
Organización de Datos

Trabajo Práctico 1

Análisis exploratorio de datos de Trocafone



Integrante	Padrón	Mail
ROBLES, Gabriel	95897	gabyrobles93@gmail.com
ALVAREZ WINDEY, Ariel Justo	97893	arieljaw12@gmail.com
DIZ MARTIN, Gonzalo	98546	dizmartin.gonzalo@gmail.com
RAMUNDO, Matías	100327	matiramundo@gmail.com

Contenido

1. Objetivo.....	3
2. Análisis Exploratorio.....	3
Forma del set de datos.....	3
Rango temporal del set de datos.....	3
Transformaciones de tipos de datos.....	3
Preguntas interesantes.....	3
1. ¿En qué proporción son nulos los atributos?.....	4
2. ¿Existe alguna correlación entre atributos nulos y tipos de evento?.....	5
3. ¿Todos los nulos que existen obedecen a la regla de "no todos los atributos se corresponden con todos los eventos", o son por dato faltante?.....	6
4. ¿Cual es la proporción entre distintos tipos de eventos en este set de datos?.....	6
5. ¿Cuál es el rango temporal de los eventos y cómo se reparten en el calendario?.....	7
3. Análisis según dispositivos y regiones.....	8
4. Análisis sobre usuarios viejos y nuevos.....	22
4.1 De todos los eventos generados ¿Cuántos fueron de usuarios nuevos y cuántos de usuarios viejos?.....	22
4.2 ¿Cuales son los canales por los que llegaron mas usuarios nuevos a lo largo de los meses?.....	23
4.3 ¿ Por que canal vuelven los usuarios que ya visitaron la pagina anteriormente?.....	23
Conclusiones.....	24
5. Analisis campañas de anuncios.....	26
5.1 ¿De todas las campañas de anuncios cuales son las que mas visitas generaron?.....	26
5.2 ¿ Cuales fueron las campañas que mas conversiones y checkouts generaron?.....	26
6. Análisis de las marcas y sus dispositivos.....	31
6.1 ¿Cuáles son las marcas de los dispositivos en la página Trocafone?.....	31
6.2 Analisis de las marcas.....	31
6.2.5 ¿Podría influir positivamente la cantidad de productos de cada marca para dar alternativas a los usuarios y así incrementar las conversiones aunque haya falta de stock de algunos dispositivos?.....	36
6.2.6 ¿Tiene alguna relación la falta de stock de dispositivos de las distintas marcas con la cantidad de conversiones realizadas?.....	37
6.2.7 Búsquedas de las distintas marcas en Trocafone.....	38
6.2.8 Búsquedas de las distintas marcas en Trocafone por mes.....	39
Apple.....	43
6.3 Análisis de dispositivos individuales.....	49
6.3.1 Top 20 de dispositivos con más eventos.....	50
6.3.2 Top 20 de dispositivos con más eventos “checkout”.....	51
.....	51
6.3.3 Top 20 de dispositivos con más conversiones.....	51
6.4 Conclusiones.....	52
7. Análisis de ventas según la condición de los dispositivos.....	54
7.1 ¿Cuántos checkouts se generaron dependiendo la condición de los dispositivos?.....	54
7.2 ¿En que condición estaban la mayor cantidad de celulares vendidos?.....	55
7.3 ¿Que celular lidera en mayor cantidad de checkouts y en que condición?.....	56
7.4 ¿Que celular lidera en mayor cantidad de conversiones y en que condición?.....	57
8. Conclusiones.....	58
8.1 Sobre el análisis exploratorio del set de datos.....	58
8.2 Sobre el análisis según los dispositivos y las regiones.....	58
8.3 Sobre el análisis según los canales de ingreso a la plataforma por parte de usuarios nuevos y por los que retornan a la plataforma.....	58
8.4 Sobre el analisis de las campañas de anuncios.....	59

8.5 Conclusiones sobre la cantidad de ventas y checkouts de celulares según condición...	59
9. Links.....	59
Kaggle.....	59
Github.....	59

1. Objetivo

El objetivo de este trabajo es realizar un análisis exploratorio del set de datos de eventos de web analytics de usuarios que visitaron www.trocafone.com. Queremos ver qué cosas podemos descubrir sobre los datos que puedan resultar interesantes.

2. Análisis Exploratorio

En este apartado se pretende lograr un entendimiento básico de los datos con los que se cuenta y se va a trabajar. Exploraremos los datos sin (aún) hacerse preguntas sobre los mismos.

Forma del set de datos

El dataset cuenta con 1011288 datos (observaciones) con 23 atributos con un tamaño total de 109 megabytes.

Rango temporal del set de datos

Se tienen datos de comienzos de 2018 hasta mediados del mes de Junio.

Transformaciones de tipos de datos

La fuente de datos csv fue cargada sin forzar tipo de datos. Los tipos de datos fueron asignados a mano. Creemos importante transformar el atributo *timestamp* dado que como *object* no es amigable de trabajar. Convirtiéndolo a tipo *datetime* se podrán aprovechar features de pandas sobre este tipo de datos. Además los atributos *condition*, *channel*, *new_vs_returning*, *country* y *device type* pueden ser categorizados.

Preguntas interesantes

En este apartado intentaremos responder las siguientes preguntas:

1. ¿En qué proporción son nulos los atributos?
2. ¿Existe alguna correlación entre atributos nulos y tipos de evento?
3. ¿Todos los nulos que existen obedecen a la regla de "no todos los atributos se corresponden con todos los eventos", o son por dato faltante?
4. ¿Cual es la proporción entre distintos tipos de eventos en este set de datos?
5. ¿Cuál es el rango temporal de los eventos y cómo se reparten en el calendario?

1. ¿En qué proporción son nulos los atributos?

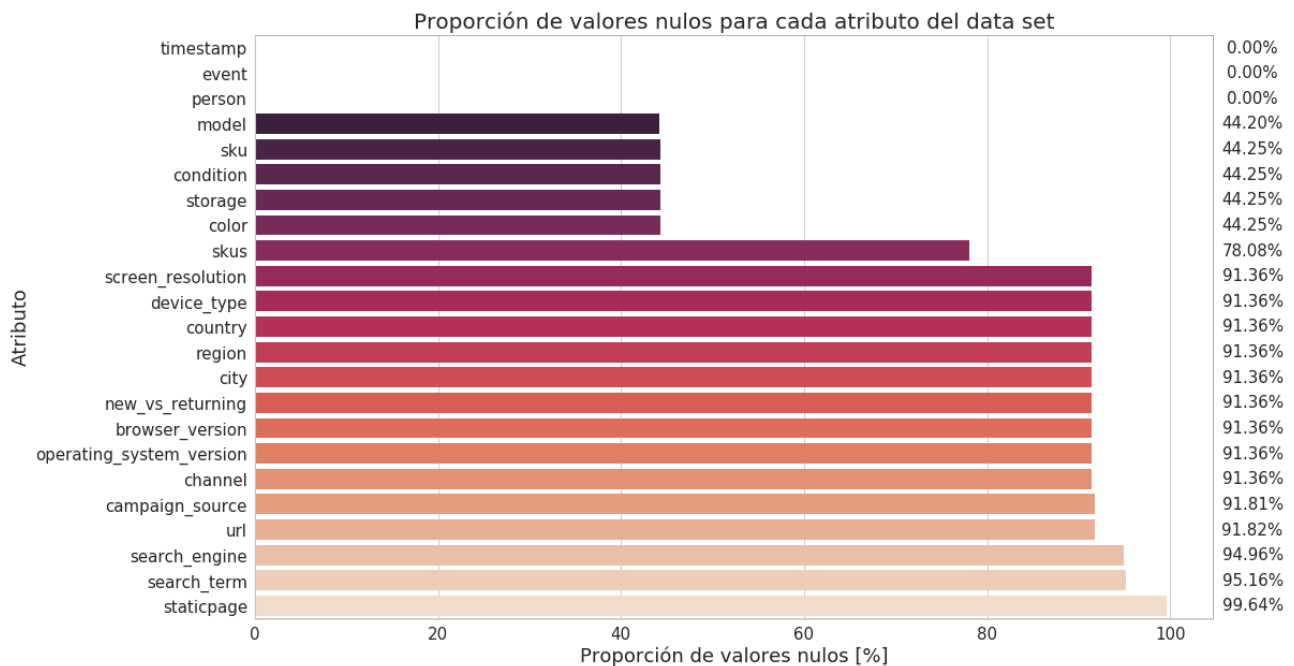


Figura 1

Como puede observarse en la figura 1, hay tres atributos que nunca son nulos: *timestamp*, *event* y *person*. Además, existen grupos de atributos con una proporción idéntica de valores nulos. Esta última observación estará relacionada a la siguiente pregunta a contestar.

Se puede observar en la figura 2 que casi el 70% de todos los atributos del data set son nulos.



Figura 2

2. ¿Existe alguna correlación entre atributos nulos y tipos de evento?

Todo indica que sí, el enunciado aclara:

"Algo a tener en cuenta es que no todos los datos descritos en las columnas corresponde a todos los tipos de eventos."

Lo que intentaremos responder entonces es ¿Qué atributo se corresponde con qué tipo de evento?

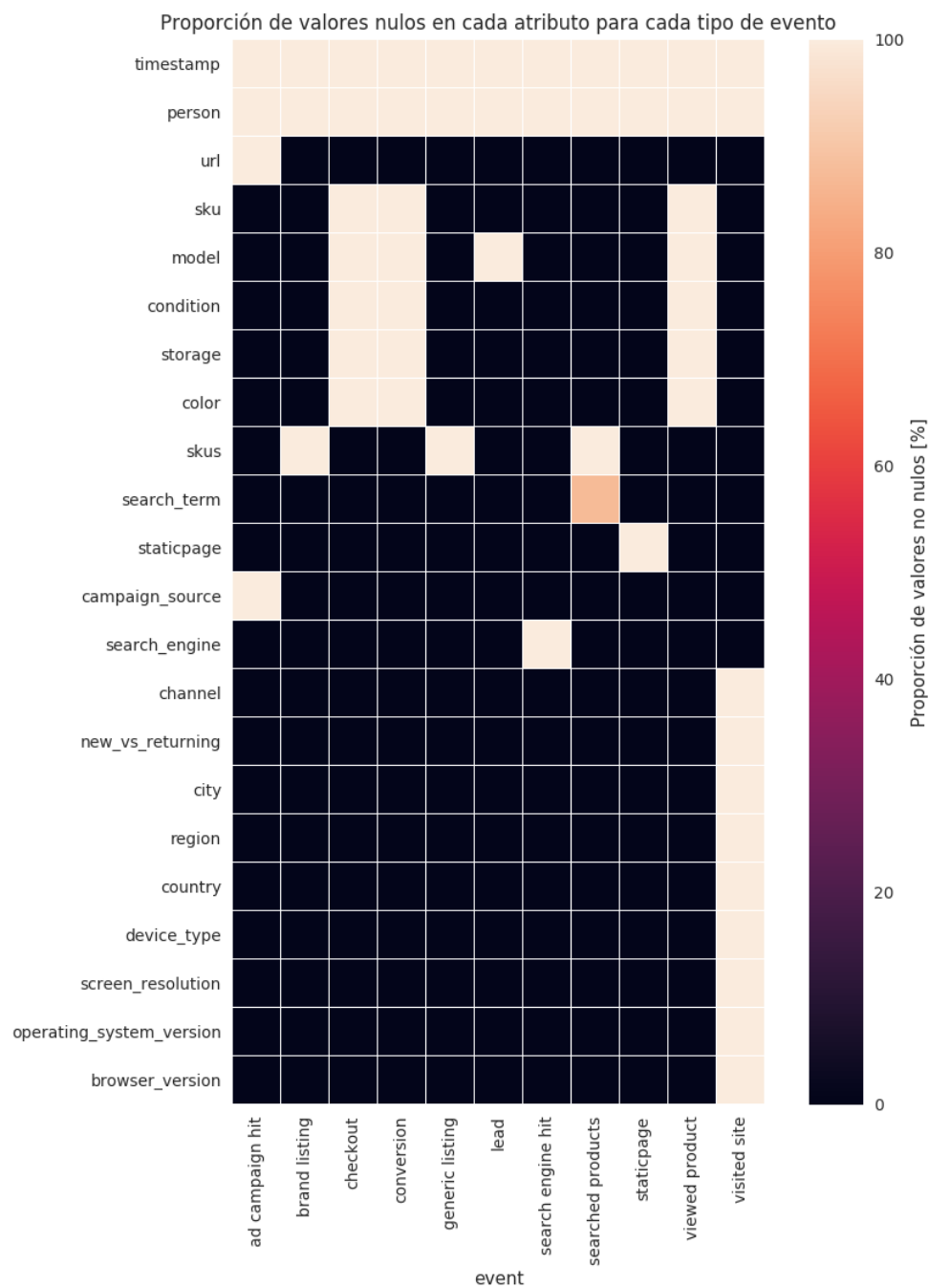


Figura 3

En el anterior gráfico se respode la pregunta de ¿Qué atributo se corresponde con qué tipo de evento?

Haciendo una lectura horizontal, puede observarse que:

1. Puede observarse que los atributos timestamp y person son comunes a todos los eventos
2. El atributo url sólo corresponde al evento ad_campaign_hit
3. Los atributos sku, condition, storage y color son compartido entre los eventos checkout, conversion y viewed product.
4. El atributo model es compartido entre checkout, conversion, lead y viewed product.
5. El atributo skus le corresponde a los eventos brand listing, generic listinc y searched products
6. El atributo search term sólomente le corresponde al evento searched products. Cabe destacar que el mismo no aparece como valor no nulo en un 100% de las observaciones.
7. El atributo static page le corresponde al evento staticpage únicamente.
8. El atriuto campaign source le corresponde únicamente al evento ad campaign hit.
9. El atriuto search engine le corresponde únicamente al evento sarch engine hit.
10. Los atributos channel, new_vs_returning, city, region, country, device_type, screen resolution, operating system version, y browser version le corresponden únicamente al evento visited site.

3. ¿Todos los nulos que existen obedecen a la regla de "no todos los atributos se corresponden con todos los eventos", o son por dato faltante?

En el anterior gráfico puede observarse claramente que la regla de "no todos los atributos se corresponden con todos los eventos" se cumple. Esto explica que haya tanta cantidad de atributos nulos. Sin embargo, existen atributos que se corresponden a un tipo de evento que tienen una pequeña proporción de observaciones nulas.

El caso mas notorio es para el evento 'searched products' con su atributo 'search term' con un aproximado de 13% de oservaciones nulas. Dicho evento significa "El usuario realiza una búsqueda de productos en la interfaz de búsqueda del site." por lo que podemos asumir que el usuario hizo una búsqueda sin un string particular.

El resto de atriutos nulos para ciertos tipos de eventos puede ser despreciable, ya que no superan el 0,1%.

4. ¿Cual es la proporción entre distintos tipos de eventos en este set de datos?

En la Figura 4 puede observarse que el evento predominante por sobre el resto es 'viewed product'.

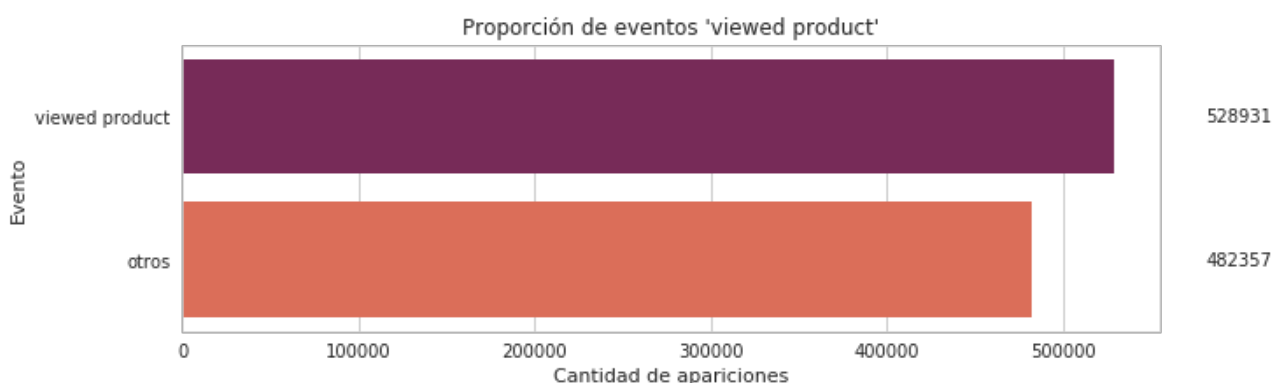


Figura 4

El resto de los eventos se reparten en la siguiente proporción:

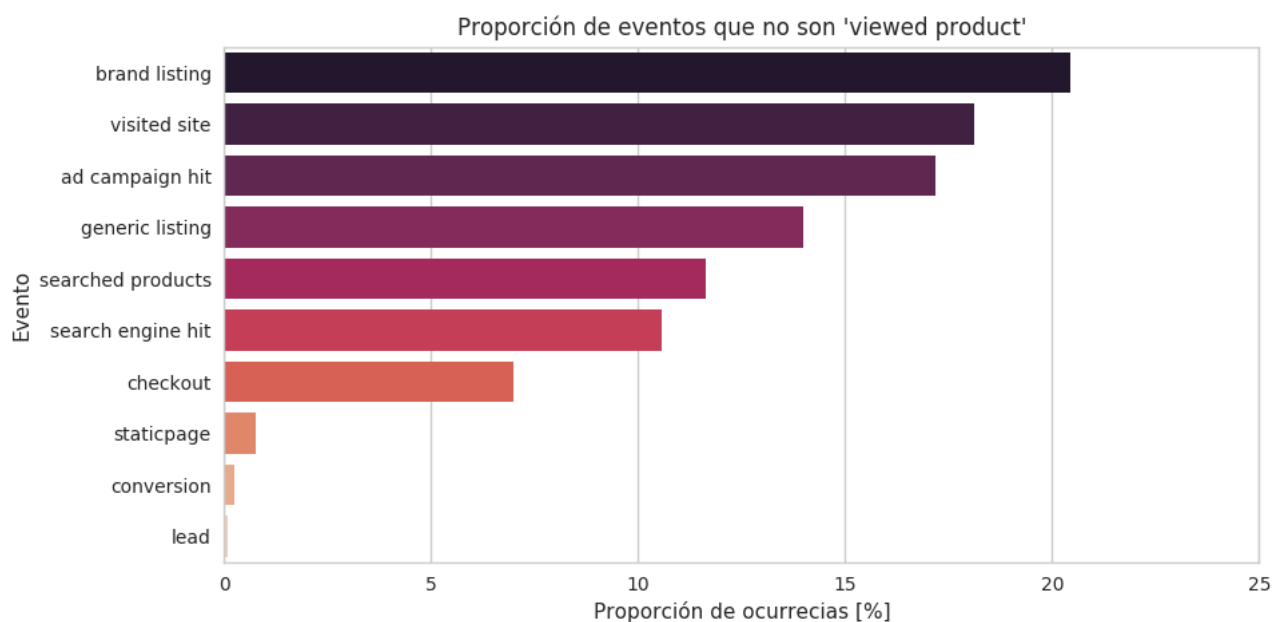


Figura 5

5. ¿Cuál es el rango temporal de los eventos y cómo se reparten en el calendario?

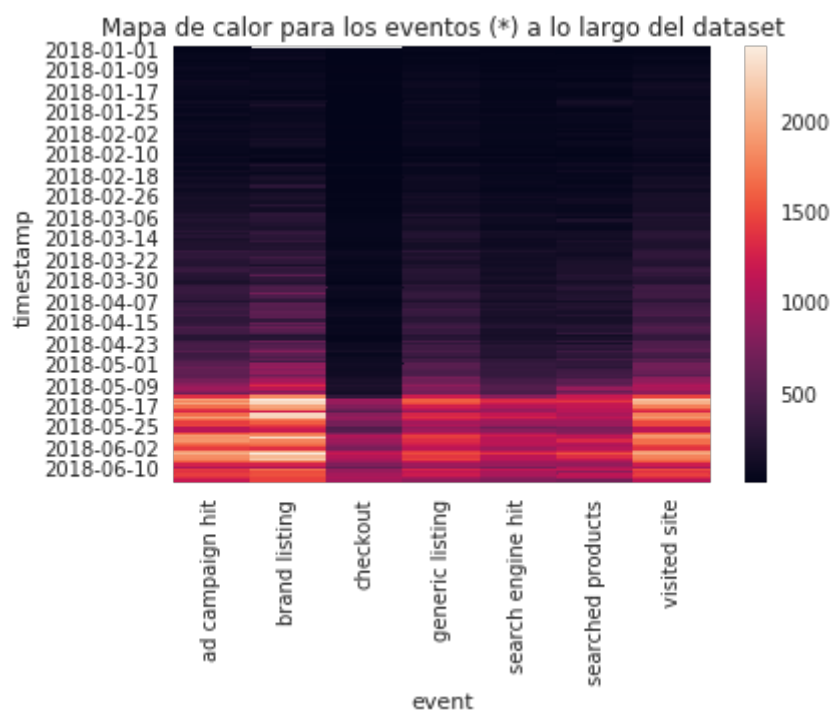


Figura 6

Para la Figura 6 se descartaron los eventos 'viewed product' (por ser altamente predominante) y los eventos 'conversion', 'lead' y 'staticpage' por tener una presencia muy inferior al resto. Puede observarse que la cantidad de eventos se disparó hasta tres veces a partir de mediados de mayo en forma pareja para cada tipo de evento.

3. Análisis según dispositivos y regiones

En esta sección se buscará analizar el comportamiento de los usuarios según desde el dispositivo en el que acceden y su ubicación geográfica. Para llevar a cabo esto, se filtraron los accesos al sitio por el país de Brasil, como los datos corresponden al eCommerce de este país, se consideraron accesos desde otro sitio como datos irrelevantes para este análisis. En particular, el evento que registra las regiones y los dispositivos es el de **visited site**, del cual se tienen 87.378 registros. El 96% de ellos (84.308) provienen de Brasil, por esto tiene sentido enfocar el estudio a este país en particular.

Especificando aún más, se investigó de qué regiones eran los registros. En la Figura 7 se puede observar la cantidad de eventos según la región de Brasil:

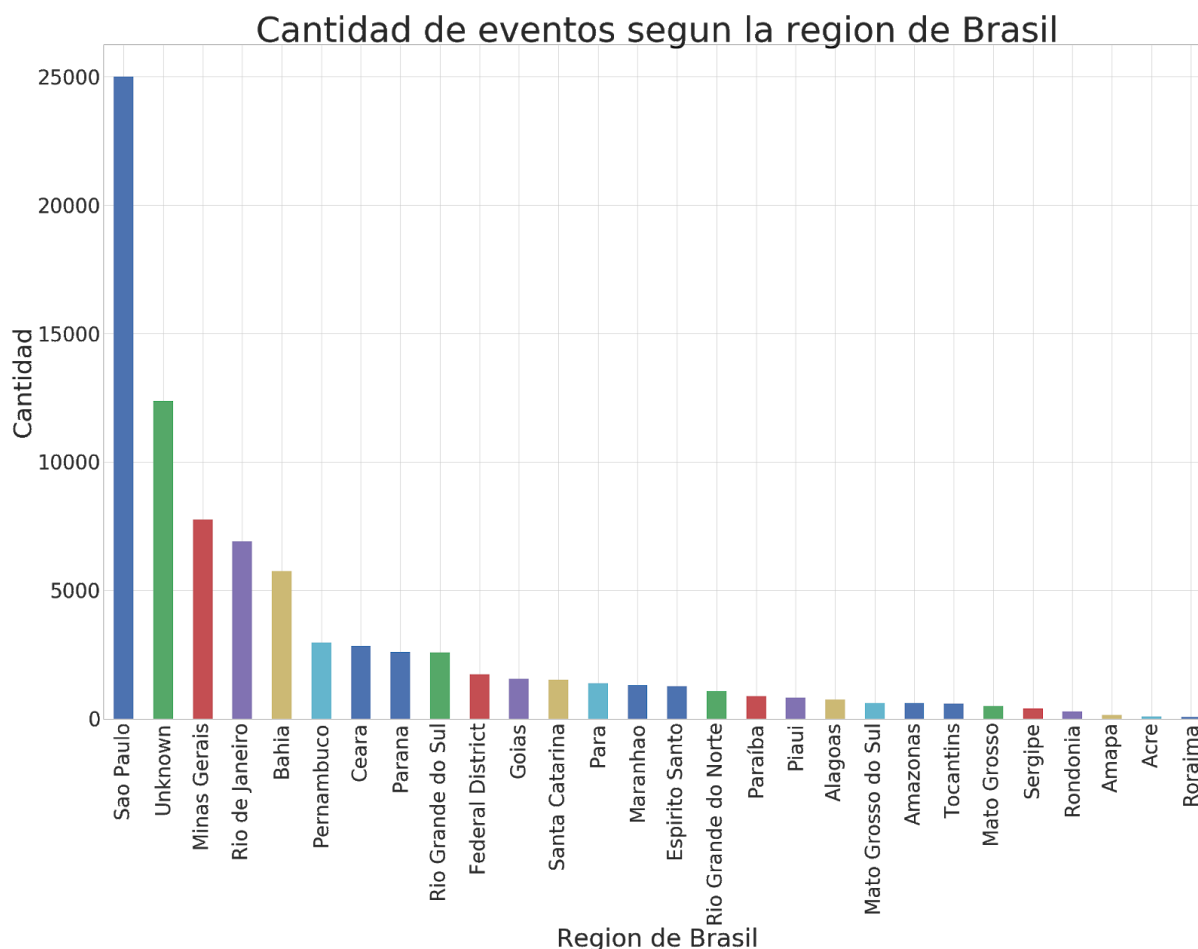


Figura 7: Cantidad de visitas al sitio según la región de Brasil

Como se puede observar hay demasiadas categorías para realizar un análisis genérico, además que el segundo valor más común es 'Unknown', es por esto que en análisis posteriores se filtraron los registros de los cuales se desconocía el origen.

Para reducir la categoría de las regiones se realizó un agrupamiento de las mismas según su posición geográfica. Se tomaron las cinco regiones geográficas que se pueden ver a continuación en la Figura 8:



Figura 8: Distintas regiones geográficas de Brasil

Haciendo esta especialización se analizó la cantidad de visitas al sitio según la región geográfica de Brasil, obteniéndose el resultado de la Figura 9:

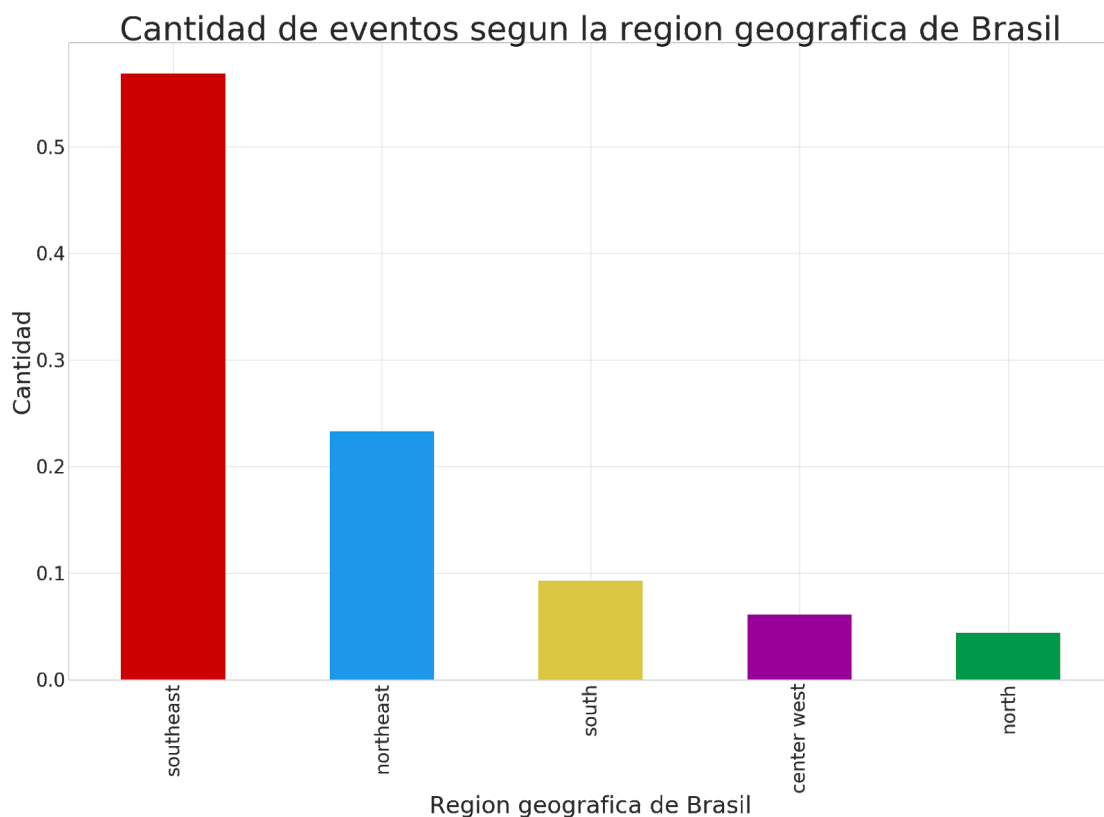


Figura 9: Cantidad de visitas al sitio según la región geográfica de Brasil

Aquí podemos observar que, sobre un total de 71.930 visitas al sitio, que predominan ampliamente la zona del sureste y noreste de Brasil, con un 57% y 23% respectivamente. Mientras que la zona del sur, centro oeste y norte tienen un 9%, 6% y 4% respectivamente.

A partir de esta información se buscó investigar el comportamiento de estas visitas al sitio según el tipo del dispositivo desde el que accedían los usuarios. Los distintos tipos se detallan en la Tabla 1 a continuación:

Dispositivo	Cantidad de visitas al sitio
Smartphone	42.775
Computer	40.239
Tablet	1.180
Desconocido	114

Tabla 1: Cantidad de visitas al sitio según el dispositivo con el que se ingresó

Dado que los dispositivos predominantes son celulares y computadoras, los análisis posteriores se especificaron para estos tipos.

A fin de realizar un estudio sobre **la experiencia del usuario** en la plataforma de Trocafone según esté accediendo desde la interfaz mobile o de escritorio se analizó la columna ***new_vs_returning*** que indica si un usuario accede por primera vez a la plataforma (new) o si ya había accedido previamente (returning).

En la Figura 10 que se muestra a continuación podemos observar desde qué dispositivo accedieron por primera vez los usuarios a Trocafone:

Cantidad de usuarios que entraron por primera vez al sitio segun su dispositivo

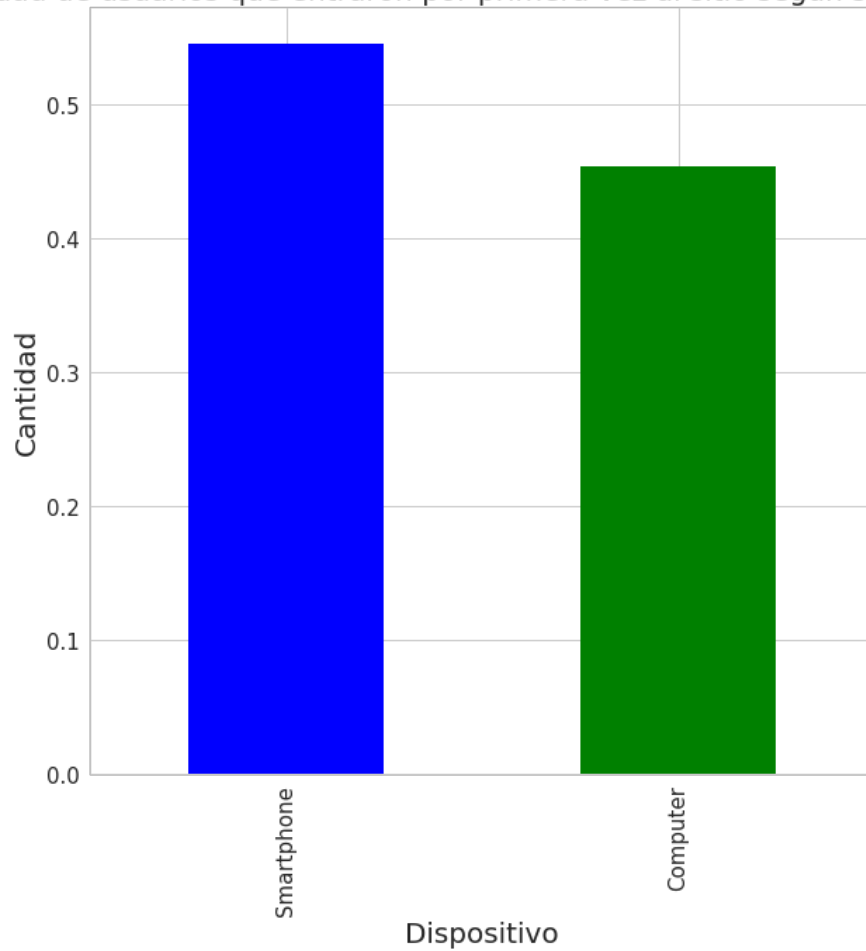


Figura 10: Cantidad de usuarios que entraron por primera vez al sitio según dispositivo

En primera instancia se ve que los accesos por primera vez fueron en su mayoría por celular, pero cuando analizamos el caso de los **returning** vemos que se emparejan al 50% ambos dispositivos, con lo que no se puede afirmar que haya una diferencia entre la experiencia del usuario dependiendo del dispositivo del cual accede:

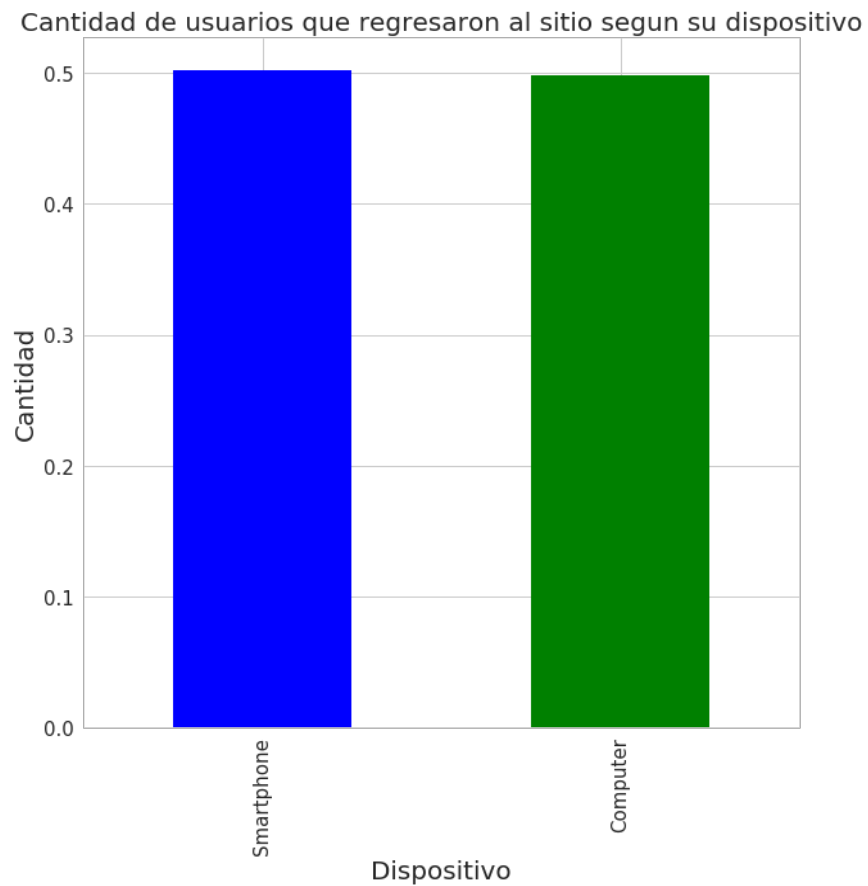


Figura 11: Cantidad de usuarios que regresaron al sitio según dispositivo

A continuación se buscó ver posibles relaciones entre los dispositivos de los que acceden los usuarios y la región geográfica a la que pertenecen. Agrupando por la región geográfica previamente definida y si el acceso fue con celular o computadora, se obtuvieron los siguientes resultados:

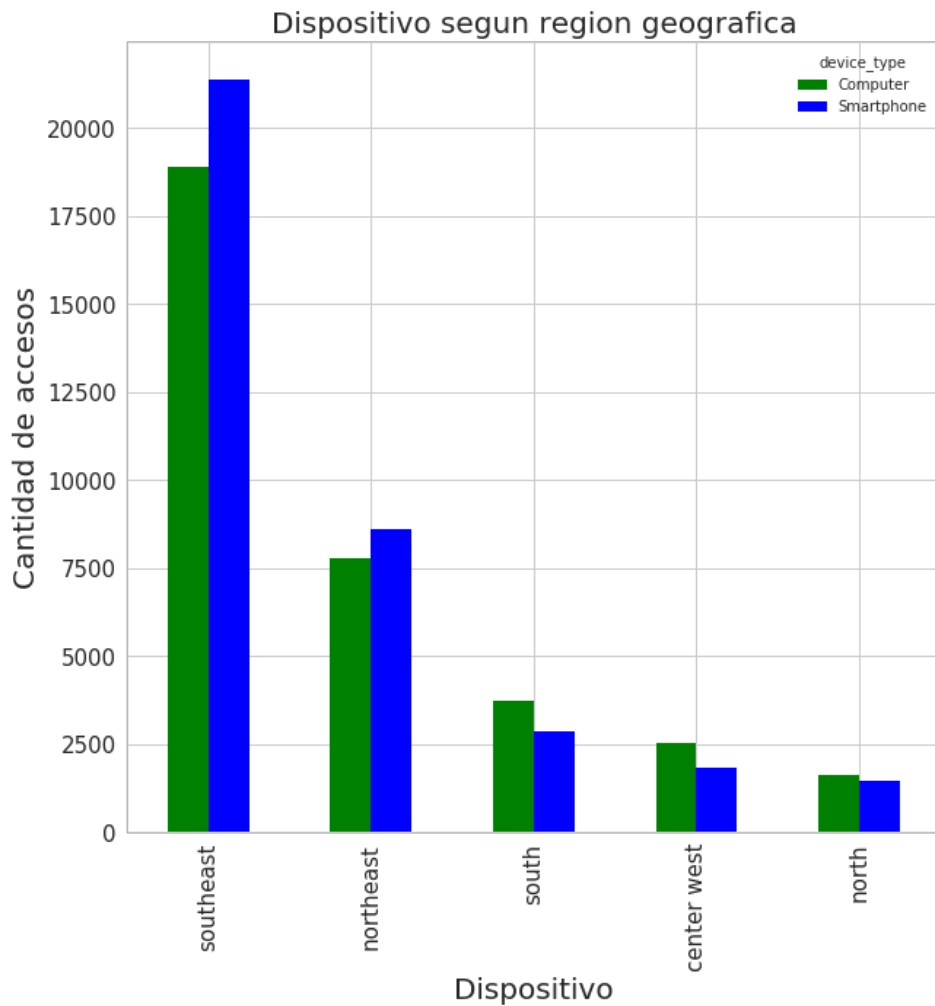


Figura 12: Accesos según el dispositivo y la región geográfica

En la Figura 12 se puede observar que el uso de los dispositivos es bastante parejo pero para las regiones que más visitas generan (sureste y noreste) predomina por poco el uso del celular, mientras que para las otras regiones es un poco mayor el uso de la computadora frente al dispositivo móvil.

Habiendo analizado esto, ahora buscamos ver cómo se distribuye el comportamiento de los usuarios según el dispositivo que usan dependiendo la hora del día. En un primer análisis se vio que es mayor el flujo de eventos en los días de semana, mientras que para los fines de semana la actividad es menor. Por ello para el análisis siguiente se separaron los datos entre estos dos casos:

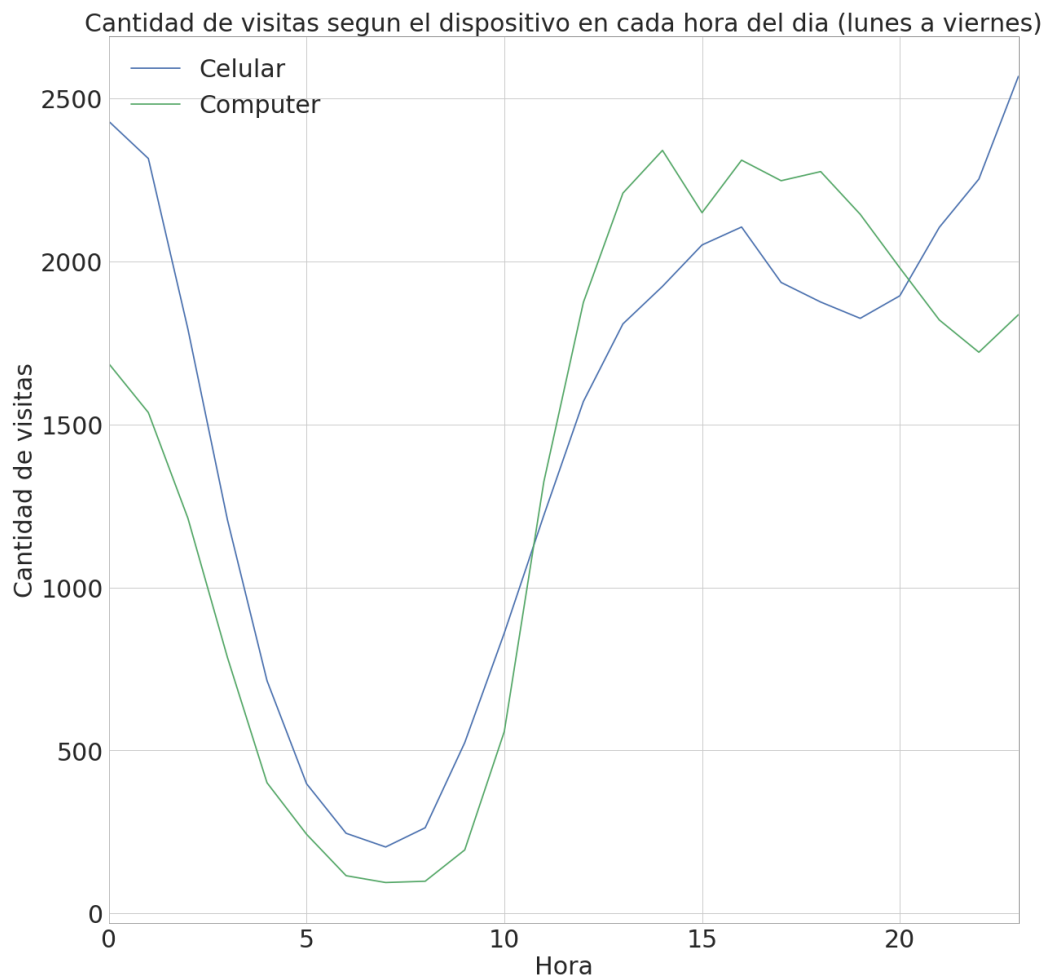


Figura 13: Visitas al sitio los días de semana según el horario y el dispositivo

Para los días de semana podemos observar un comportamiento bastante parejo entre ambos dispositivos. El comportamiento lógico que se observa en esta Figura y también en las siguientes es la disminución del flujo de visitas en horario nocturno y la mañana. Para este caso de 02:00 hs hasta las 11:00 hs aproximadamente la cantidad de visitas por hora es menor a 1.000. La franja en la que más se utiliza la plataforma es desde las 12:00 / 13:00 hs hasta las 00:00 / 01:00hs. Otro detalle a tener en cuenta es cómo en la franja de mayor actividad toma protagonismo el acceso por computadoras desde las 11:00 hs hasta las 20:00hs donde el acceso se vuelve más frecuente por celular.

En la Figura 14 que se ve a continuación se ve el comportamiento para los días sábado y domingo:

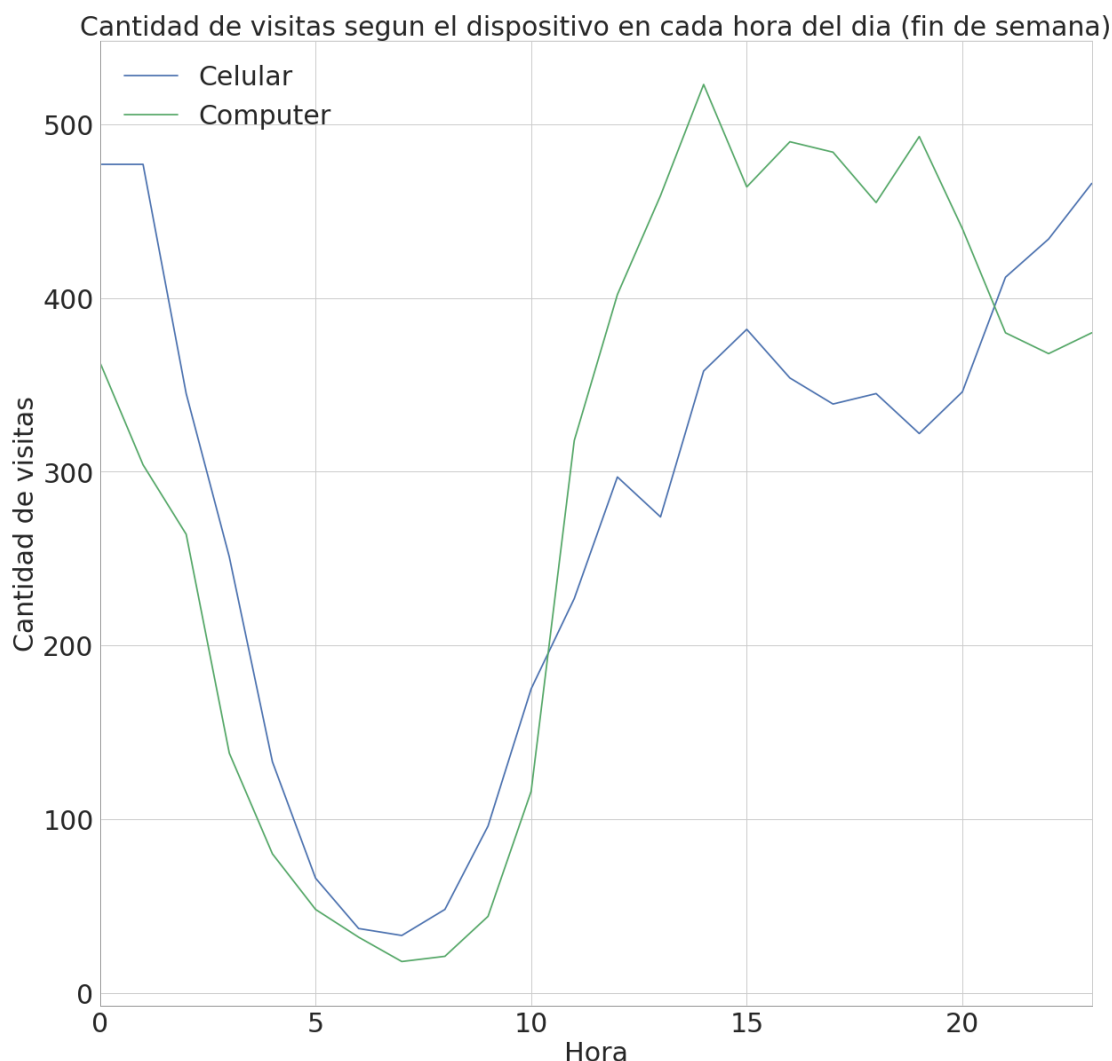


Figura 14: Visitas al sitio los días sábado y domingo según el horario y el dispositivo

Para el fin de semana se observa un comportamiento más errático entre las distintas horas pero las franjas de mayor y menor actividad son prácticamente las mismas. También se observa una mayor predominancia en el uso de la computadora en la franja de 12:00 hs a 20:00hs. De todos modos es importante notar que el pico de este gráfico es de aproximadamente 500 visitas por hora (sin contar la de celulares), que es un número cercano a la cantidad de visitas por hora que ocurren los días de semana en la franja de menor actividad (03:00 hs a 10:00 hs).

A fin de hacer un estudio más específico, en las siguientes figuras se observa el mismo análisis de las visitas según el dispositivo (semana y fin de semana) pero particularizando para cada región geográfica:

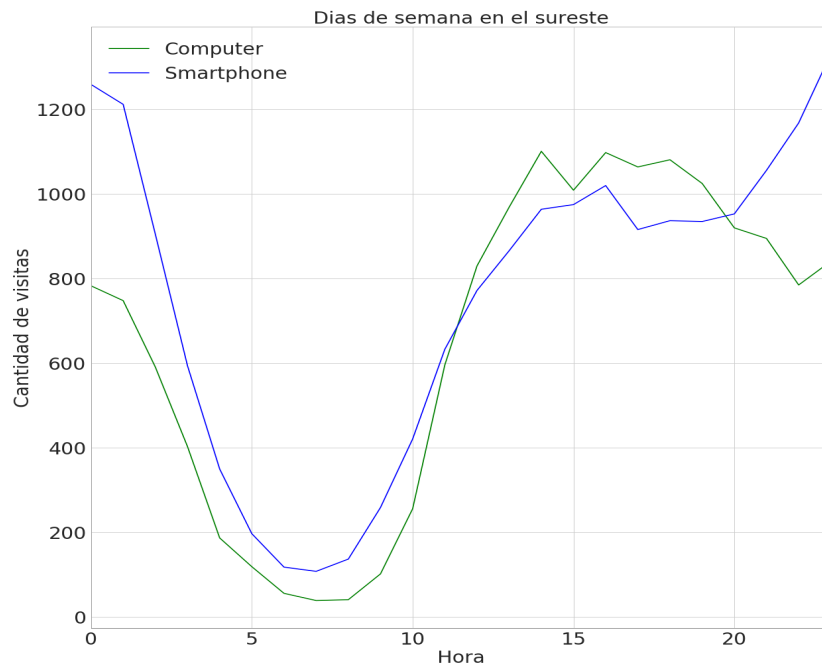


Figura 15: Visitas al sitio según dispositivo y hora del día en la semana (sureste)

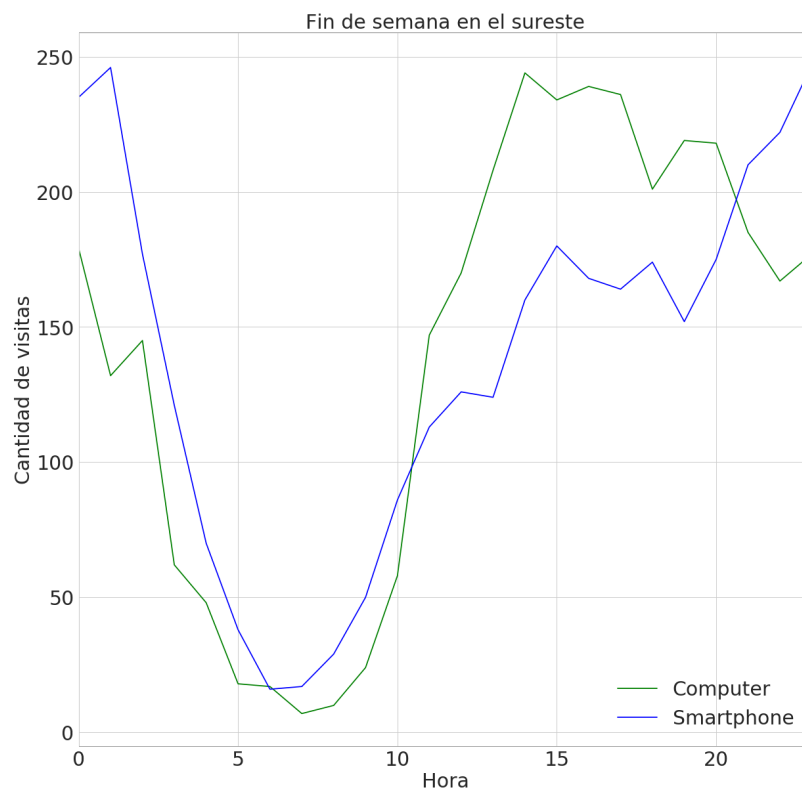


Figura 16: Visitas al sitio según dispositivo y hora del día en el fin de semana (sureste)

Para la región sureste de Brasil se observa un comportamiento similar al visto en el general que agrupaba todas las regiones, esto es ya que es la región con mayor cantidad de eventos y por lo tanto la que mayor peso en el caso genérico. Se puede observar una menor diferencia entre celular y computadora en la franja de mayor actividad.

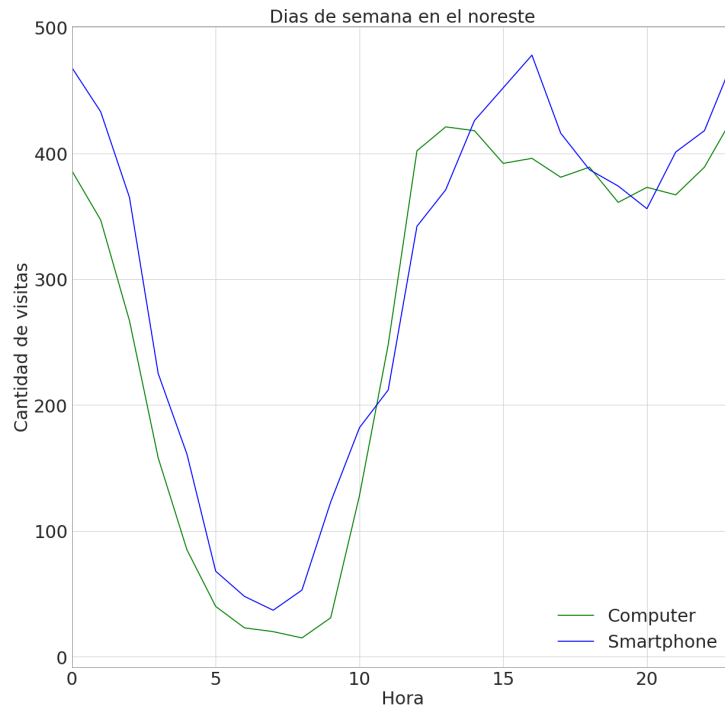


Figura 17: Visitas al sitio según dispositivo y hora del día en la semana (noreste)

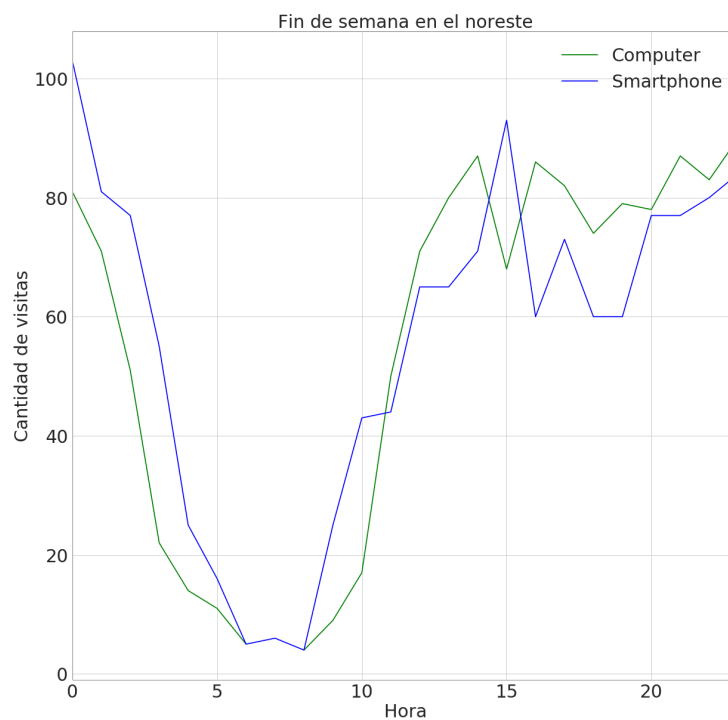


Figura 18: Visitas al sitio según dispositivo y hora del día en el fin de semana (noreste)

Para el noreste se puede observar un comportamiento aún más parejo entre el celular y la computadora para los horarios que presentan mayor diferencia, incluso en los fines de semana. En estas franjas por momentos predomina el celular y por momentos la computadora.

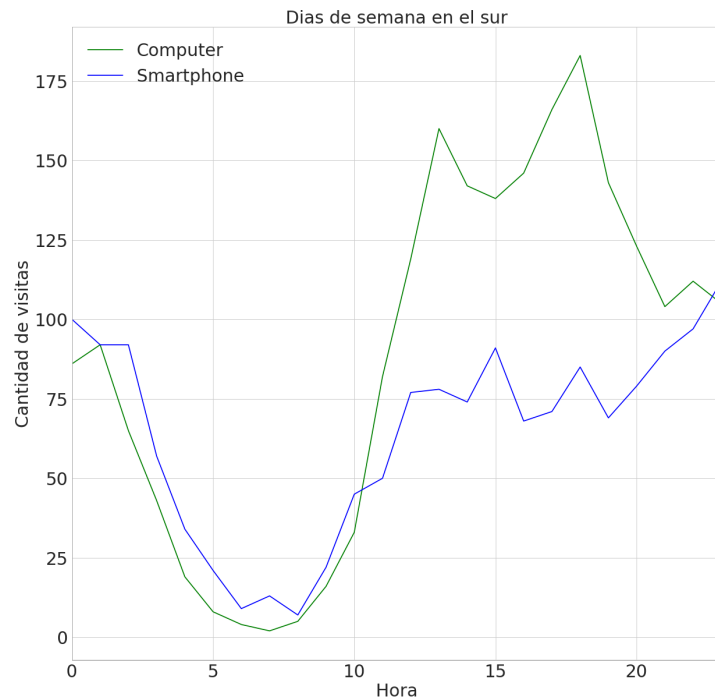


Figura 19: Visitas al sitio según dispositivo y hora del día en la semana (sur)

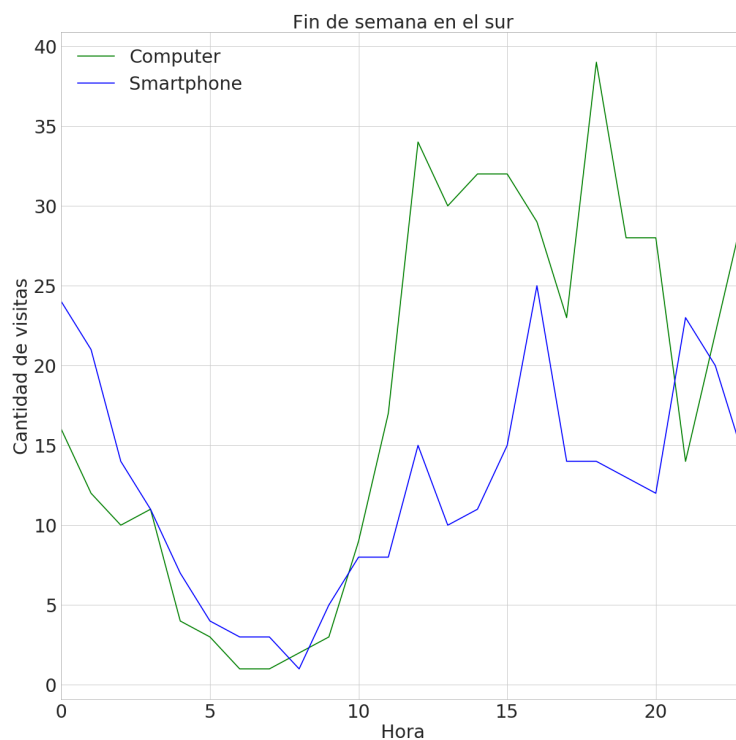


Figura 20: Visitas al sitio según dispositivo y hora del día en el fin de semana (sur)

Para la región sur se ve, durante la semana, mayor diferencia distancia entre el uso de celular y computadora en la franja de mayor actividad. En el fin de semana también con la observación de que aproximadamente a las 16:00 hs el celular empieza a tener mayor actividad y la computadora decrece, luego termina nuevamente predominando esta última en lo que queda de la franja.

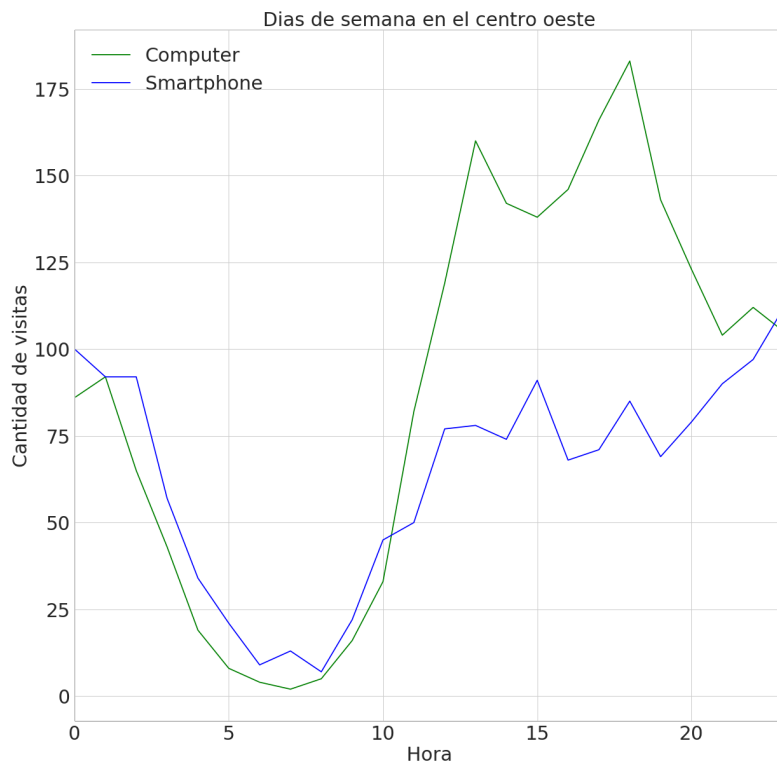


Figura 21: Visitas al sitio según dispositivo y hora del día en la semana (centro oeste)

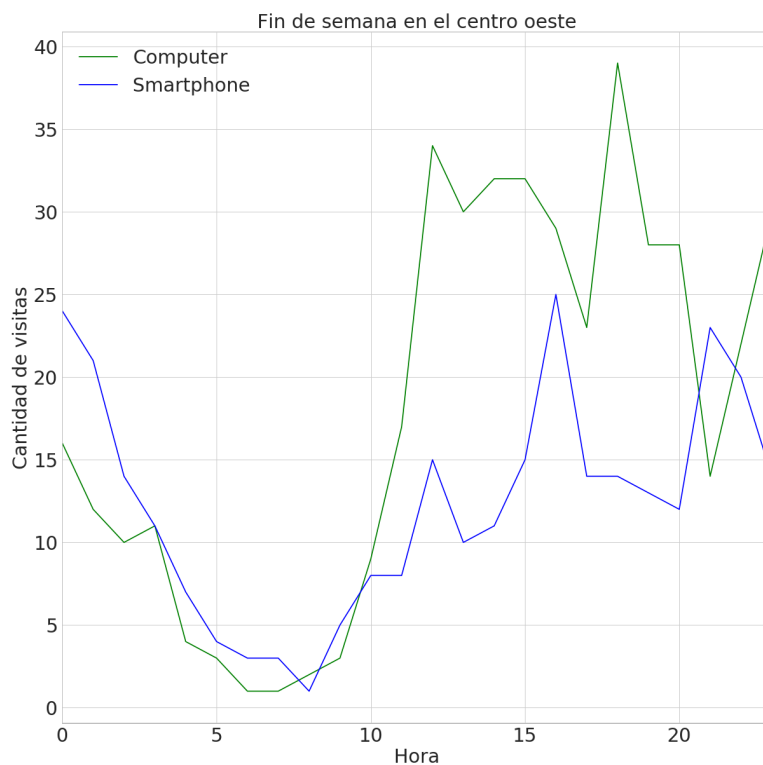


Figura 22: Visitas al sitio según dispositivo y hora del día en el fin de semana (centro oeste)

El comportamiento del centro oeste es muy similar al del sur, con el mismo máximo y mínimo a las 16:00 hs para celular y computadora respectivamente.

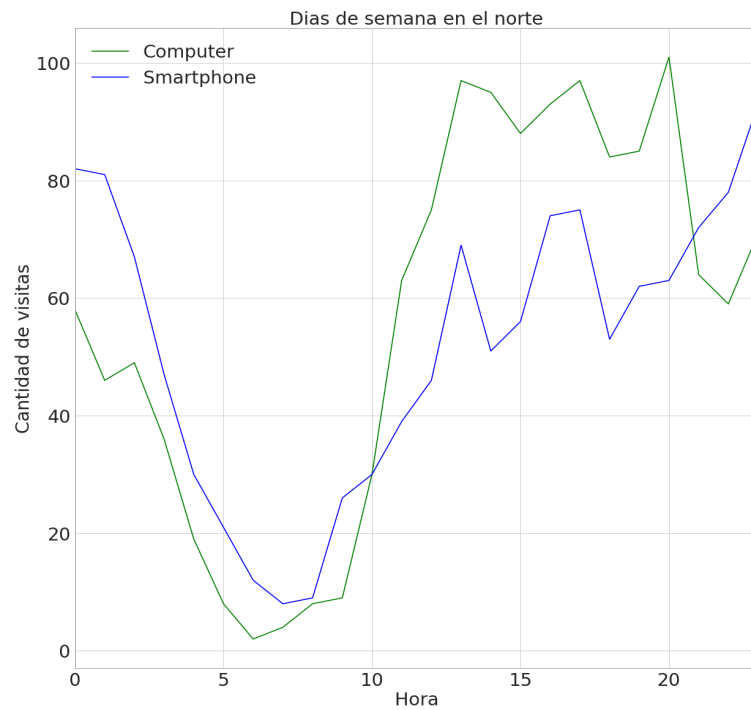


Figura 23: Visitas al sitio según dispositivo y hora del día en la semana (norte)

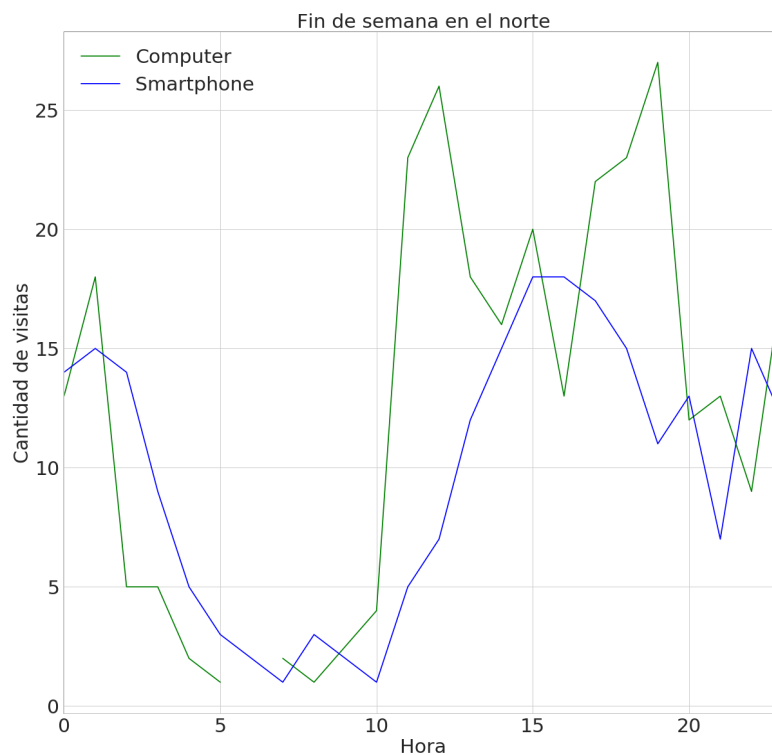


Figura 24: Visitas al sitio según dispositivo y hora del día en el fin de semana (norte)

Para el caso del norte el comportamiento es similar a los dos anteriores con la salvedad que la diferencia entre los dispositivos durante la semana en proporción es menor y que durante el fin de semana en el pico de las 16.00 hs el aumento del celular es más uniforme y llega incluso a superar a la computadora.

4. Análisis sobre usuarios viejos y nuevos

Se entiende como usuario nuevo a las personas que no habían accedido a la plataforma anteriormente y como usuario viejo a las personas que si lo habían hecho.

4.1 De todos los eventos generados ¿Cuántos fueron de usuarios nuevos y cuantos de usuarios viejos?

Como se puede observar en la Figura 25 , entre los meses de enero y junio, mas del doble de las visitas fueron generadas por usuarios que ya habían accedido a la plataforma anteriormente.

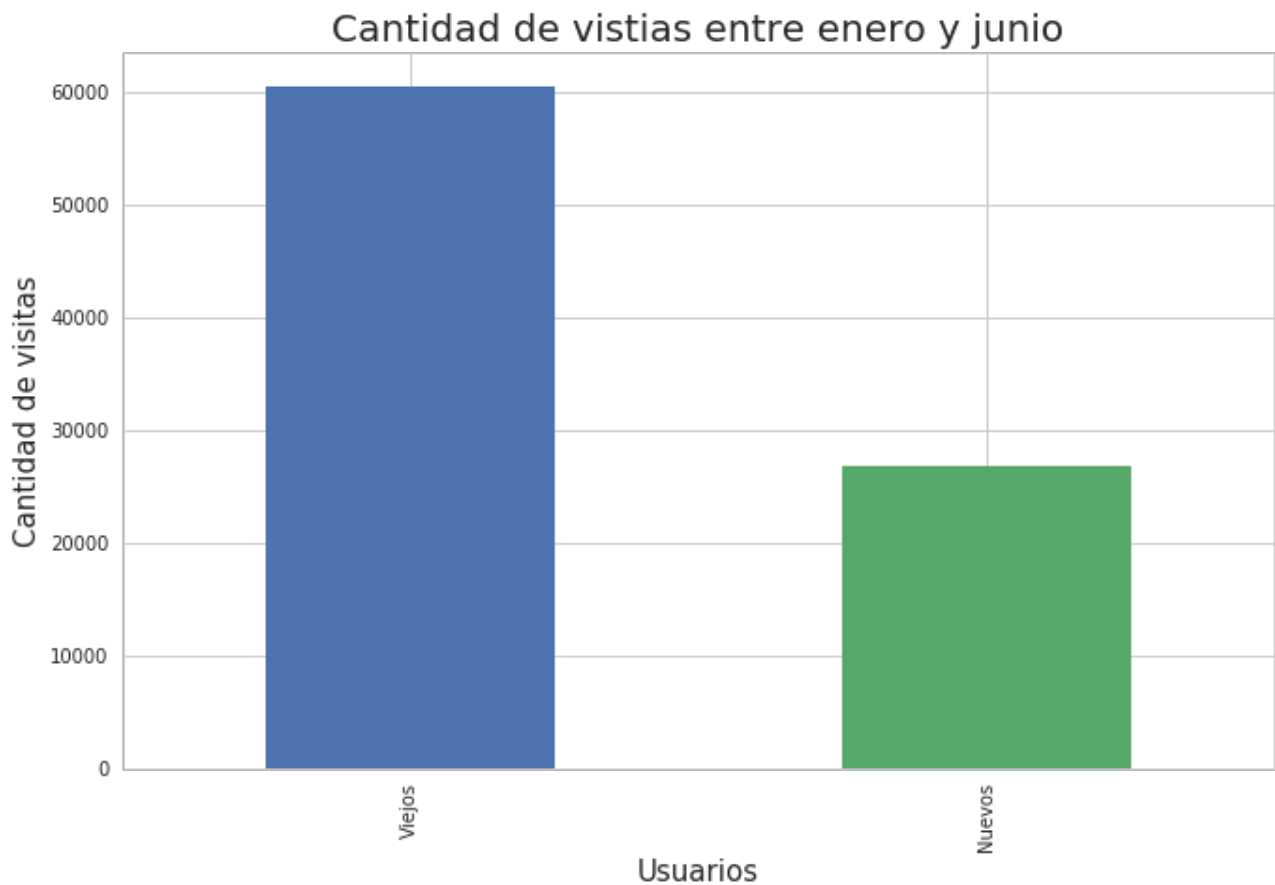


Figura 25

4.2 ¿Cuales son los canales por los que llegaron mas usuarios nuevos a lo largo de los meses?

En la figura 26 se puede observar como a lo largo de los meses los canales que mas usuarios nuevos traen a la plataforma son los canales Paid, Organic y luego Direct en ese orden.

Cantidad de visitas de usuarios nuevos por canal y mes

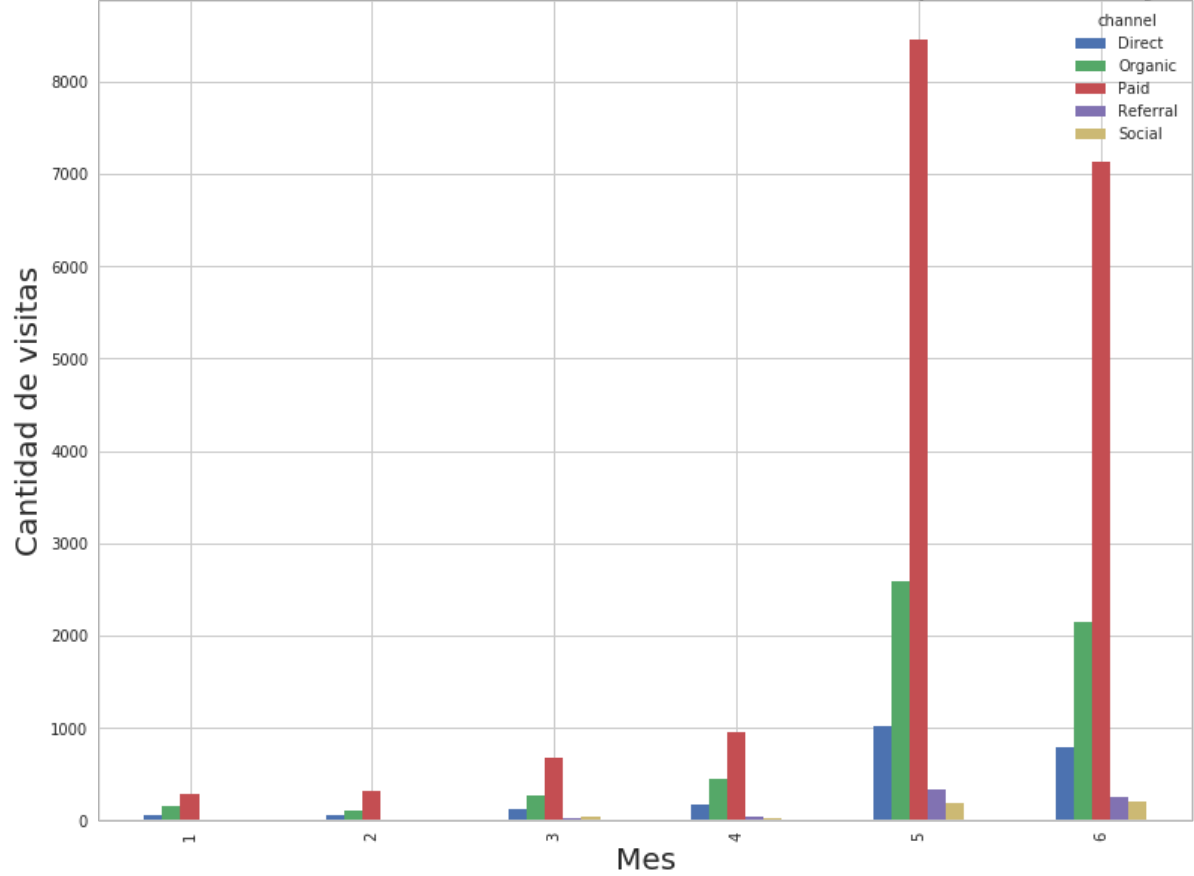


Figura 26

4.3 ¿ Por que canal vuelven los usuarios que ya visitaron la pagina anteriormente?

A diferencia de la figura anterior en la figura 27, se puede ver como el canal Pago sigue predominando (exceptuando el mes de enero) y el canal directo que era el tercer canal predominante en el análisis anterior, ahora es el segundo, seguido por Referral y luego por Organico.

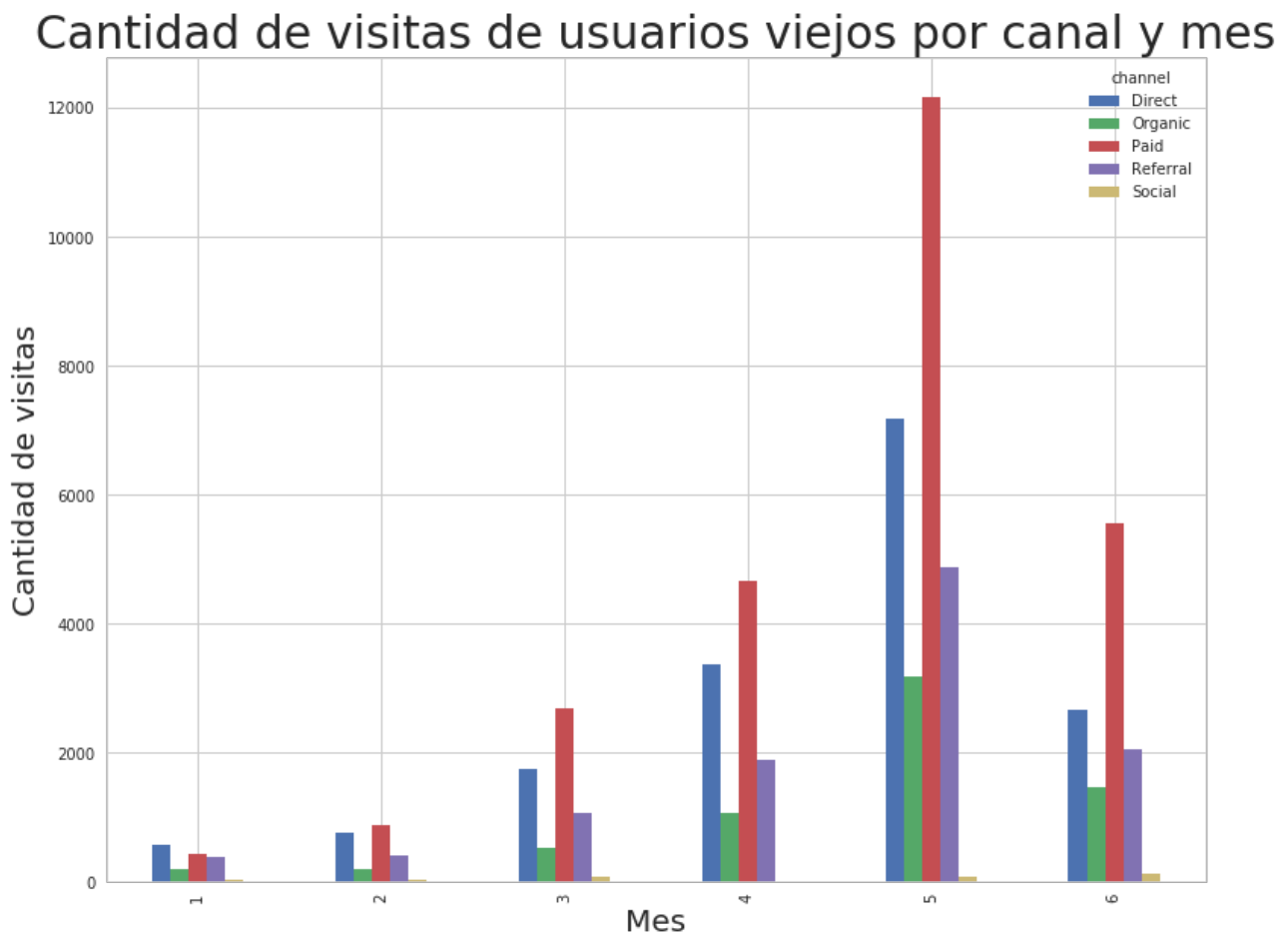


Figura 27

Conclusiones

Luego de analizar los distintos canales por los cuales ingresaron los usuarios viejos y nuevos a la pagina a lo largo de los meses, se puede concluir:

- El canal pago predomina durante los seis meses analizados tanto para usuarios nuevos como para los que vuelven a la plataforma.
- Como se observa en la figura 28 , la cantidad de usuarios viejos que acceden de forma directa a la plataforma es ampliamente superior que los nuevos que acceden de esa manera, lo mismo sucede con el canal referral.
- Una vez que los usuarios llegan a la plataforma, suelen volver a la misma de forma directa u organica.

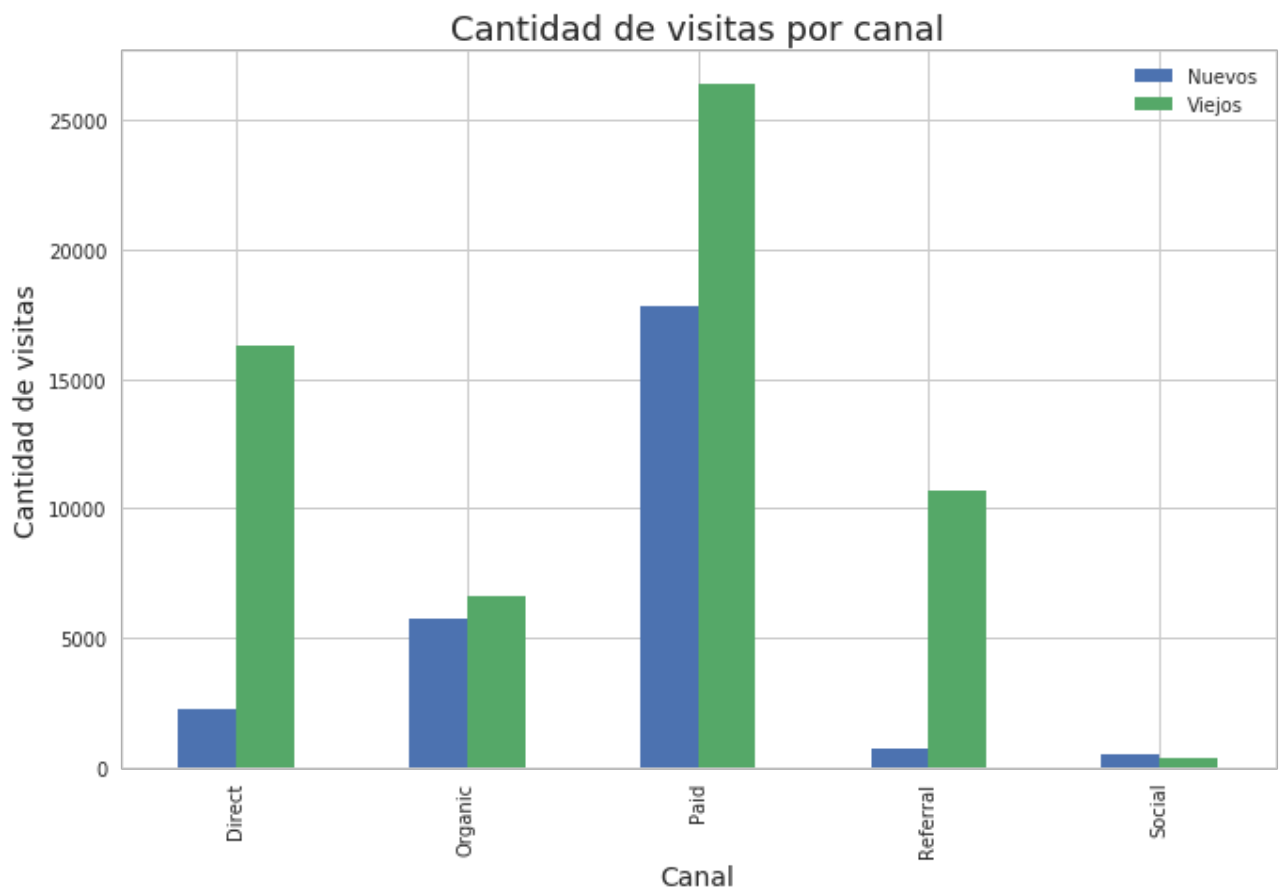


Figura 28

5. Analisis campañas de anuncios

5.1 ¿De todas las campañas de anuncios cuales son las que mas visitas generaron?

En la figura 29 se pueden ver las 7 campañas de anuncios que mas visitas generaron en la plataforma.

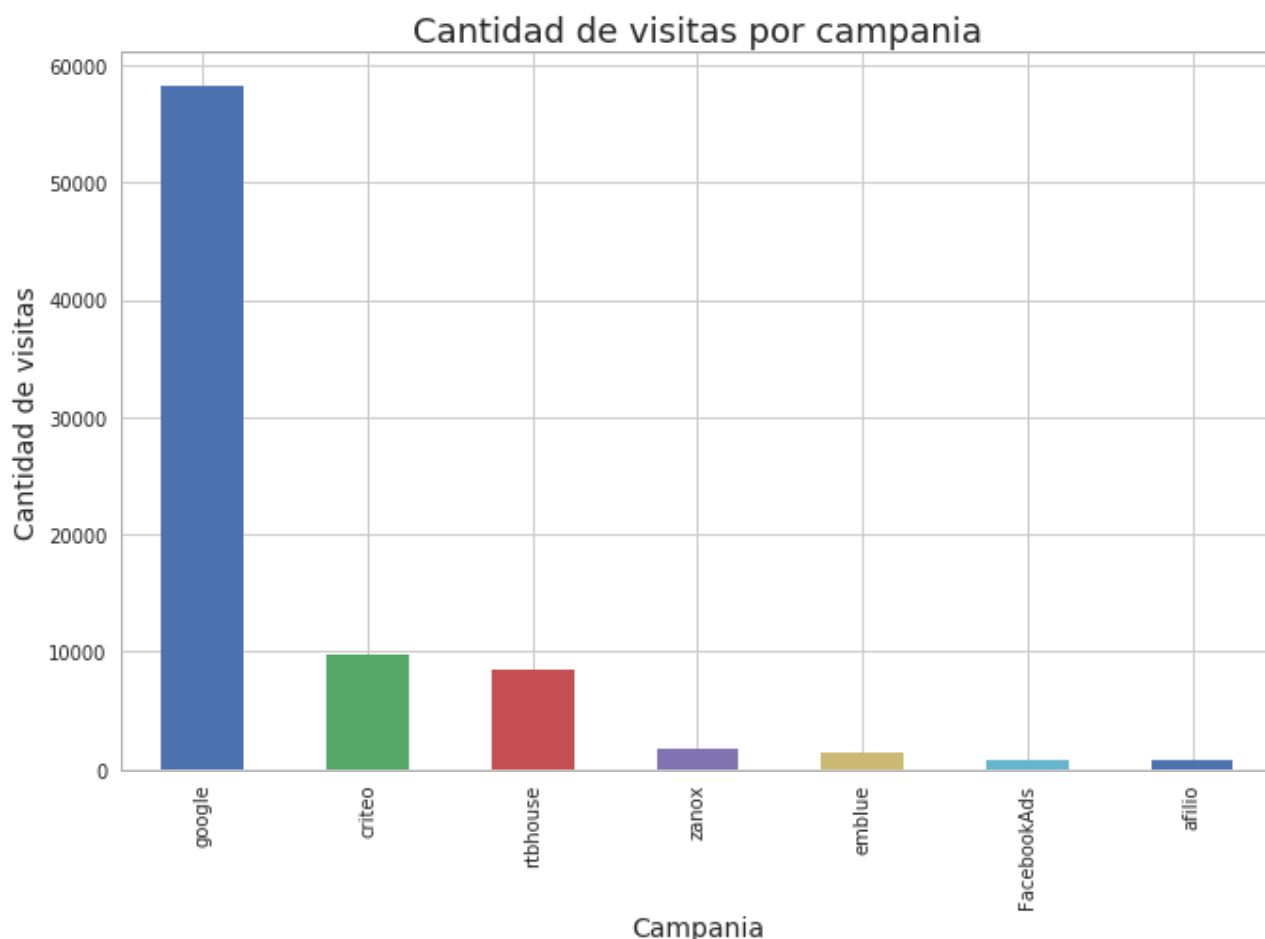


Figura 29

5.2 ¿ Cuales fueron las campañas que mas conversiones y checkouts generaron?

Para asociar los checkouts y las conversiones con las campañas de anuncios, lo que hicimos fue filtrar a todas las personas que hayan ingresado a la plataforma por alguna campaña y a cada persona la categorizamos con la primer campaña con la que hayan ingresado, finalmente asociamos los checkouts y conversiones con los usuarios categorizados por campaña.

De esta manera pudimos obtener de que campaña provenian usuarios que generaron checkouts y conversiones.

En la siguiente figura se puede ver como la mayoría de los checkouts y conversiones realizadas fueron generadas por usuarios que llegaron a la plataforma a partir de un anuncio de google.

En la siguiente figura se puede observar como google lidera la cantidad de conversiones generadas, seguido por critico y rtbhouse.

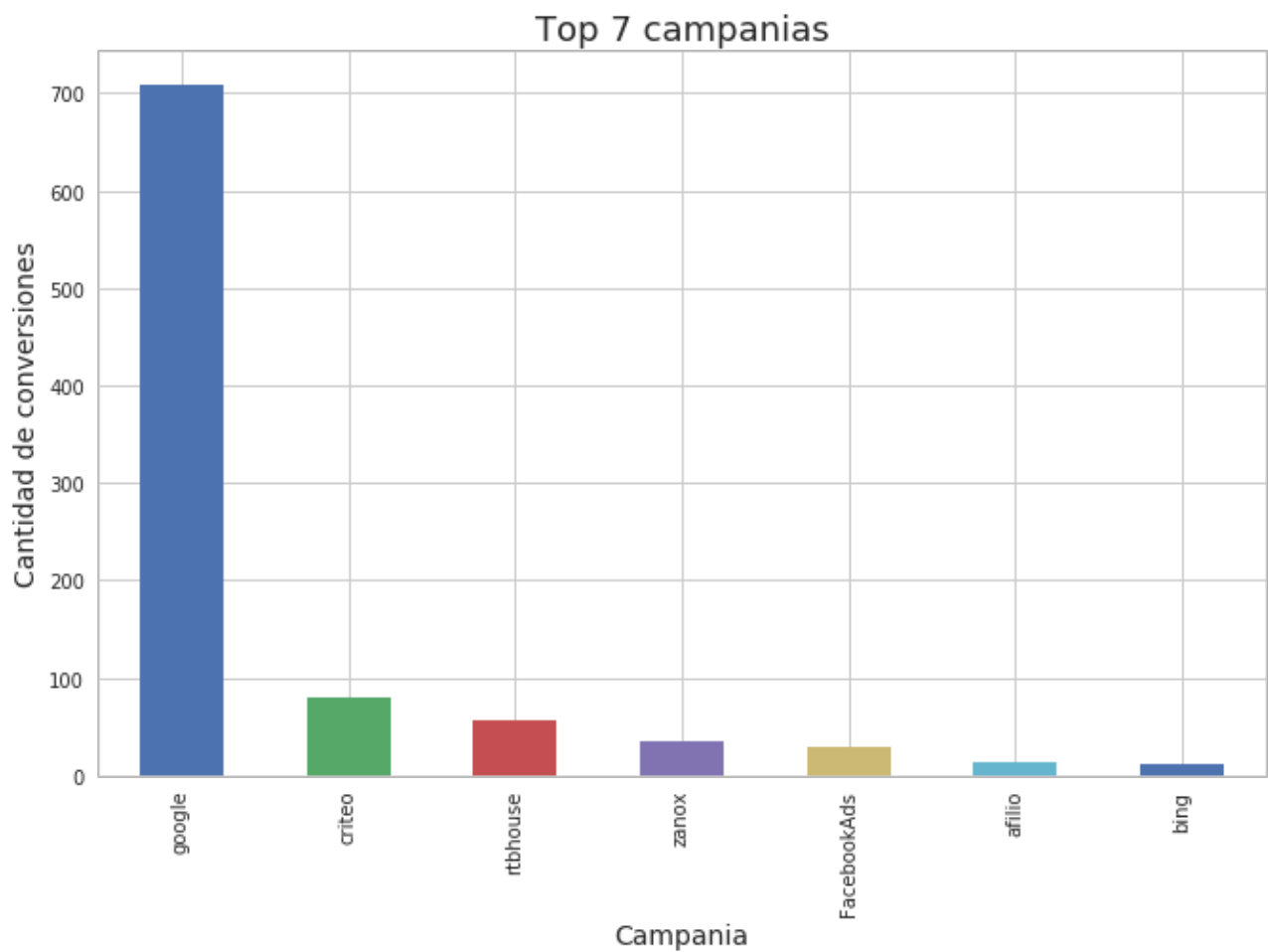


Figura 30

A lo largo del semestre google siendo el lider tanto en checkouts como en conversiones:

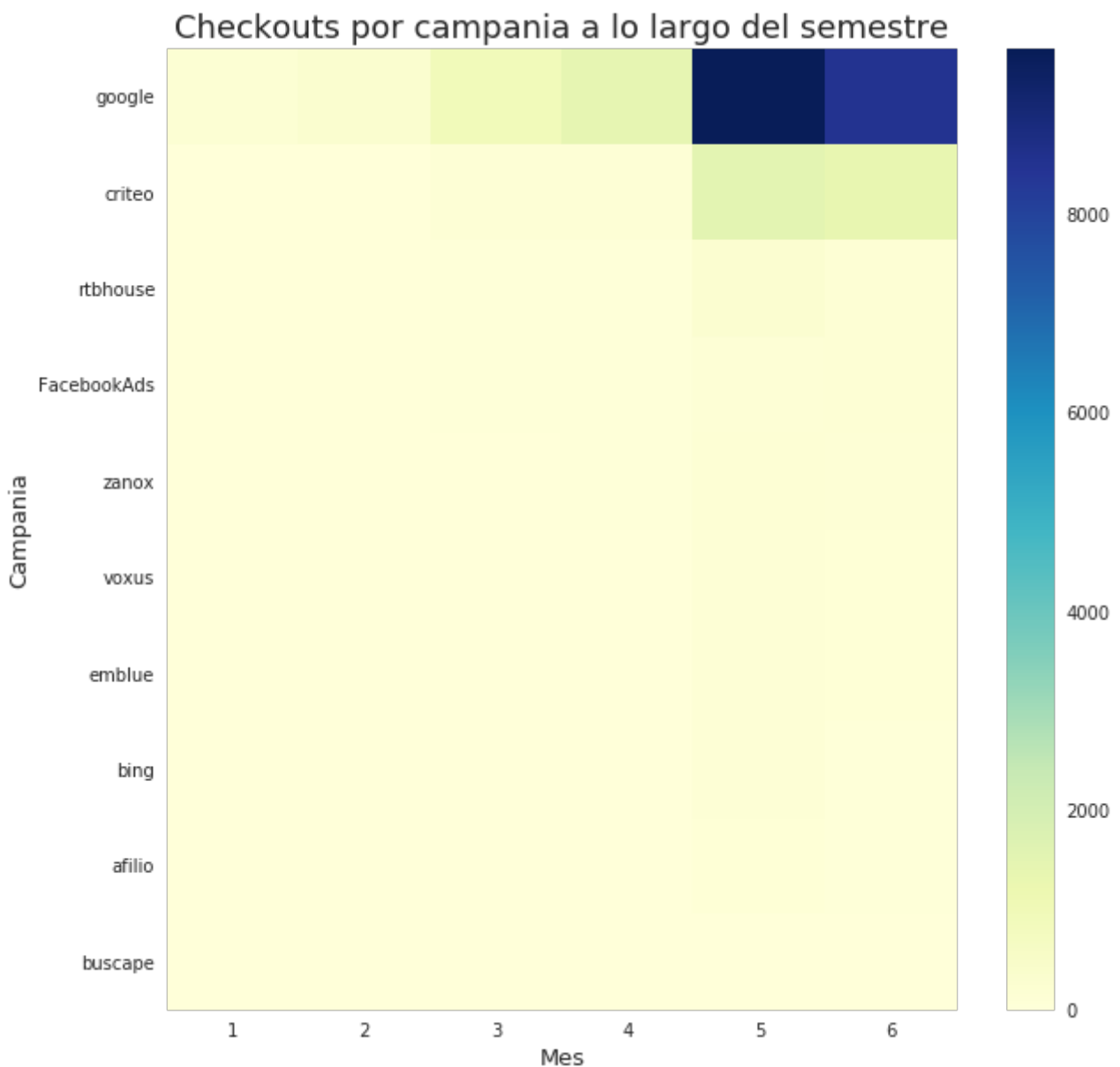


Figura 31

A medida que pasan los meses se puede observar como la cantidad de checkouts y conversiones aumenta, esto se puede deber a un aumento la cantidad de anuncios publicados por google.

Las cantidad del mes de junio son inferiores a los anteriores meses por que no se cuentan con datos del mes completo.

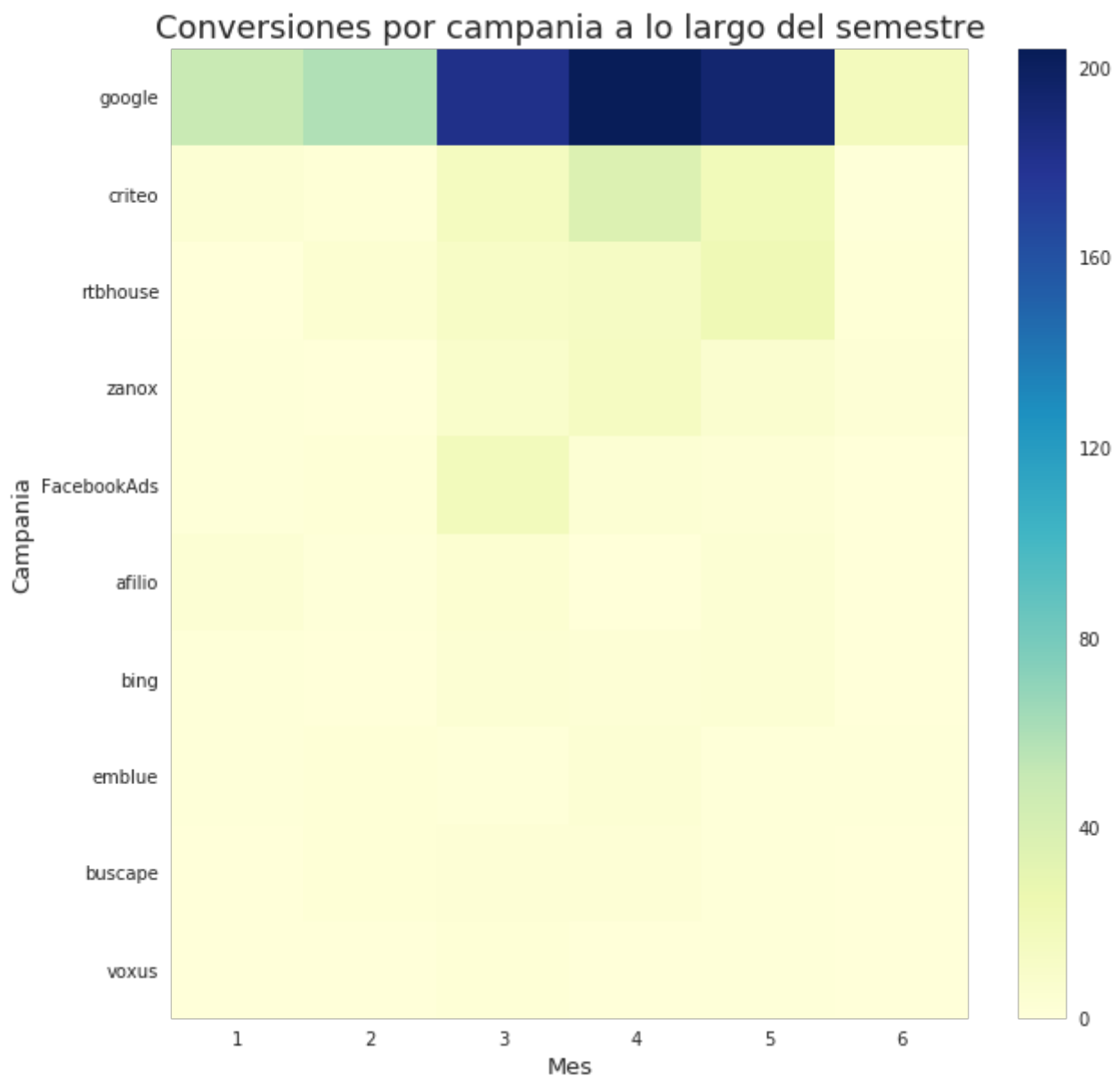


Figura 32

6. Análisis de las marcas y sus dispositivos

6.1 ¿Cuáles son las marcas de los dispositivos en la página Trocafone?

- Se pudo observar que las marcas de dispositivos que comercializan en su página son ocho: **Iphone, Samsung, Motorola, Sony, Lenovo, Asus, LG y Blackberry.**

6.2 Analisis de las marcas

En las siguientes cuatro visualizaciones se van a poder apreciar los eventos más interesantes sobre las marcas.

6.2.1. Eventos de cada marca.

Se pueden observar los eventos totales sobre las diferentes marcas.

Los eventos son todos los eventos posibles que se pueden contabilizar en cada producto, como, “conversion”, “checkout”, “lead”, “view product”, etc.



Figura 33

6.2.2. Checkout de cada marca

Se pueden observar los eventos “checkout” sobre las marcas.

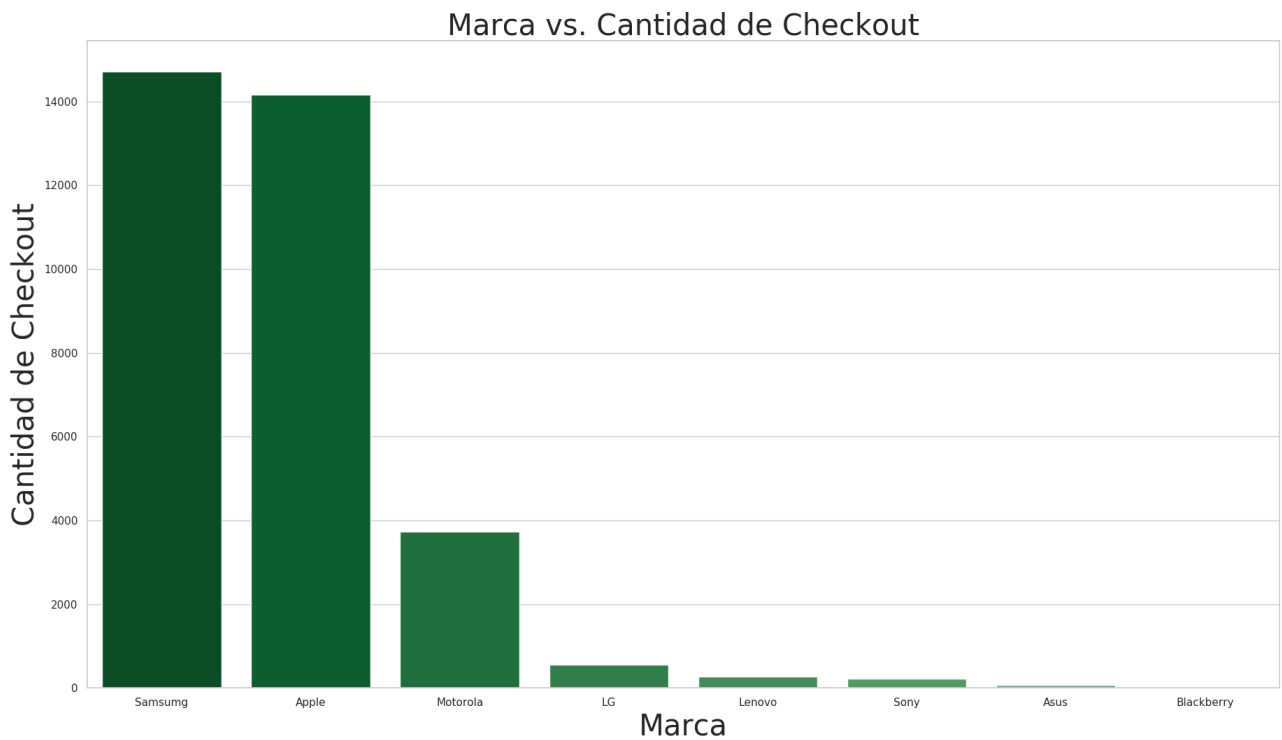


Figura 34

6.2.3 Cantidad de conversiones

Se puede observar los eventos “conversion” sobre todas las marcas, teniendo como la marca más vendida a Samsung.

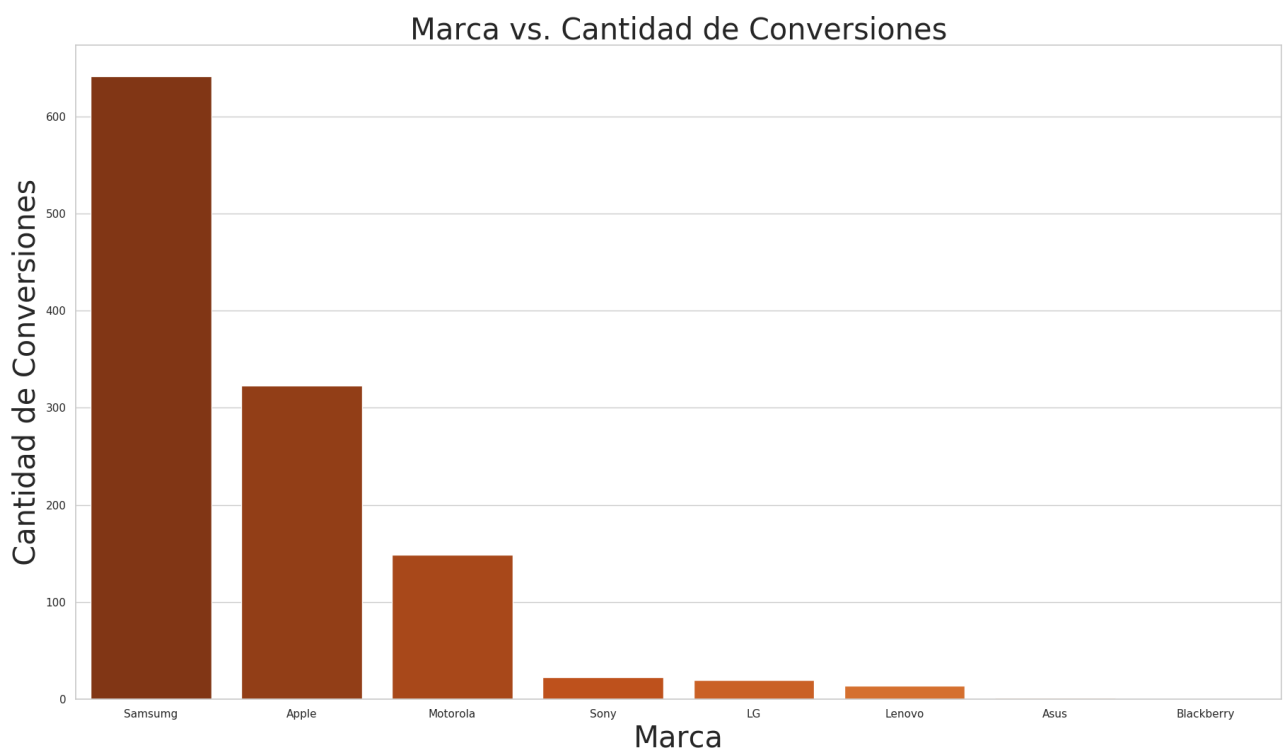


Figura 35

6.2.4 Cantidad de eventos “lead”

Se puede observar los eventos “lead” sobre los productos de las marcas, tales eventos son suscripciones de los usuarios sobre un producto específico para recibir notificaciones de stock disponible.

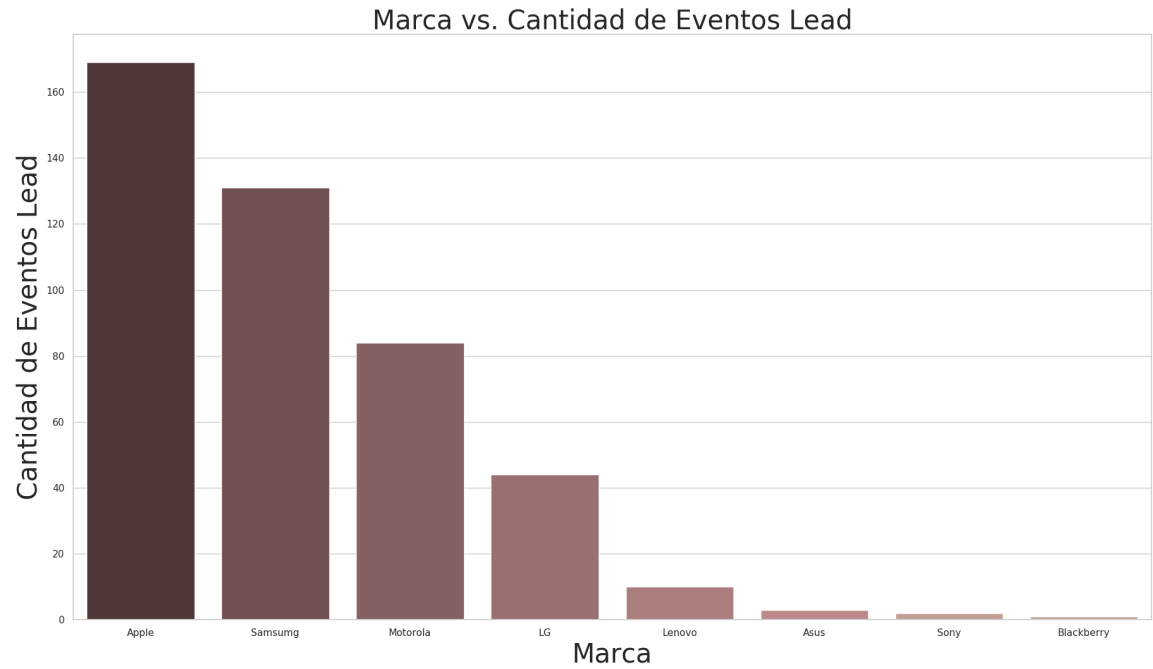


Figura 36

6.2.5 ¿Podría influir positivamente la cantidad de productos de cada marca para dar alternativas a los usuarios y así incrementar las conversiones aunque haya falta de stock de algunos dispositivos?

Se muestra la cantidad total de diferentes dispositivos de cada marca. Por ejemplo: Samsung cuenta con más de ochenta productos comercializados en la página web.

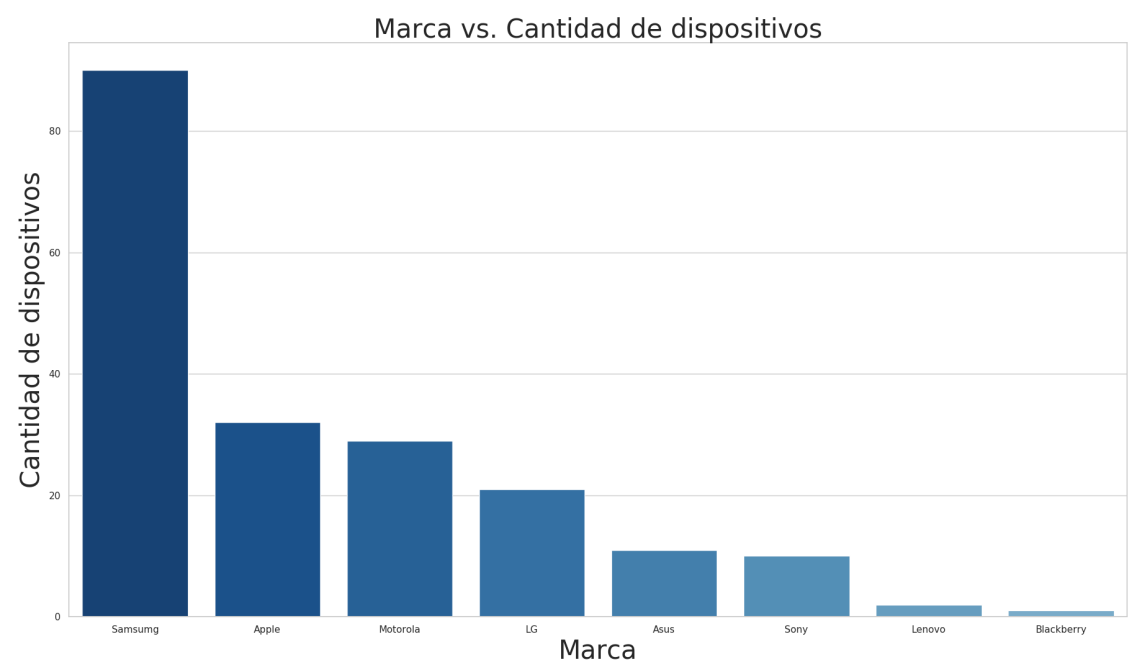
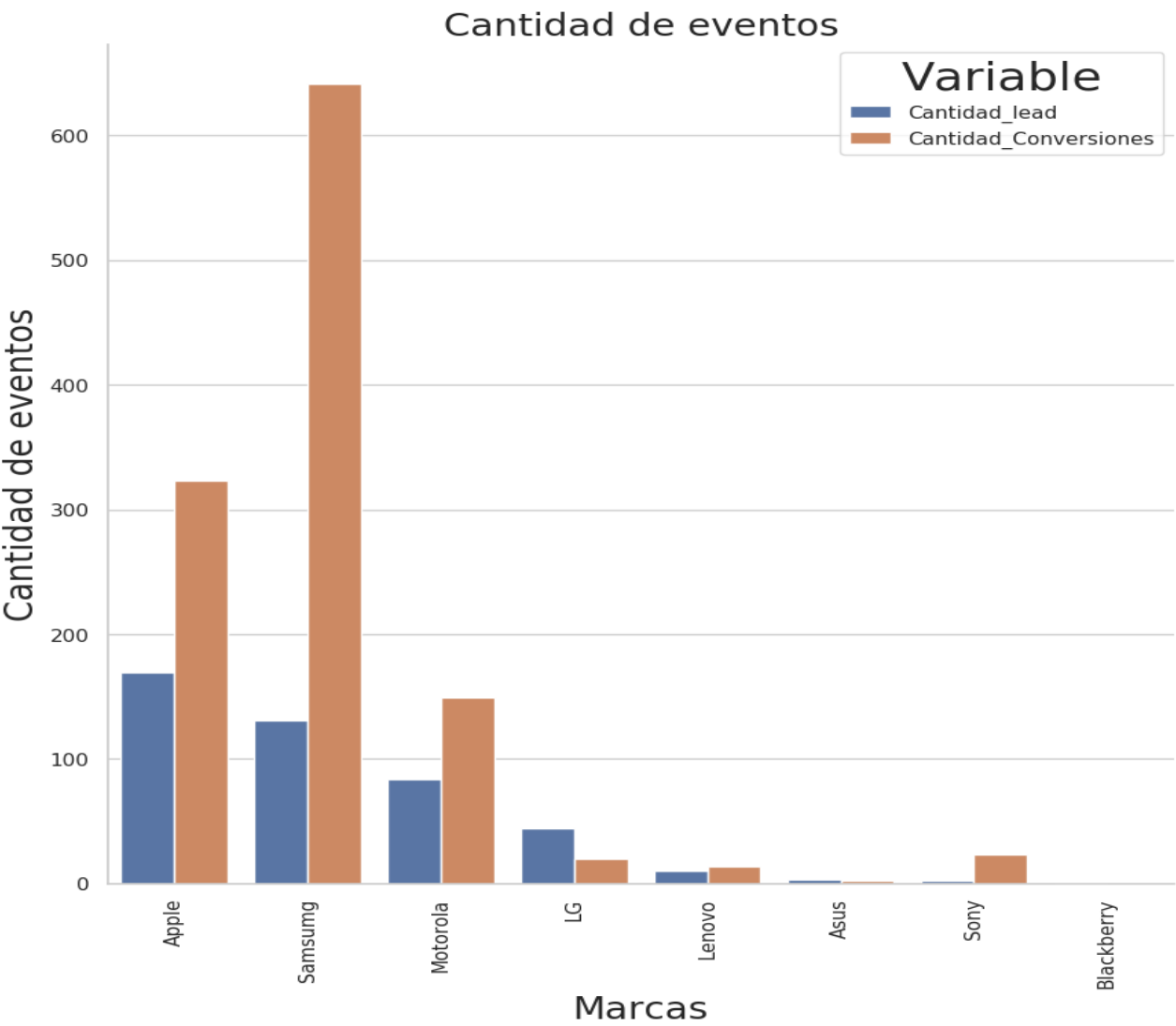


Figura 37

6.2.6 ¿Tiene alguna relación la falta de stock de dispositivos de las distintas marcas con la cantidad de conversiones realizadas?

En la siguiente visualización se muestran las diferencias entre los eventos “conversion” y los eventos “lead” de las distintas marcas.



Texto

Figura 38

6.2.7 Búsquedas de las distintas marcas en Trocafone.

En la siguiente visualización se va a poder apreciar las búsquedas de dispositivos agrupadas con sus respectivas marcas, estas búsquedas se generan desde la pagina web de Trocafone.

La marca Apple es la más buscada por los usuarios, y en un escalón debajo las marcas Samsung y Motorola. En paginas siguientes se podrán ver los términos mas buscados de cada marca para respaldar esta información.

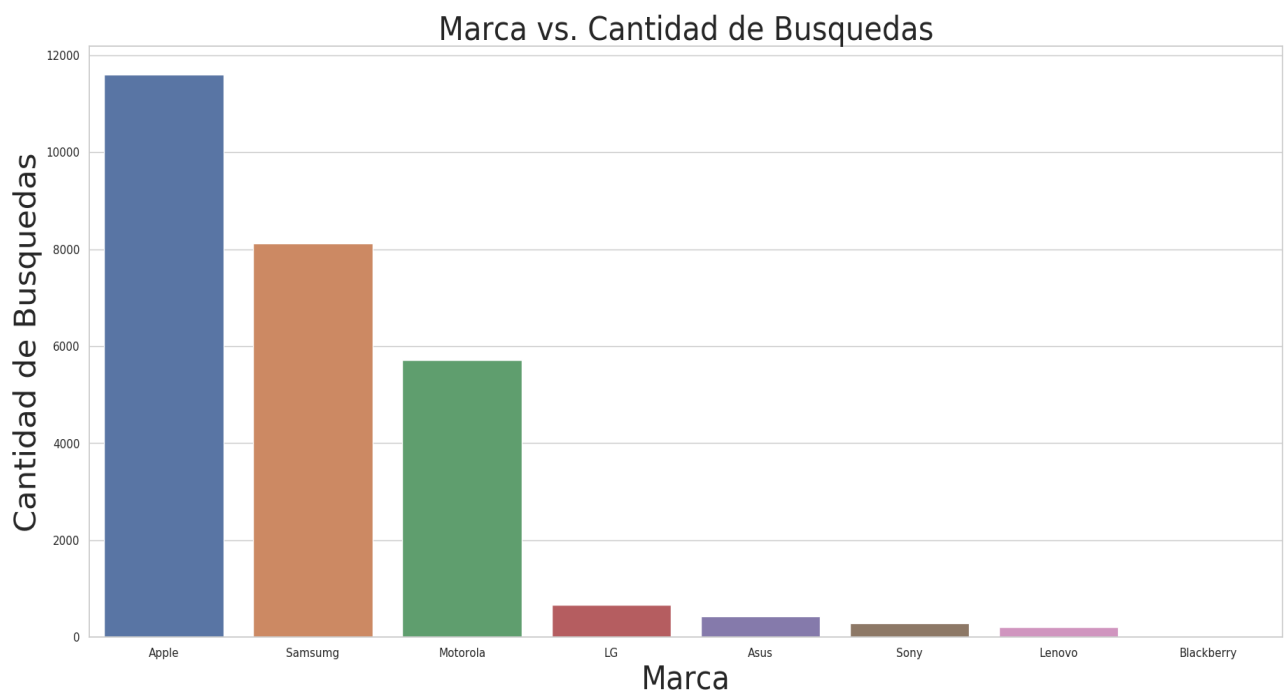


Figura 39

6.2.8 Búsquedas de las distintas marcas en Trocafone por mes.

Predominan las marcas Samsung y Apple en las búsquedas de Trocafone en todos los meses, estando un escalón por detrás se observa a los dispositivos de Motorola. Las demás marcas tienen mucho menor peso a la hora de buscar dispositivos de interés.

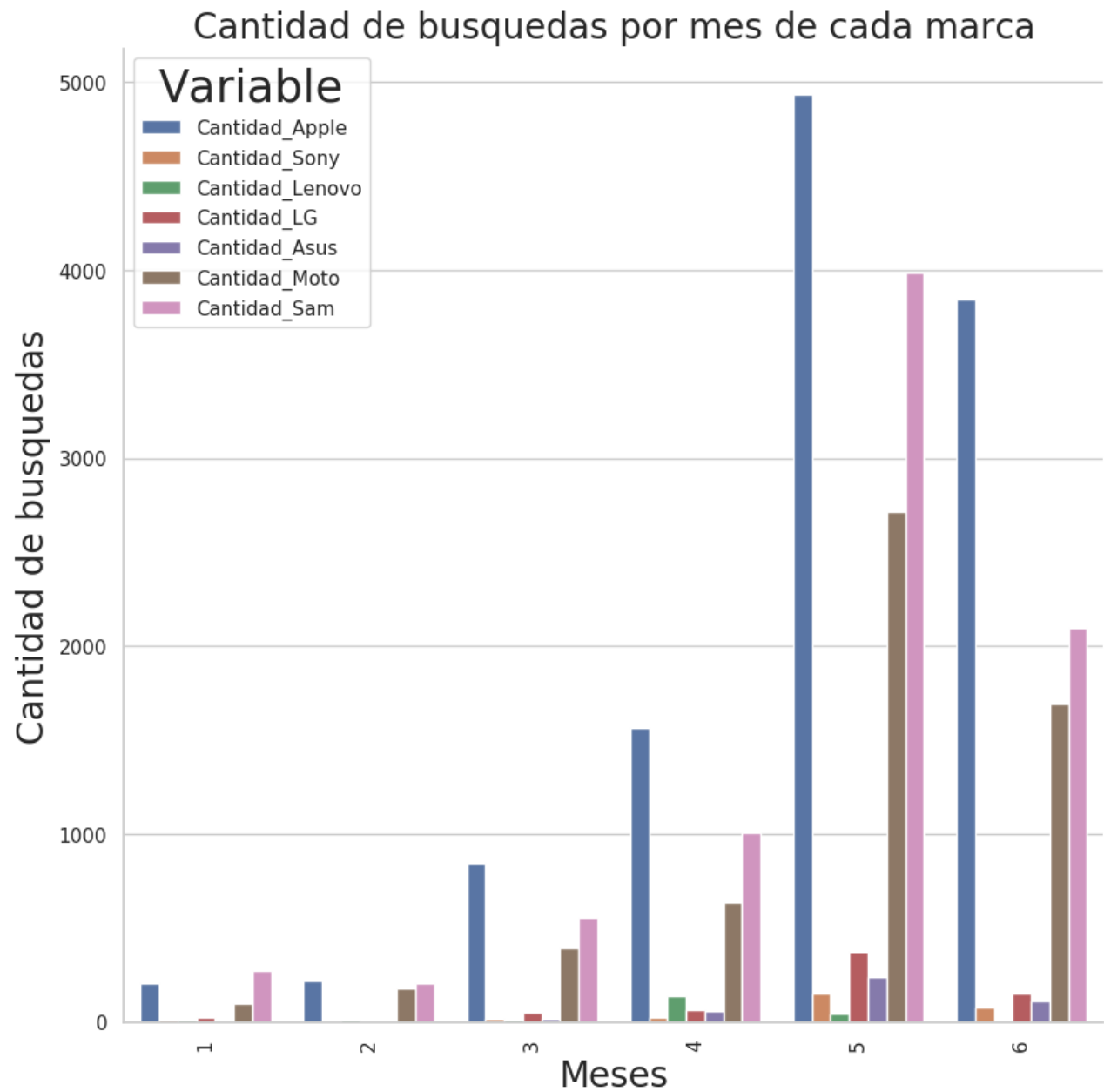


Figura 40

6.2.9 Los términos mas buscados en cada marca.

Se van a poder apreciar los términos que más se buscaron en cada marca, teniendo en cuenta las diferentes posibilidades a la hora de buscar en la web. No son datos sumamente exactos, ya que no se pudieron extraer todos los términos de búsqueda pero al observar e investigar los dispositivos más comprados y los de mayor eventos se pudo hacer un filtro y relevar los términos de mayor consideración.

Apple

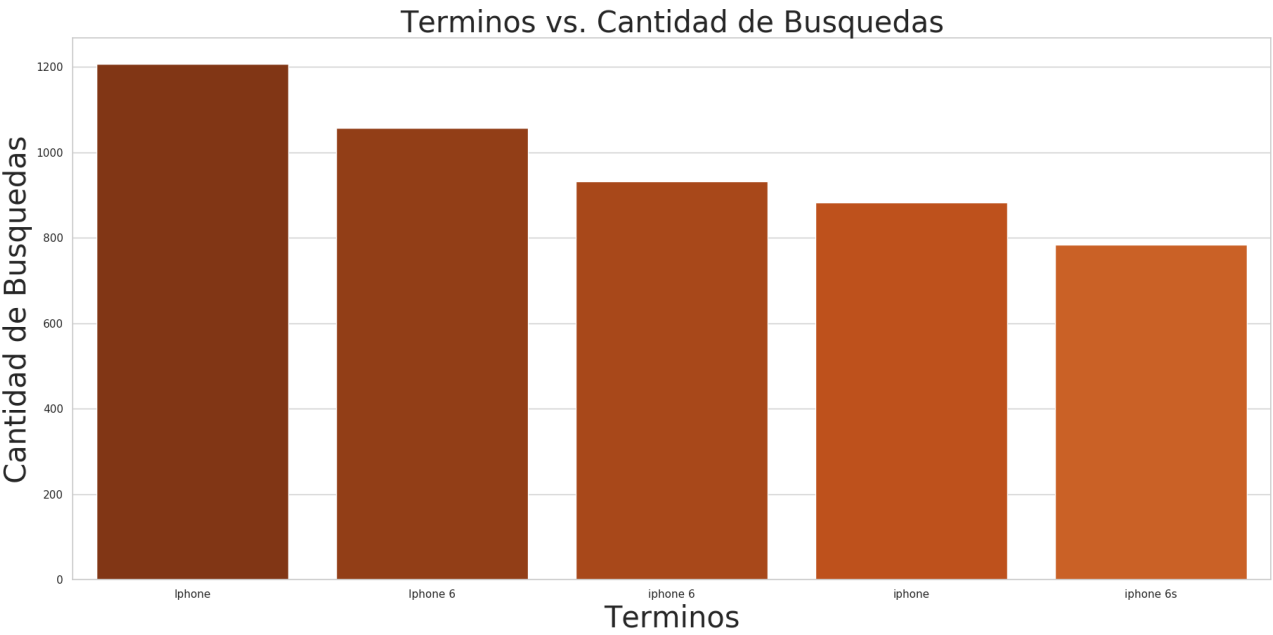


Figura 41

Samsung

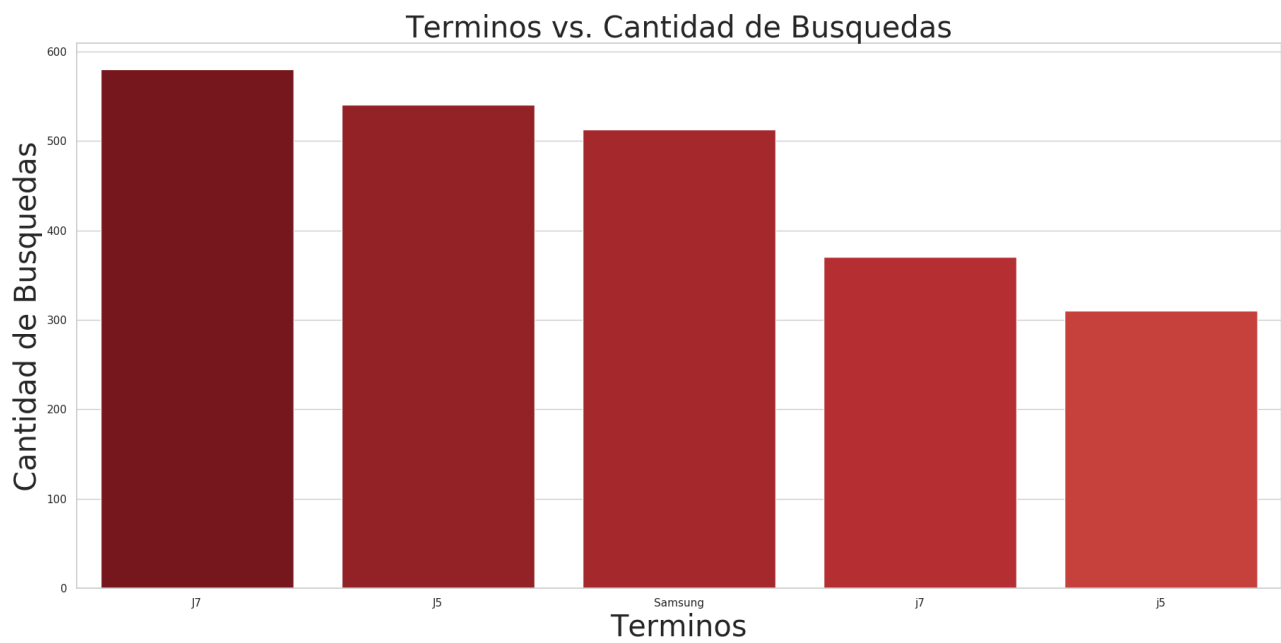


Figura 42

Motorola

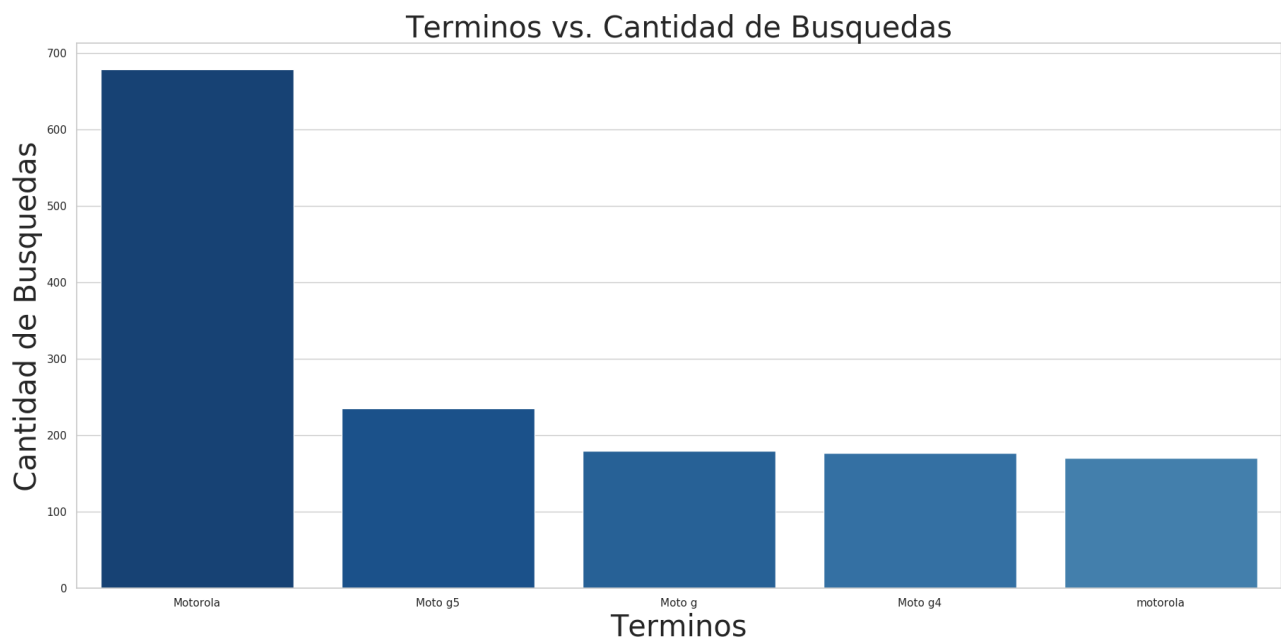


Figura 43

Asus

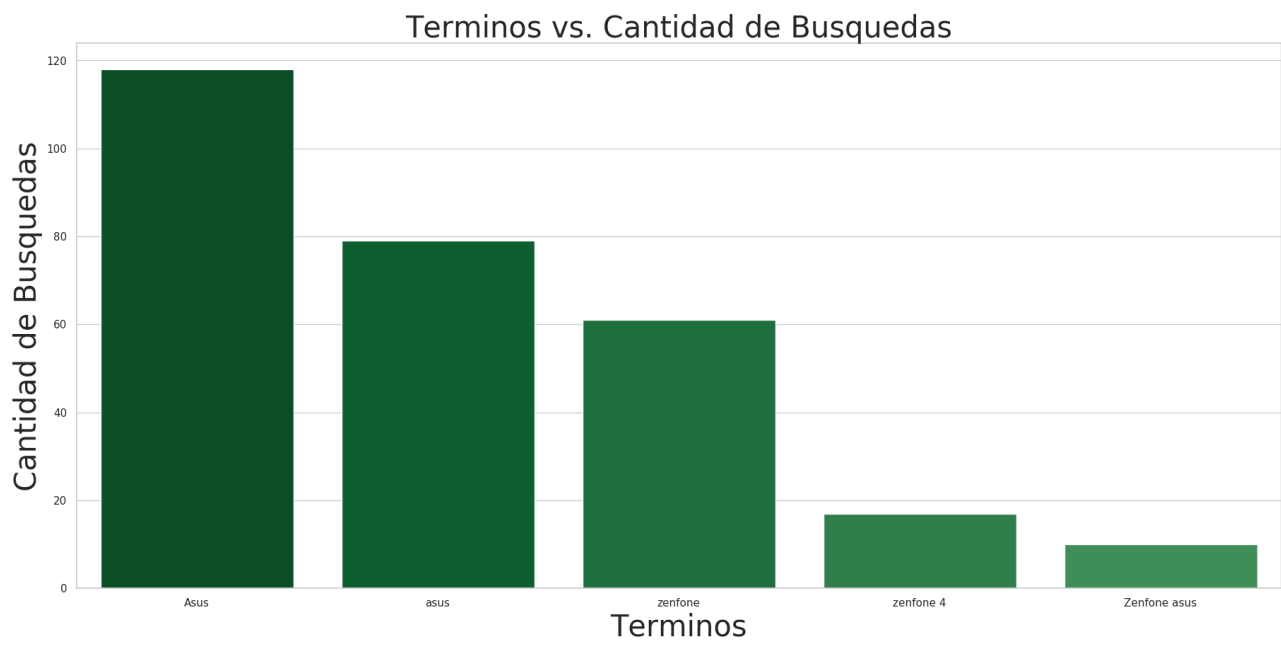


Figura 44

LG

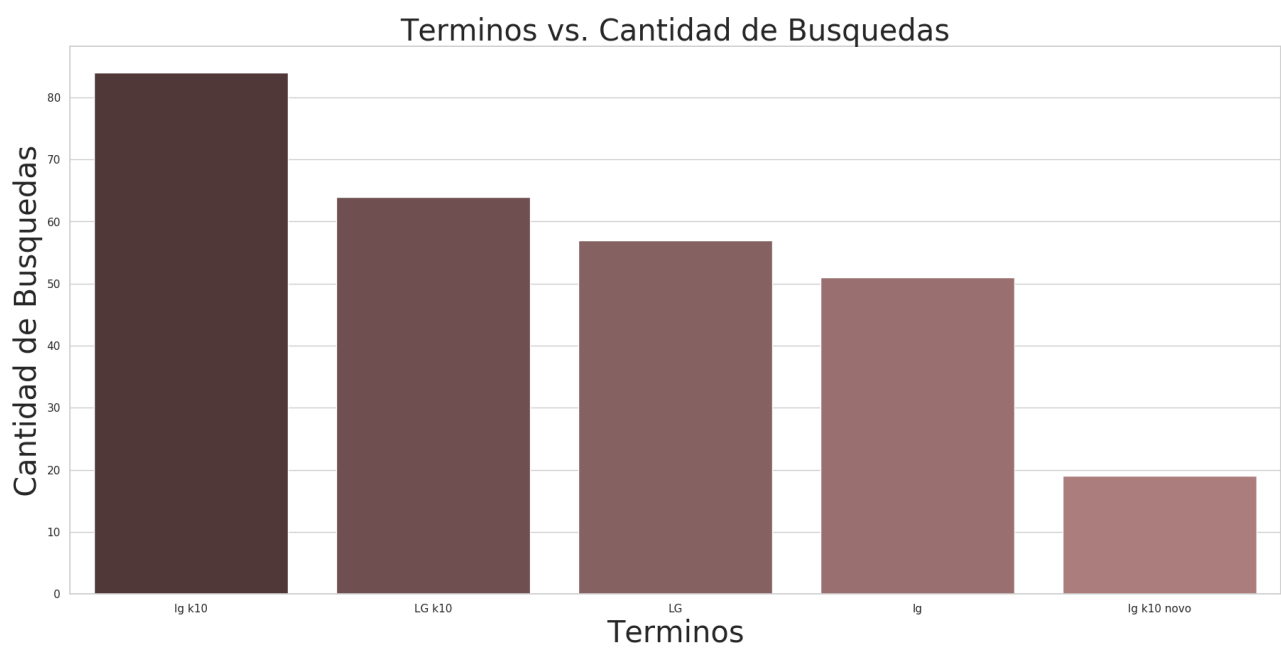


Figura 45

Sony

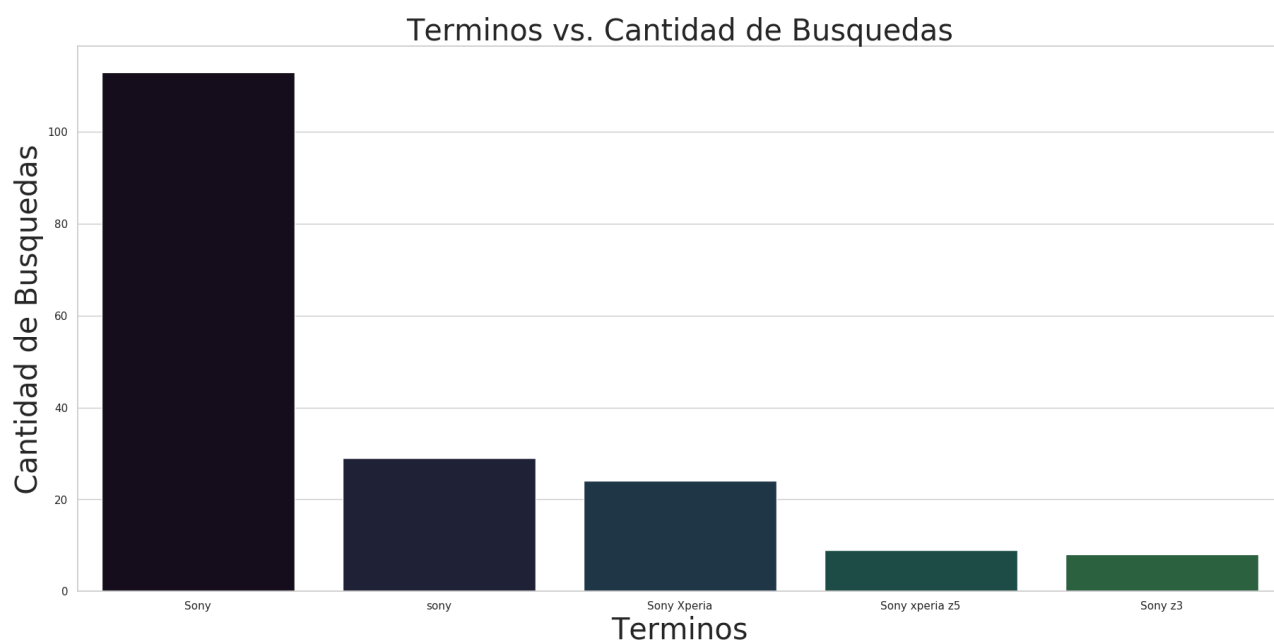


Figura 46

Lenovo

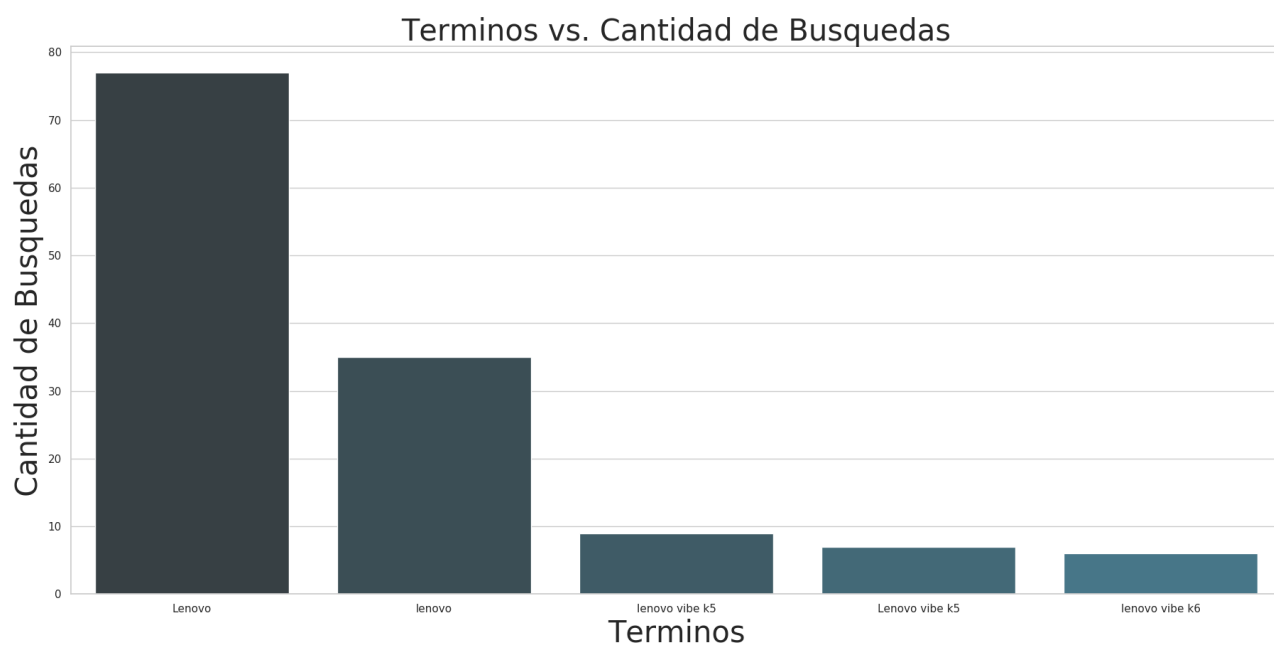


Figura 47

6.3 Análisis de dispositivos individuales

En las siguientes visualizaciones se van a poder apreciar los eventos más relevantes sobre los dispositivos individuales de cada marca.

6.3.1 Top 20 de dispositivos con más eventos

Se puede ver que los dispositivos con más eventos son de las marcas Samsung y Apple.

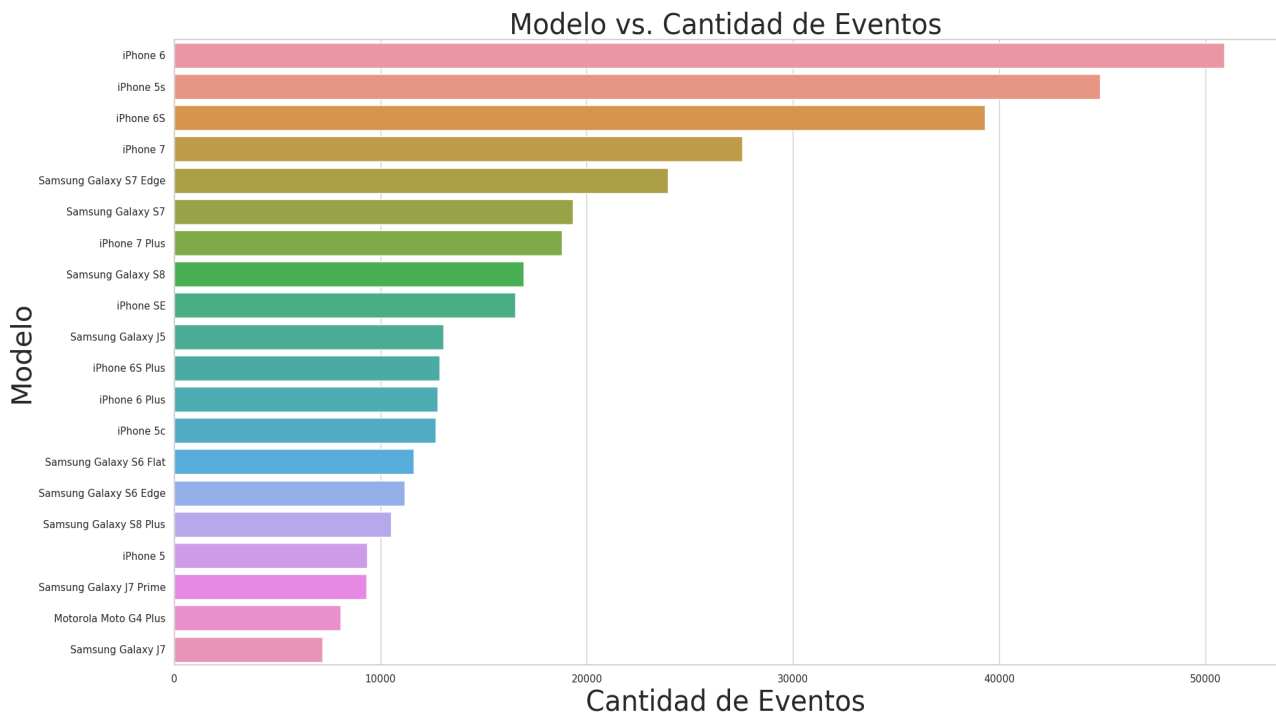


Figura 48

6.3.2 Top 20 de dispositivos con más eventos “checkout”.

Siguen predominando los dispositivos de las marcas Samsung y Apple con mayores eventos “checkout”.

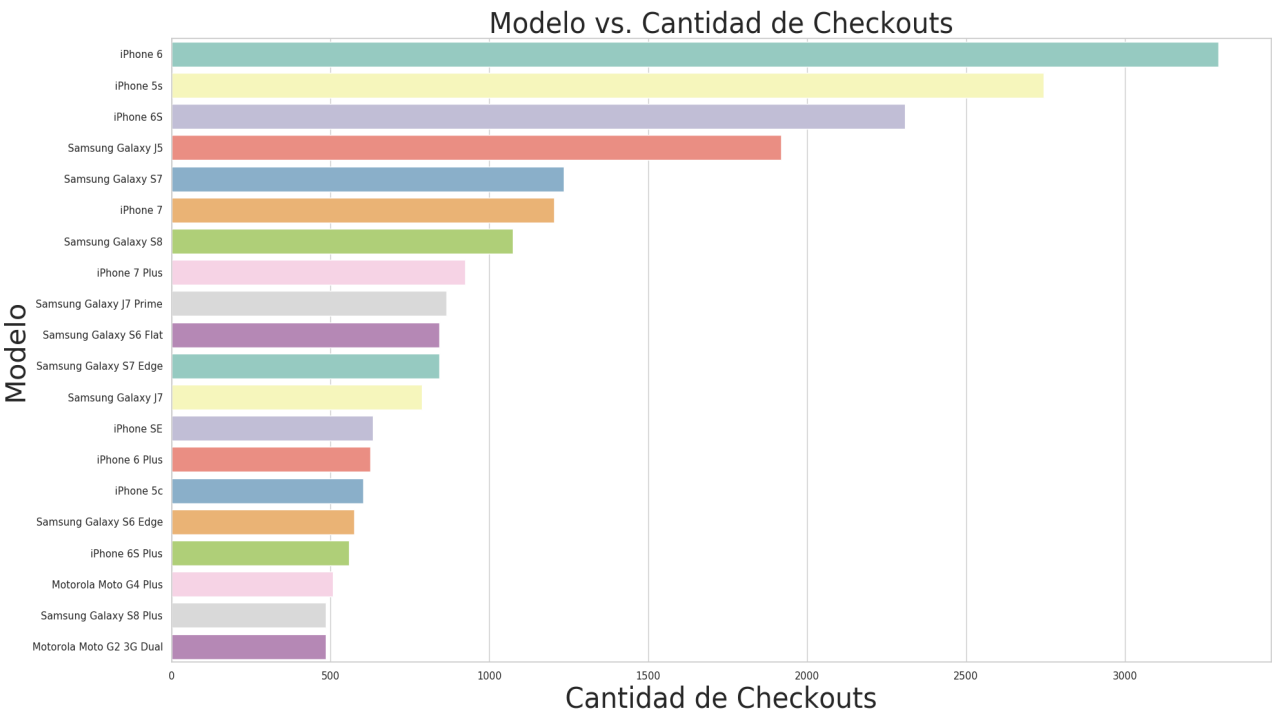


Figura 49

6.3.3 Top 20 de dispositivos con más conversiones.

Se puede observar en este gráfico que los dispositivos con más eventos “conversion”, o sea compras efectivas de los usuarios siguen siendo las marcas Apple y Samsung aunque se ve la aparición de dos dispositivos de la marca Motorola.

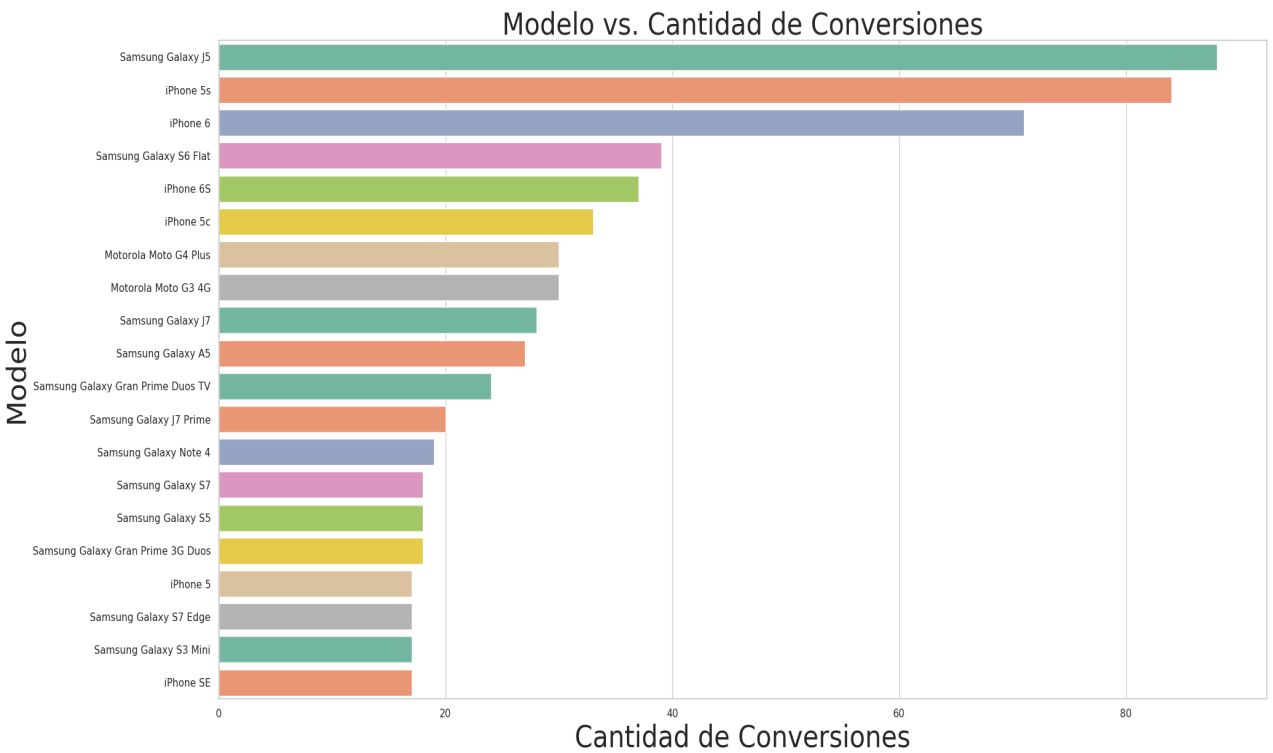


Figura 50

6.3.4 Top 20 de dispositivos con más eventos “lead”

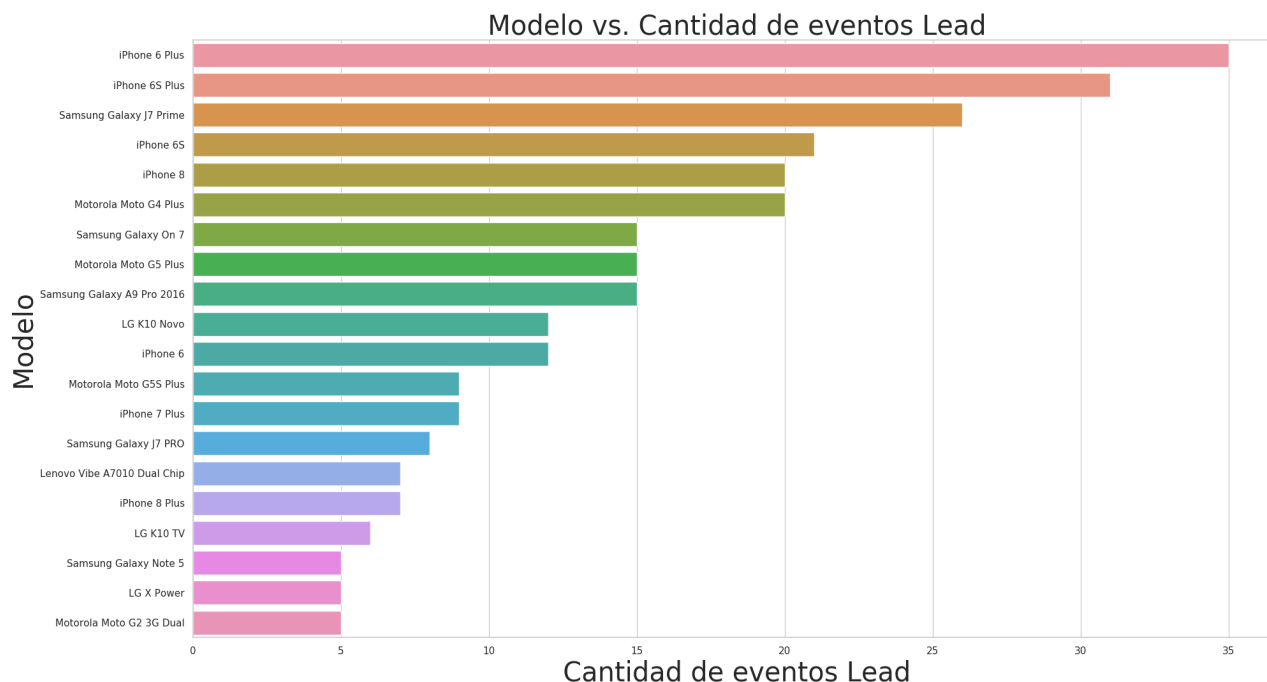


Figura 51

6.4 Conclusiones

Figuras de los puntos: 6.2.1, 6.2.2, 6.2.3, 6.2.4

- Al ver la cantidad de distintos eventos que se realizan por marca, se pueden dividir las marcas en tres bloques. En el primer bloque, las marcas Iphone y Samsung, como las más confiables y las que tiene más prestigio en el mercado.
- En el segundo bloque, se posiciona Motorola, como punto intermedio entre las marcas más confiables, y las marcas del tercer bloque como las menos solicitadas en el mercado.
- En el tercer bloque se encuentran las marcas Sony, LG, Asus, Lenovo como las marcas menos confiables y con menos prestigio en el mercado.
- Blackberry tiene un solo dispositivo y un solo evento, por lo tanto, es una marca con poco poder de análisis. Además, en la actualidad la página de Trocafone no cuenta con dispositivos de esta marca.

Figuras de los puntos: 6.2.5 y 6.2.6

- Al ver el gráfico del punto 6.2.6 con la cantidad de eventos “conversion” y cantidad de eventos “lead”, se podría decir, que en las marcas más vendidas no cambiaría mucho la tendencia al comprar, porque tampoco se puede decir con certeza que los usuarios que pidieron recibir notificación de stock disponible, harían o no, la compra.
- En el caso de LG, donde hay más pedidos de stock que la cantidad de conversiones, se podría decir que podría influir un poco a la hora de generar ventas, ya que los usuarios demandan más stock de productos. Estos no pueden ser suplidos, aunque cuenten con varios dispositivos alternativos dentro de la misma marca.

- La cantidad de dispositivos en cada marca, como se indica en la figura 1.6, también ayudan a poder elegir otra alternativa. Por lo que todavía se hace mas incierto si la falta de stock influye en las compras de dispositivos de estas dos grandes marcas.
- También se puede observar que la cantidad total de ventas de dispositivos dentro la marca Samsung se ve influenciada por su amplitud de opciones a la hora de elegir, ya que cuenta con el doble de variantes en comparación a los diferentes dispositivos de las otras marcas.

Figuras de los puntos: 6.3.1, 6.3.2, 6.3.3, 6.3.4

- A la hora de analizar los eventos generales y los eventos “checkout” de los dispositivos individuales, se puede observar el mismo patrón de marcas en el escalón más alto, siendo Apple y Samsung los más buscados y requeridos por los usuarios.
- En el caso de las conversiones, entran en juego los dispositivos de la marca Motorola, posicionándose entre las opciones más vendidas del mercado.
- Y por último, en los eventos “lead”, se observan varios dispositivos de la marca LG, haciendo referencia a lo mencionado con anterioridad, sobre el bajo stock con respecto a la demanda.
- En las marcas con menos cantidad de variantes, la falta de stock puede influir directamente a la hora de elegir una marca en la cual confiar.

7. Análisis de ventas según la condición de los dispositivos

7.1 ¿Cuántos checkouts se generaron dependiendo la condición de los dispositivos?

Como se puede ver en el siguiente gráfico (Figura 43), la mayor cantidad de checkouts implican celulares en buen estado, seguido por excelente y muy bueno.

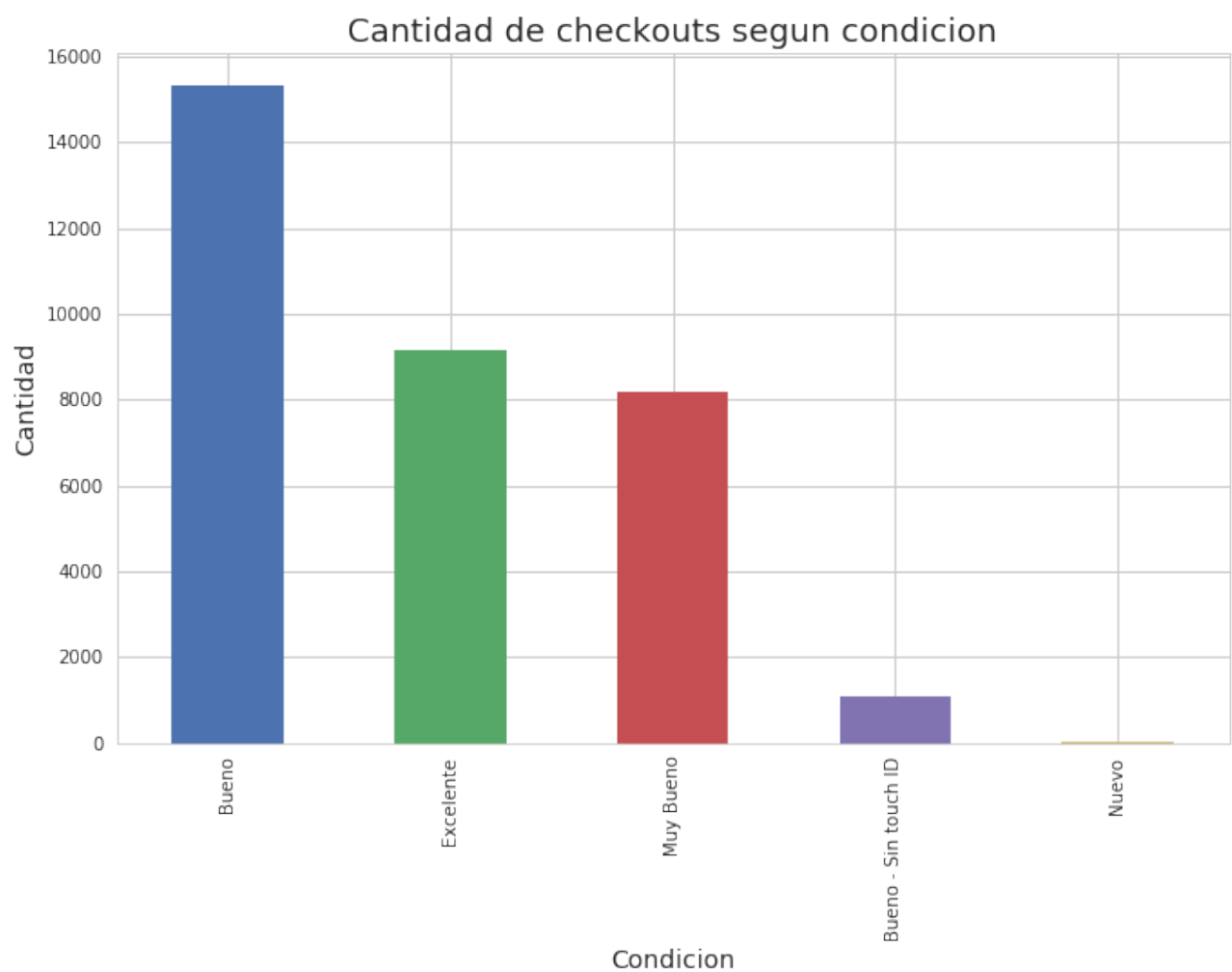


Figura 52

7.2 ¿En que condición estaban la mayor cantidad de celulares vendidos?

Como se puede ver en la Figura 44, al igual que con la cantidad de checkouts, la mayor cantidad de celulares convertidos fueron celulares en buen estado.

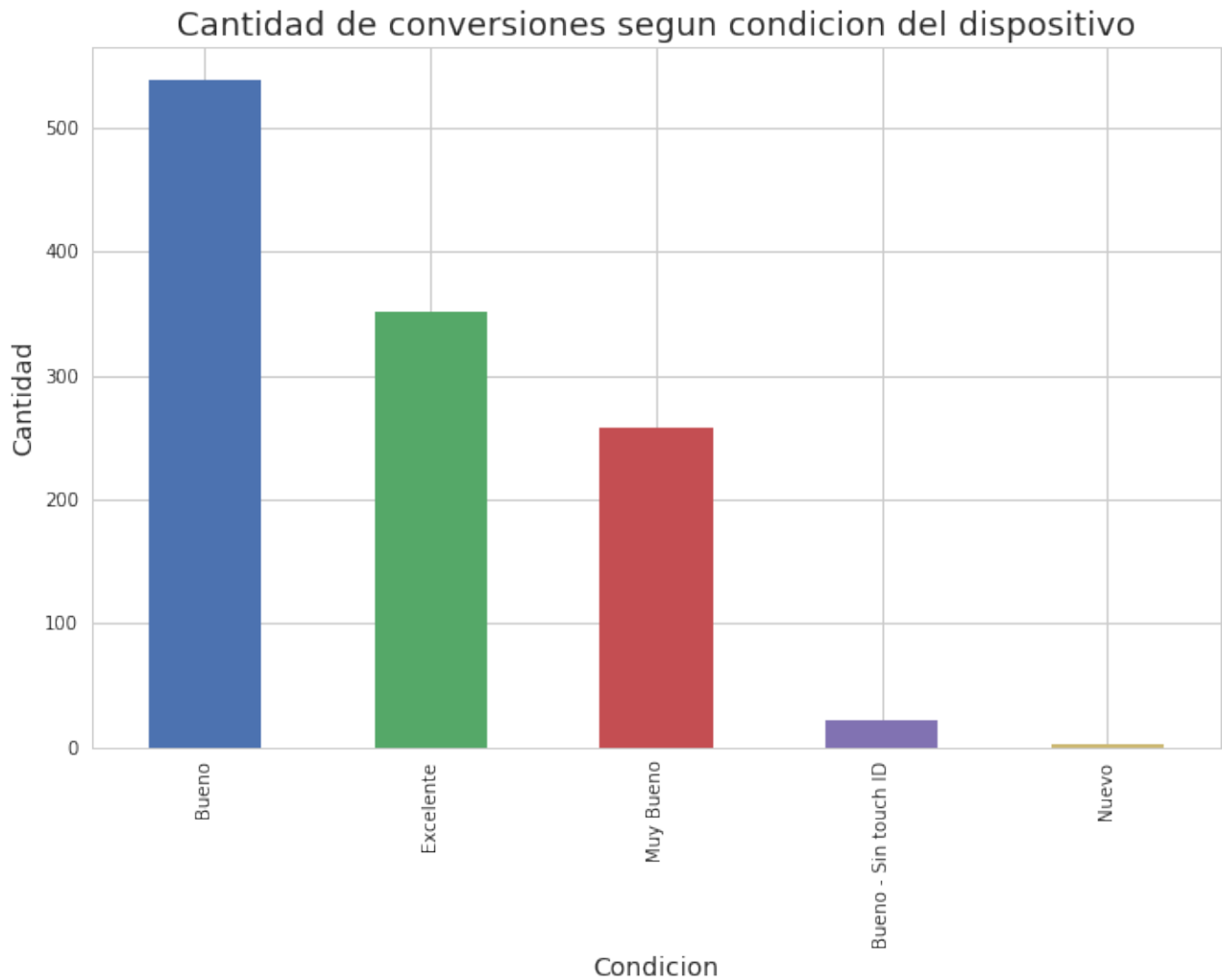


Figura 53

7.3 ¿Que celular lidera en mayor cantidad de checkouts y en que condición?

En la Figura 45 se mantiene la tendencia de que la mayor cantidad de celulares de los checkouts estaban en buen estado.

Los Iphone 6, 5s y 6s lideran las primeras 3 posiciones seguidos por los Samsung Galaxy J5 y S7.

Se puede observar como la gente busca iphones en estado Excelente y muy bueno, en mayor cantidad que los samsung en el mismo estado.

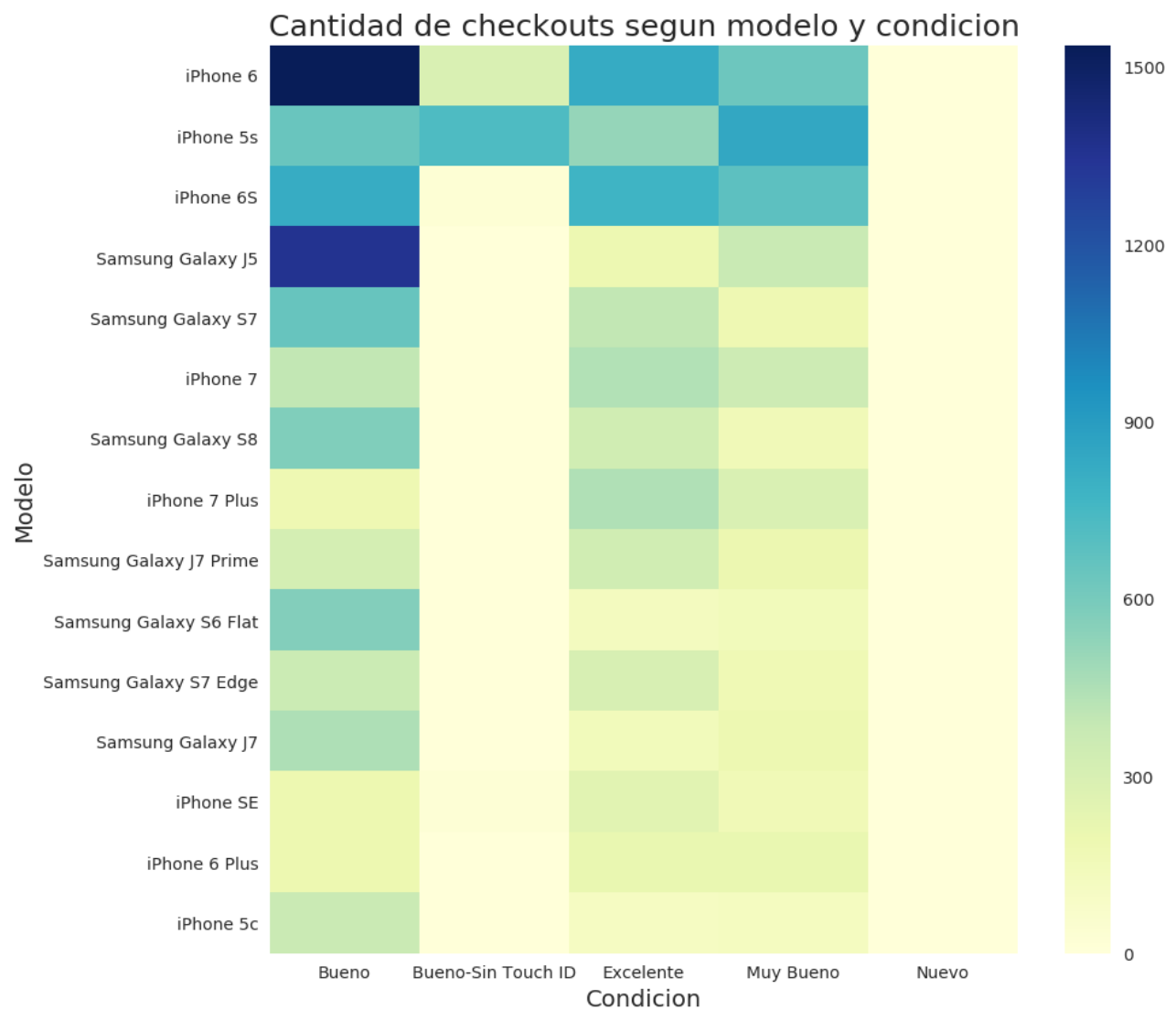


Figura 54

7.4 ¿Que celular lidera en mayor cantidad de conversiones y en que condición?

A diferencia de los checkouts, el celular que lidera la cantidad de conversiones es el Samsung Galaxy J5 en estado Bueno, seguido por el iPhone 5s que estaba segundo en cantidad de checkouts, y podemos encontrar al celular con mayor cantidad de checkouts tercero en cantidad de conversiones.

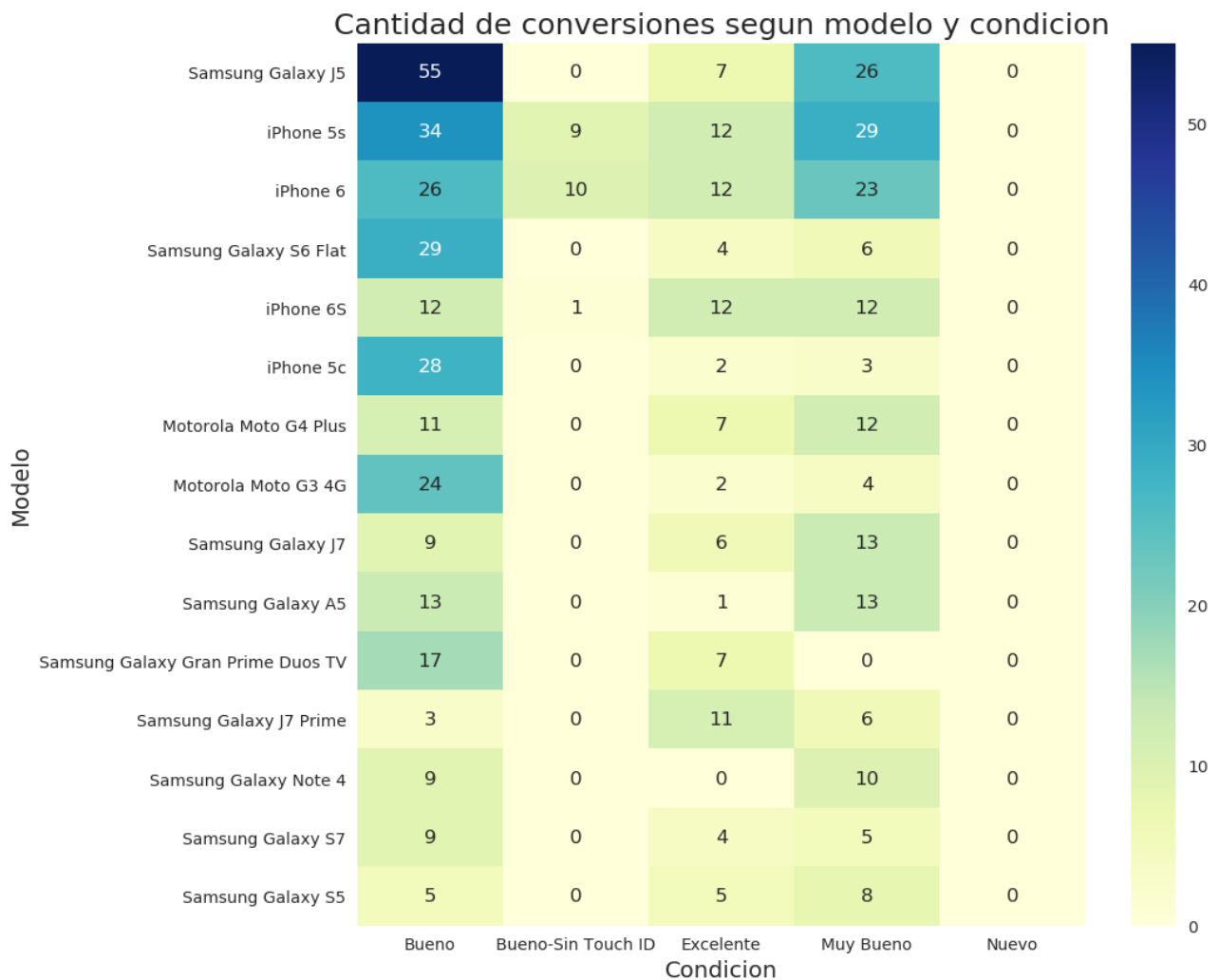


Figura 55

8. Conclusiones

Mediante el presente trabajo se pudo llegar a las siguientes conclusiones:

8.1 Sobre el análisis exploratorio del set de datos

Un dato relevante es que aproximadamente el 70% de todos los valores del data set son nulos, pero en su gran mayoría no son por dato faltante, sino porque sólo algunos atributos se corresponden a cada tipo de evento. Una manera mas eficiente de guardar esta información sería separar en distintas tablas para registrar cada tipo de evento.

8.2 Sobre el análisis según los dispositivos y las regiones

Con el análisis puntual en el evento de **visited site**, estudiando el comportamiento de los usuarios dependiendo su región geográfica y el dispositivo del que acceden se llegó a las siguientes conclusiones:

- Las regiones predominantes que utilizan Trocafone son el sureste y el noreste de Brasil. Esto puede implicar que una posible campaña de marketing de la empresa tendrá mayor peso en estas regiones. Esto tiene sentido dado que éstas regiones son las de mayor densidad de población de Brasil.
- En cuanto a experiencia de usuario frente al dispositivo del cual se accede se llegó a la conclusión que es de igual calidad para los dos dispositivos principales (celulares y computadoras), esto se debe a que hay casi igual cantidad de usuarios que accede por primera vez al sitio por un dispositivo u otro y hay prácticamente la misma cantidad de usuarios que vuelven a visitar el sitio por uno u otro dispositivo.
- En cuanto a las regiones geográficas se observó que las regiones más activas para Trocafone (sureste y noreste) predomina el uso del celular, mientras que para las otras tres regiones predomina el uso de la computadora. Con esto podemos concluir que en caso de realizar publicidad, para ciertas regiones, sería conveniente efectuarla en aplicaciones mobile que sean de uso frecuente en Brasil y para otras regiones sería preferible efectuarla en aplicaciones desktop de uso frecuente de esa zona.

8.3 Sobre el análisis según los canales de ingreso a la plataforma por parte de usuarios nuevos y por los que retornan a la plataforma

- Las cantidad de visitas a la pagina web es predominada por los usuarios viejos sobre los nuevos.
- La mayoría de los usuarios nuevos ingresan a la plataforma por medio de canales pagos, esto se repite para los usuarios que retornan, sin embargo se observa un aumento considerable en la cantidad de usuarios que retornan de manera directa.

8.4 Sobre el analisis de las campañas de anuncios

- A medida que pasaron los meses las campañas fueron cada vez mas masiva.
- Google es la campaña de anuncios que mas rindio en nivel de visitas, checkouts y conversiones.

8.5 Conclusiones sobre la cantidad de ventas y checkouts de celulares según condición

- La mayor cantidad de celulares vendidos y en checkouts se encuentran en Buen estado, esto se puede deber a que las personas encuentran una mejor relación precio-calidad.
- El iphone 6 lidera claramente la cantidad de checkouts generados, sin embargo el celular mas vendido es el Samsung Galaxy J5 que esta 4 en cantidad de checkouts, esto se puede deber a que las personas se encuentran indecisas a realizar una compra de un celular a un precio tan alto y prefieren optar por un celular mas barato que implica un menor riesgo de compra si la persona desconfiá de la plataforma.

9. Links

Kaggle

<https://www.kaggle.com/gaabrobles/tp1-trocafone/notebook>

Github

<https://github.com/gabyrobles93/7506-TP1-Trocafone>