

# Trabajo Práctico 1

[7506/9558] Organización de Datos  
Segundo cuatrimestre de 2018

Grupo Random Name

Alumno	Número de Padrón	Email
Peña, Alejandro	98529	alee.pena.94@gmail.com
Blázquez O., Sebastián A.	99673	sebastian.blazquez96@gmail.com
Cavazzoli, Federico	98533	fedecavazzoli@gmail.com
Ruiz, Francisco	99429	fran7ruiz9@gmail.com

Repositorio GitHub: <https://github.com/alepenaa94/TP1-Datos>

# Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Procesamiento y análisis de los datos</b>	<b>2</b>
2.1. Primera aproximación a los datos . . . . .	2
2.2. Visitantes según su lugar de residencia . . . . .	3
2.2.1. Países . . . . .	3
2.2.2. Región . . . . .	4
2.2.3. Ciudades . . . . .	6
2.3. Visitantes según el dispositivo . . . . .	10
2.3.1. Sistema Operativo utilizado . . . . .	11
2.4. Visitantes según tipo de usuario . . . . .	12
2.5. Eventos . . . . .	13
2.5.1. Comparación eventos Brasil-Resto del Mundo . . . . .	13
2.5.2. Eventos en Brasil . . . . .	14
2.5.3. Relación entre eventos . . . . .	16
2.5.4. Eventos y su relación según la hora . . . . .	18
2.5.5. Tipo de evento por producto . . . . .	19
2.5.6. Relación entre eventos según la marca . . . . .	20
2.6. Canales de ingreso . . . . .	20
2.6.1. Origen de Campaña . . . . .	21
2.7. Ventas . . . . .	22
2.7.1. Celulares más vendidos . . . . .	23
2.7.2. Celulares en carrito de compras . . . . .	23
2.7.3. Compras según dispositivo . . . . .	24
2.7.4. Ventas según S.O . . . . .	25
2.7.5. Ventas según el mes . . . . .	26
2.7.6. Ventas según el día . . . . .	26
2.7.7. Ventas según el horario . . . . .	27
2.7.8. Ventas según tipo de usuario . . . . .	28
2.7.9. Ventas según búsqueda . . . . .	29
2.7.10. Ventas según Lead's . . . . .	30
2.7.11. Ventas según apariciones en la Homepage . . . . .	31
2.7.12. Ventas según apariciones en listado específico de una marca . . . . .	33
<b>3. Conclusiones</b>	<b>35</b>

## 1. Introducción

En el siguiente informe se detalla un estudio de los eventos que suceden en el sitio [www.trocafone.com](http://www.trocafone.com), encargada de la compra, reacondicionamiento y venta de celulares usados y venta de dispositivos nuevos.

El objetivo es realizar un análisis de los datos proveídos por esta empresa para luego desarrollar un algoritmo de machine learning capaz de recomendar publicaciones a los usuarios maximizando la probabilidad de que este compre el producto, suponiendo que el objetivo principal de la empresa es maximizar las ventas de los celulares brindándoles a sus clientes las mejores opciones.

El análisis presentado se hizo utilizando Python3 con las siguientes librerías:

- Pandas
- Numpy
- Matplotlib
- Seaborn
- Folium
- Gmaps

A lo largo del informe se pueden ver distintas visualizaciones de los datos, las cuales fueron hechas con la intención de facilitar la comprensión de la información disponible.

Para la integración del trabajo se utilizó un repositorio GitHub: <https://github.com/alepenaa94/TP1-Datos> donde se pueden encontrar los notebooks utilizados tanto para la limpieza de los datos como para el análisis exploratorio.

## 2. Procesamiento y análisis de los datos

### 2.1. Primera aproximación a los datos

El set de datos brindado por Trocafone es sobre una muestra representativa de usuarios que generaron eventos en la plataforma y fueron registrados por la web analytics del sitio. Se trabajó con el archivo **events.csv** que nos brindaba diferente información dependiendo del evento que se haya registrado.

A partir de una limpieza del set, observamos que en general los datos seleccionados estaban en "buen estado" permitiéndonos trabajar casi con la totalidad de los eventos registrados, exceptuando con algunos que contenían datos nulos o que no nos daban información importante para el análisis.

Sin embargo para realizar un buen análisis de los datos fue necesario examinar en detalle con que tipo de datos se estaban trabajando (cadenas de caracteres, números, categóricos, etc.). El análisis completo del mismo se puede ver en detalle en <https://github.com/alepenaa94/TP1-Datos/blob/master/limpieza.ipynb> y a modo de resumen mostramos el resultado para cada uno de los atributos:

atributo	tipo de dato	con nulos
event	categórico (13 categorías)	no
person	categórico (27624 categorías)	no
url	categórico (227 categorías)	sí
sku	categórico (2205 categorías)	sí
model	categórico (202 categorías)	sí
condition	categórico (5 categorías)	sí
storage_gb	categórico (8 categorías)	sí
color	categórico (61 categorías)	sí
staticpage	categórico (14 categorías)	sí
campaign_source	categórico (24 categorías)	sí
search_engine	categórico (4 categorías)	sí
channel	categórico (6 categorías)	sí
new_vs_returning	categórico (2 categorías)	sí
city	categórico (1938 categorías)	sí
region	categórico (92 categorías)	sí
country	categórico (45 categorías)	sí
device_type	categórico (3 categorías)	sí

operating_system_version	categórico (121 categorías)	sí
browser_version	categórico (343 categorías)	sí
screen_resolution_width	numérico	sí
screen_resolution_height	numérico	sí
skus	objeto (cadena de caracteres)	sí

## 2.2. Visitantes según su lugar de residencia

### 2.2.1. Países

Aquellos registros donde el evento era la visita al sitio se almacenaba, además de la fecha y el usuario, el lugar desde donde se realizaba la acción.

Cuando se comenzó a analizar los países donde vivían los visitantes de Trocafone, nos dimos cuenta que tiene una gran proporción de los mismos en su país de origen, Brasil, donde aproximadamente residen el 96 % de los visitantes.



Figura 1: Visitantes según país de residencia

El 4 % restante se reparte, principalmente, entre aquellos que no tienen información precisa de donde proviene el visitante y Estados Unidos. Muy por detrás se puede observar a Argentina como el tercer país donde se registraron más visitas. En este caso, se consideró dejar a aquellos registros donde su paradero es desconocido porque supusimos que eran eventos por fuera de Brasil, ya que también hemos observado que había muchos eventos de este país en los cuales no se conocía su ciudad y región pero sí que pertenecían al país latinoamericano.

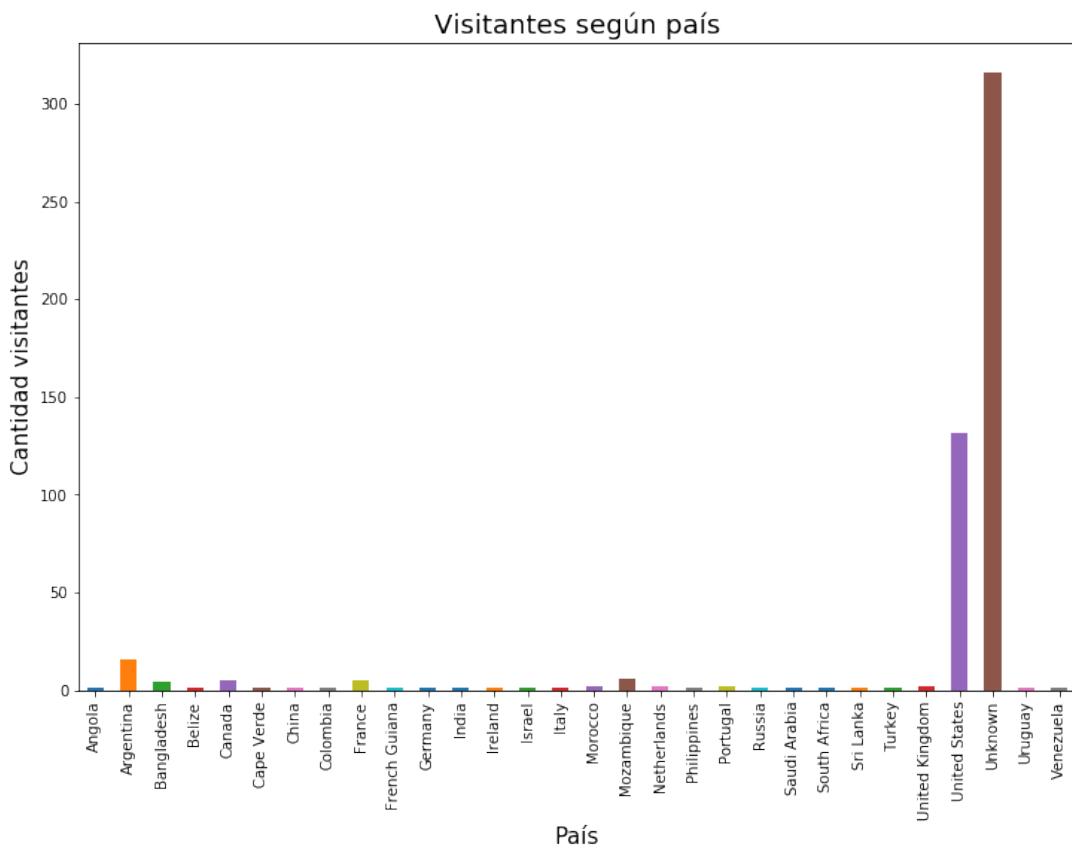


Figura 2: Visitantes según país de residencia, exceptuando Brasil

Además se puede observar que Trocafone tiene muchos países en los cuales su sitio tuvo visitas, se pueden considerar pocas respecto a las cantidades que manejan Brasil, Estados Unidos, e incluso, Argentina. Pero pueden ser lugares donde ya se conoce a la empresa pero capaz falta explotar esos posibles mercados.

### 2.2.2. Región

Como se vio en la sección anterior el 96 % de las visitas corresponden a Brasil, por lo tanto al analizar las regiones de las cuales provienen dichas visitas se considerarán únicamente las de dicho país. Esto es, en parte, para simplificar las visualizaciones y para realizar un análisis más profundo del principal mercado de *Trocafone*. Para este tipo de análisis realizamos un mapa coroplético que nos permite visualizar desde qué regiones hay más visitas:



Figura 3: Visitantes según región de residencia, dentro de Brasil

De esta forma se puede apreciar que claramente São Paulo es la región con más visitas de todas, seguida de Rio de Janeiro y Minas Gerais. Es notorio que estos tres estados también son los más poblados de Brasil (<https://www.worldatlas.com/articles/brazilian-states-by-population.html>).

Debido a esto resulta interesante analizar qué porcentaje de la población visita la página (por cada región), lo cual se muestra a continuación:

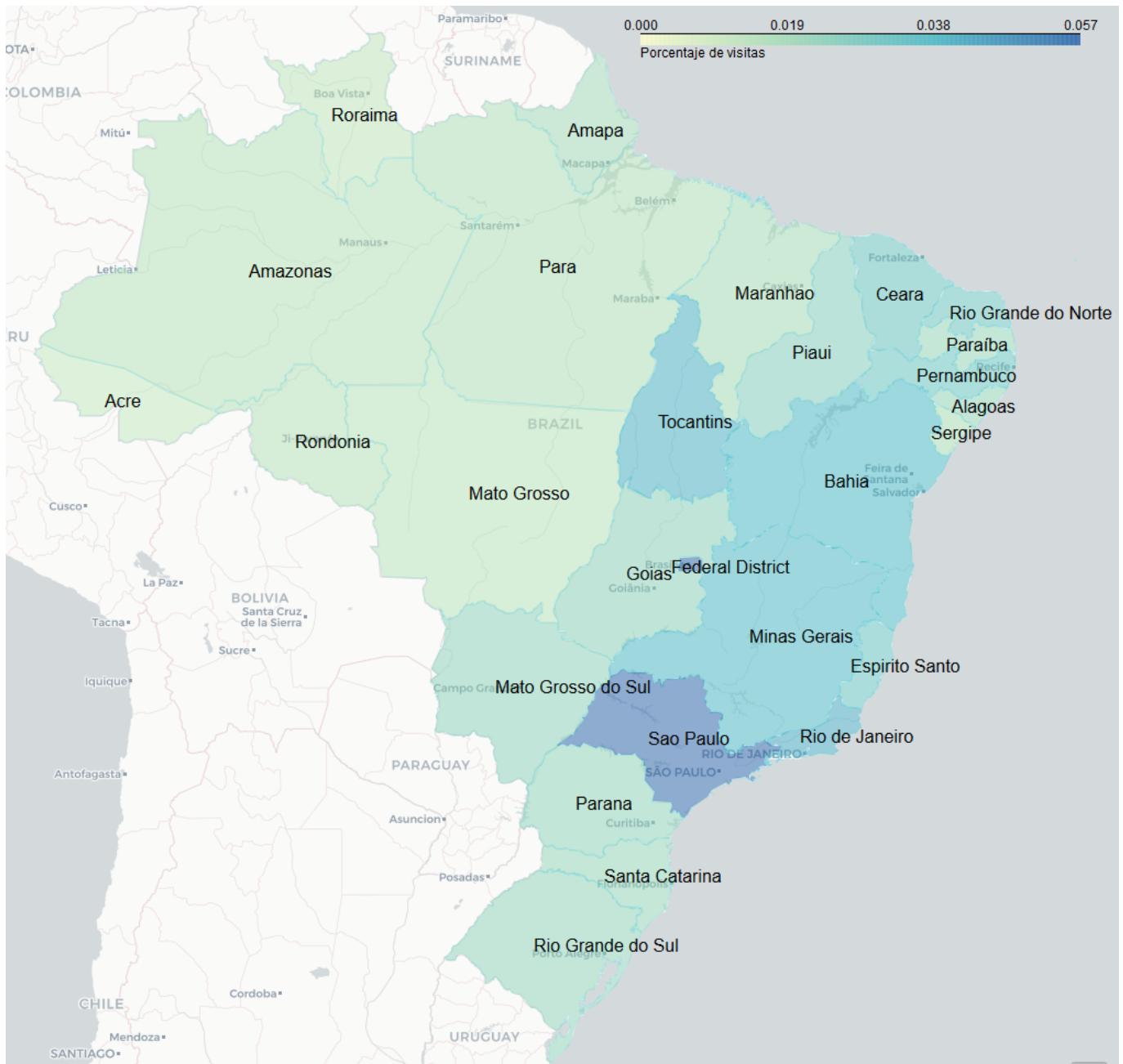


Figura 4: Porcentaje de visitantes según región de residencia, dentro de Brasil

Lo primero que debemos notar es que se modificó la escala para que el valor máximo sea el de mayor porcentaje (aproximadamente 0.057 %), debido a que los mismos eran muy bajos y si se utilizaba una escala de 0 a 100 no se podría apreciar nada.

En primera medida mirando ambos mapas notamos que São Paulo es la región tanto con mayor porcentaje como con mayor cantidad de visitas, pero si analizamos el segundo puesto nos encontramos con una sorpresa. Del primer mapa se ve que Minas Gerais y Rio de Janeiro tienen una cantidad similar de visitas, pero en porcentaje es bastante mayor la de Rio de Janeiro.

A su vez notamos que claramente las regiones del sur y este tienen un mayor porcentaje, lo cual no quedaba claro en el primer mapa. Esto puede explicarse dado que la zona norte del país, al ser la que mayor porcentaje de pobreza tiene, representa una baja proporción de los visitantes a *Trocafone*.

### 2.2.3. Ciudades

Debido a que los datos también proveían las ciudades se pudo realizar un heatmap. Para ello primero se buscó las coordenadas de cada ciudad en un set de datos de <http://www.geonames.org> (hay más información sobre qué crite-

rios se utilizaron en [https://github.com/alepenaa94/TP1-Datos/blob/master/visualizaciones\\_geograficas.ipynb](https://github.com/alepenaa94/TP1-Datos/blob/master/visualizaciones_geograficas.ipynb), y luego se procedió a realizar la visualización con gmaps (que utiliza la api de google maps). El resultado obtenido fue el siguiente:

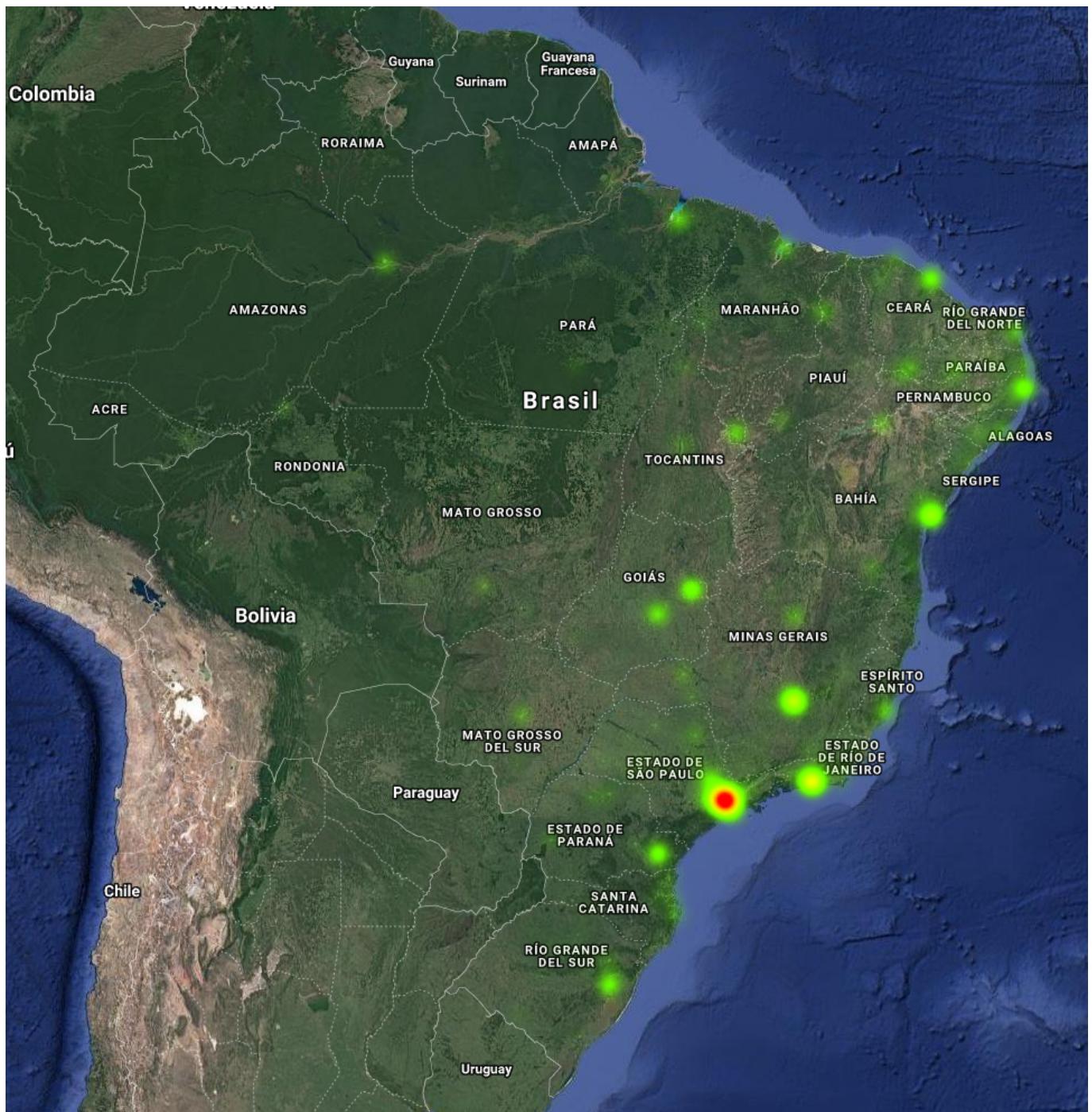


Figura 5: Heatmap de las visitas en Brasil

Claramente vemos que en el Estado de Sao Paulo existe la mayor concentración, lo cual es consistente con los análisis anteriores. Pero lo interesante es que vemos claramente que dentro del estado de Sao Paulo las visitas se concentran en una zona costera. A su vez notamos que esto ocurre en la mayoría de los estados costeros, con lo cual nuestra conclusión previa (dónde observábamos que la mayor parte de las visitas se encontraban en las regiones este y sur) puede refinarse viendo que la mayor concentración de visitas se encuentran en las ciudades costeras. Esto puede entenderse dado que el mayor comercio y desarrollo social de Brasil se encuentra cerca del océano Atlántico.

Hay algunas excepciones interesantes, como en el estado de Minas Gerais y Goias. Debido a que el heatmap es interactivo podemos acercarnos y ver qué ciudades son:



Figura 6: Heatmap Goias



Figura 7: Heatmap Minas Gerais

Al hacer esto encontramos una explicación: en la figura 6 podemos ver que se encuentra la capital de Brasil (Brasilia) y en la figura 7 vemos la ciudad Belo Horizonte, la capital de Minas Gerais, lo cual explica la gran cantidad de visitas.

Luego resulta interesante acercarnos a la ciudad con mayor cantidad de visitas, Sao Paulo:



Figura 8: Heatmap de São Paulo

Aquí notamos que ciertas ciudades aledañas a São Paulo (Osasco, Guarulhos y São Bernardo do Campo) también

aportan a la cantidad de visitas del estado de São Paulo en buena medida. Podemos medir la distancia desde São Paulo hasta ellas:



Figura 9: Distancias desde Osasco, Guarulhos y São Bernardo do Campo a São Paulo

De esta manera notamos que están realmente cerca (la distancia desde la sede de la Facultad de Ingeniería en Paseo Colón a Martínez es mayor), por lo cual es lógico contarlas como una cuando evaluamos las visitas a un nivel más general aunque en el set de datos vienen por separado. Esto muestra las bondades del heatmap, que agrupa los datos cuando los vemos desde lejos y los separa cuando queremos ser más específicos.

Por último mostramos una de las visualizaciones más interactivas, que muestra las ciudades con al menos una visita:

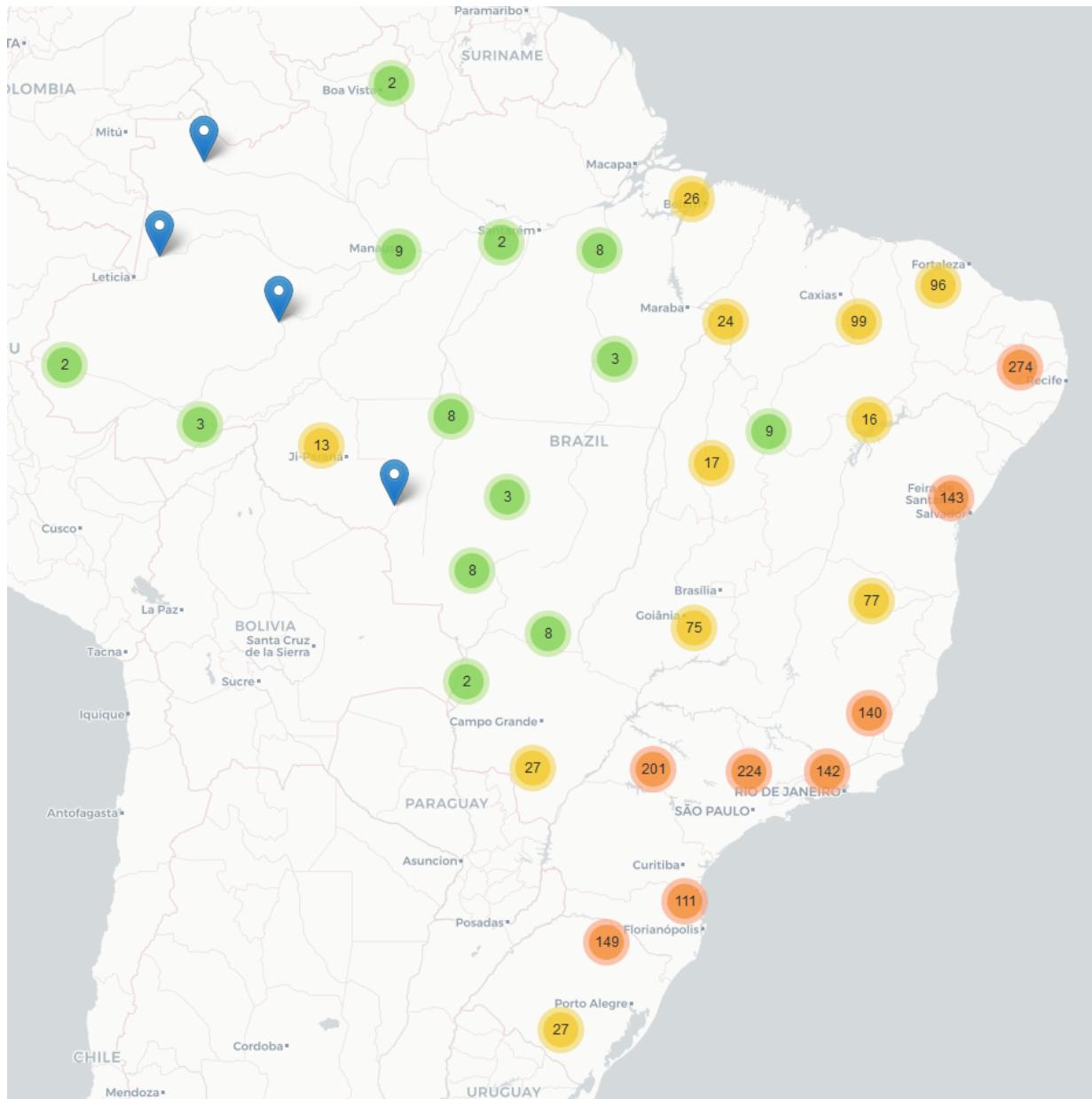


Figura 10: Ciudades con al menos una visita

En las regiones donde hay muchas ciudades con visitas los marcadores se agrupan formando clusters (los círculos de colores) que muestran la cantidad de ciudades que agrupan. Los verdes agrupan pocas ciudades, los amarillos una cantidad media y los rojos muchas. A su vez a medida que hacemos zoom los clusters se separan y podemos ver los marcadores.

De esta manera notamos nuevamente que las regiones costeras agrupan muchas más ciudades con visitas, corroborando las conclusiones anteriores. Para más detalles es posible interactuar directamente con el mapa y ver como se separan los clusters en [https://github.com/alepenaa94/TP1-Datos/blob/master/visualizaciones\\_geograficas.ipynb](https://github.com/alepenaa94/TP1-Datos/blob/master/visualizaciones_geograficas.ipynb), dado que al clickear en los marcadores podemos ver el nombre de la ciudad y la cantidad de visitas de la misma.

### 2.3. Visitantes según el dispositivo

Ya vimos de que parte del mundo son los visitantes y de que manera estos llegan al sitio de Trocafone. Debido a que tenemos información acerca de qué tipo de dispositivos utilizan los visitantes, nos pareció interesante analizarlo:

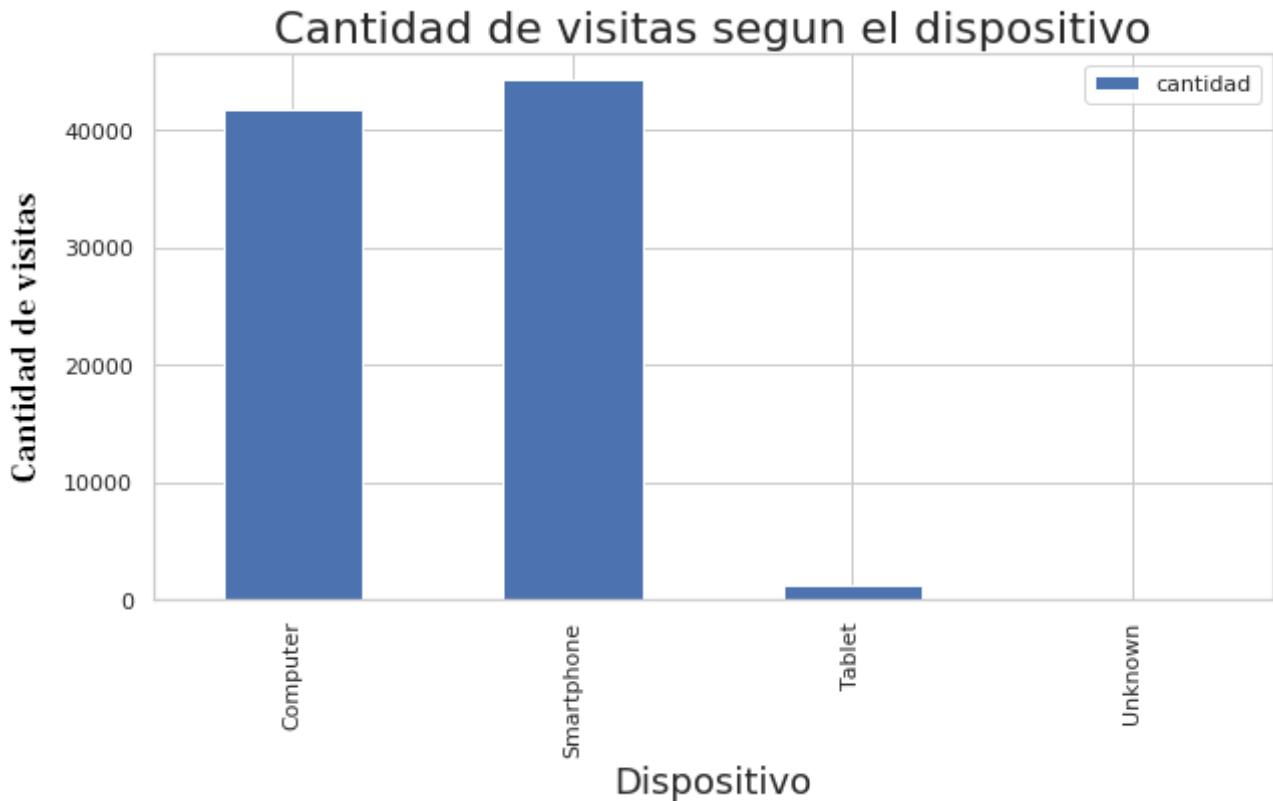


Figura 11: Cantidad de visitas según el dispositivo

Como vemos, los smartphones son los que más visitantes le brindan a la página por una poca diferencia por sobre las computadoras, siendo estas dos plataformas tecnológicas las que brindan casi la totalidad de los visitantes. Por último, las tablets son el dispositivo menos utilizado para ingresar a la sitio web. Igualmente a pesar de lo observado, más adelante en la sección donde se analizan las ventas de la página, se podrá observar que la mayoría de las compras es realizada por medio de una computadora.

### 2.3.1. Sistema Operativo utilizado

Resulta interesante saber con qué sistemas operativos se conecta la gente con el sitio. En el siguiente gráfico de torta puede observarse una amplia mayoría por parte de sistemas Android y iOS lo cual implica que este servicio es utilizado en gran medida desde dispositivos móviles, seguido por el uso con ordenadores de escritorio predominando en estos el sistema operativo de Mac.

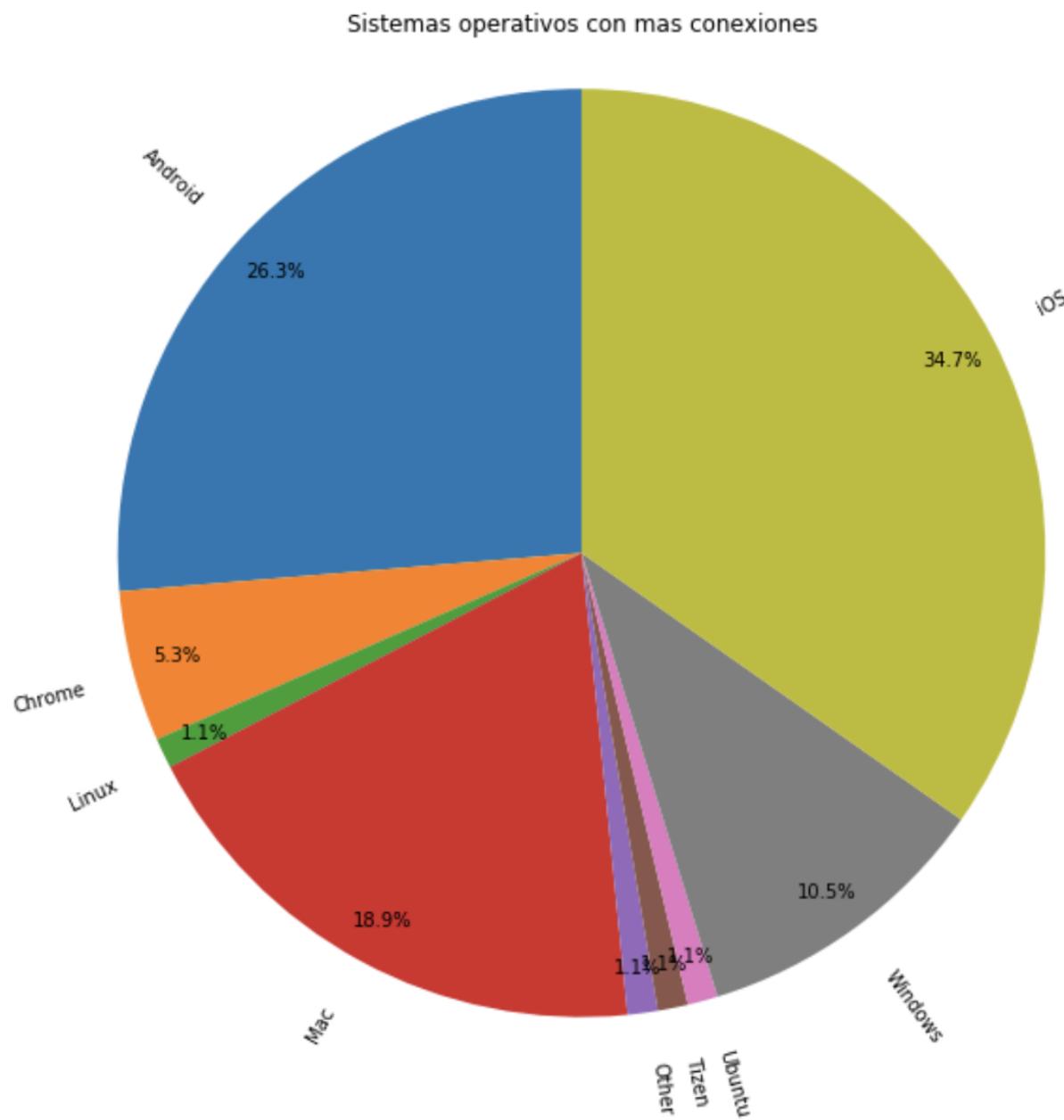


Figura 12: Sistemas operativos

#### 2.4. Visitantes según tipo de usuario

En esta sección se buscó analizar una característica particular de los usuarios, que define si es un usuario nuevo o no. Buscamos analizar que sucedía con esta característica durante los días de semana, de forma que podríamos ver cuantos usuarios nuevos por día ingresan aproximadamente al sitio.

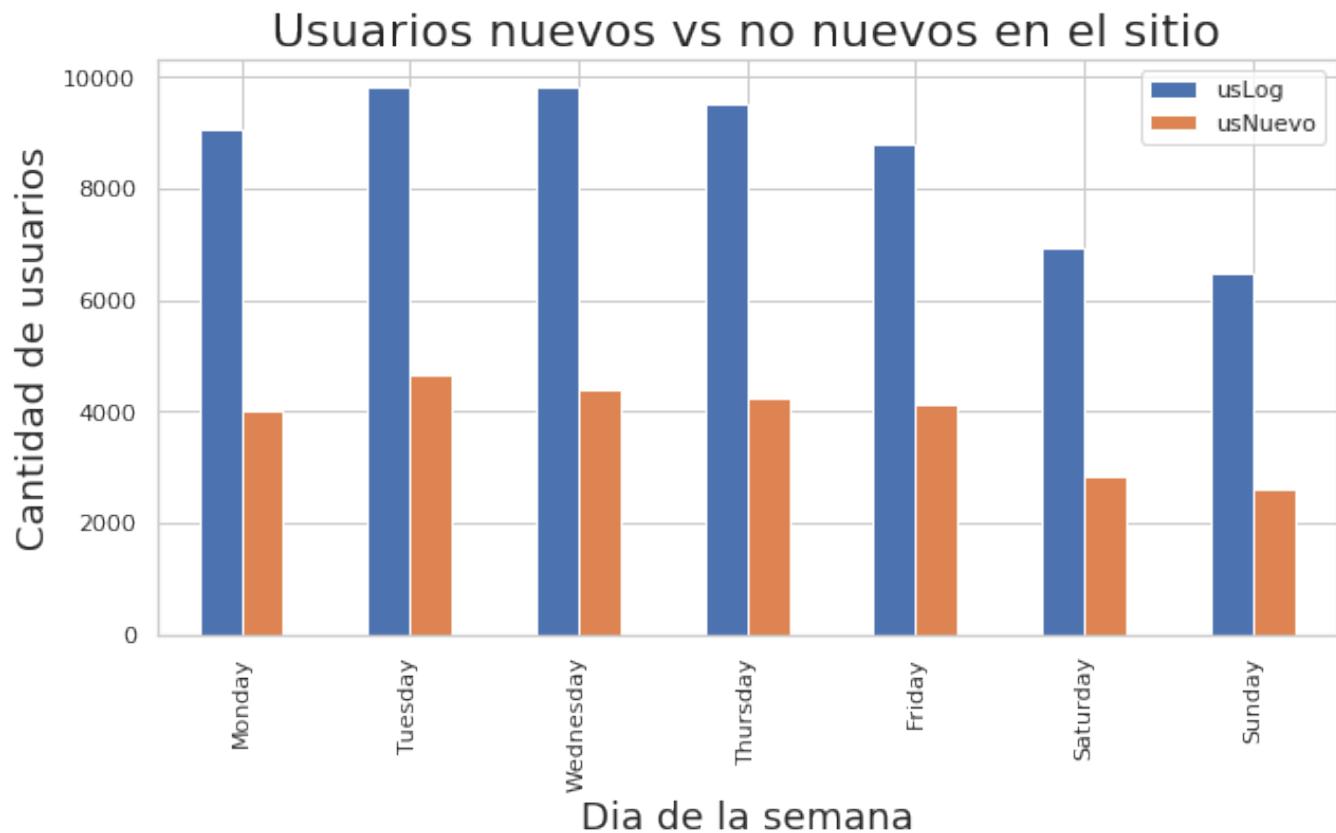


Figura 13: Comparación de usuarios nuevos y que retornan durante la semana

Como anteriormente se ha observado, la actividad en fin de semana disminuye, incluyendo al registro de usuarios nuevos. Por otro lado, vemos que suelen ser (los usuarios nuevos) por día aproximadamente la mitad de usuarios ya existentes. Esto da un indicio de que la plataforma crece y se está haciendo mas conocida por otros usuarios.

## 2.5. Eventos

### 2.5.1. Comparación eventos Brasil-Resto del Mundo

Como vimos en la Figura 1, la cantidad de visitantes de Brasil contra el resto del mundo es abrumadora. Pero analicemos si también esta mayor cantidad de visitantes se ve reflejada en los distintos tipos de eventos que se registran en la página:

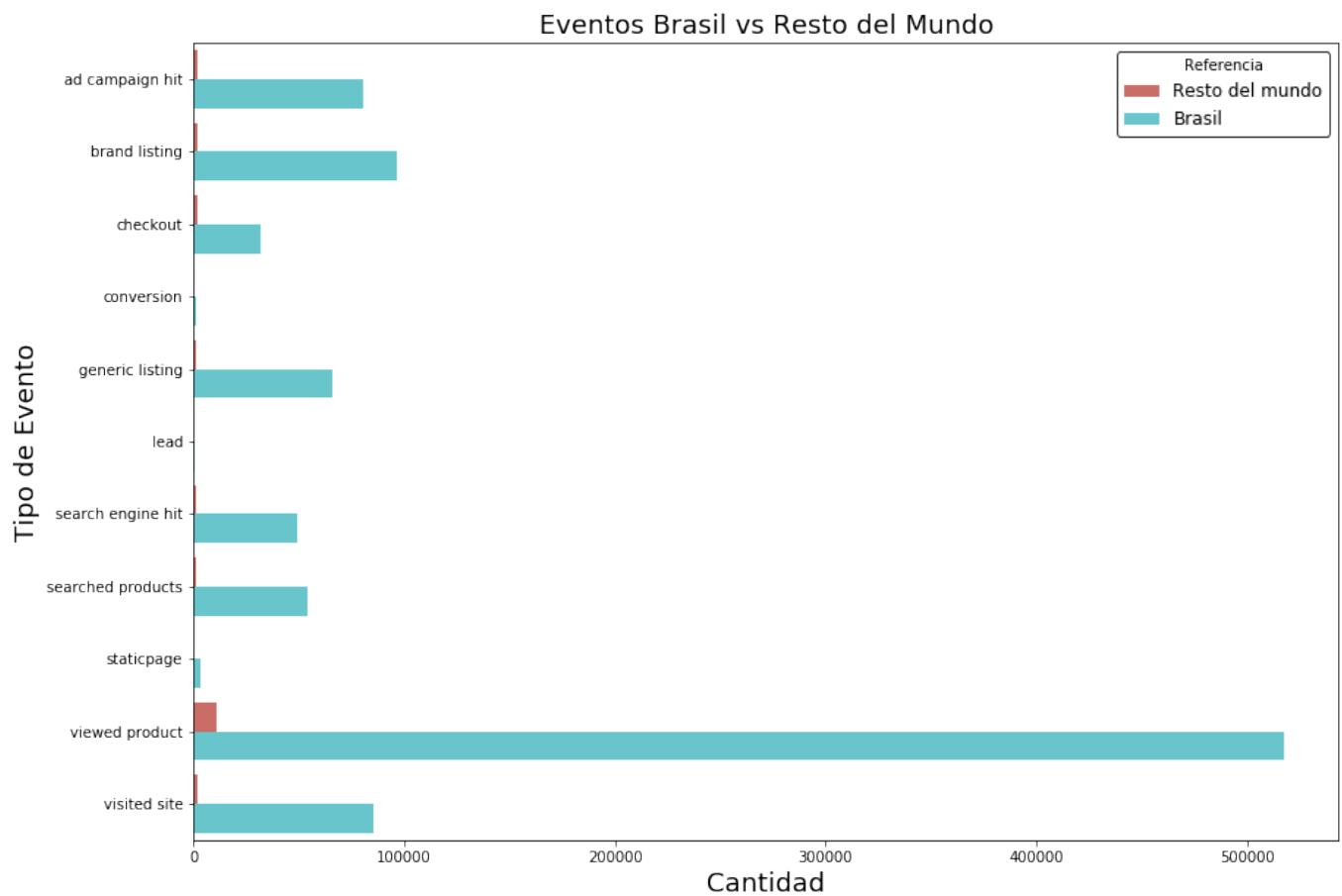


Figura 14: Cantidad de eventos Brasil vs Resto del Mundo

Fácilmente, se puede observar que en absolutamente todas las categorías la mayoría de los eventos son de origen brasileño, exceptuando los eventos *lead* y *conversion* donde hay tan pocos registros que no se pueden apreciar.

Por otro lado, las vistas de un producto tienen una cantidad de ocurrencias mucho mayor que a cualquier otro tipo de evento tanto para aquellos generados en Brasil como para los que se encuentran en el resto del mundo.

### 2.5.2. Eventos en Brasil

La sección anterior nos muestra que el principal nicho de mercado de Trocafone se encuentra en el gigante de América Latina y que en el resto del mundo no tiene un peso importante. Por lo tanto, veamos la comparación de los eventos que suceden en Brasil de manera ordenada:

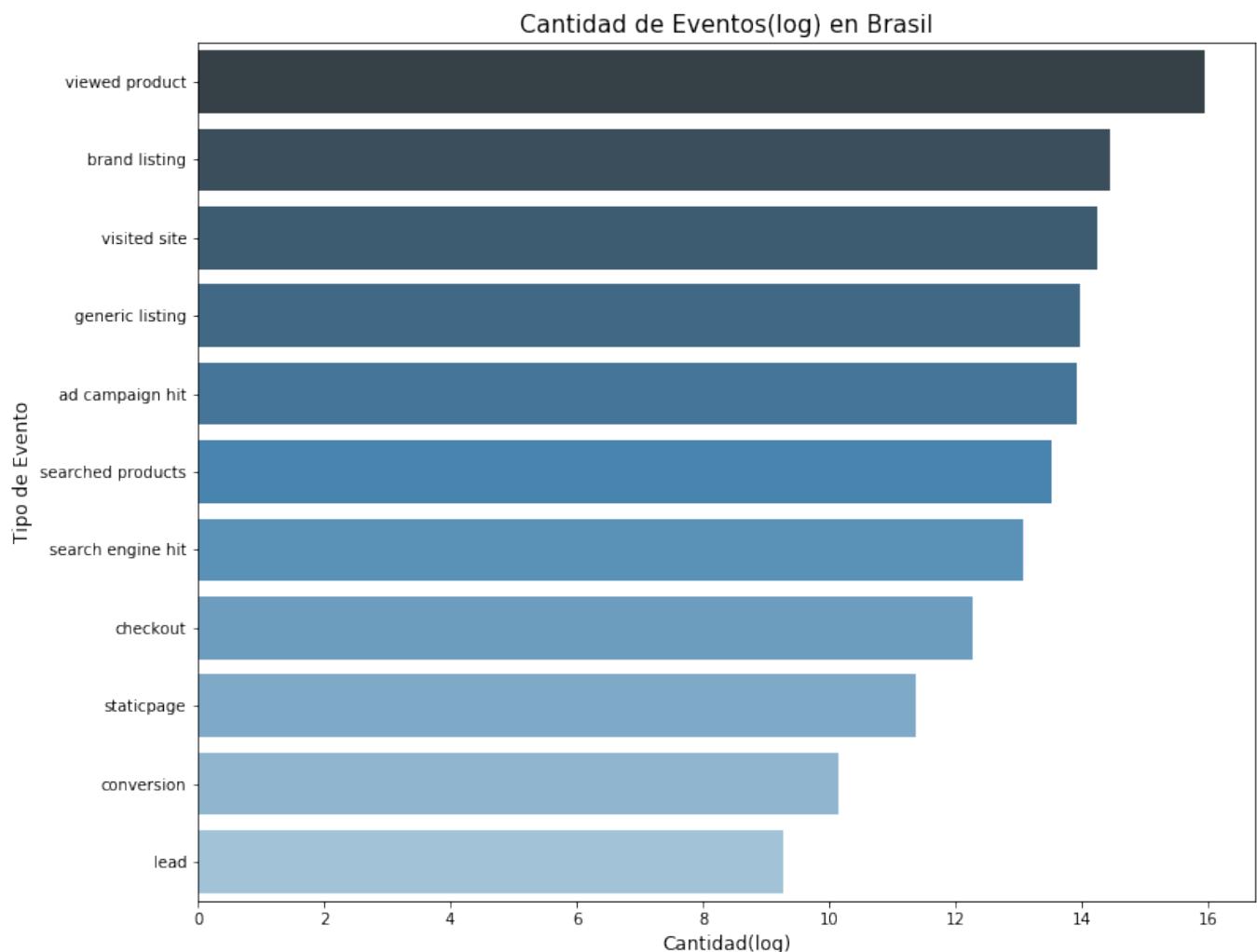
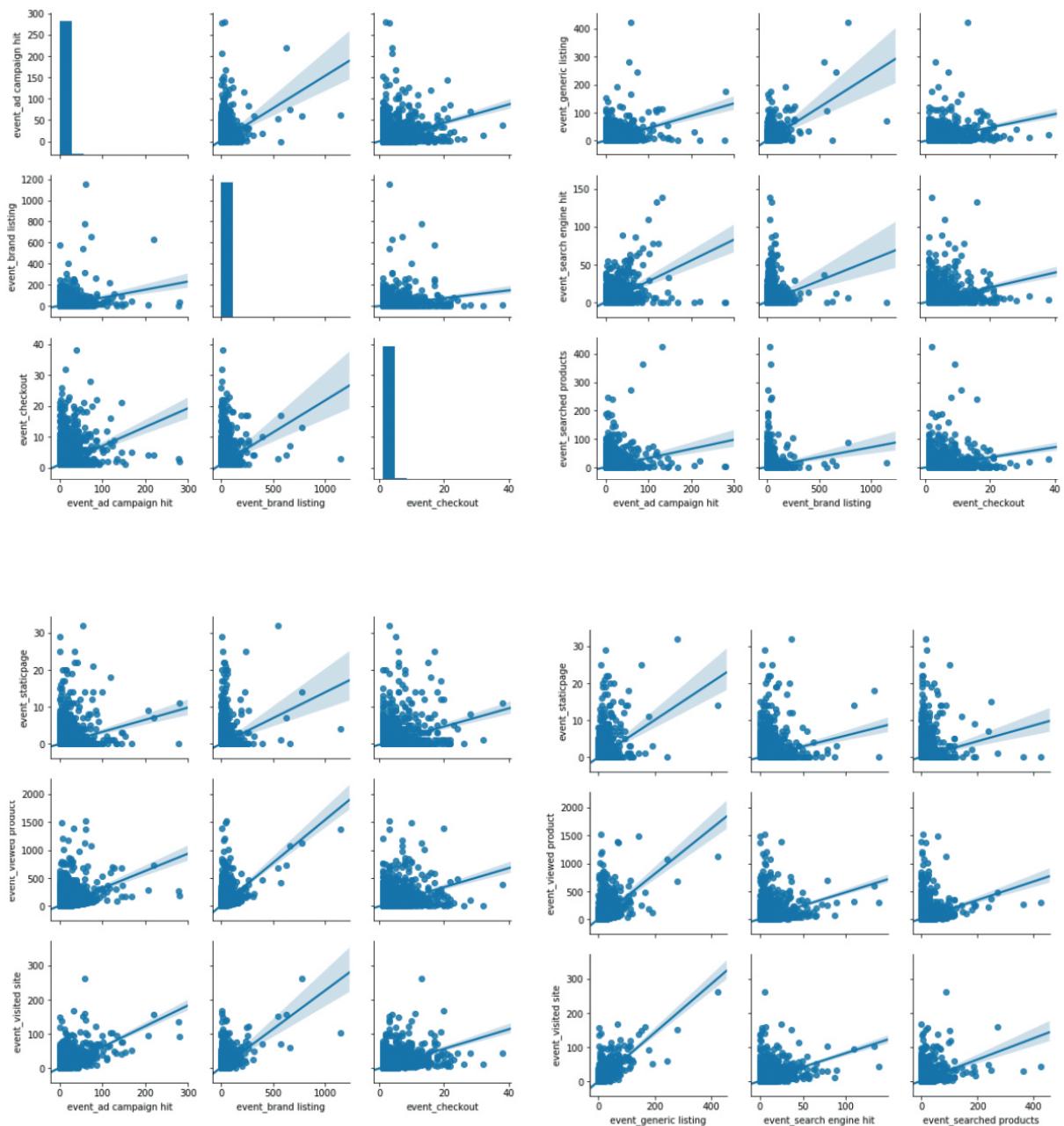


Figura 15: Cantidad de eventos en Brasil

Con el gráfico superior, podemos concluir que los usuarios de Brasil ven más de un producto cada vez que ingresan al sitio de Trocafone. E incluso que para ver los productos visitan un listado específico de una marca debido a que el evento *brand listing* se encuentra por encima de las visitas a la página. Otro aspecto importante a destacar, es que muchas personas ingresan a Trocafone a una determinada url, sin pasar por la homepage (esto último está representado por el evento *generic listing*).

### 2.5.3. Relación entre eventos



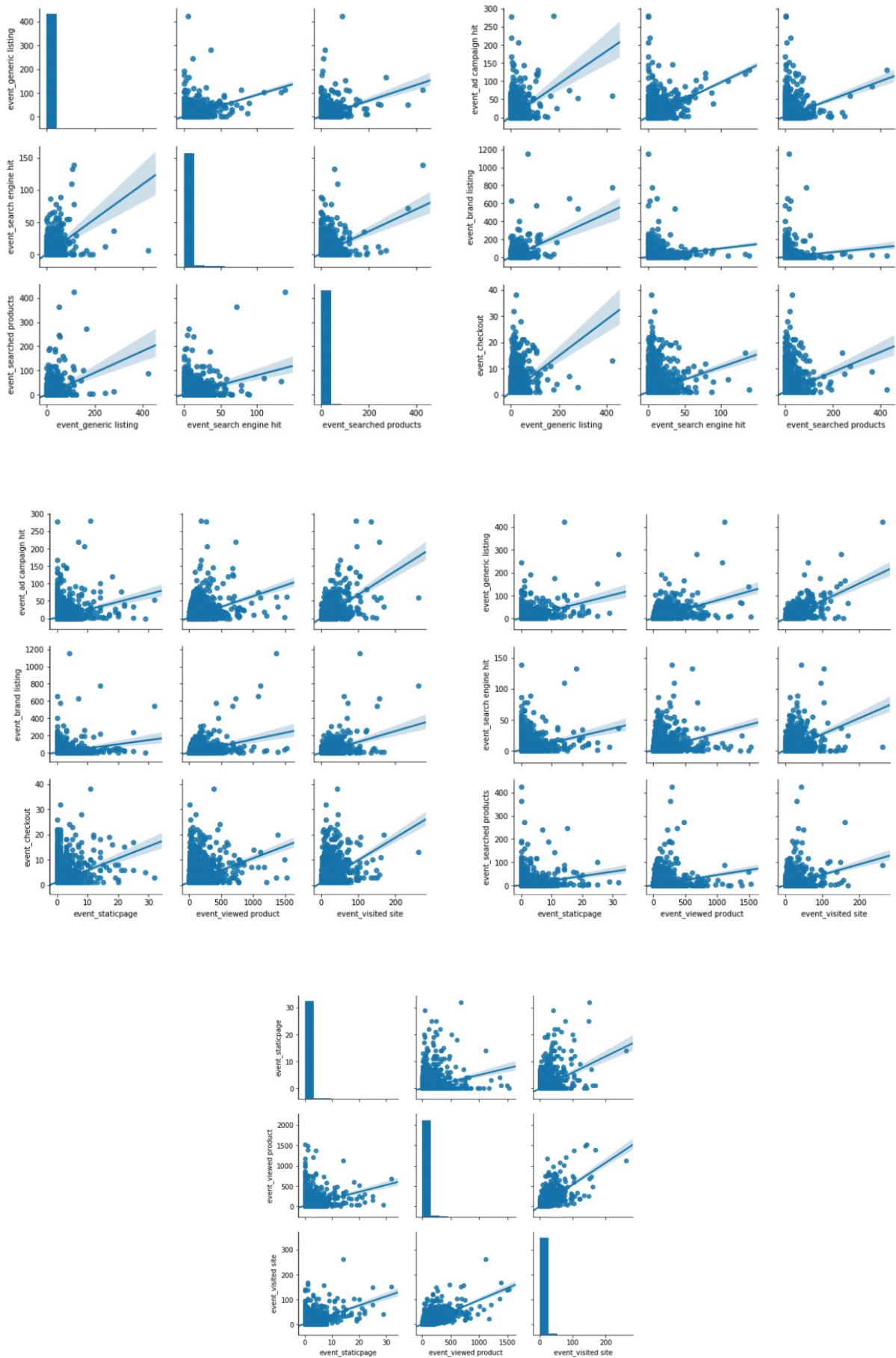


Figura 16: Scatter's plot de la relación entre eventos

En el gráfico Scatter mostrado anteriormente podemos ver una relación entre los tipos de eventos generados por el uso del sitio. Se ignora en el análisis el histograma generados entre eventos del mismo tipo dado que no aporta información. Se utilizó una regresión lineal para una mayor apreciación de la variación entre los datos. Dicho esto puede observarse como el evento *visited site* tiene una relación lineal con pendiente positiva muy marcada con los eventos *brand listing*, *generic listing* y *viewed product*. También existe una relación lineal de pendiente positiva entre los eventos *checkout* y *viewed site*. De esto podemos concluir como cuanto mas veces un sitio es visitado, mas eventos de *checkout* se generan, característica que resulta esperable. Un comportamiento esperable que también se evidencia es que una mayor cantidad visitas al sitio produce una mayor cantidad de visitas a través de una pagina de una empresa y su lista de productos, además genera una mayor cantidad de visitas a productos en general.

#### 2.5.4. Eventos y su relación según la hora

Ya se analizó previamente que los eventos mayormente son de Brasil, pero ahora nos interesa saber que actividad tiene *Trocafone*. El siguiente análisis se enfoca en la actividad según el horario de cada día de la semana, sin importar año y mes.

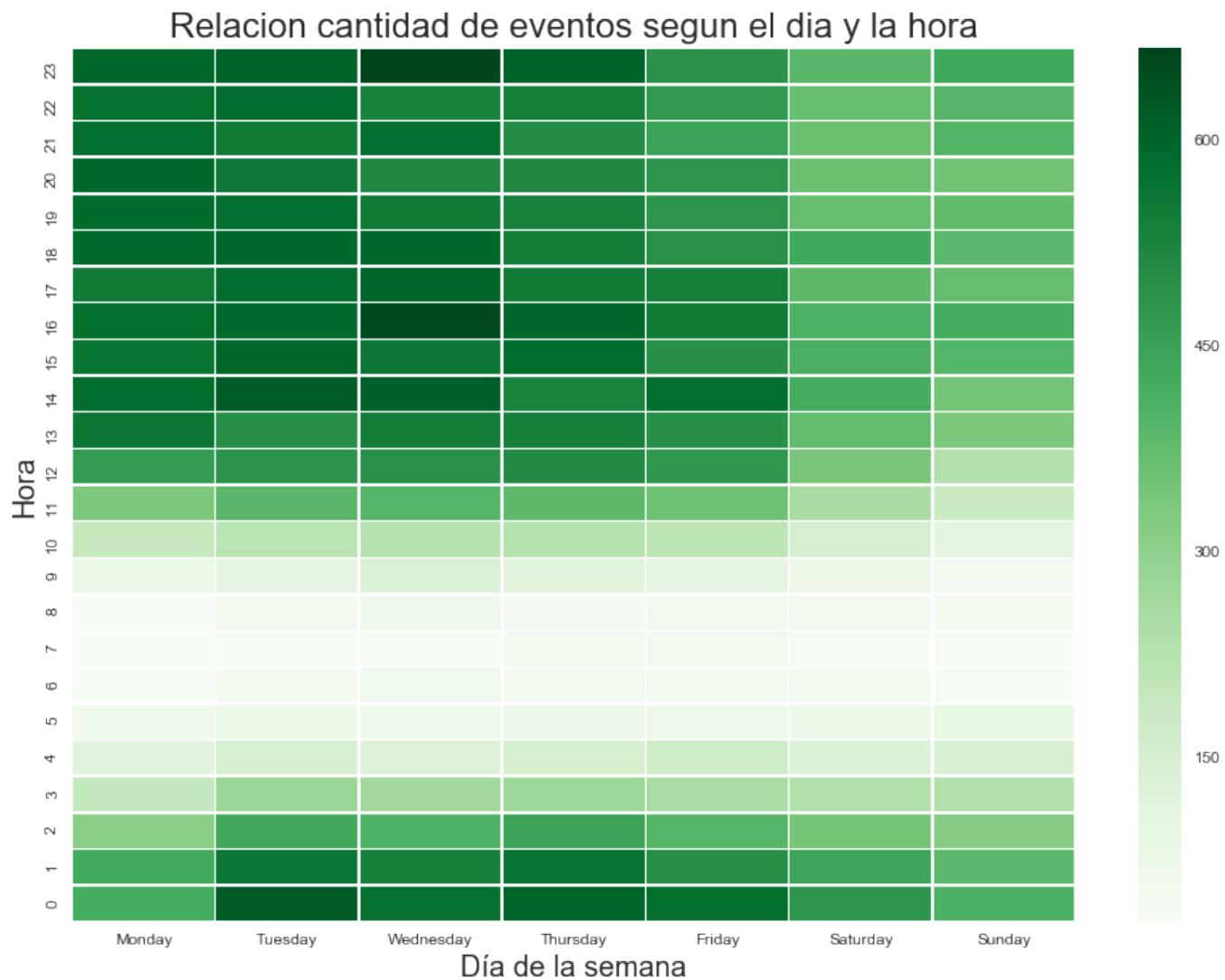


Figura 17: Cantidad de eventos según la hora y día de la semana

Observando detenidamente, vemos que la mayor concentración de actividad se encuentra entre las 10 a 23hs de cada día y que principalmente la menor actividad es cuando las personas suelen dormir. A su vez los picos de actividad se dan los miércoles a las 16 y 23hs, donde podría deberse al momento en el cual uno usa mas el celular, ya sea al salir de trabajar o antes de acostarse. Por otro lado se ve que a lo largo de la semana la actividad nocturna se incrementa

y, por último, se observa que durante fines de semana la pagina tiene menos actividad durante la mañana, lo cual es lógico ya que es común que se duerma más.

#### 2.5.5. Tipo de evento por producto

En la siguiente figura se puede observar el tipo de evento mayoritario para los distintos productos que se publican en el sitio.

Para este análisis se agruparon todos los modelos posibles de un mismo producto con el fin de reducir la varianza de los datos, por ejemplo, se creó un producto *iPhone* el cual incluye todos sus posibles modelos. Esto mismo fue realizado con las distintas marcas posibles.

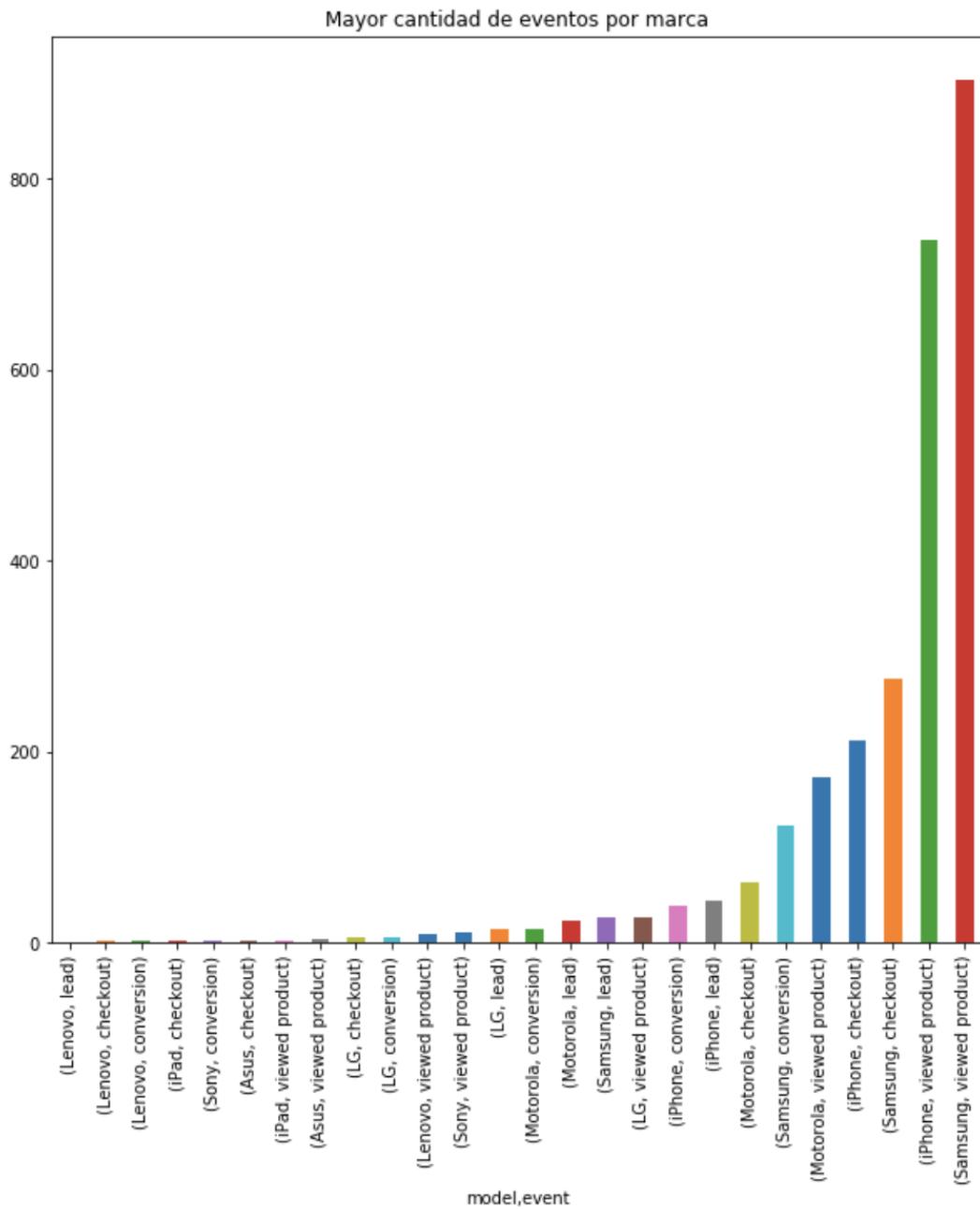


Figura 18: Cantidad de eventos por marca

Vemos que indistintamente del producto, una mayor cantidad de visitas a la publicación viene asociado con una mayor cantidad de compras, aunque esta última siendo muy inferior, siendo *iPhone* y *Samsung* las dos compañías con mayor cantidad de eventos.

### 2.5.6. Relación entre eventos según la marca

En el siguiente Scatter Plot se buscó encontrar una relación entre la cantidad de ocurrencias de los eventos y la marca del producto generador de los mismos. Al igual que en el gráfico anterior, fueron agrupados los distintos modelos de una marca en la compañía única con el fin de disminuir la varianza en los datos.

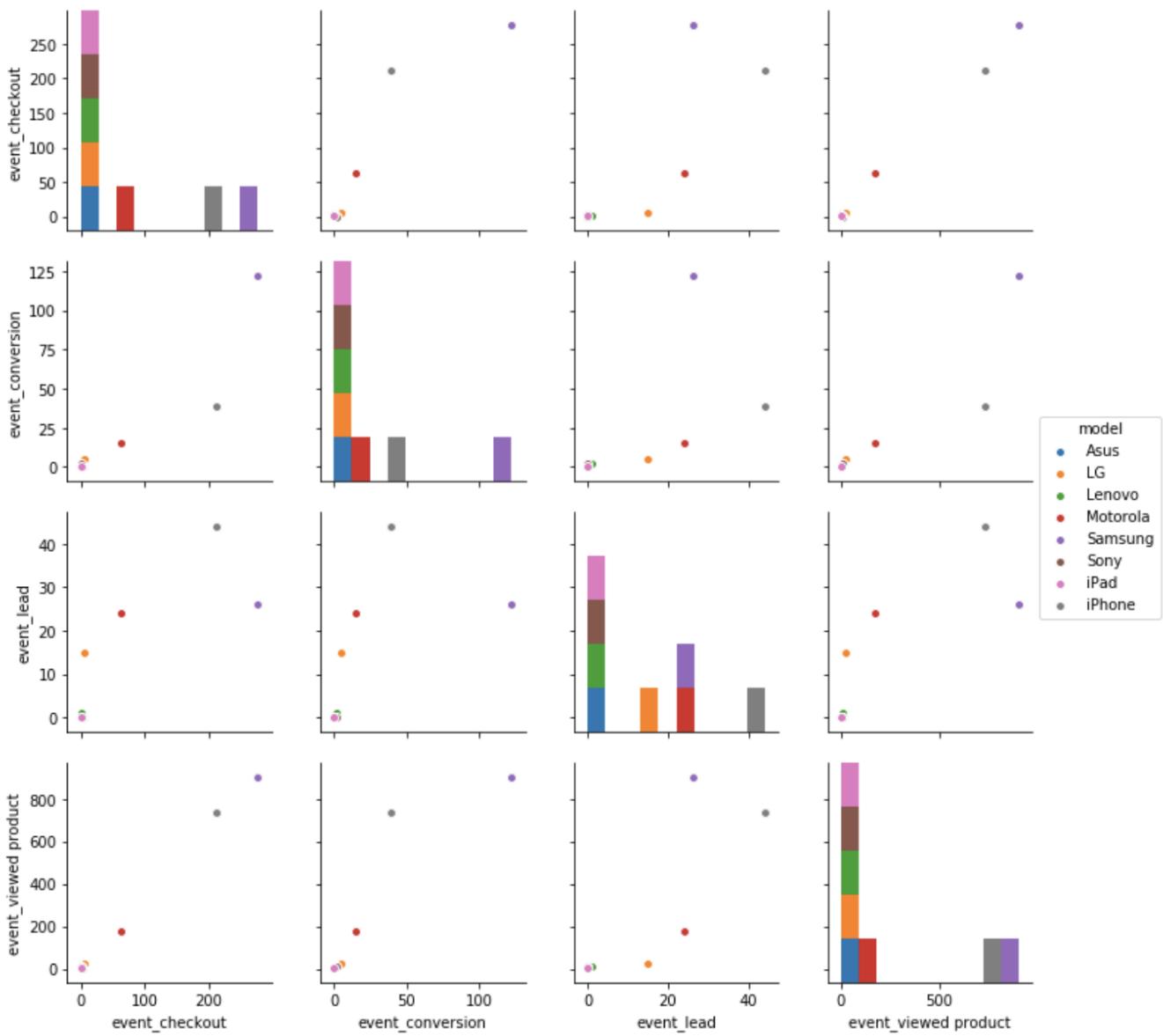


Figura 19: Relación entre eventos según la marca del producto

Siendo *Samsung*, *iPhone* y *Motorola* las marcas más comunes se puede observar que existe una relación perfectamente lineal entre *Viewed product* y *checkout* esto afirma las conclusiones obtenidas en la sección 2.5.3, sumando la característica de que esto ya es independiente del producto del cual se esté hablando. También se observa un comportamiento casi logarítmico entre *conversion* y *checkout* dando como resultado que cuantos más eventos *conversion* se producen más *checkout* pero en este caso si depende de que producto se este analizando.

### 2.6. Canales de ingreso

El ingreso a una página de web puede generarse a través de diferentes canales. Estos pueden ser:

- Pagos: se ingresa por medios por la cual se invirtió, campaña de marketing, por ejemplo.
- Orgánicos: se ingresa por un medio por el cual la empresa no puso dinero.

- Directos: se ingresa directamente a la página.
- Sociales
- Referidos: se ingresa por recomendaciones que se pueden encontrar en otros sitios, por ejemplo páginas web.
- Email

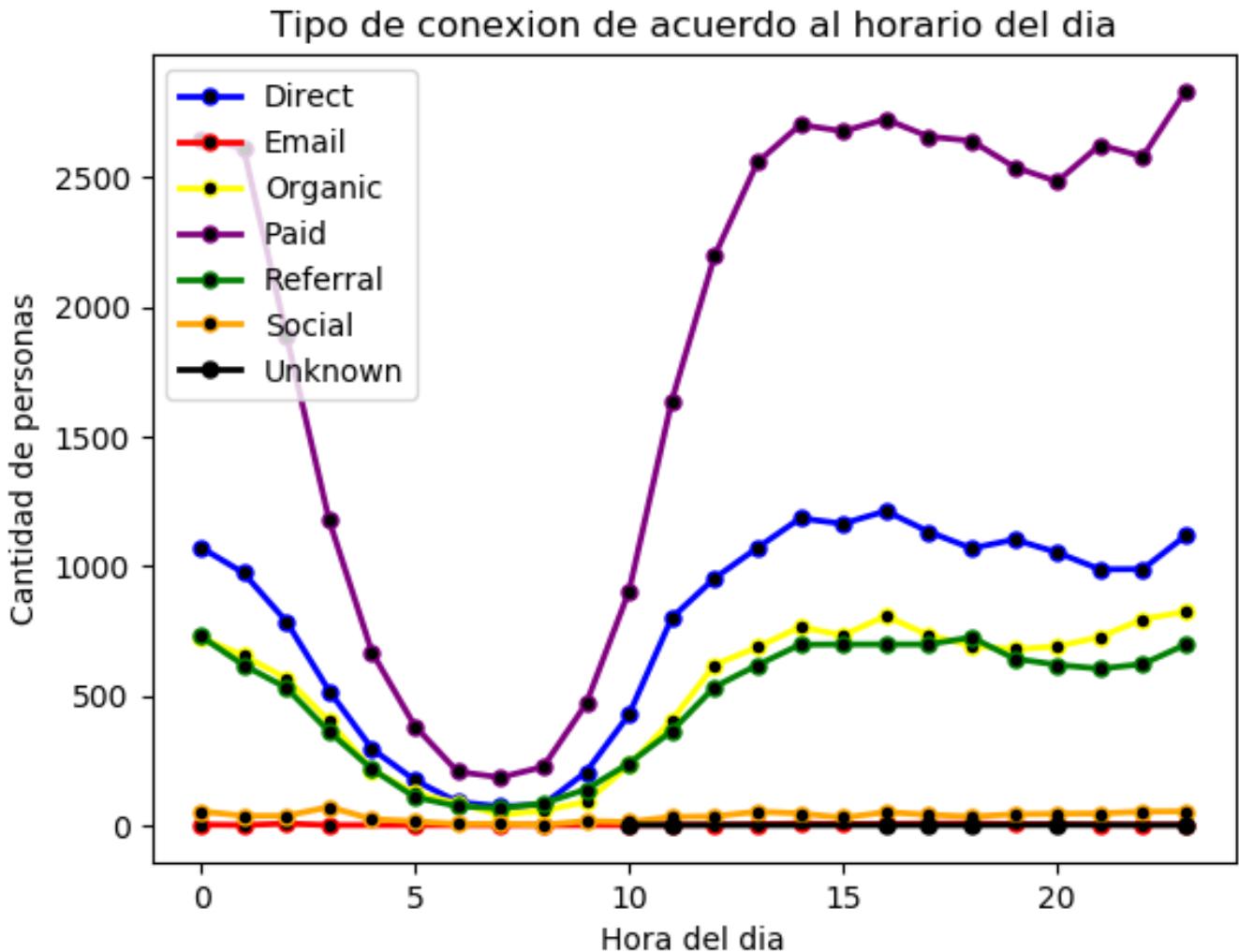


Figura 20: Tipo de conexión de acuerdo al horario del día

Comencemos analizando que hay dos canales, social y email, que durante todo el día se mantienen con un flujo casi nulo de visitantes. Por otro lado, los canales pagos y los canales directos tienen bastante actividad durante toda la jornada.

En la madrugada, entre las 1 y 6 de la mañana, se puede observar una gran baja en las visitas a la página en absolutamente todos los tipos de canales. A partir de las 8, se puede observar una gran alza en los dos canales principales y una menor en los canales orgánicos y referidos, que tienen una actividad baja en relación a los mencionados anteriormente.

Es interesante ver que durante todo el día hay ingresos a la página de Trocafone, incluso en la madrugada, donde los canales pagos son los que la mantienen activa. Luego, en el horario matutino y vespertino la actividad se vuelve mucho más grande, la mayoría se concentraba en los canales pagos.

### 2.6.1. Origen de Campaña

Los canales pagos, como ya se dijo, representan el flujo de la página a través de medios pagos por la empresa. Son muy utilizados para que las páginas web tengan anuncios y se encuentren bien posicionados dentro del sistema de recomendación de los buscadores.

Como ya se ha mostrado en la sección anterior, los canales pagos son los que generan mayor actividad durante todo el día en el sitio de Trocafone. Ahora analicemos como se distribuyen entre los distintos orígenes:

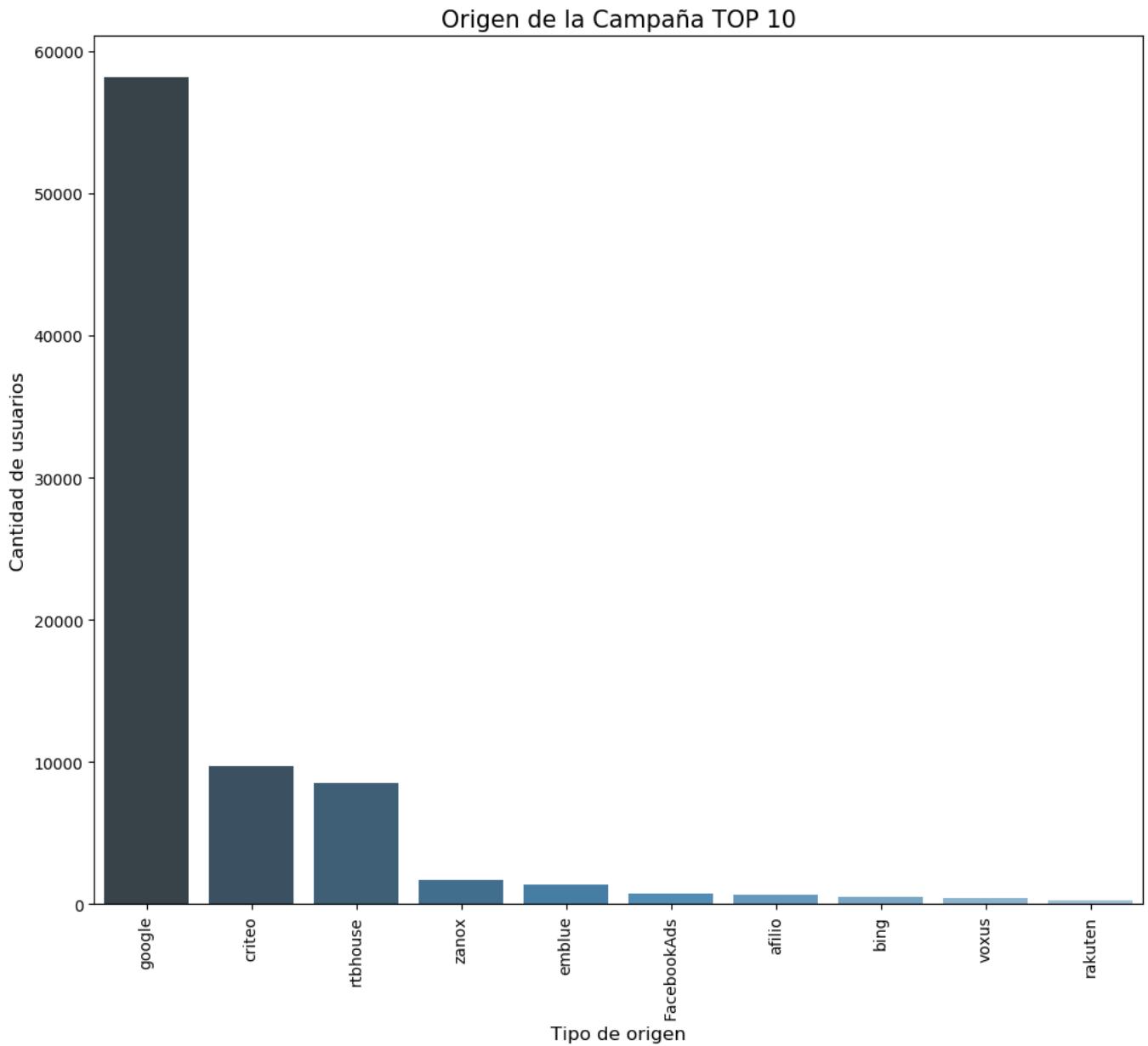


Figura 21: Origen de las Campañas

Claramente, Google es el principal origen de los canales pagos llevándose un alto porcentaje de los eventos. Luego, la plataforma de publicidad web, Criteo, y RTB House son las alternativas a Google. Por último, podemos ver que el resto de los orígenes de las campañas no tienen una influencia importante en los canales pagos, aunque la mayoría registran eventos, se trata de un volumen de tráfico ínfimo comparado a los tres canales mencionados anteriormente.

## 2.7. Ventas

Algunas de las preguntas que nos surgieron al analizar los datos fueron: ¿Cuántas ventas por día, mes y según horario posee la plataforma? ¿Las compras desde dónde se realizan mayormente? ¿Cuáles son los celulares más vendidos? ¿De los celulares ingresados al carrito de compras cuáles finalizan el proceso? ¿De las ventas cuáles son usuarios nuevos y cuáles no? ¿Tienen alguna relación las ventas con las palabras más buscadas? La aparición en la homepage, ¿acarrea mayor cantidad de ventas? La idea es tratar de abordar todas estas preguntas y responderlas mediante el análisis realizado.

### 2.7.1. Celulares más vendidos

La idea de esta parte del análisis es variado. Principalmente y más importante es el análisis de los celulares mas vendidos.

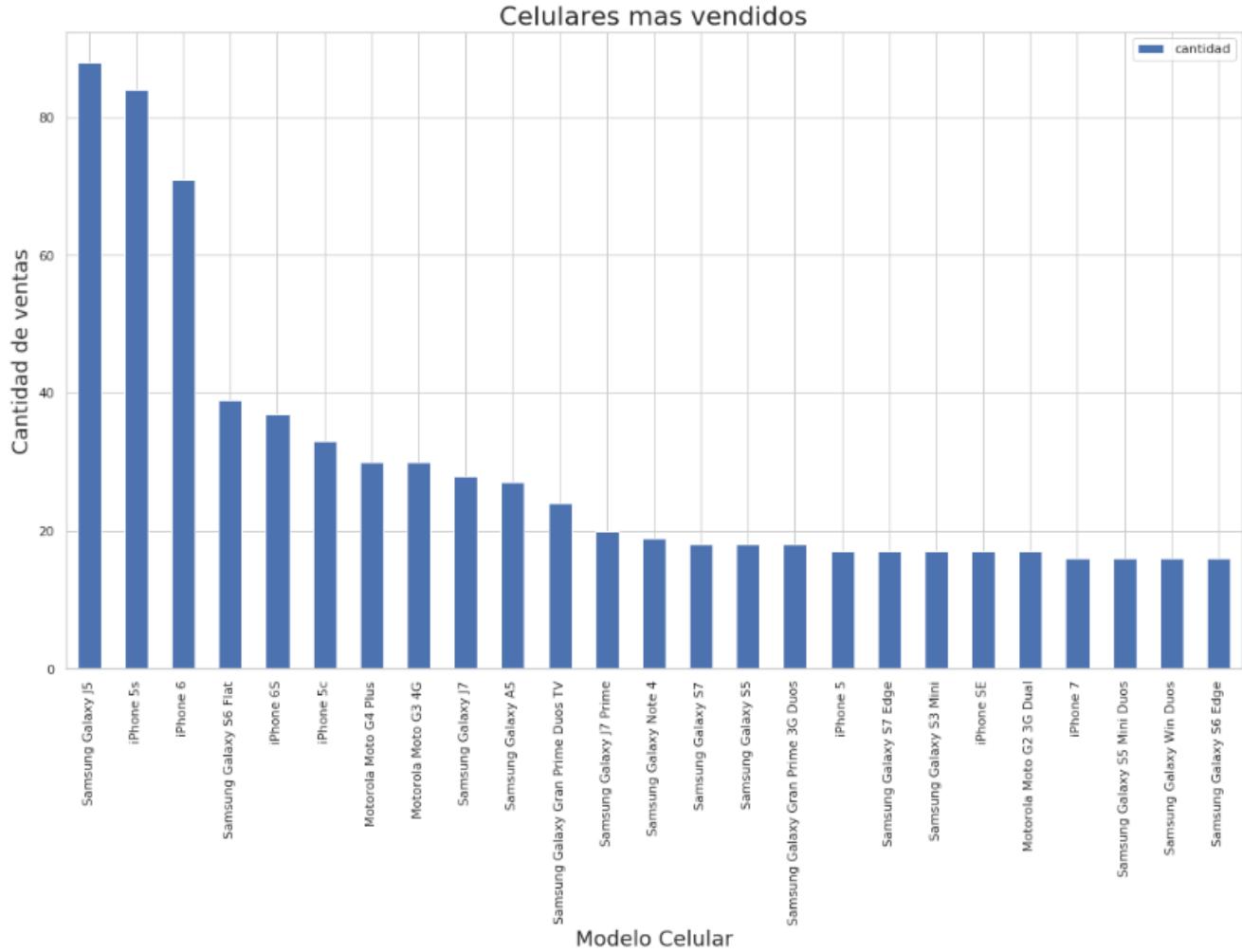


Figura 22: Celulares más vendidos

En la figura 22 no se muestran todos los celulares vendidos, seleccionamos los más vendidos, con el fin de evaluar cuales son los modelos que más se consumen en el mercado.

El set de datos entregado abarca información desde enero hasta junio (inclusive) del 2018, lo cual todo el análisis es durante este mismo año. Si tenemos en cuenta esto y los 4 celulares más vendidos, se puede pensar que en el mercado se compran celulares que no son del 2018, sino que se prefieren más viejos que aún permanecen a la venta. Esto puede deberse a que los modelos más actuales suelen tener un costo más elevado y a veces con poca diferencia entre su modelo anterior en cuanto a prestaciones.

### 2.7.2. Celulares en carrito de compras

Celulares *en carrito* (a.k.a *checkout*) son celulares que el usuario puso en cola para en otro momento realizar la compra, mientras tanto poder seguir viendo otros productos y una vez finalizado poder hacer un único movimiento de compra de todos los productos agregados.

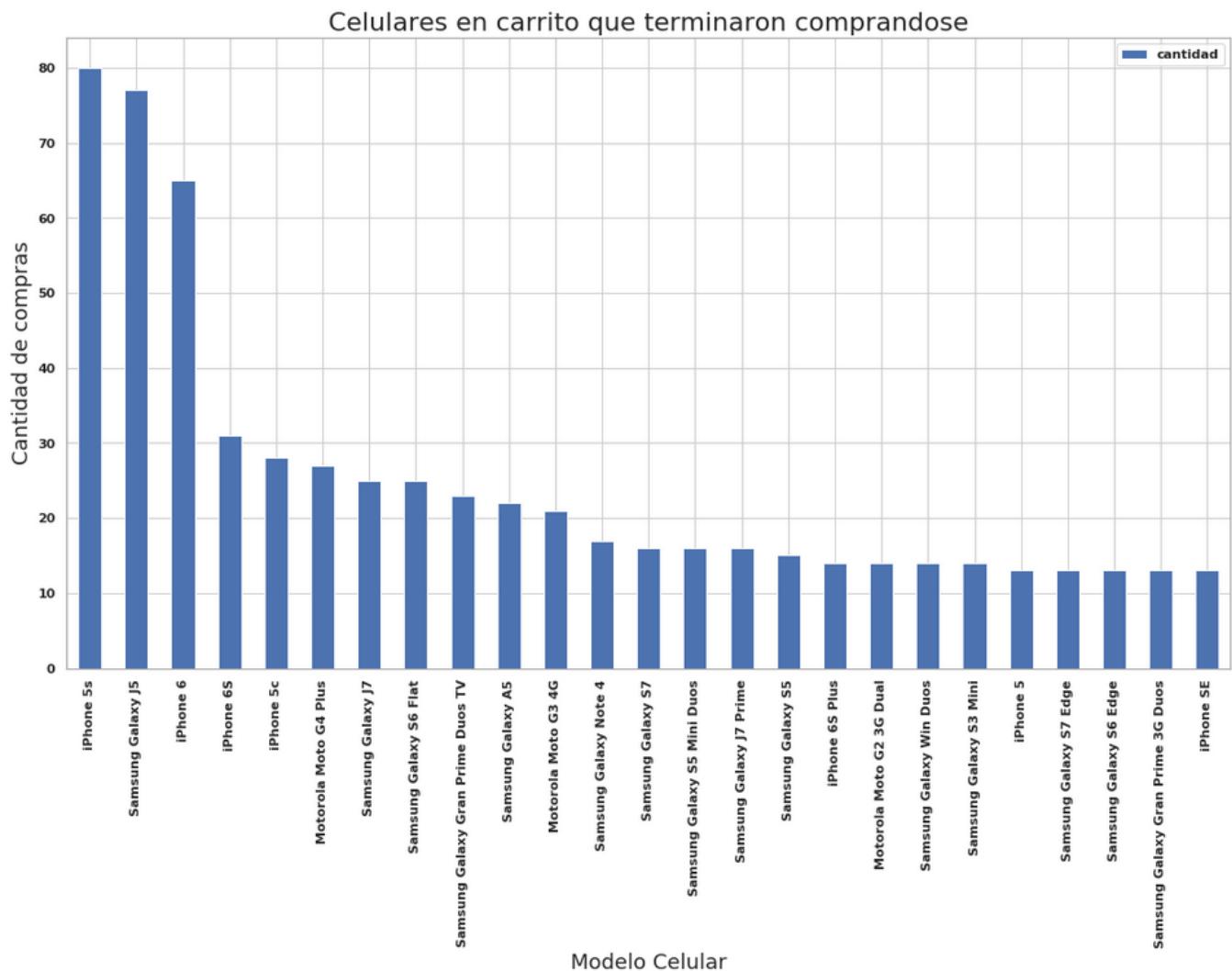


Figura 23: Celulares vendidos luego de estar en carrito

El objetivo del gráfico anterior es mostrar los celulares que fueron agregados al carrito y luego resultaron en una venta concreta. Como se puede observar los primeros coinciden, aunque en diferentes orden, con los de la figura 22, los cuales son los más vendidos de la plataforma.

### 2.7.3. Compras según dispositivo

En esta sección se busca analizar desde qué tipo de dispositivo se realizan las compras. En la figura siguiente se representan las ventas de la plataforma según el dispositivo con el cual se lograron previamente a la compra.

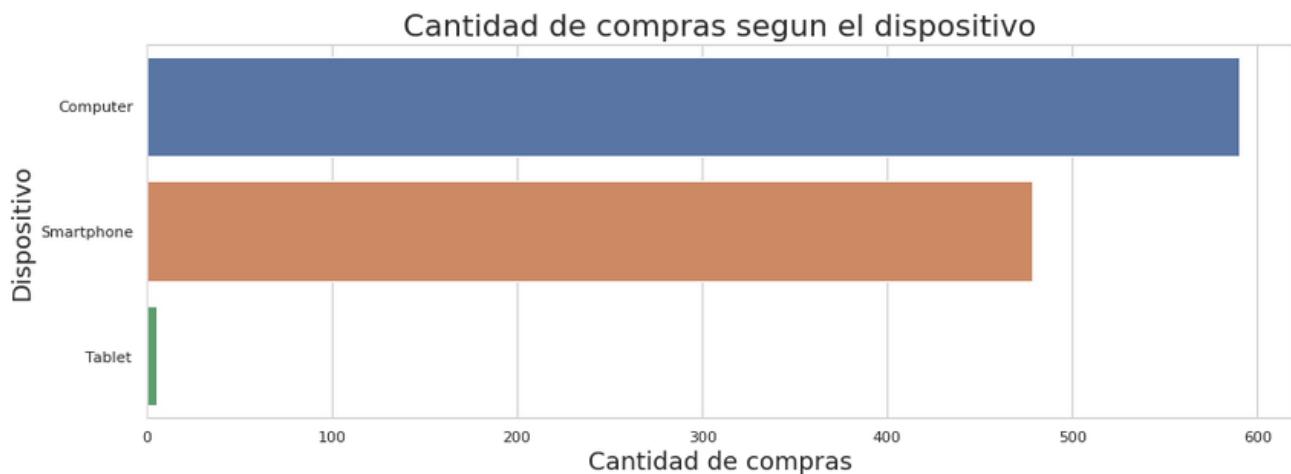


Figura 24: Cantidad de compras según dispositivo

Como se mencionó previamente, las visitas al sitio son mayormente realizadas mediante dispositivos móviles. Sin embargo, las ventas se concretan en mayor medida desde una computadora, prefiriéndola por sobre un celular, muy posiblemente debido a la desconfianza que genera colocar datos de tarjeta en el mismo y luego correr el riesgo de extraviarlo o que se lo hurten.

#### 2.7.4. Ventas según S.O

Lo que se buscó fue mostrar cuales son los sistemas operativos con los cuales se realizan más compras. En cierta forma tienen estos una relación directa a las ventas según el tipo de dispositivo.

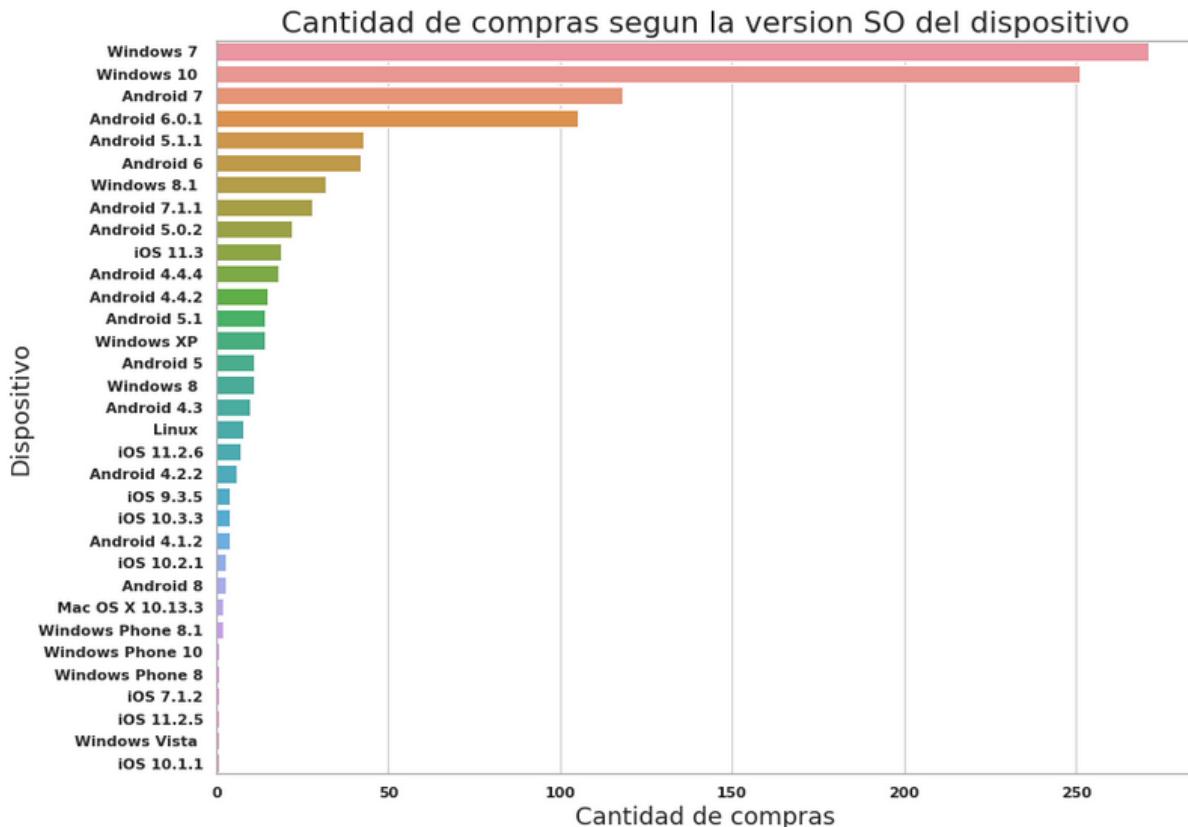


Figura 25: Ventas según S.O

En la figura anterior se observa, cómo era de esperarse, que el S.O predominante sea de una computadora, Windows, y luego el de un celular, Android, debido a lo que habíamos analizado en ventas según el tipo de dispositivo.

#### 2.7.5. Ventas según el mes

En el siguiente gráfico se representan las ventas de *Trocafone* por cada mes.

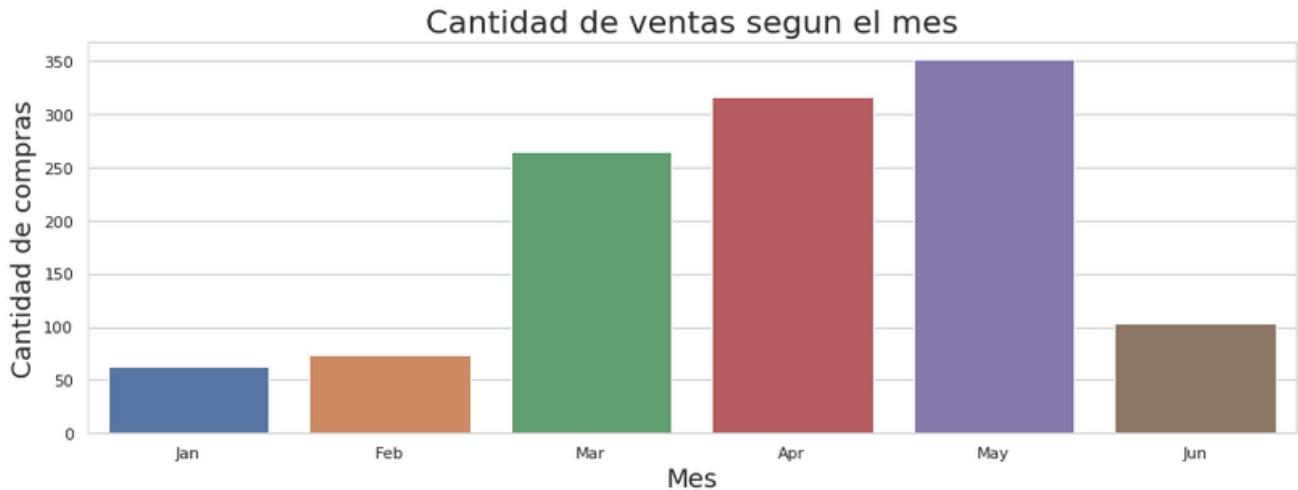


Figura 26: Ventas según el mes

Podemos concluir de la figura anterior que durante los meses de enero y febrero no hay tantas ventas como en los siguientes meses, esto puede venir a colación de que en Brasil, principal nicho de mercado, es época de vacaciones y generalmente los ahorros se ven destinados a otras actividades.

En junio es un caso muy particular ya que los datos son incompletos, la fecha mas reciente que se tiene de junio es el día 15, lo que hace que tengamos aproximadamente la mitad de datos para analizar en dicho mes, justificando así la posibilidad de baja en ventas.

#### 2.7.6. Ventas según el día

Ya se analizó previamente cómo es la actividad de la plataforma durante los días de la semana y ahora buscamos analizar, específicamente, que sucede con las ventas:

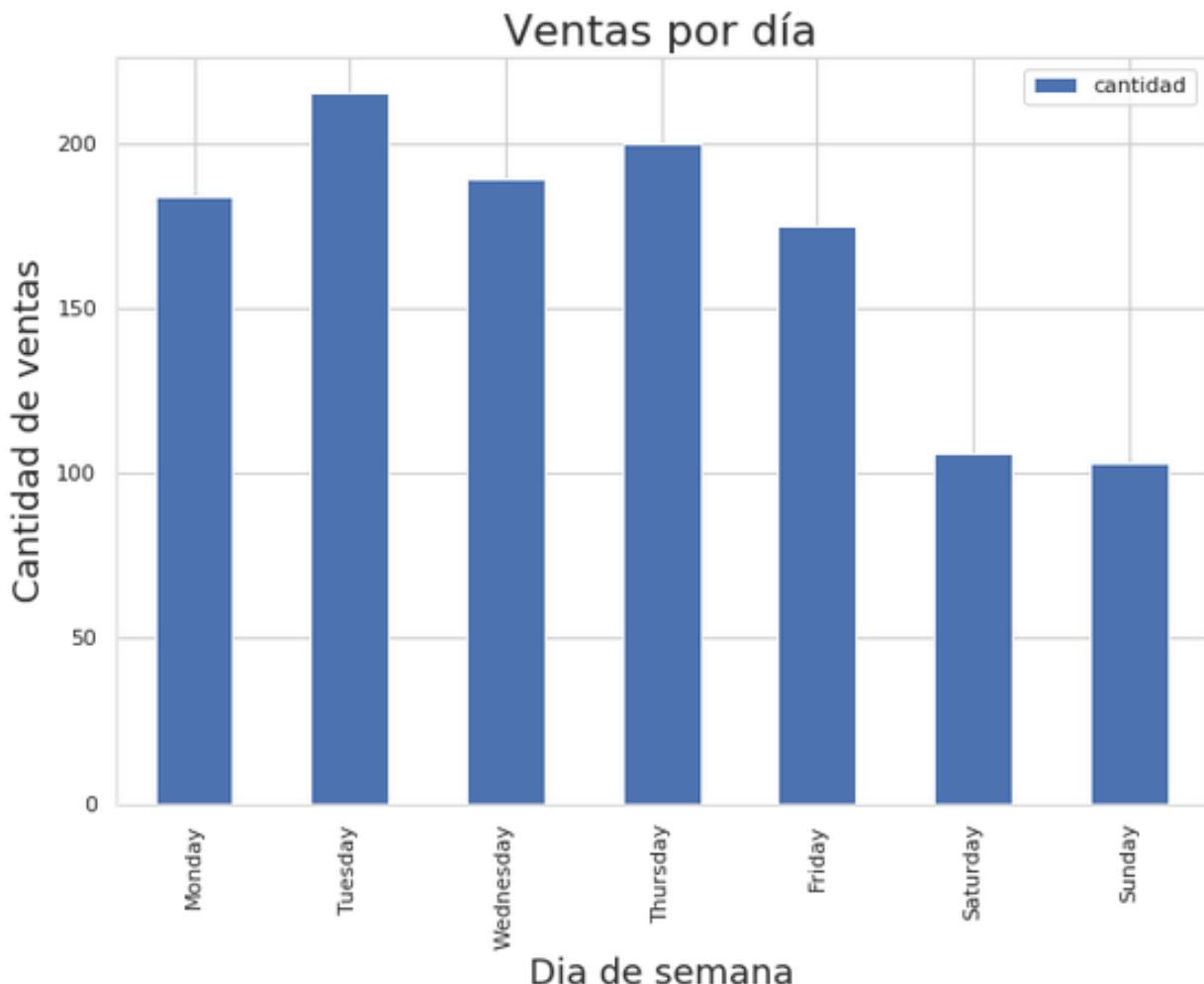


Figura 27: Ventas según el día

En relación con la actividad de la empresa, se había podido observar como en los fines de semana la misma disminuía. Ahora analizando la figura anterior, vemos que sucede lo mismo, esto tiene relación con lo mencionado anteriormente, debido a que las ventas son un tipo de actividad de la plataforma.

Por último, se puede ver que el día martes y jueves son los días con mayores ventas, los cuales están dentro de los principales días con mayor actividad. Aunque sorprendentemente, los picos de actividad que se veían los días miércoles, no se ven traducidos a la cantidad de ventas que hay en este día de la semana.

#### 2.7.7. Ventas según el horario

Si bien ya se analizaron las ventas tanto por mes como por día, nos interesa ver como son en las horas del día.

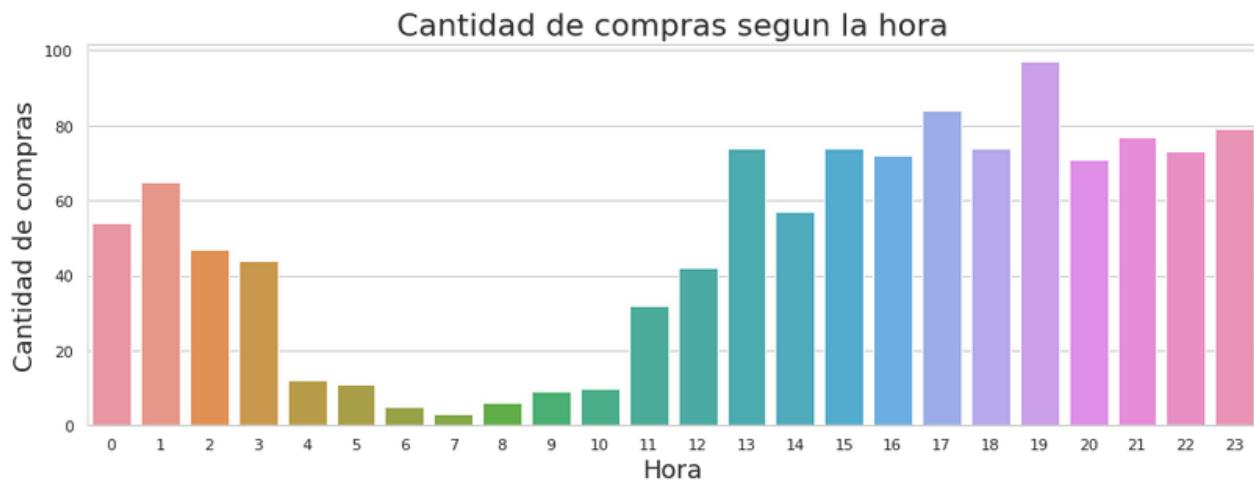


Figura 28: Ventas según el horario

Si se compara la figura 28 con la figura 17 del análisis de actividad según el horario, podemos ver exactamente el mismo comportamiento, en el cual en las horas más típicas de sueños la actividad baja. A su vez, teniendo en cuenta cómo son los registros de actividades en la plataforma y cómo son los registros de ventas, ambos según el horario, sería recomendable que cualquier aviso publicitario o notificación al usuario, sea en la franja horaria entre las 12 a 23hs.

#### 2.7.8. Ventas según tipo de usuario

Como ya vimos los usuarios pueden retornar o ser nuevos, esto nos llevó a analizar que sucedía con las ventas, si había alguna que se concretara para un usuario nuevo, ya que es poco probable que en su primera visita ya concrete la compra de un producto.

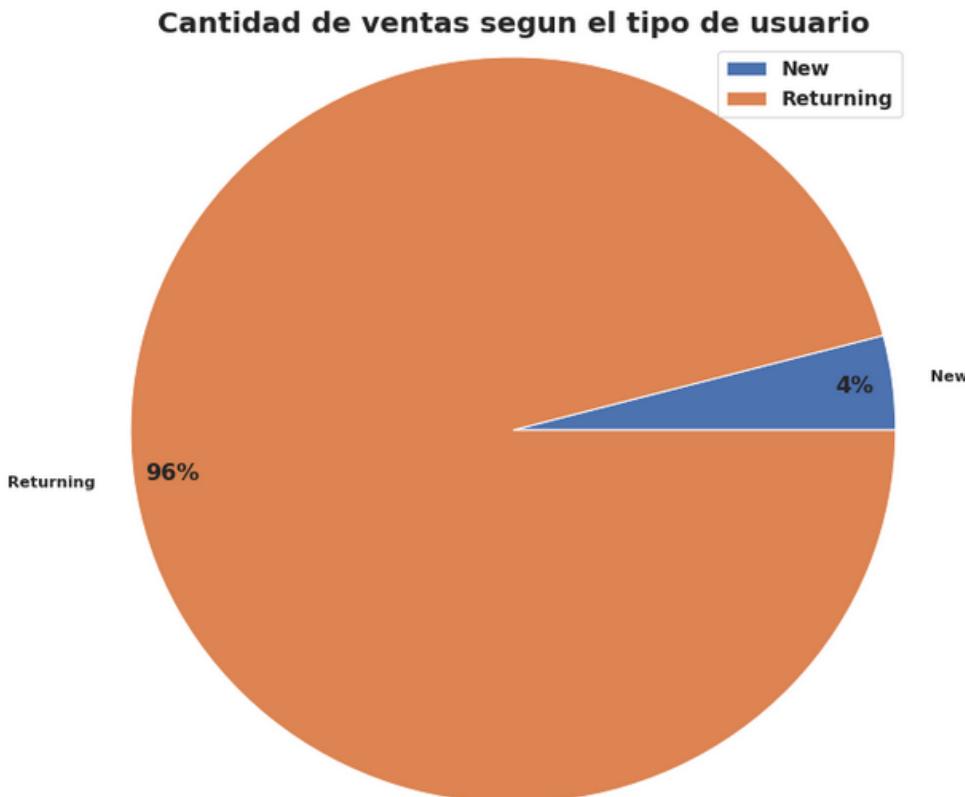


Figura 29: Ventas según el tipo de usuario

Observando el gráfico anterior, se ve claramente que el 96 % de las compras son de usuario no nuevos, y simplemente

4 % de compras concretadas por usuarios nuevos.

Esto es muy lógico, porque cuando uno tiene que comprar algo que requiere tanto dinero como un celular, el cliente quiere estar seguro de lo que está comprando y que el proveedor genere confianza para brindarle los datos de su tarjeta de crédito o débito.

### 2.7.9. Ventas según búsqueda

En esta sección la idea es visualizar las ventas que tienen relación con las palabras buscadas en el sitio. Para realizar este análisis, primero es necesario analizar cuáles son las palabras más buscadas en el sitio.

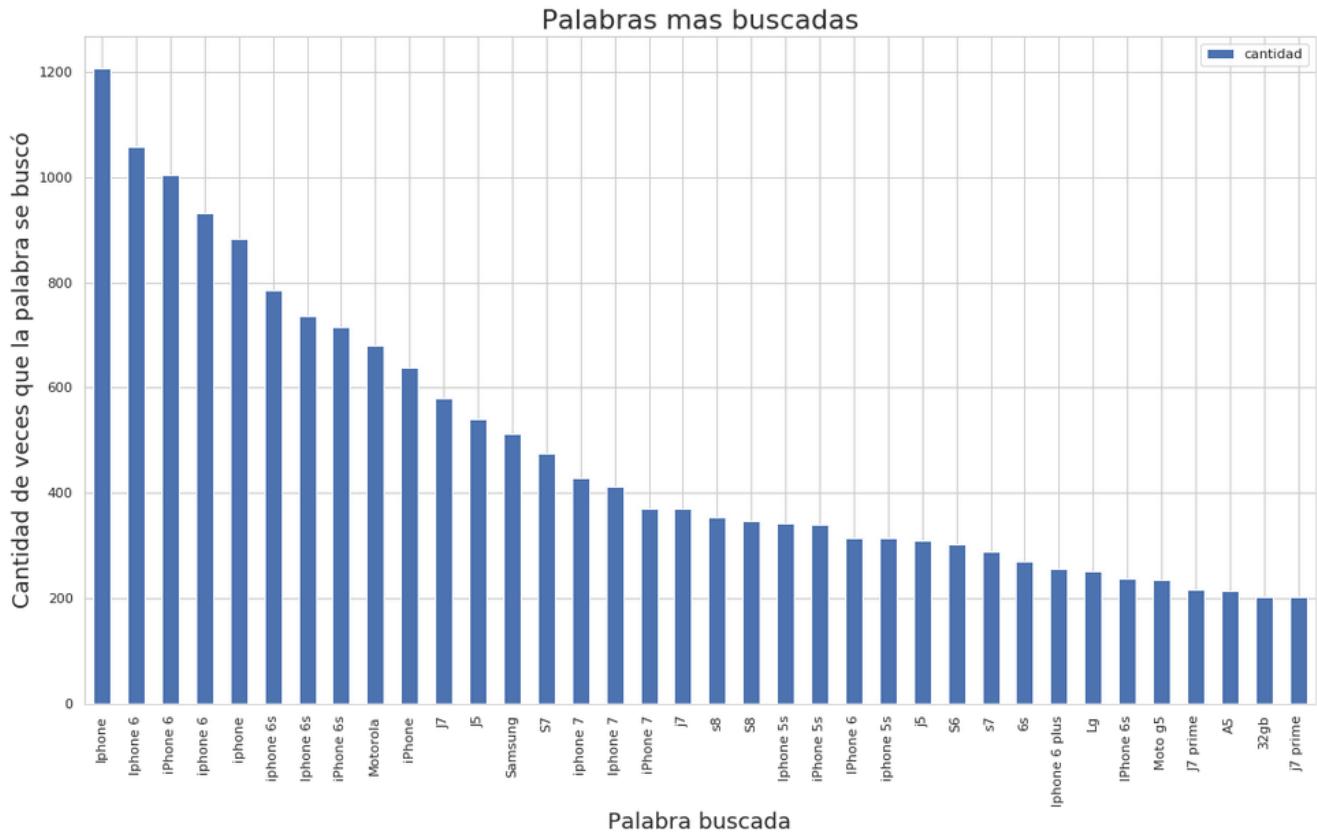


Figura 30: Palabras más buscadas

En el gráfico anterior se muestran algunas de las palabras más buscadas en la plataforma. Claramente hay una búsqueda muy importante de los iPhone, sin importar su modelo, y de los Samsung, que se pueden suponer por las palabras J7,S7,etc.

Ahora la idea es ver cuáles de estas palabras generan alguna compra:

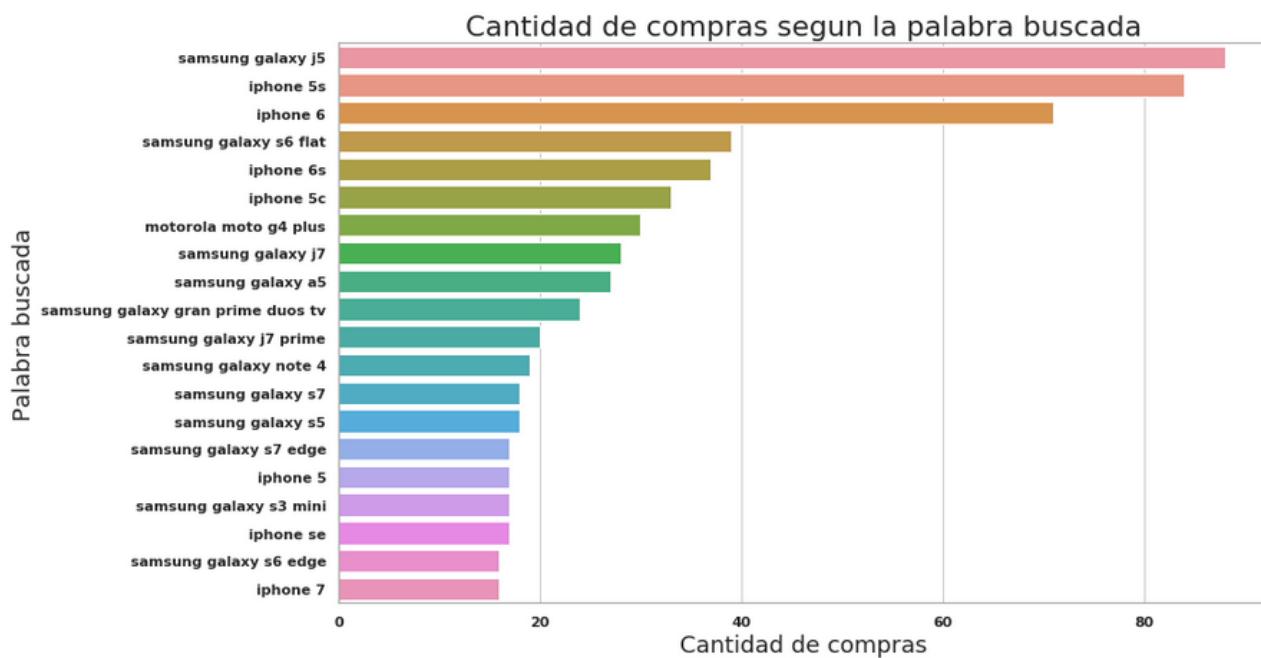


Figura 31: Ventas según la búsqueda

Vemos que palabras buscadas se relacionan directamente con una compra de celular, la cual se puede comparar con los celulares más vendidos en ésta plataforma; y claramente coinciden con las ventas de celulares.

#### 2.7.10. Ventas según Lead's

El lead de un celular en el sitio, básicamente es registrarse a *Trocafone* para recibir una notificación para cuando el dispositivo vuelva a tener stock en la plataforma. Con esta información nos enfocamos en analizar cuales de los lead's en el sitio terminan concretando una compra, pero para realizar esto, primero vamos a analizar los lead's por modelo.

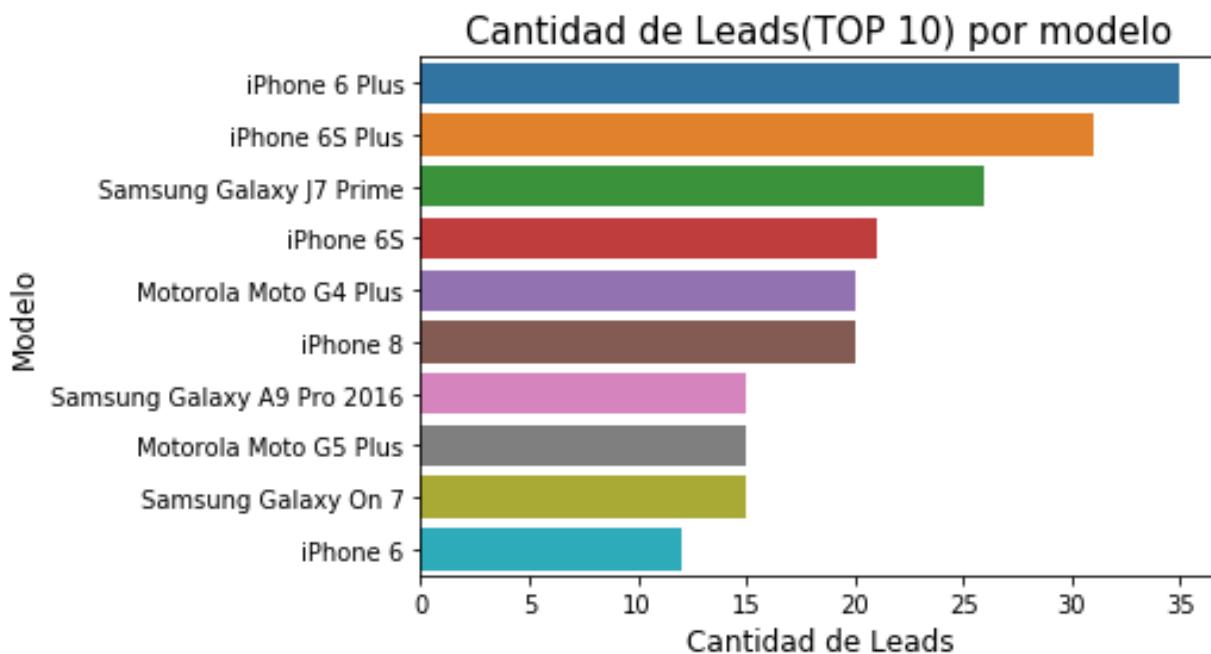


Figura 32: Cantidad de Lead's por modelo

En la figura anterior se puede ver cuales son los celulares que no tienen stock y son demandados por los usuarios, con intención de ser avisados cuando tengan stock para continuar con la compra del mismo. Ahora la idea es poder ver que es lo que sucede con estos lead's, si terminan en una compra concretada o no y ver en que modelos se da más.

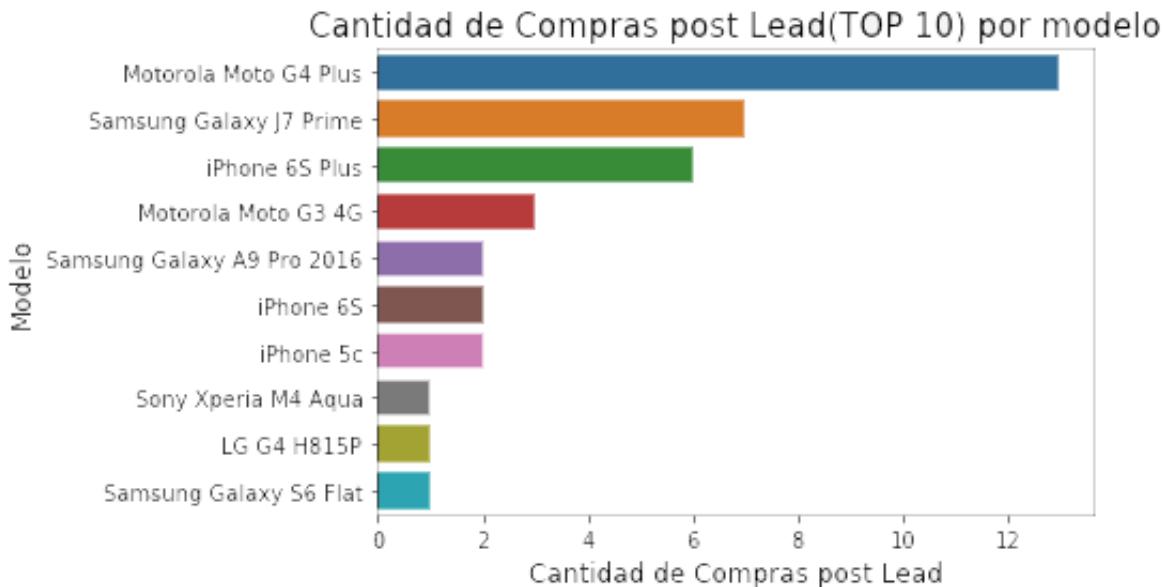


Figura 33: Cantidad de compras post lead's por modelo

Lo obtenido de este análisis es que son muy pocas las compras que se concretan post lead del celular, generando una relación mucho más débil de lo que se suponía porque esperábamos que si una persona desea recibir una notificación sobre el stock del celular es porque tiene intenciones de comprarlo, pero esto no fue así.

Como siempre, hay una excepción a la regla. El modelo que mayor compras tiene es el Motorola Moto G4 Plus, en el cual la cantidad de compras post lead es muy cercana a la cantidad de lead's para dicho modelo. Además si se observa la cantidad de celulares vendidos de este modelo y de estos cuantos fueron vendidos post lead, vemos que corresponde a casi un tercio de la totalidad de ventas para este modelo.

#### 2.7.11. Ventas según apariciones en la Homepage

Los eventos *generic listing* son aquellos que representan cuando el usuario visita la homepage y tienen asociado el atributo *skus* que es una lista de los identificadores de los celulares que aparecen en la página principal. ¿Tiene alguna relación la aparición de celulares en la homepage con la cantidad de ventas de los mismos? ¿Se venden más?

Para determinar esto, buscamos cuales eran los 10 modelos que más ocurrencias tenían dentro de la página principal de Trocafone:

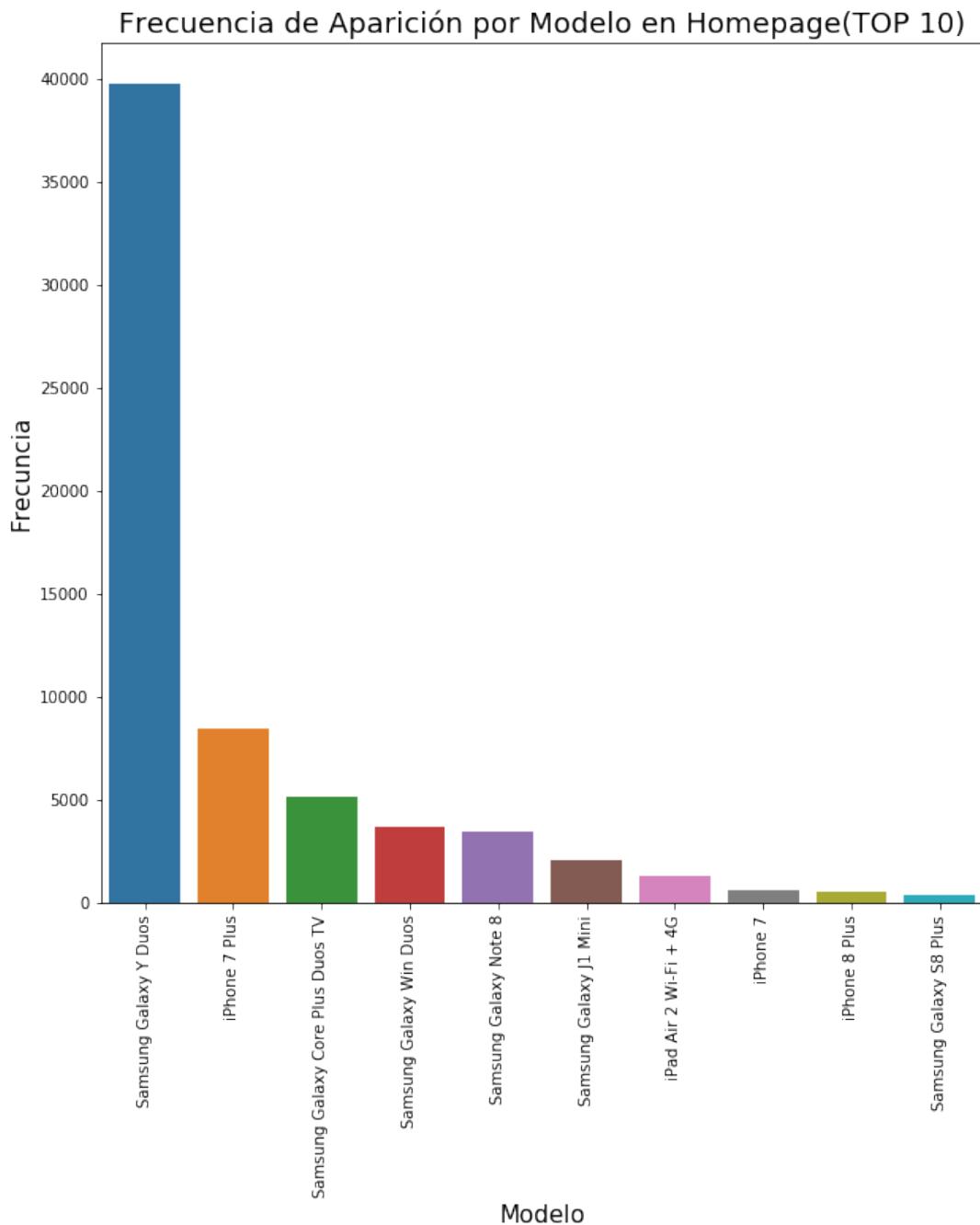


Figura 34: Frecuencia de aparición por modelo en la homepage

En la siguiente tabla, observaremos como esta relación que se suponía, en un principio, fuerte, termino siendo solamente una suposición porque no se tiene ningún tipo de relación entre ambos eventos.

Modelo	Apariciones en Homepage	Ventas
Samsung Galaxy Y Duos	39794	3
iPhone 7 Plus	8478	6
Samsung Galaxy Core Plus Duos TV	5143	3
Samsung Galaxy Win Duos	3717	16
Samsung Galaxy Note 8	3454	3
Samsung Galaxy J1 Mini	2118	11
iPad Air 2 Wi-Fi + 4G	1344	0
iPhone 7	655	16
iPhone 8 Plus	592	0
Samsung Galaxy S8 Plus	386	7

Ninguno de los 10 modelos con más apariciones tiene una gran cantidad de ventas, de hecho ninguno supera las 20 ventas de manera individual. Podemos concluir que se tratan simplemente de celulares de alta gama, modelos relativamente nuevos, y de las dos marcas líderes del mercado como lo son iPhone y Samsung. Estos pueden ser utilizados como anzuelo para atrapar a los usuarios con curiosidad y que a medida que vayan navegando por la web vean otros celulares.

### 2.7.12. Ventas según apariciones en listado específico de una marca

Al igual que como hicimos con los *generic listing* que nos indicaban los celulares que aparecían en la página principal de *Trocafone*, ahora analizaremos los *brand listing* que son los eventos que representan que el usuario visitó un listado específico de una marca viendo un conjunto de productos.

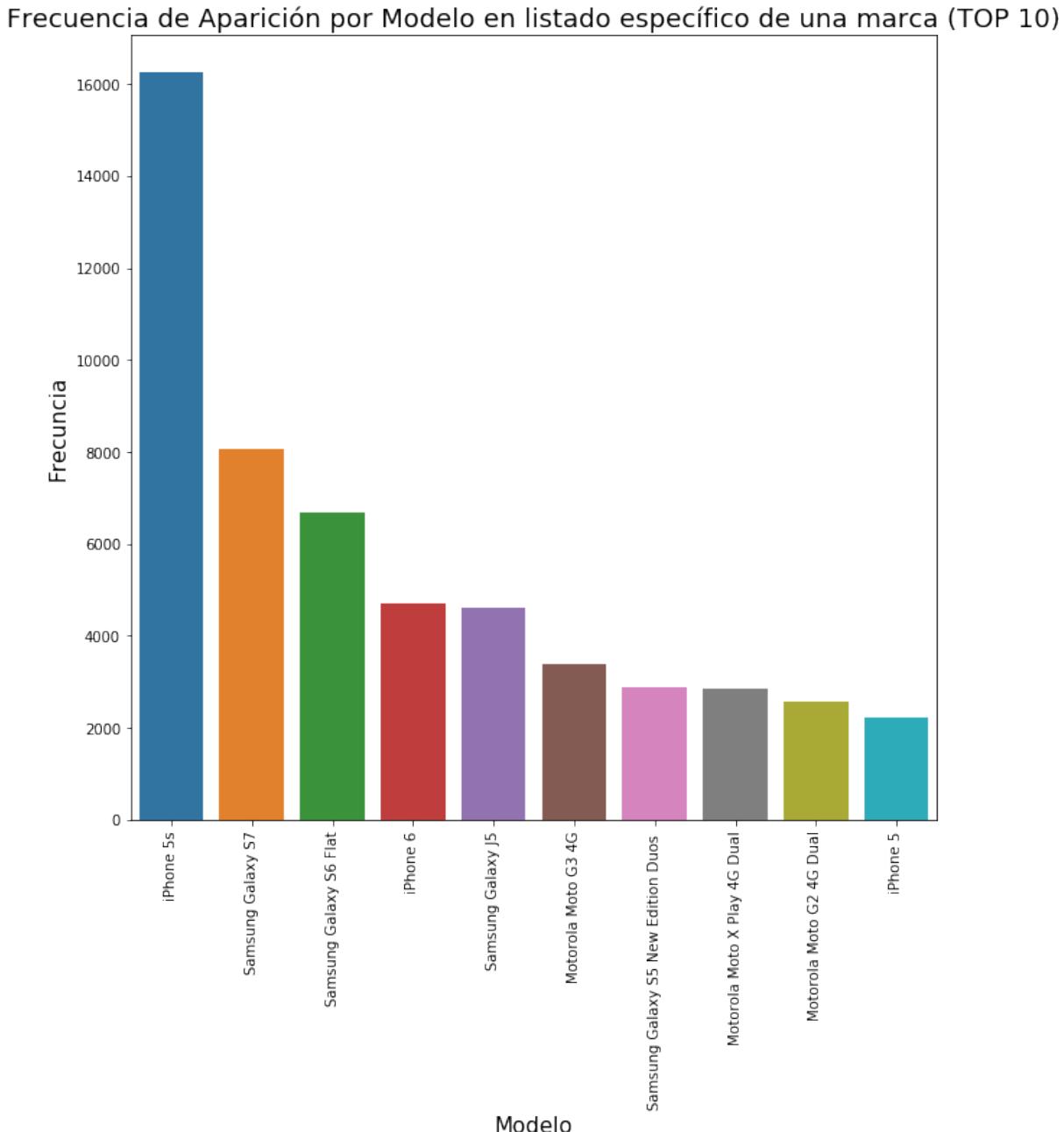


Figura 35: Frecuencia de Aparición por Modelo en listado específico de una marca (TOP 10)

Se puede observar que de los 10 modelos con más apariciones, tenemos 3 iPhones, 3 Motorolas y 4 Samsungs. Con esto, podemos determinar que las tres marcas más buscadas, por ejemplo, a través de algún filtro son las mencionadas anteriormente.

Ahora, comparemos estas apariciones con la cantidad de ventas que se hayan registrado para cada modelo. ¿Sucederá lo mismo que las apariciones en la homepage?

Modelo	Apariciones en listado específico de una marca	Ventas
iPhone 5s	16257	84
Samsung Galaxy S7	8062	18
Samsung Galaxy S6 Flat	6667	39
iPhone 6	4696	71
Samsung Galaxy J5	4613	88
Motorola Moto G3 4G	3389	30
Samsung Galaxy S5 New Edition Duos	2886	8
Motorola Moto X Play 4G Dual	2856	12
Motorola Moto G2 4G Dual	2551	5
iPhone 5	2220	17

A comparación con las apariciones en la homepage, podemos concluir que esta relación es un poco más fuerte, debido a que la cantidad de ventas de casi todos los modelos supera las 20 unidades y el número de ocurrencia es mucho menor.

Incluso, al sumar las ventas de los 10 productos que frecuentemente aparecen en la página de inicio tenemos un total de 65 ventas contra las 372 ventas de los productos con mayor ocurrencia en las apariciones en listados específicos de una marca. Esto se puede dar a que los usuarios, generalmente, buscan el celular que sea acorde a sus gustos y al dinero que está dispuesto a gastar. Por eso, los celulares que aparecen en esta última tabla son modelos más viejos y con menores prestaciones que los de alta gama que aparecen en la pagina principal.

### 3. Conclusiones

Aunque en cada sección fuimos haciendo algún comentario o conclusión respecto a lo que se estaba tratando, es hora de darle un carácter más general y abarcativo a nuestras conclusiones respecto al análisis que recién les presentamos y que este tenga cierta utilidad para la empresa *Trocafone*.

Al principio del informe nos enfocamos en ver donde *Trocafone* tenía mayor actividad y pudimos observar que Brasil era su principal fuente de eventos, más específicamente en las regiones costeras y del sur del país, que tenían un importante contraste con el norte del mismo. Esto se debe a que las zonas con menos visitas se tratan de aquellas que tienen mayor porcentaje de pobreza, e incluso dentro de los estados con más eventos, la mayoría se concentran en la ciudad capital del mismo. Esto genera que en definitiva, los visitantes se encuentren, mayoritariamente, en las capitales de los estados más importantes de Brasil, siendo estas las zonas con mayor poder adquisitivo en promedio.

De acuerdo a esto, dependiendo de la zona en la que se esté, y conociendo las situaciones económicas globales de la región, se podría mostrar en la homepage, celulares más acorde a la situación y no los celulares de alta gama que generalmente aparecen ya que tienen baja probabilidad de venta, según lo analizado. Y de esta manera, se podría aumentar la influencia de la página de inicio de *Trocafone* en las ventas de los celulares que allí se muestran.

Con respecto al resto del mundo, una influencia mayor en el mercado ya sería suficiente con que aumenten la cantidad de visitas y a medida que vaya pasando el tiempo, estas se vayan convirtiendo en ventas.

Luego, habría que aprovechar más el uso de los dispositivos móviles porque éstos tienen una gran cantidad de visitas pero su porcentaje de compra es muchísimo menor a la que tiene la computadora, es verdad que en parte es algo que excede a la empresa y que es una costumbre en la gente preferir comprar en una computadora que en un celular, y más si se trata de una página web. La implementación de una app para dispositivos móviles, en especial para Android, puede llegar a ser una alternativa que genere comodidad y confianza en el cliente.

Una observación que consideramos destacar también es con respecto a las compras realizadas por usuarios nuevos, la cual presenta un 4% del total de las compras que se hicieron y simplemente podría mejorarse con algún posible descuento por ser nuevo.

Por otro lado se pudo ver que durante los meses de Enero y Febrero no hubieron casi ventas comparado al resto de los meses analizados, tal vez con tan solo promociones podrían generar más ventas en dichos meses.

Respecto a los canales de ingreso al sitio, es abrumador el tráfico que se genera por los medios pagos, primordialmente de la mano de Google, Criteo y RTB House, por sobre los canales restantes. De hecho, es aquel que durante el transcurso del día la mantiene con constante actividad. Por lo tanto, es una buena idea seguir invirtiendo en las campañas de marketing que tengan a estos medios para generar mayor publicidad y alcance en los clientes. En cuanto al resto de los orígenes de los canales pagos, generan un tráfico de visitantes casi despreciable frente a lo mencionado con anterioridad.

Otro aspecto que nos llamó la atención fue que los celulares más vendidos en el período Enero 2018- Junio 2018 (inclusive) fueron todos modelos presentados en años anteriores y ninguno se trata de los últimos modelos de las marcas líderes, como son Samsung, iPhone y Motorola. Por lo tanto, estos modelos pueden ser grandes candidatos a ser los que más frecuencia tengan en la homepage de *Trocafone*.

También se pueden repetir muchas otras conclusiones ya mencionadas a lo largo del informe, pero acá seleccionamos algunas que nos generaron más inquietudes y que creemos que pueden ser llevadas a la práctica.