* 87378
- *city* 87378
- *region* 87378
- *country* 87378
- *device_type* 87378
- *screen_resolution* 87378
- *operating_system_version* 87378
- *browser_version* 87378

Todas estan completas en el mismo cantidad de veces. Del mismo modo podemos encontrar la siguiente relación.

- *sku* 563838
- *model* 564284
- *condition* 563836
- *storage* 563836
- *color* 563836

3b - Columna 'Events'

Generamos una vista rápida donde categorizamos los eventos y la cantidad de repeticiones de cada uno en el dataset:

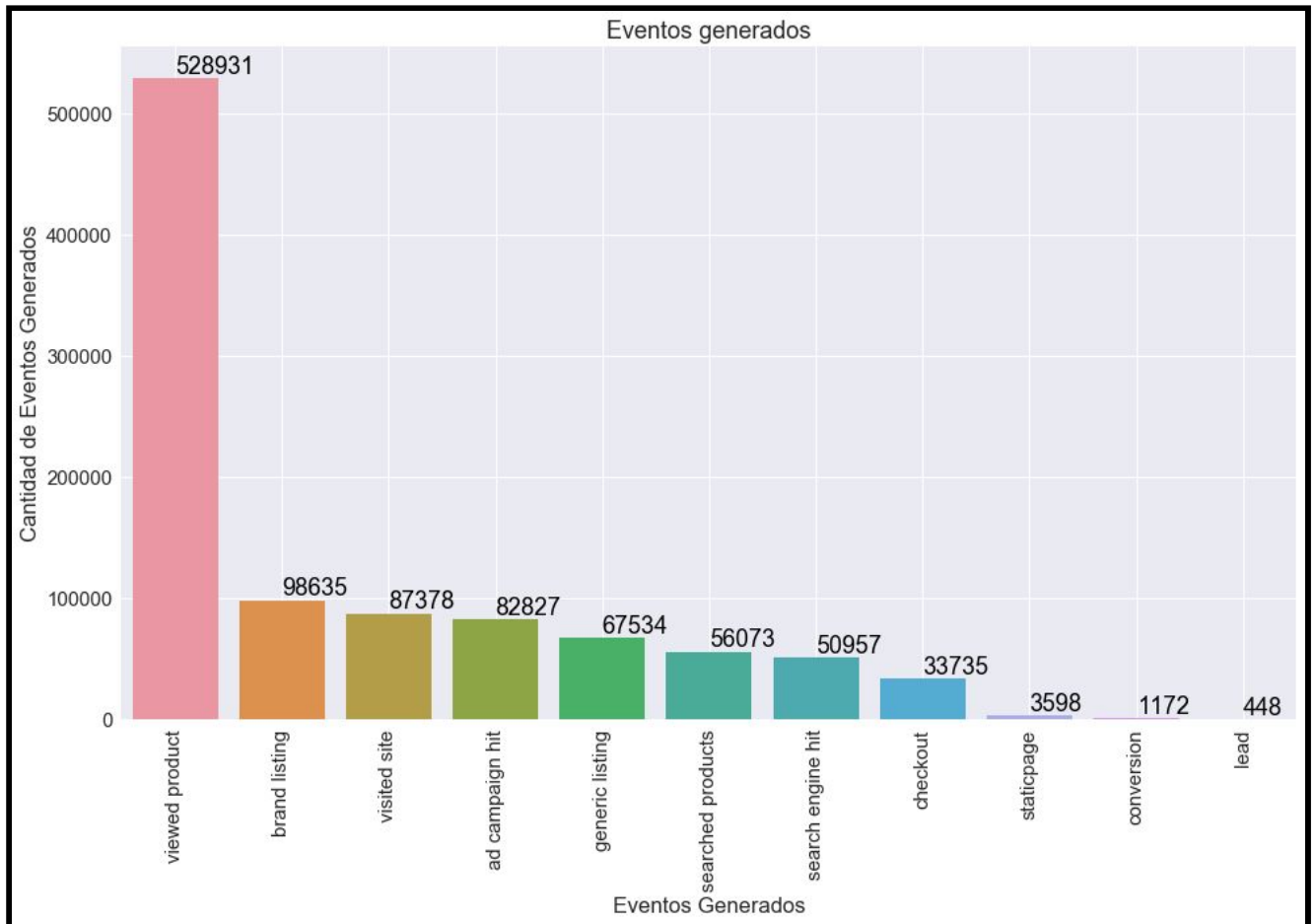


Gráfico 3b.1

Inmediatamente notamos que '**visited site**' tiene **87378** apariciones que es el mismo número que los campos que se encontraban en el primer grupo de registros:

- *channel* 87378
- *new_vs_returning* 87378
- *city* 87378
- *region* 87378
- *country* 87378
- *device_type* 87378
- *screen_resolution* 87378
- *operating_system_version* 87378
- *browser_version* 87378

Estos nos lleva a pensar que el evento Visite Site se genera cuando un usuario ingresa al sitio.

3ba - Evento 'visited site'

Revisamos brevemente los campos, para ver qué información obtenemos

channel	new_vs_returning	city	region	country	device_type	screen_resolution	operating_system_version	browser_version
Paid	New	Camaragibe	Pernambuco	Brazil	Smartphone	360x640	Android 6	Chrome Mobile 39
Paid	New	Rio de Janeiro	Rio de Janeiro	Brazil	Smartphone	360x640	Android 5.1.1	Android 5.1
Paid	New	Rio de Janeiro	Rio de Janeiro	Brazil	Computer	1920x1080	Windows 10	Chrome 65.0
Paid	Returning	Rio de Janeiro	Rio de Janeiro	Brazil	Computer	1920x1080	Windows 10	Chrome 66.0
Organic	New	Ribeirão Preto	Sao Paulo	Brazil	Computer	1366x768	Windows 10	Chrome 65.0
Direct	Returning	Ribeirão Preto	Sao Paulo	Brazil	Computer	1366x768	Windows 10	Chrome 66.0
Paid	New	Unknown	Unknown	Brazil	Smartphone	360x640	Android 6	Chrome Mobile 66.0
Organic	New	Unknown	Unknown	Brazil	Computer	1366x768	Windows 7	Firefox 52

Tabla 3bb.1

La información que nos da la Tabla 3bb.1 es:

Geográfica:

- **País** (country)
- **Región** (region)
- **Ciudad** (City)

Dispositivo:

- **Tipo** (device_type)
- **Resolución de pantalla** (screen_resolution)
- **Sistema Operativo** (operating_system_version)
- **Navegador** (browser_version)

Usuario:

- **Canal de conexión** (channels)
- **Si es usuario recurrente** (new_vs_returning)

3bb - ¿El evento 'visited site' indica un inicio de sesión?

Tomando como conjetura que cada vez que se inicia una sesión se genera el evento 'visited site' lo que hicimos fue agrupar los datos por usuarios, y contabilizar la cantidad de eventos generado dentro del día por ese usuario. Quedando la información de la siguiente manera:

person	periodo	ad_campaign_hit	brand_listing	checkout	conversion	generic_listing	lead	search_engine_hit	searched_products	staticpage	viewed_product	visited_site
0004b0a2	531	1.0	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0
0006a21a	529	1.0	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0
000a54b2	49	1.0	2.0	NaN	NaN	1.0	NaN	1.0	NaN	NaN	1.0	1.0
	524	1.0	4.0	1.0	NaN	4.0	NaN	1.0	NaN	NaN	18.0	1.0
00184bf9	46	NaN	5.0	2.0	NaN	NaN	NaN	2.0	NaN	NaN	32.0	1.0
	517	NaN	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	54	NaN	2.0	1.0	NaN	NaN	NaN	NaN	NaN	NaN	4.0	1.0
0019c395	517	2.0	NaN	1.0	NaN	NaN	NaN	2.0	NaN	NaN	2.0	1.0
001bb7eb	66	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	1.0	1.0
	68	1.0	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0
001f1653	529	1.0	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0
00204059	118	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN	NaN	1.0

La cantidad de filas generadas: 67056

Tabla 3bb.2

Donde 'período' es la concatenación del número de mes y número de día en que se generaron los eventos. Si observamos rápidamente podemos ver que los campos más recurrentes son 'visited site'.

Si agrupamos las apariciones de los eventos, es decir, contamos cada aparición por una unidad (no sumamos) y desplegamos un gráfico obtenemos lo siguiente:

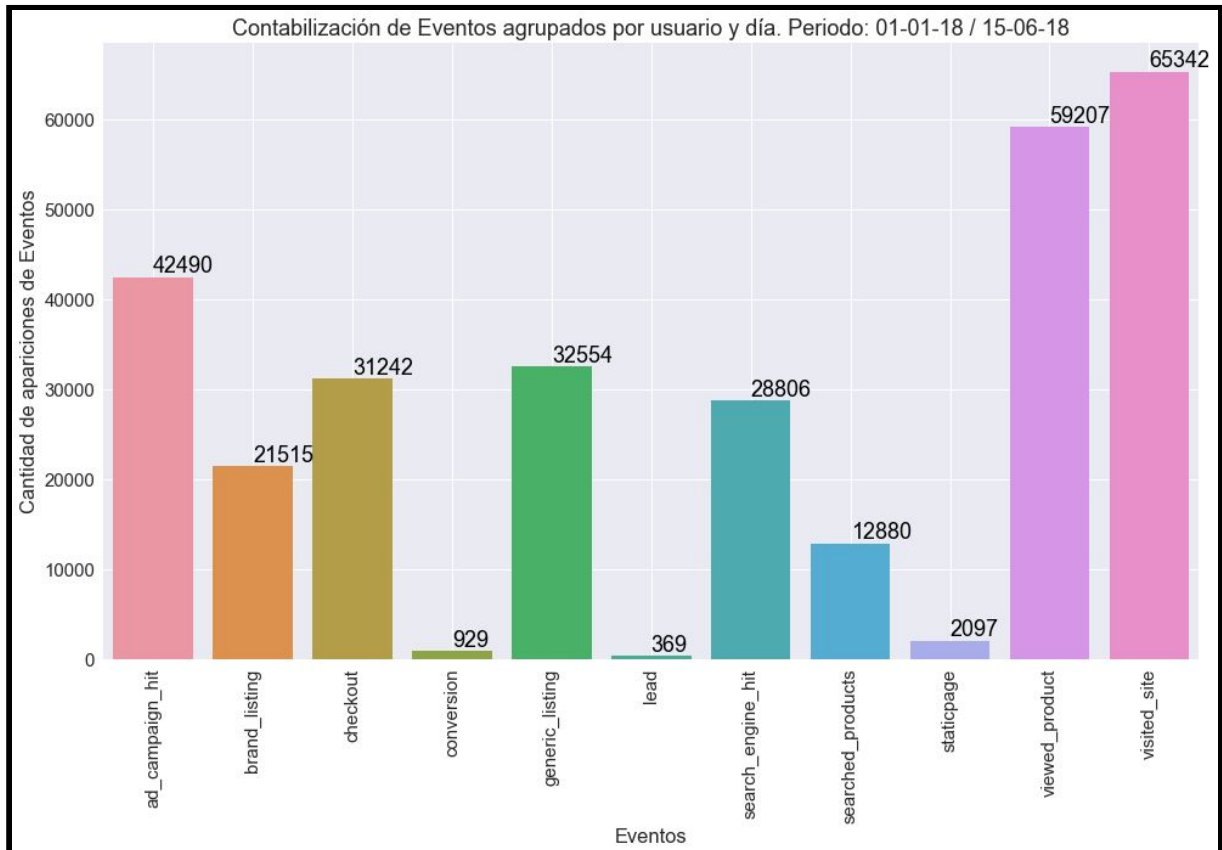


Gráfico 3bb.1

Vemos que el Evento que más veces se puede contabilizar, es decir que aparece en la mayor cantidad de registros es 'Visited Site' con 65342 sobre un total de 67056 (Tabla 3bb.2), es decir, un 97% de apariciones. Por lo que podemos estimar que por regla general el inicio de sesión de un usuario coincide con el evento 'Visited Site'. Veamos qué ocurre cuando no aparece:

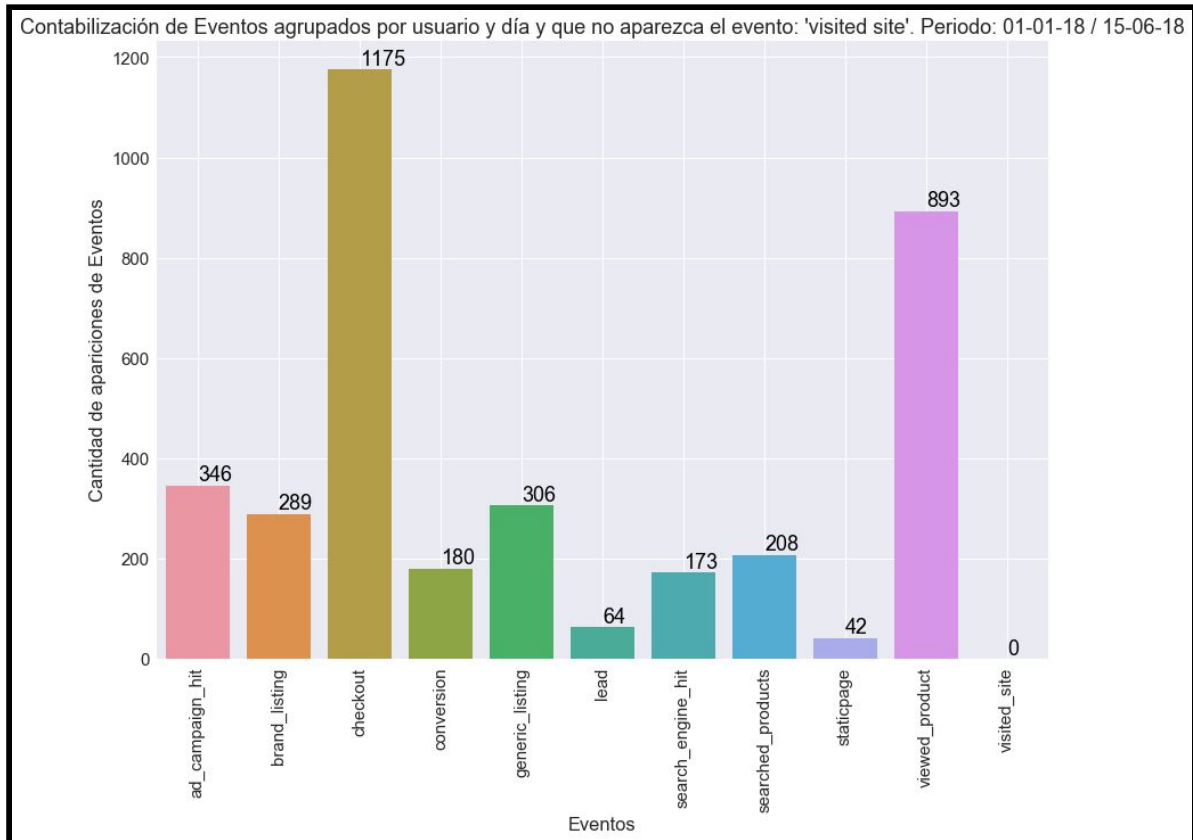


Gráfico 3bb.2

En el gráfico 3bb.2 podemos notar que lo el evento que mayor participación tiene (cuando no aparece el evento 'Visited Site') es checkout, esto puede ser explicado si el usuario recibe una notificación o un enlace directo para llegar a esa instancia y no ser registrado por la plataforma de analytics. El otro evento que está por sobre el resto es 'Viewed Product', lo que podría suceder si se entra por historial o se activa alguna pagina o algun caso similar. Podemos concluir que el evento 'visited site' si corresponde a un inicio de sesión de un usuario.

3bc - ¿Cómo se relaciona los inicio de sesión y las conversiones?

Un dato interesante que sugiere la comparación del Gráfico 3bb.2 con 3bb.1 son las conversiones, es decir la formalización de una compra.

Podemos visualizar a modo de resumen:

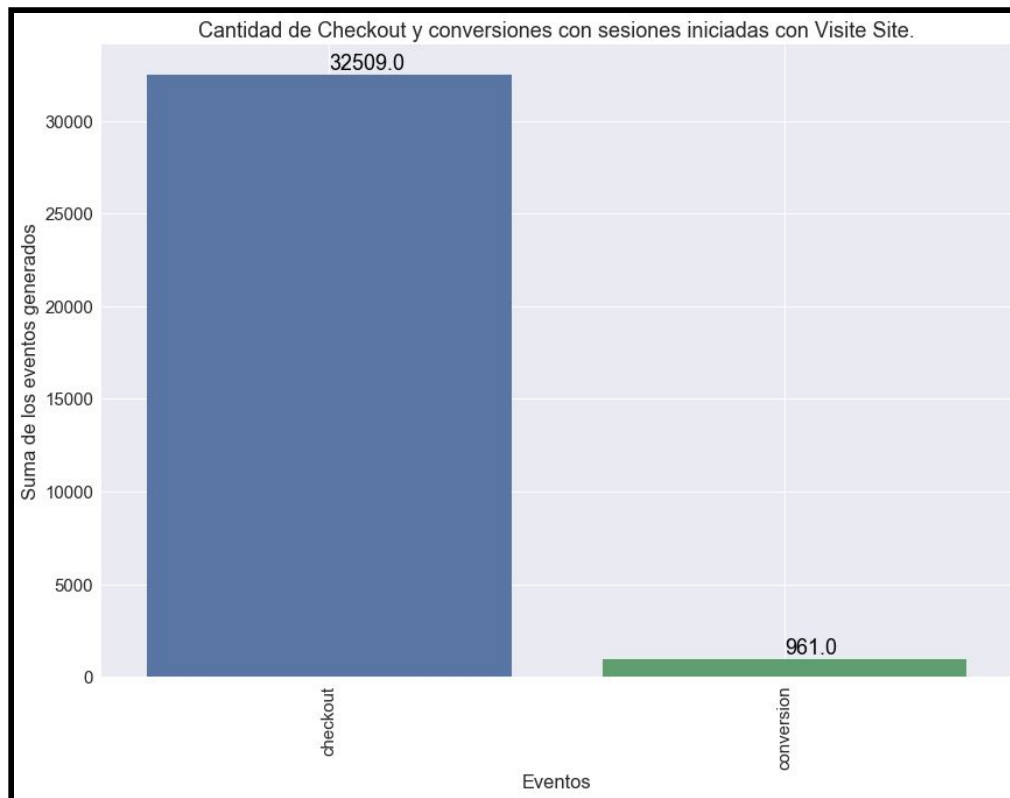


Gráfico 3bb.3

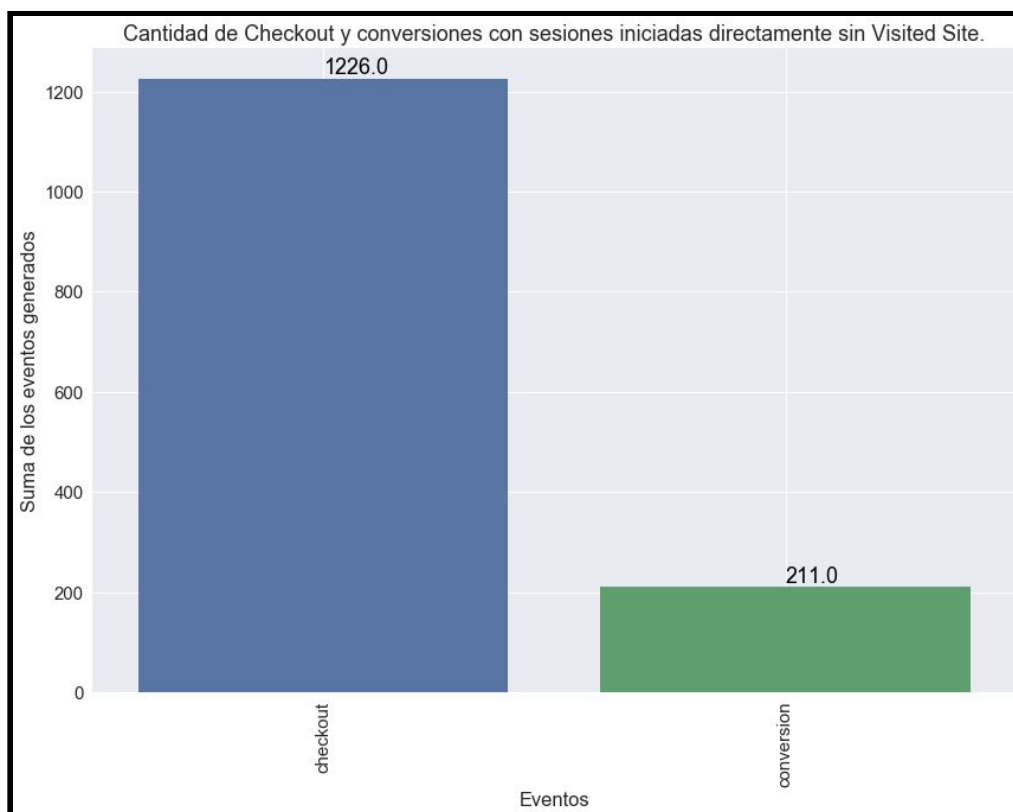


Gráfico 3bb.4

De lo anterior podemos realizar el siguiente cuadro:

Gráfico	Cantidad Máximo de Sesiones ²	Checkouts ³	Conversiones ⁴	Participación.
Gráfico 3bb.3	65342	32509	961	81.99%
Gráfico 3bb.4	1714	1126	211	18.01%
Total	67056	33635	1172	100%

Lo que se ve en el cuadro es una mayor efectividad de conversiones cuando el usuario ingresa directamente al sitio. Aunque hay que tener en cuenta que la cantidad de sesiones generadas de una y otra manera son muy dispares, lo que nos lleva a la siguiente pregunta:

3bd - ¿Los usuarios que ingresaron directamente tuvieron actividad anterior en la página o es el único registro que existe de ellos?

Para comprobarlo lo que hacemos fue listar las personas que hayan realizado un 'conversion' directamente, sin 'visited site'. Luego contabilizamos aquellos que hayan realizado algún visited site y obtuvimos: 93 campos. Por lo que 93 personas, que realizaron compras, tuvieron contacto con la plataforma analytics, antes o después.

² La cantidad de sesiones está considerado de la agrupación de usuarios por día. Como se describió en el principio del apartado 3bb. El cálculo es el residual del total menos la cantidad de sesiones que son excluidas ('visited site').

³ Suma de la Cantidad de Checkouts y/o Conversiones Generados.

⁴ Suma de la Cantidad de Checkouts y/o Conversiones Generados.

3be - Comparación de Evento por usuario.

Vamos a observar qué cantidad de usuarios realizaron cada evento al menos una vez.

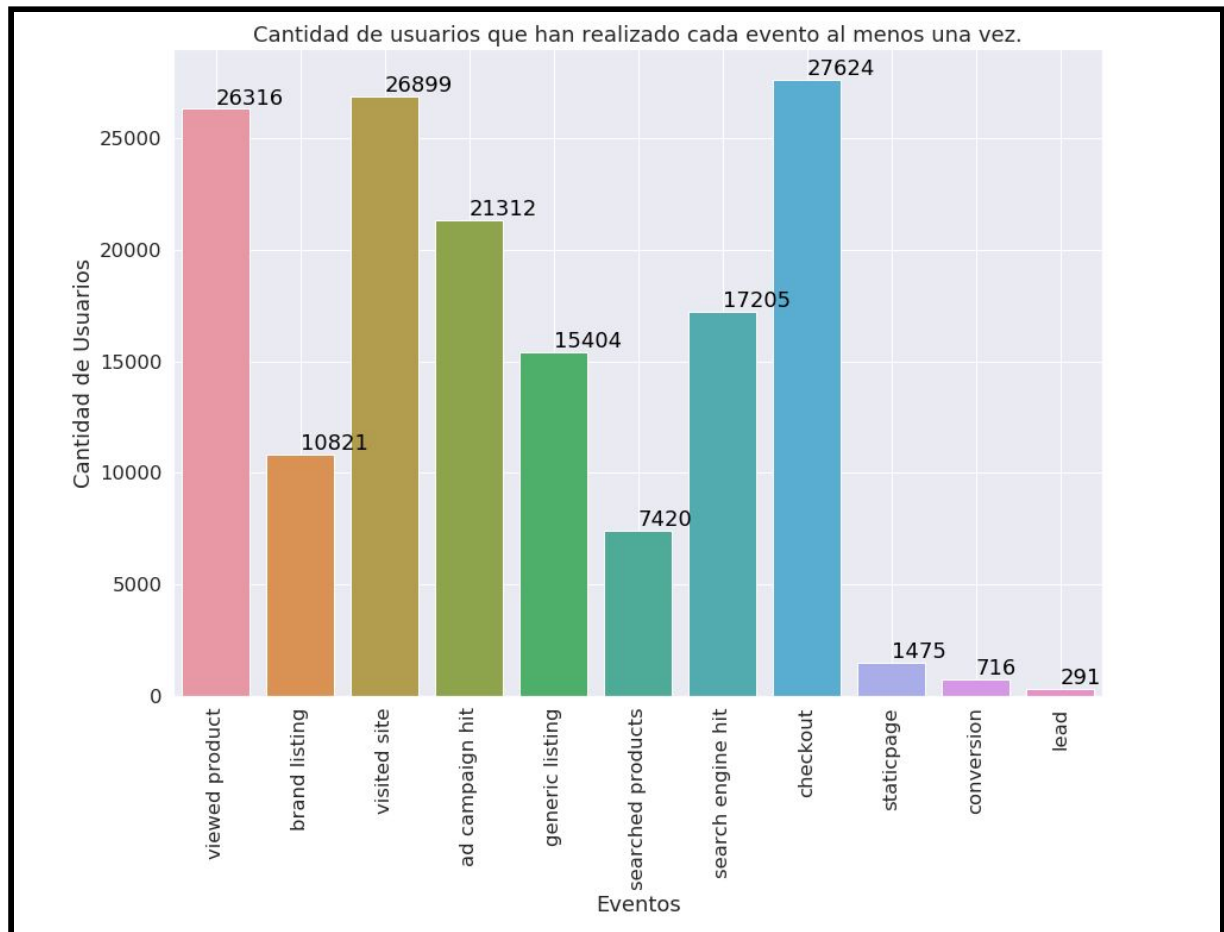


Gráfico 3be.1

Si recordamos que la cantidad de usuarios es 27624 vemos que coincide con la cantidad de Checkout realizados. Como también podemos ver algo que era previsible las 1172 ventas se observan 716 usuarios.

3c - Análisis temporal

Primero vemos que el Dataset contiene datos desde el: 01/01/2018 y llega hasta el 15/06/2018. Por lo cual para determinados análisis quitaremos este último mes. Para otros donde la temporalidad lo tendremos en cuenta.

Cantidad de eventos generados por mes:

Mes	Cantidad de Eventos.
Enero	19270
Febrero	27886
Marzo	77293
Abril	130008
Mayo	474949
Junio	281882

Cuadro 3c.1

3ca - ¿Cómo se distribuye la generación de eventos a lo largo del día?

Vamos a comenzar a ver el comportamiento en la generación de eventos primero a partir de la hora del día.

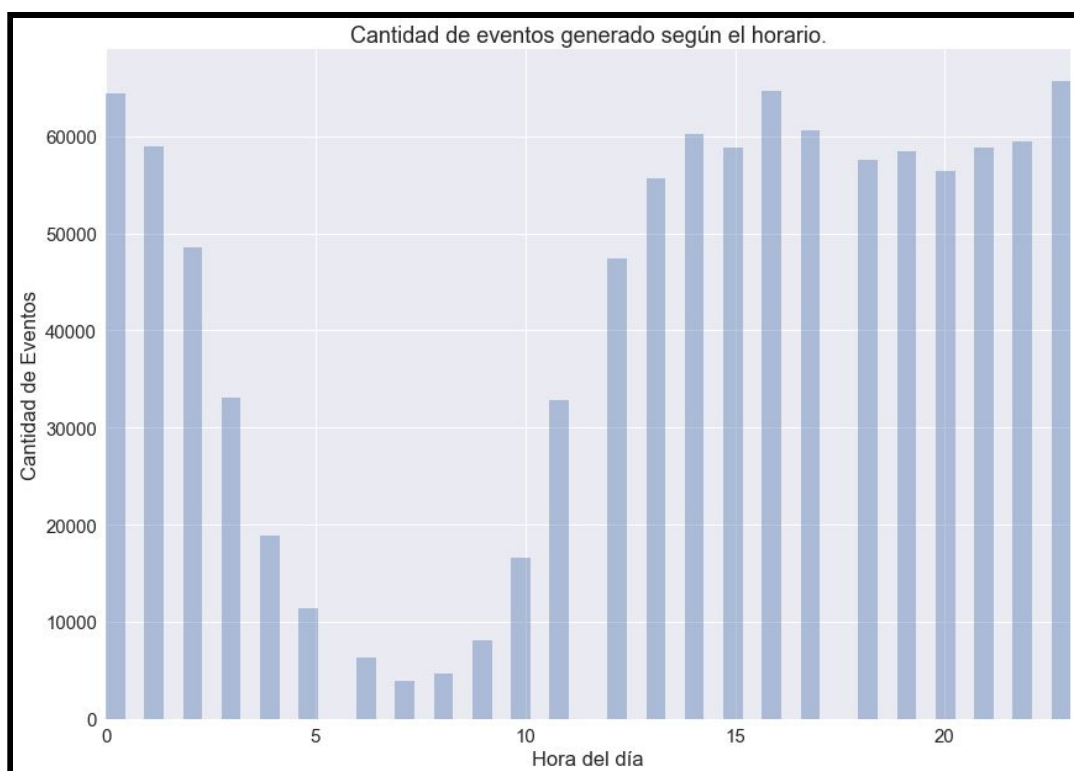


Gráfico 3ca.1

En el gráfico 3c.1 podemos notar un comportamiento bastante lógico que se ve en una caída a partir de la 1 de la madrugada hasta las 10 de la mañana. Quizá lo llamativo es que los picos de generación de eventos se den a las 16 horas y a las 23 horas.

3cb - ¿Cómo se distribuye la generación de eventos a lo largo del mes?

En este gráfico si quitaremos el mes de Junio para que no pese en la totalidad de los primeros 15 días dando un resultado falso.

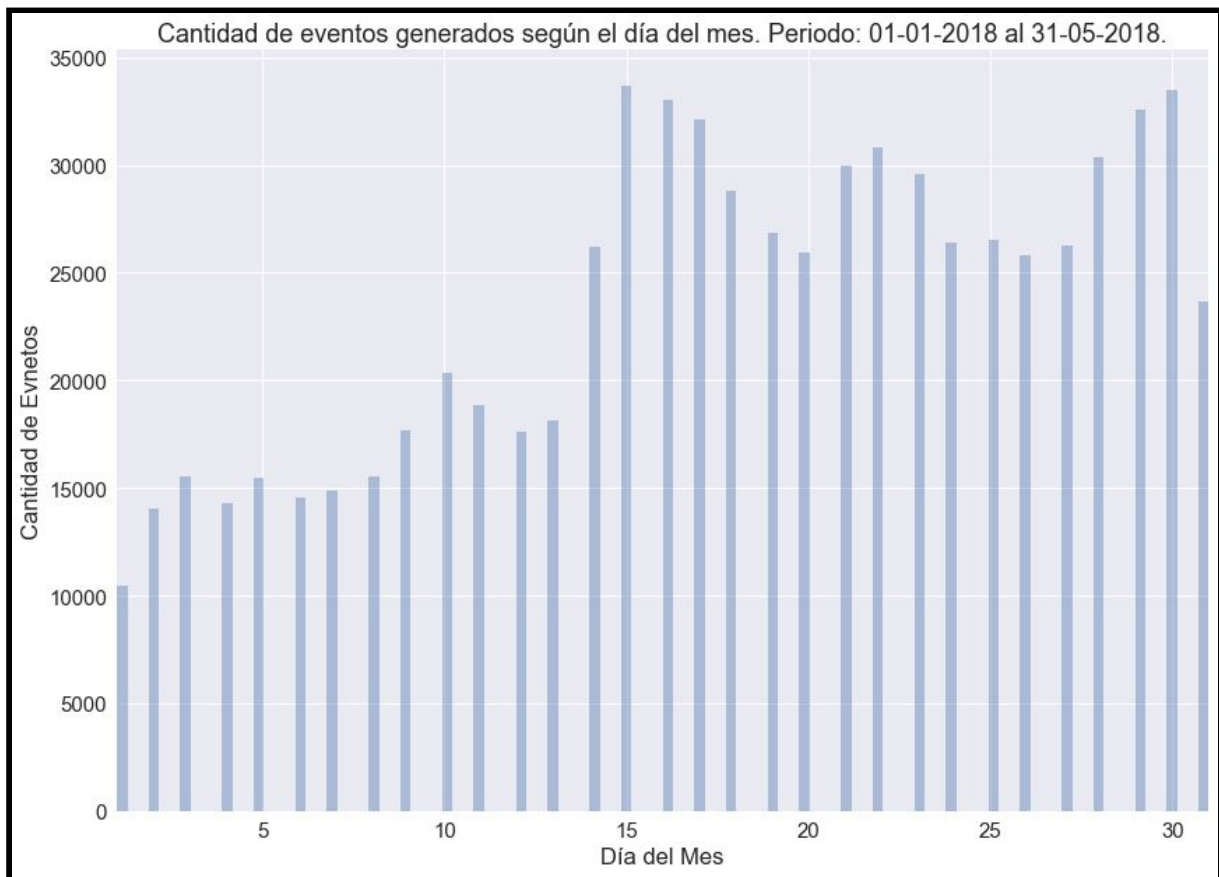


Gráfico 3cb.1

Obtenemos de esto que tanto a mitad de mes como a fin de mes se dan la mayor cantidad de compras. Incluso si consideramos que no todos los meses tienen 31, el número es más importante que el primero de mes.

3cc - ¿Cómo se distribuye la generación de eventos por mes?

Ahora vemos la distribución de eventos según los meses. Continuamos trabajando con los datos hasta Mayo para que no de una idea confusa.

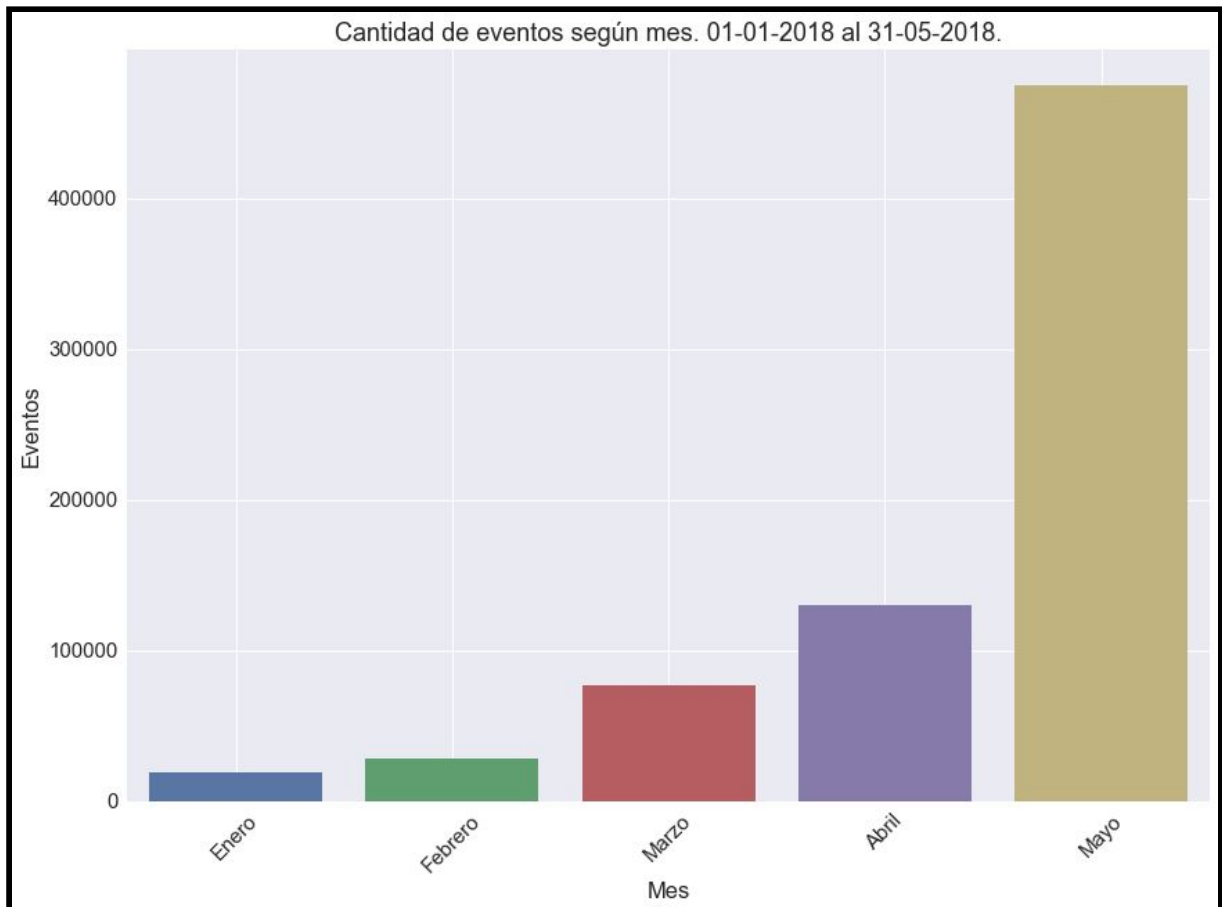


Gráfico 3cc.1

Lo que se observa en el gráfico 3cc.1 es que se va incrementando la cantidad de eventos generados en la página a través de los meses.

3cd - ¿Cómo se distribuyen la generación de eventos a los largo del día?

Vamos a generar gráficos comparativos para ver si la distribución es similar, o algunos de los eventos, tiene un comportamiento anómalo con respecto a los otros. Para esto utilizaremos la Estimación de Densidad de Kernel (KDE).

En estos informes volvimos a tomar hasta el mes de Junio ya que no afecta lo que se quiere observar.

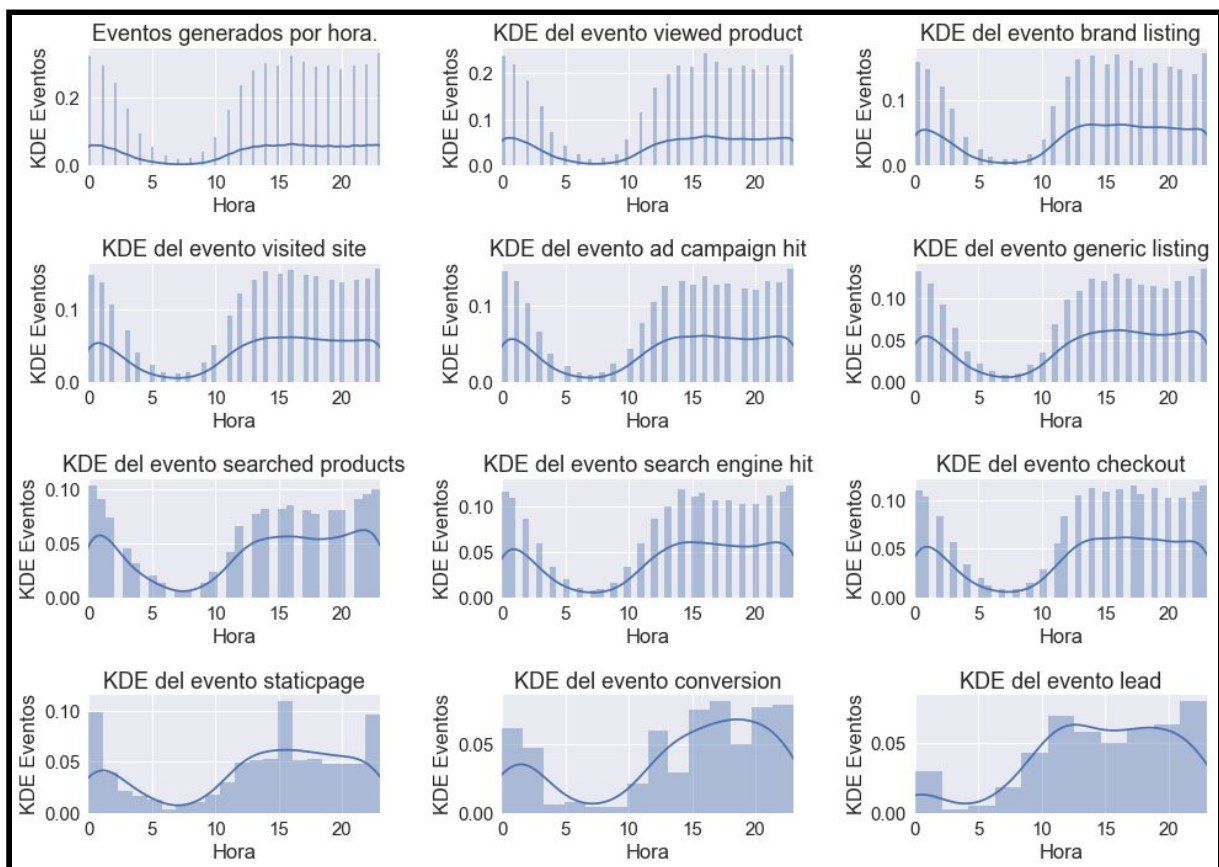


Gráfico 3cd.1

Podemos observar que los eventos totales tiene una curva muy similar al evento 'viewed product' lo cual tiene sentido ya que el 50% de los eventos generados son de esta categoría. Podemos afirmar que hay bastante relación entre los eventos generados en total con los 'checkout', 'conversiones' y 'lead', en la distribución a lo largo de las horas del día.

3ce - ¿Cómo se distribuyen los eventos a lo largo de los meses?

Ahora vamos a ver cómo se distribuyen a lo largo de los meses los eventos. Con una lógica similar a la anterior. Al ser mensual se vuelve a utilizar los datos excluyendo a junio.

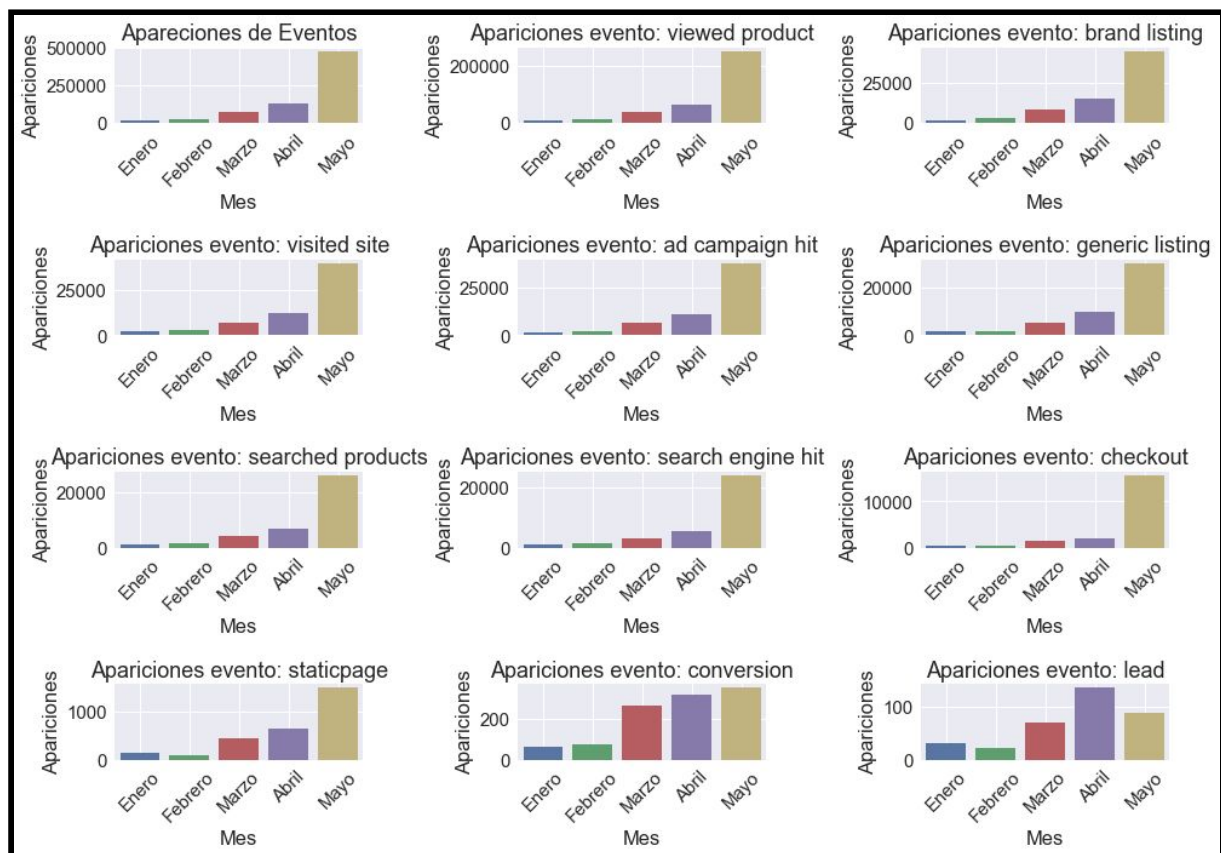


Gráfico 3ce.1

En este gráfico si podemos ver ya una diferencia en el comportamiento de los distintos eventos en cada mes. Por ejemplo: excepto 'conversion' y 'lead', tenemos un comportamiento similar entre los distintos eventos. Queda para más adelante analizar si el peso que tiene el evento lead sobre las conversiones tiene importancia. Ya que el porcentaje de 'lead' (reservas) permitiría predecir algún valor inicial para 'conversion'.

3cf - ¿Cómo se distribuyen los eventos a los largo de los meses?

Nuevamente utilizaremos una Estimación de Densidad de Kernel para obtener la distribución de eventos a lo largo del mes.

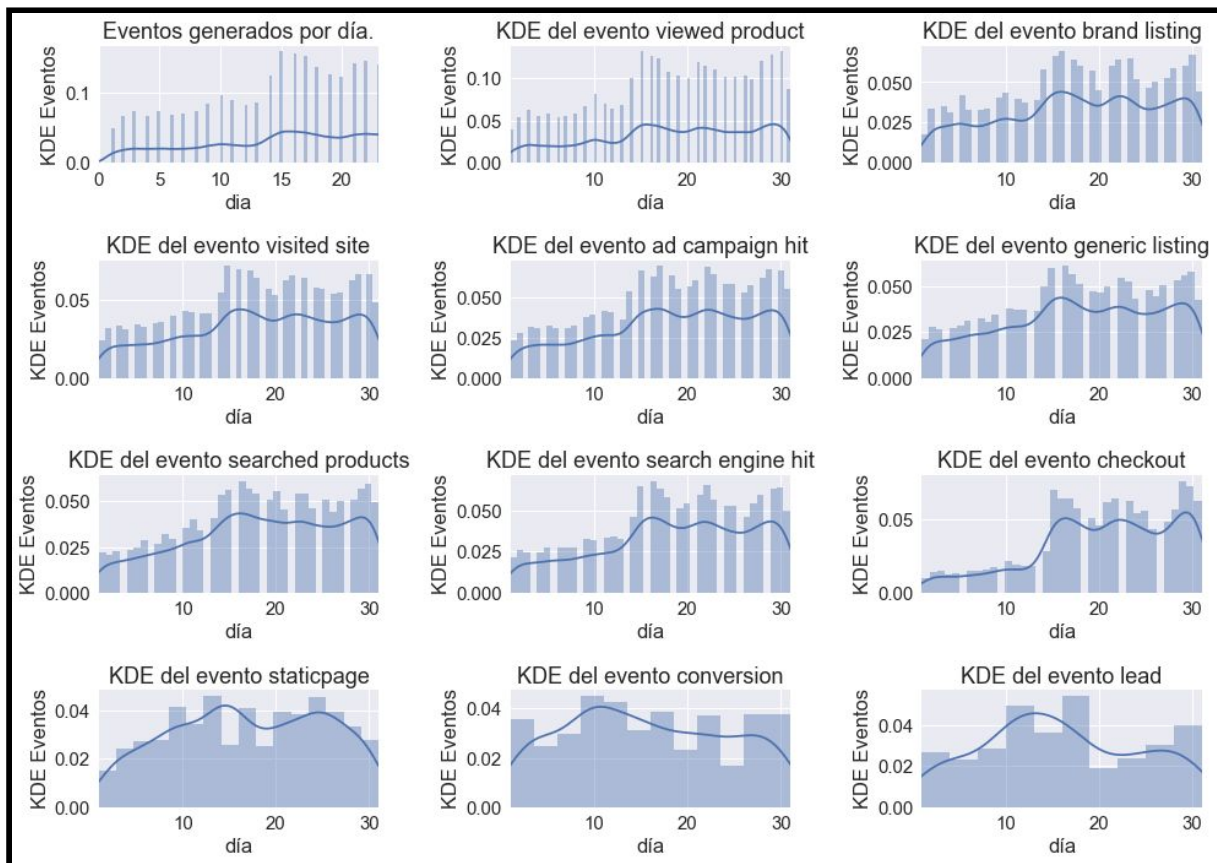


Gráfico 3cf.1

En esta comparativa si obtenemos datos distintos, ya que no existe una relación directa entre los eventos generados por día (ver en detalle el gráfico: 3cb.1) y el del evento 'conversion' ni 'lead'. Incluso es interesante notar que la cantidad no van de la mano la cantidad de checkout con 'conversion', ya que tienen una curva bastante diferente, si bien necesariamente están relacionados.

3cg - ¿Cómo se distribuyen los eventos en la semana?

En los siguientes gráficos vemos la distribución de los eventos en los días de la semana.

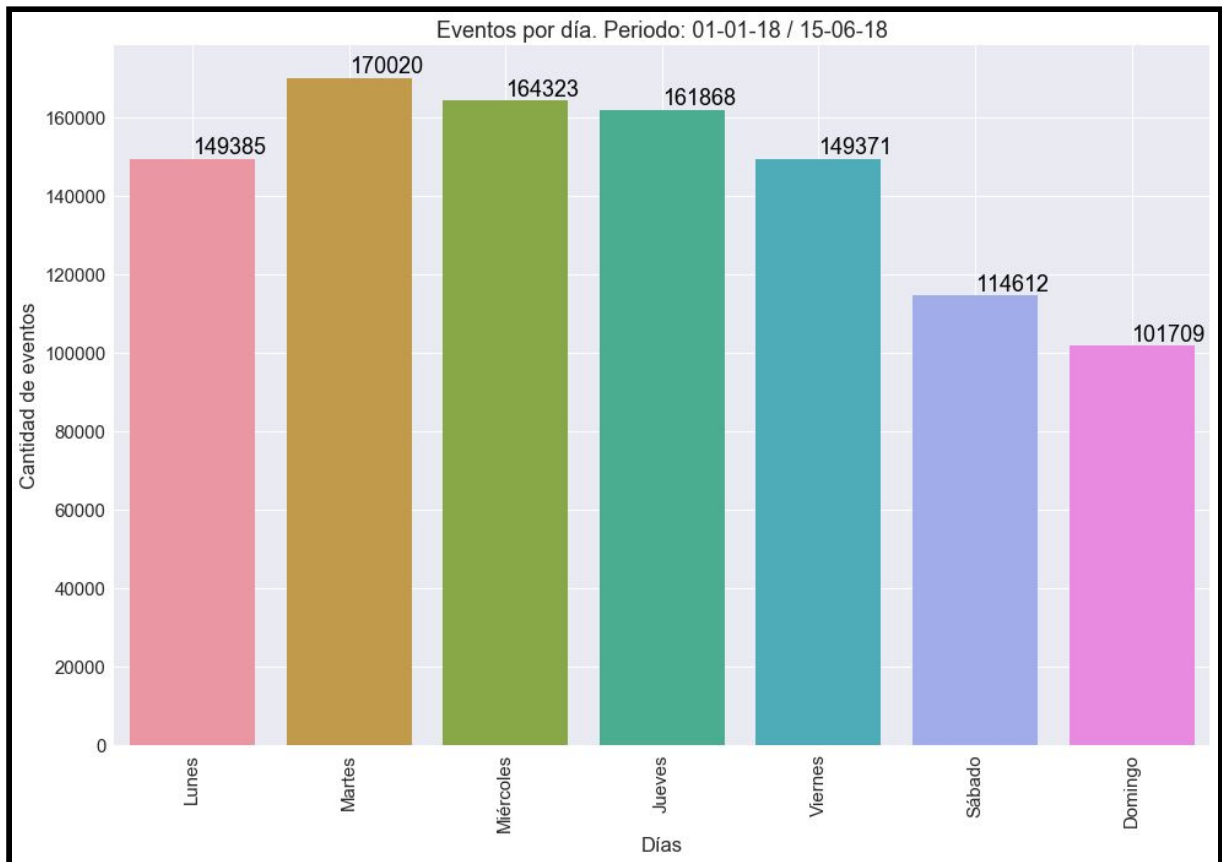


Gráfico 3cg.1

Vemos que los fines de semana la cantidad de eventos que se generan en la página caen, siendo el Martes el día con más eventos generados.

3ch - ¿Y los eventos 'conversiones'?

Si hacemos la comparación de la distribución de las conversiones en la semana contra el grafico 3cg.1 vemos que es bastante similar.

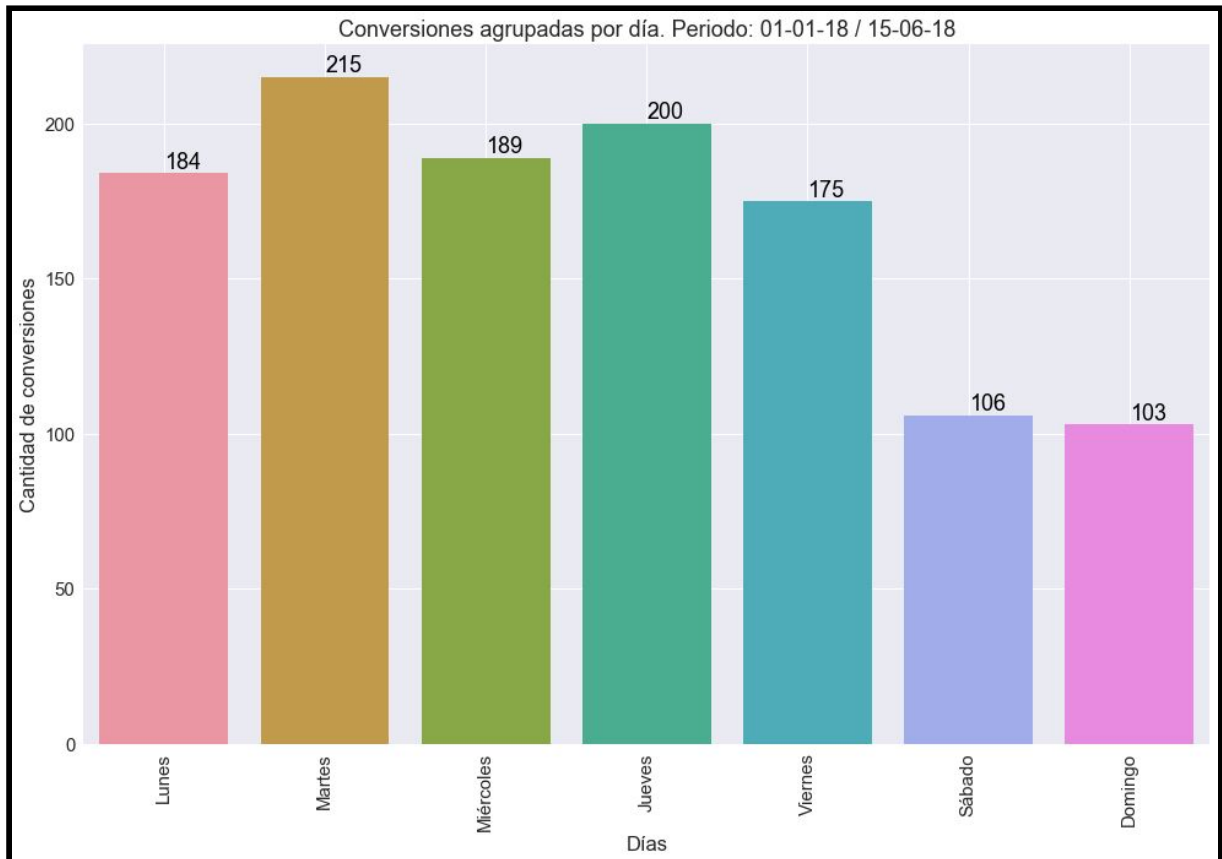
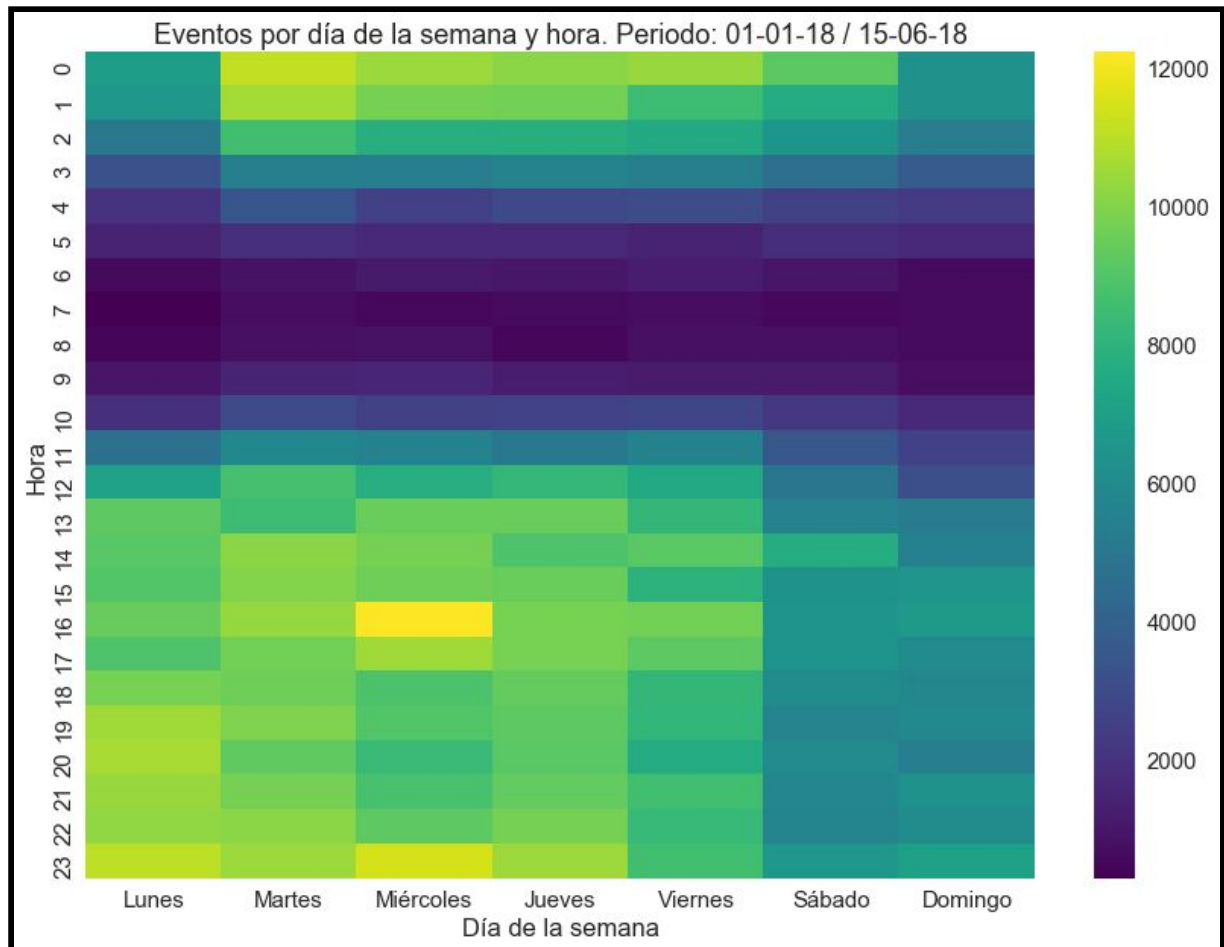


Gráfico 3ch.1

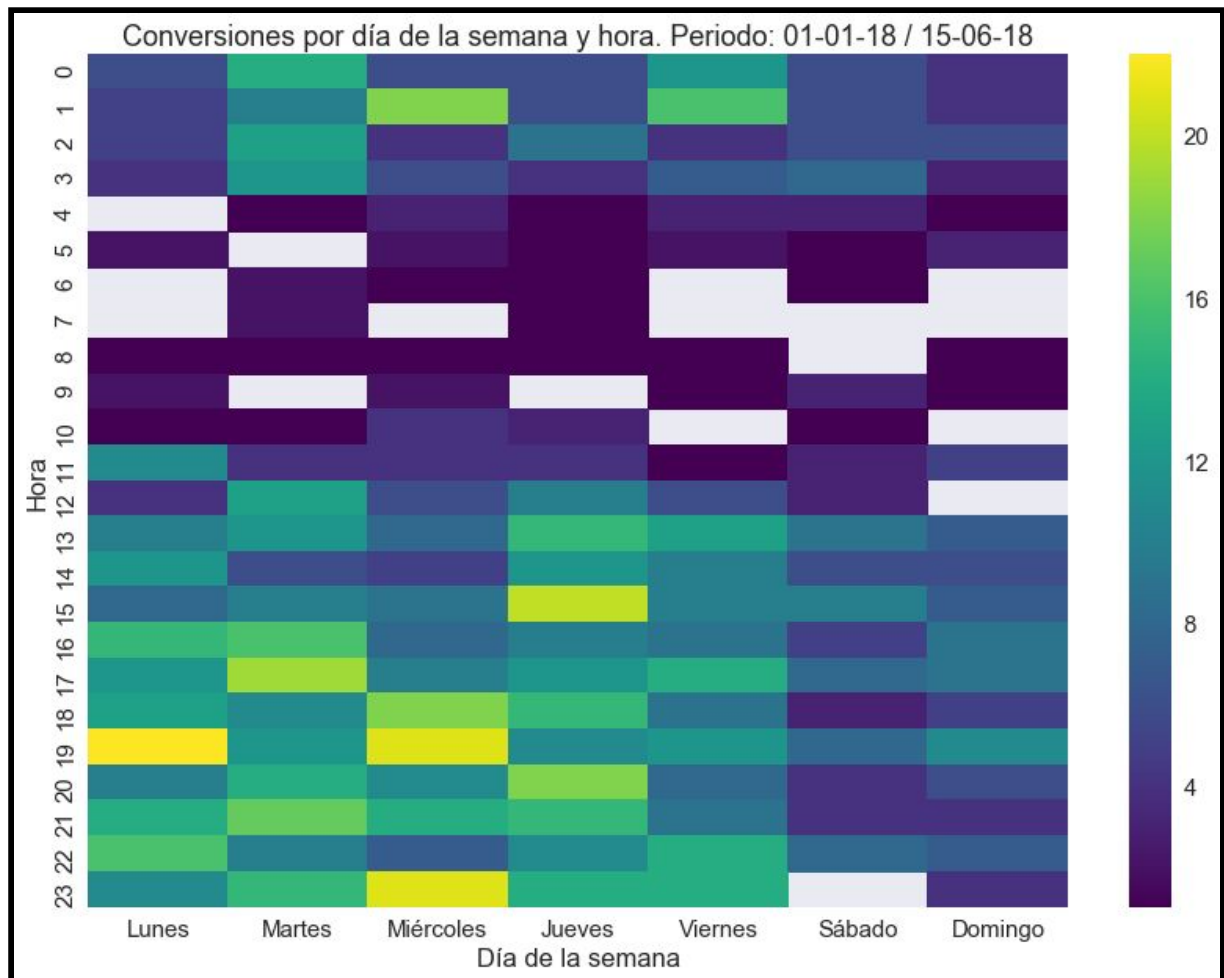
Los días que no muestra una diferencia entre la relación generación de eventos y conversiones, son los miércoles y los sábados.

3ci - Eventos según horario y día de la semana.



A través de esta visualización se desprende que la actividad en la plataforma es más intensa comienza el mediodía de los días Lunes a Jueves, y finaliza alrededor de la medianoche, los viernes comienza una merma en la misma y los fines de semana la actividad es baja.

3cj - Conversiones según horario y día de la semana.



En el caso de las conversiones, el mapa de calor informa similares resultados, en el mismo se nota que hay franjas horarias en las que no hay conversiones en el período en estudio (en blanco).

3c - Análisis Geográfico

3ca - Análisis por país.

Cuando comenzamos a ver los primeros valores de los países desde donde se ingresó vemos lo siguiente:

Brazil	84308
Unknown	2313
United States	465
Argentina	94
Canada	35
France	13
Netherlands	13
United Kingdom	12
Mozambique	12
Japan	11
Poland	9
French Guiana	8
Romania	7
Italy	7
Portugal	7
Russia	6
Dominican Republic	6
Guadeloupe	6
Germany	5
Paraguay	4
Colombia	4
Bangladesh	4
Switzerland	3

El resumen nos queda:

```
Ingresos totales: 87378 | Ingresos Brazil 84308 | Otros Ingresos 3070
```

Analizamos la cantidad de usuarios que se conectaron desde otro lugar que no sea Brasil y evaluamos el tipo de ingreso que realizaron. Luego, filtramos aquellos que hayan tenido alguna conversión realizada (esto nos indica que es de Brasil) y aquellos usuarios que alguna vez se conectaron desde Brasil, porque evaluamos que probablemente sean de allí. De esto nos da como resultado que:

```
La cantidad de inicios de sesión a eliminar: 4309 | registros a eliminar: 44160.
```

3cb - Análisis por región y ciudad.

Hacemos un listado rápido de las regiones:

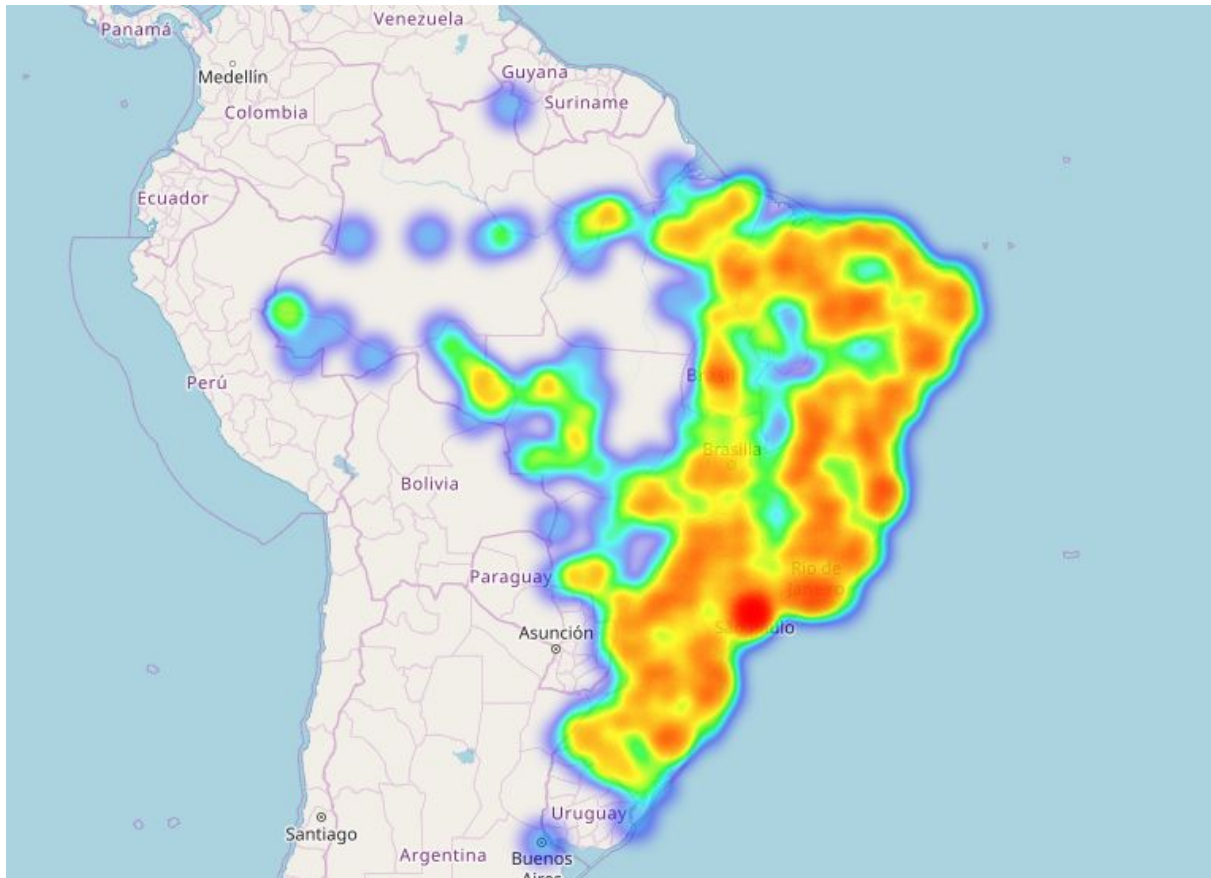
Sao Paulo	24611
Unknown	12495
Minas Gerais	7679
Rio de Janeiro	6785
Bahia	5632
Ceara	2769
Pernambuco	2755
Parana	2559
Rio Grande do Sul	2549
Federal District	1731
Goias	1552
Santa Catarina	1502
Para	1345
Maranhao	1267
Espirito Santo	1255
Rio Grande do Norte	1012
Piaui	810
Paraíba	756

Cantidad de Regiones: 40

Notamos que concentración de eventos, se encuentra en San Pablo, y hay muchísimos de origen desconocido.

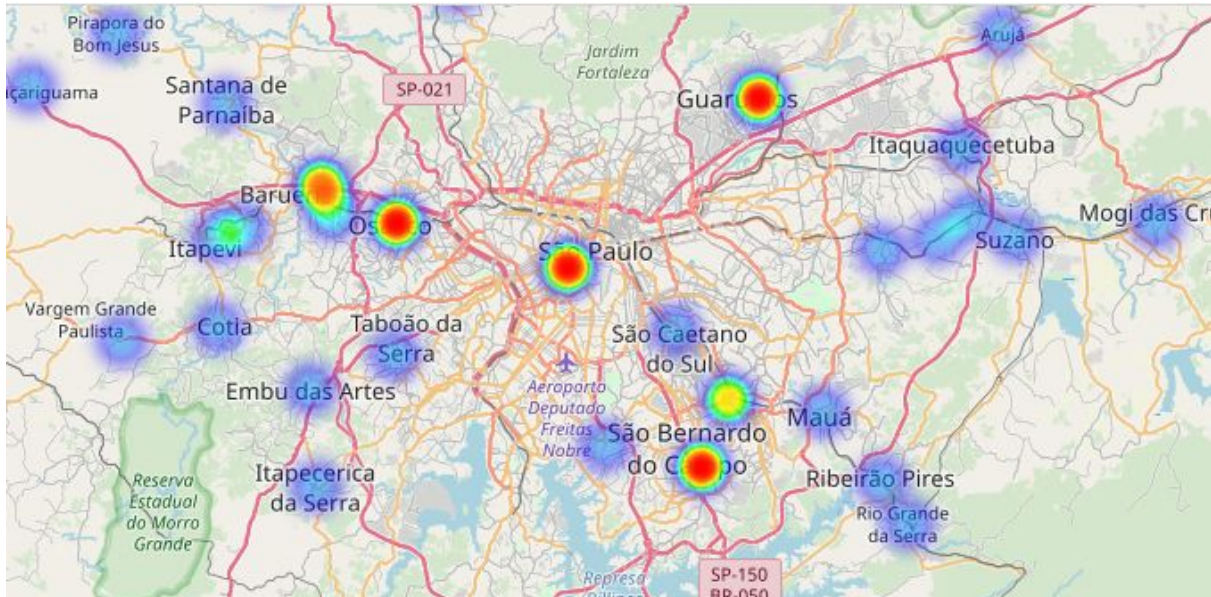
Lo primero que vamos a hacer son mapas de calor para ver si identificamos un patrón:

Mapa de calor de Generación de eventos



Vamos que la concentración está en Sao Pablo. Vemos que la mayor con respecto al resto del territorio.

Mapa de calor con Zoom sobre la zona de San Pablo. Eventos generados.



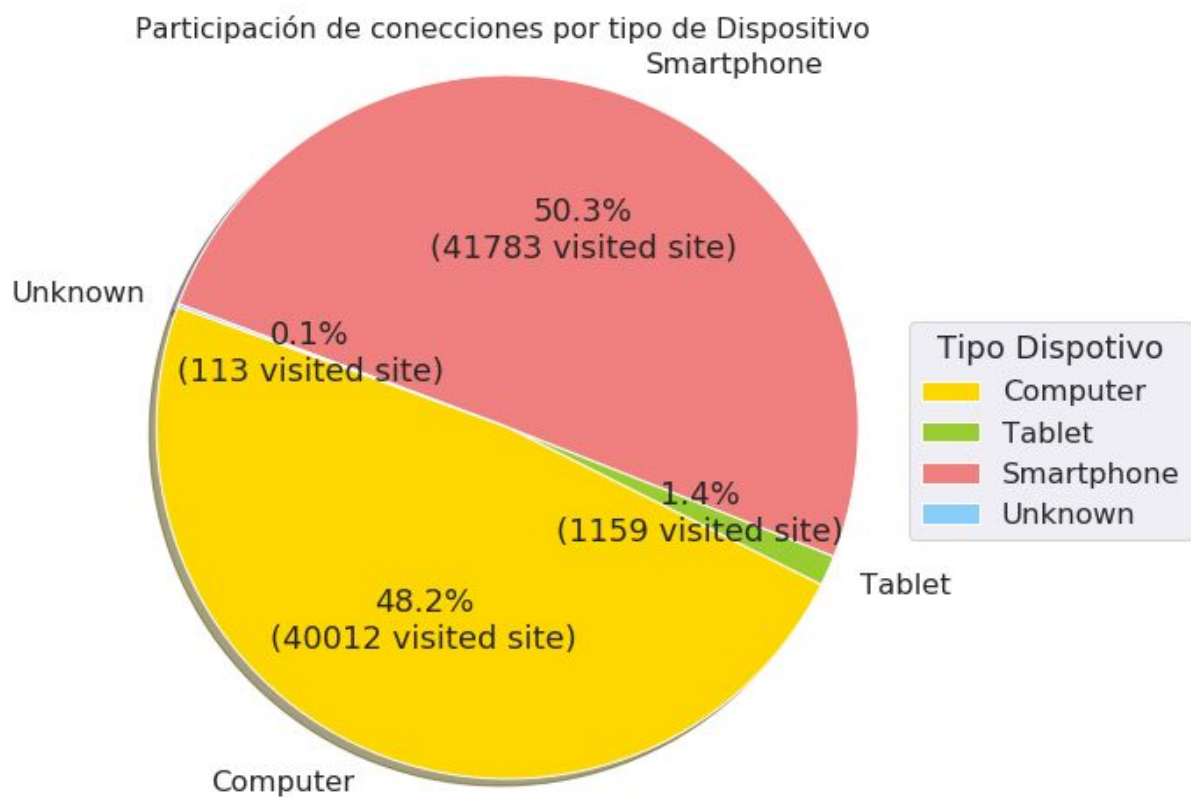
Mapa con los Quiosques (presencia física de Trocafone)



Se puede ver que existe una relación entre los lugares donde mayor cantidad de eventos se generan y los lugares físicos donde Trocafone dispuso Quiosques.

3cc - Tipo de dispositivo desde donde ingresan los usuarios.

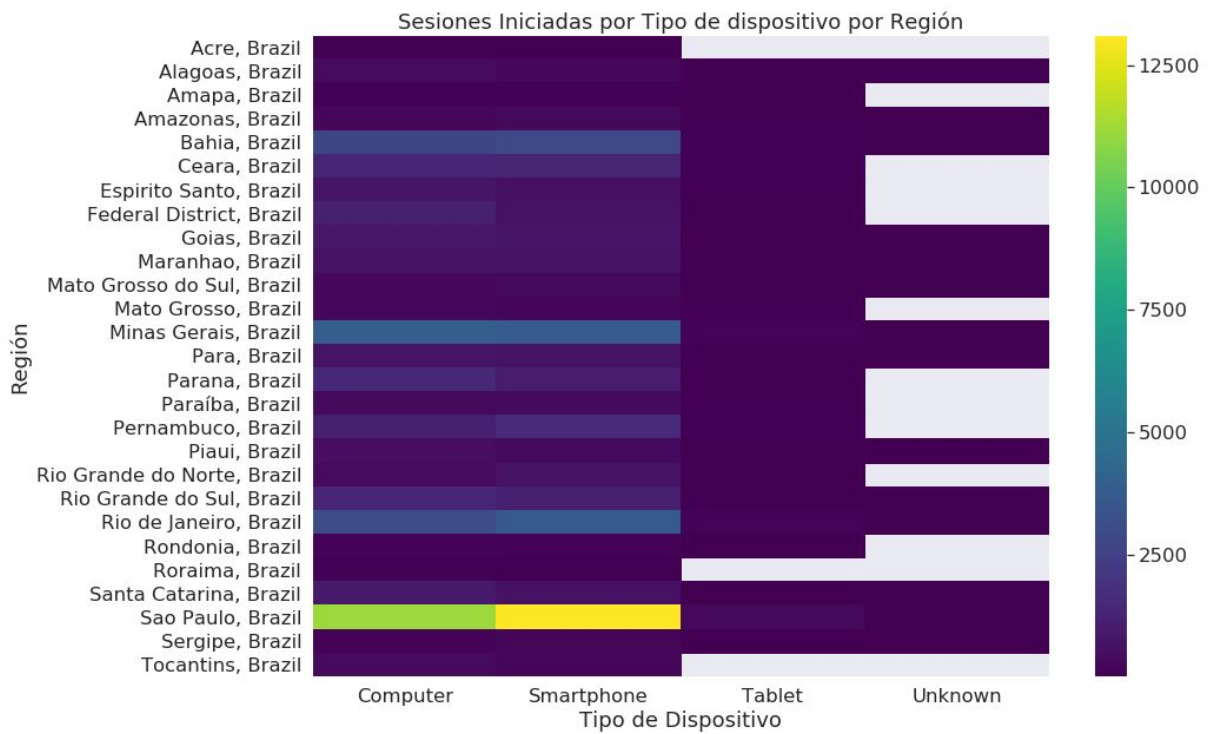
Vamos a realizar un análisis dependiendo del dispositivo desde donde se conectan.



Lo que estamos viendo es que los ingresos están muy parejos entre Computadoras y Smartphones.

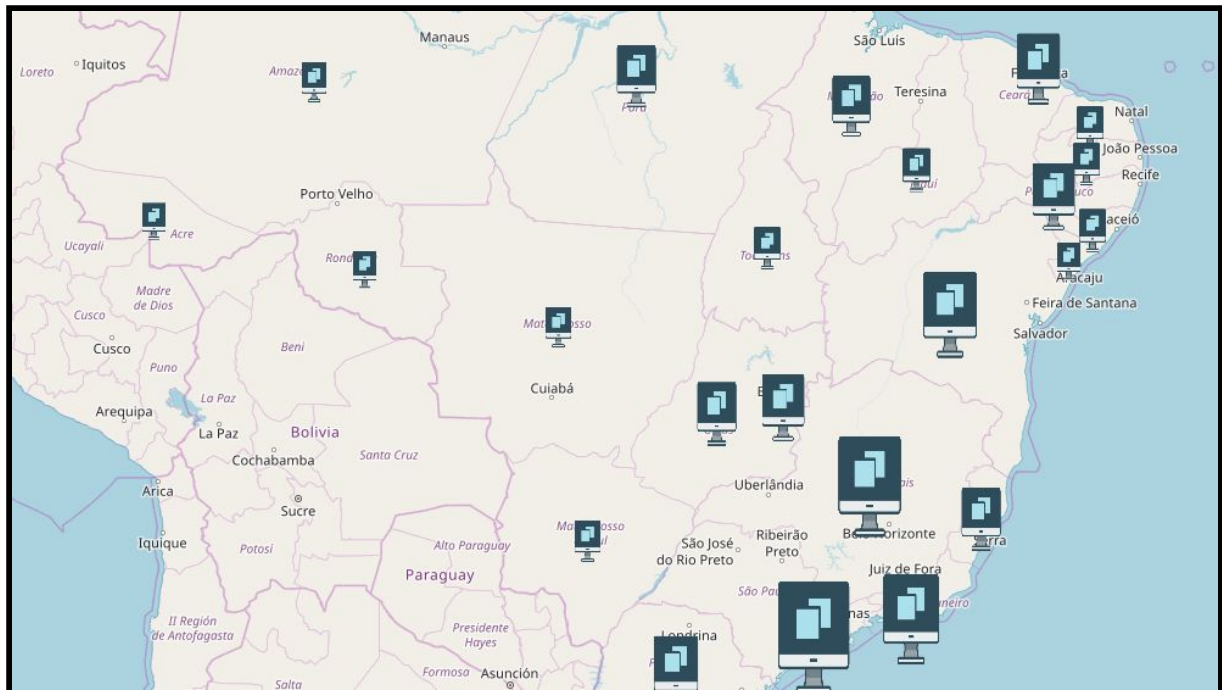
Los ingresos que se contabilizaron se realizaron a través del evento: 'visited site'. Ahora vamos a verlos desde un punto de vista geográfico.

Realizamos un gráfico HeatMap para mostrar donde estan los puntos más fuertes de conexión.



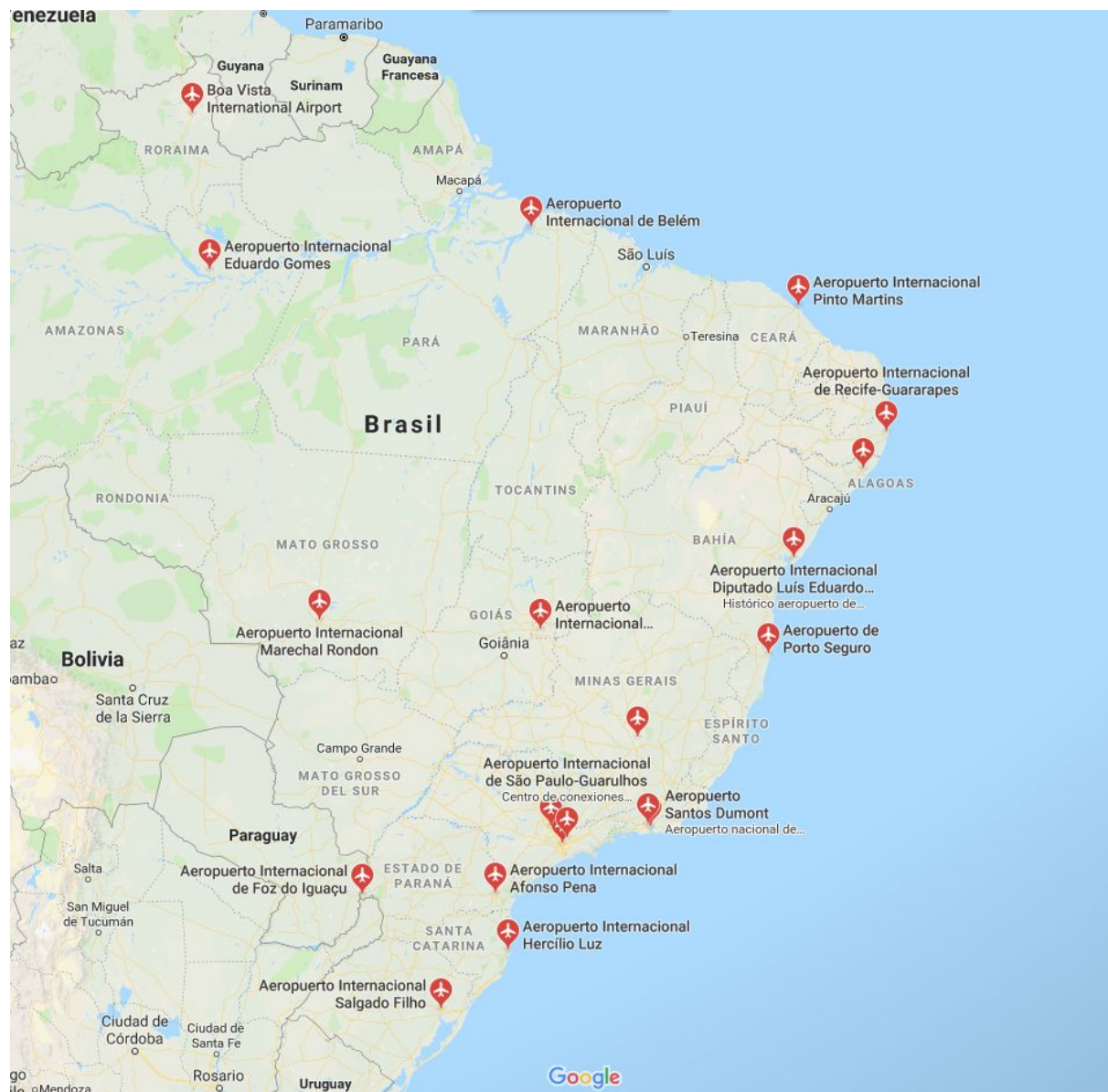
Entre las cosas interesantes podemos ver que en San Pablo un lugar donde hay mayor cantidad de conexiones y eventos generados las conexiones con smartphones es superior porcentualmente a la media. Igualmente en Rio de Janeiro. En el resto de los lugares es bastante homogéneo.

Distribución de conexiones desde Computadores en Brasil.



Distribución de conexiones desde Smartphone en Brasil.





4 - Análisis de Productos y Sesiones.

Vamos a ver en primera Instancia los productos y la cantidad de veces que aparecen:

iPhone 6	48475
iPhone 5s	43126
iPhone 6S	37646
iPhone 7	26307
Samsung Galaxy S7 Edge	22551
Samsung Galaxy S7	18478
iPhone 7 Plus	17886
Samsung Galaxy S8	16241
iPhone SE	15954
Samsung Galaxy J5	12512
iPhone 5c	12318
iPhone 6S Plus	12187
iPhone 6 Plus	12136
Samsung Galaxy S6 Flat	11204
Samsung Galaxy S6 Edge	10719
Samsung Galaxy S8 Plus	9938
iPhone 5	9014
Samsung Galaxy J7 Prime	8904
Motorola Moto G4 Plus	7789
Samsung Galaxy J7	6896
Motorola Moto X Play 4G Dual	6585
Samsung Galaxy A5 2017	5664
Samsung Galaxy A7 2017	5572
Samsung Galaxy S5	5317
iPhone 4S	5013

Ahora vamos a mostrar la cuenta de los productos que aparecen relacionados a los eventos:

	checkout	conversion	lead	viewed product
model				
Samsung Galaxy J5	1824.0	88.0	2.0	10598.0
iPhone 5s	2604.0	84.0	5.0	40433.0
iPhone 6	3135.0	71.0	11.0	45258.0
Samsung Galaxy S6 Flat	807.0	39.0	1.0	10357.0
iPhone 6S	2208.0	37.0	19.0	35382.0
iPhone 5c	571.0	33.0	4.0	11710.0
Motorola Moto G4 Plus	486.0	30.0	19.0	7254.0
Samsung Galaxy J7	749.0	28.0	2.0	6117.0
Samsung Galaxy J7 Prime	818.0	20.0	25.0	8041.0
Samsung Galaxy S5	256.0	18.0	0.0	5043.0
Samsung Galaxy S7	1166.0	18.0	0.0	17294.0
iPhone 5	450.0	17.0	3.0	8544.0
iPhone SE	595.0	17.0	4.0	15338.0
Samsung Galaxy S7 Edge	796.0	17.0	4.0	21734.0
Samsung Galaxy S6 Edge	541.0	16.0	2.0	10160.0
iPhone 7	1167.0	16.0	5.0	25119.0
Samsung Galaxy S8	1036.0	15.0	2.0	15188.0
iPhone 6S Plus	540.0	14.0	31.0	11602.0
Samsung Galaxy A5 2017	271.0	13.0	0.0	5380.0
Motorola Moto X Play 4G Dual	269.0	12.0	1.0	6303.0
iPhone 4S	323.0	10.0	2.0	4678.0
iPhone 6 Plus	598.0	8.0	33.0	11497.0
Samsung Galaxy S8 Plus	463.0	7.0	2.0	9466.0
Samsung Galaxy A7 2017	239.0	6.0	2.0	5325.0
iPhone 7 Plus	867.0	6.0	9.0	17004.0

Ahora vamos a sumar los eventos en un nuevo campo y ordenarlos a partir de esta suma de forma descendente:

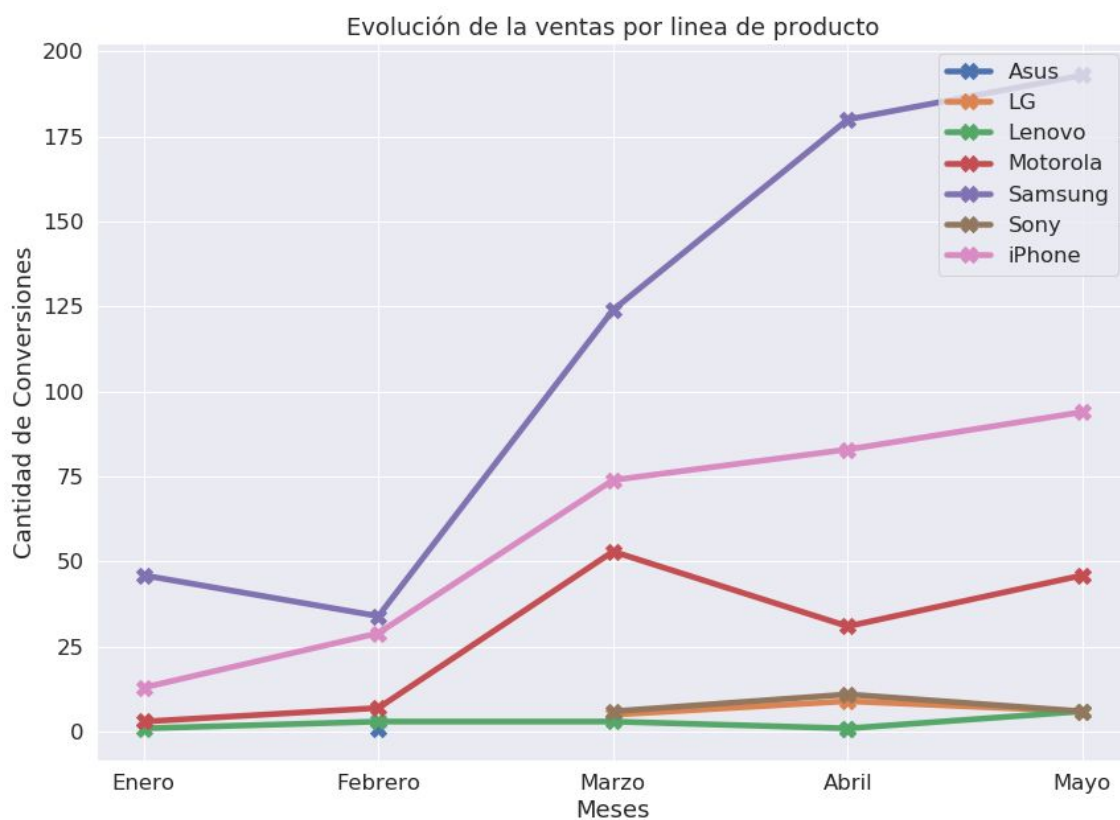
	checkout	conversion	lead	viewed product	total_events
model					
iPhone 6	3135.0	71.0	11.0	45258.0	48475.0
iPhone 5s	2604.0	84.0	5.0	40433.0	43126.0
iPhone 6S	2208.0	37.0	19.0	35382.0	37646.0
iPhone 7	1167.0	16.0	5.0	25119.0	26307.0
Samsung Galaxy S7 Edge	796.0	17.0	4.0	21734.0	22551.0
Samsung Galaxy S7	1166.0	18.0	0.0	17294.0	18478.0
iPhone 7 Plus	867.0	6.0	9.0	17004.0	17886.0
Samsung Galaxy S8	1036.0	15.0	2.0	15188.0	16241.0
iPhone SE	595.0	17.0	4.0	15338.0	15954.0
Samsung Galaxy J5	1824.0	88.0	2.0	10598.0	12512.0
iPhone 5c	571.0	33.0	4.0	11710.0	12318.0
iPhone 6S Plus	540.0	14.0	31.0	11602.0	12187.0
iPhone 6 Plus	598.0	8.0	33.0	11497.0	12136.0
Samsung Galaxy S6 Flat	807.0	39.0	1.0	10357.0	11204.0
Samsung Galaxy S6 Edge	541.0	16.0	2.0	10160.0	10719.0
Samsung Galaxy S8 Plus	463.0	7.0	2.0	9466.0	9938.0
iPhone 5	450.0	17.0	3.0	8544.0	9014.0
Samsung Galaxy J7 Prime	818.0	20.0	25.0	8041.0	8904.0
Motorola Moto G4 Plus	486.0	30.0	19.0	7254.0	7789.0
Samsung Galaxy J7	749.0	28.0	2.0	6117.0	6896.0
Motorola Moto X Play 4G Dual	269.0	12.0	1.0	6303.0	6585.0
Samsung Galaxy A5 2017	271.0	13.0	0.0	5380.0	5664.0
Samsung Galaxy A7 2017	239.0	6.0	2.0	5325.0	5572.0
Samsung Galaxy S5	256.0	18.0	0.0	5043.0	5317.0
iPhone 4S	323.0	10.0	2.0	4678.0	5013.0

Ventas totales: 1172 | Total de ventas 25 tipos productos: 640.0 | participación: 54.61%

Hacemos una agrupación por marca:

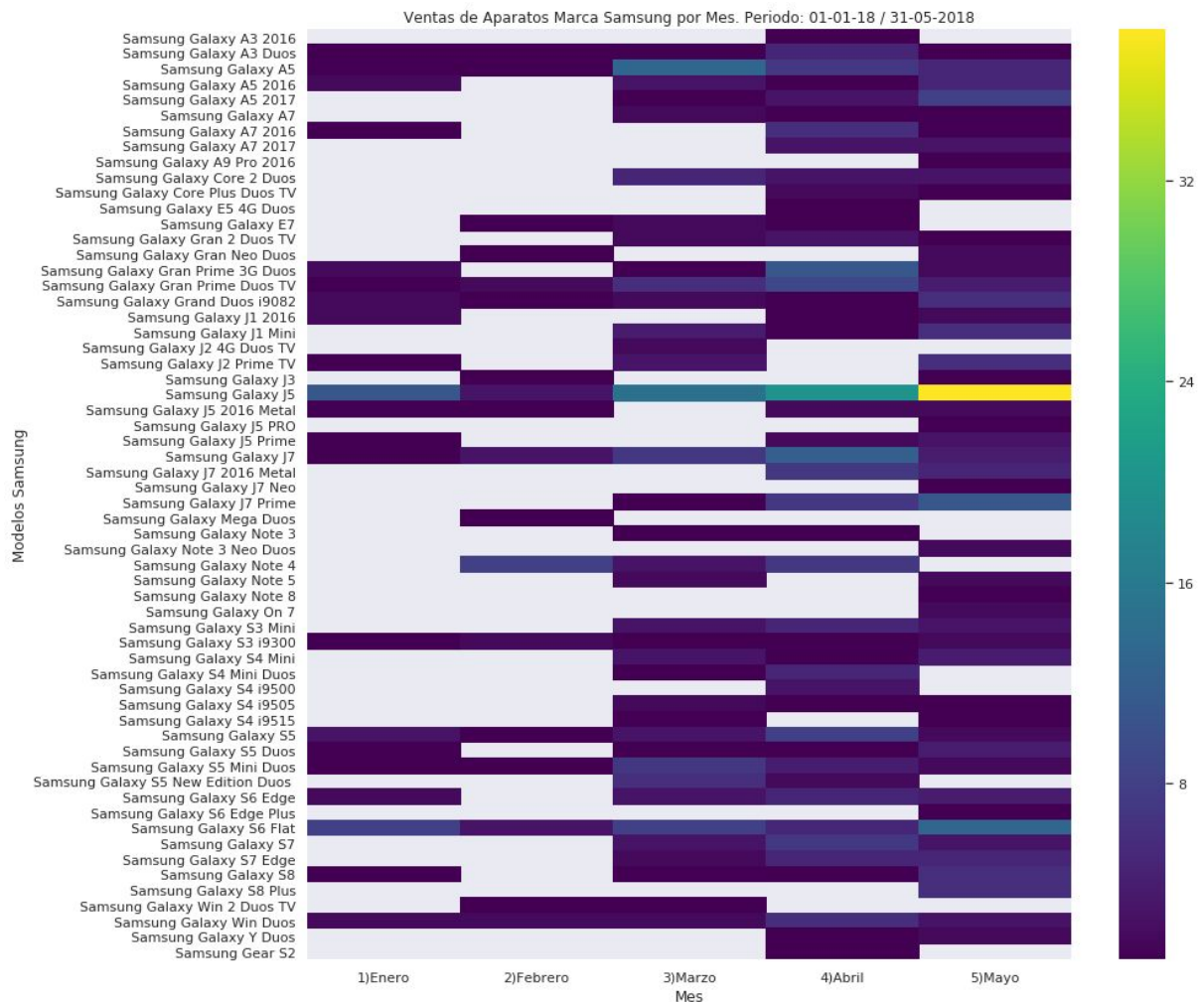
	checkout	conversion	lead	viewed product
Marca				
Samsung	13997.0	641.0	125.0	194040.0
iPhone	13415.0	323.0	157.0	231712.0
Motorola	3572.0	149.0	81.0	57431.0
Sony	212.0	23.0	2.0	6504.0
LG	533.0	20.0	43.0	8666.0
Lenovo	253.0	14.0	10.0	4265.0
Asus	60.0	2.0	3.0	1831.0
Blackberry	0.0	0.0	1.0	0.0
Quantum	15.0	0.0	4.0	468.0
iPad	70.0	0.0	3.0	1793.0

Evolución de Conversiones por Marca a lo largo de los meses:

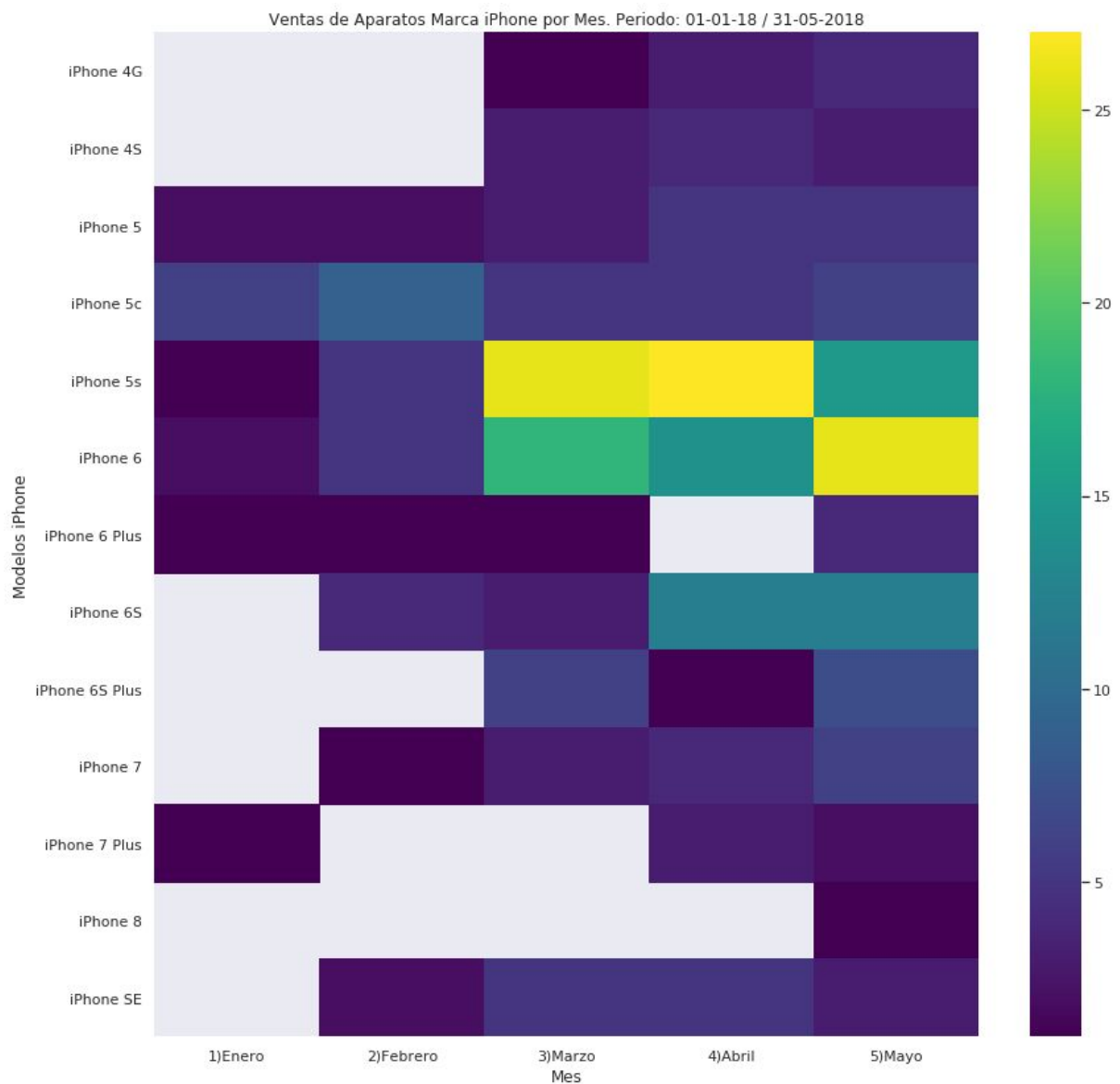


Vamos a analizar las marcas más compradas y su evolución en el tiempo según modelo.

Empezamos por Samsung. Quien claramente ha crecido mucho en las ventas. Y principalmente con su modelo Samsung Galaxy J5.



Veremos ahora el caso de la marca Apple (etiquetada como iPhone)



Vemos que predomina la venta del iPhone 5s y del iPhone 6.

Y el caso de la tercera empresa Motorola.

