



75.06/95.58 Organización de Datos - 2C 2018

# Trabajo Práctico 1

## Análisis Exploratorio

---

### Grupo 17: "Barplotters"

Integrantes:

- Andrés Pablo Silvestri: 85881 ([silvestri.andres@gmail.com](mailto:silvestri.andres@gmail.com))
- Axel Brian Erlich: 94847 ([axel.erlich@yahoo.com.ar](mailto:axel.erlich@yahoo.com.ar))
- Juan Manuel González: 79979 ([juanmg0511@gmail.com](mailto:juanmg0511@gmail.com))

Link a repositorio de GitHub: [https://github.com/silvahlaravel/Organizacion\\_Datos\\_2C](https://github.com/silvahlaravel/Organizacion_Datos_2C)

Fecha de entrega: 24/09/2018

Fecha de re-entrega: 08/10/2018

# Contenido

<b>1 - Introducción</b>	<b>3</b>
1.1 - ¿De qué se encarga Trocafone?	3
<b>2 - Estructura y manejo de los datos</b>	<b>4</b>
<b>3 - Análisis sobre visitas y campañas publicitarias</b>	<b>7</b>
3.1 - ¿Cómo evolucionó la cantidad de visitas en el año 2018?	7
3.2 - ¿Se realizó alguna campaña publicitaria fuerte en los meses de crecimiento?	8
3.3 - ¿Habrá correlación entre las campañas publicitarias y las visitas?	9
<b>4 - Análisis sobre los horarios y la actividad:</b>	<b>10</b>
4.1 - ¿Cuáles son los horarios donde se efectivizan mayor cantidad de ventas?	10
4.2 - ¿Hay correlación con los horarios donde se efectúan las campañas publicitarias?	11
4.3 - ¿Cómo varía la condición de los elementos buscados en los últimos meses?	12
<b>5 - Análisis sobre cantidades en general</b>	<b>14</b>
5.1 - ¿Cuáles son los eventos más predominantes?	14
5.2 - ¿Desde qué sistemas operativos se utilizó más el sitio?	15
5.3 - ¿Cuáles son las marcas con más checkouts hechos?	16
5.4 - ¿Cuántos eventos hacen los usuarios en relación a las marcas principales?	17
5.5 - ¿Qué podemos observar con respecto a los modelos escogidos de las primeras marcas?	18
5.6 - ¿Con qué navegador la gente accede al sitio?	20
<b>6 - Análisis sobre los dispositivos utilizados por los usuarios del sitio</b>	<b>21</b>
6.1 - ¿Cuáles son los dispositivos que se usan para ingresar al sitio?	21
6.2 - ¿Cómo se relaciona la capacidad de almacenamiento de los dispositivos según su condición de uso?	22
<b>7 - Análisis sobre la zona geográfica de los usuarios del sitio</b>	<b>24</b>
7.1 - ¿Cuáles son los principales países desde los que se ingresa al sitio?	24
7.2 - ¿Cuál es la relación de visitas entre Brasil y Argentina?	25
7.3 - ¿Para Brasil y Argentina, cuáles son las provincias o estados más populares?	26
<b>8 - Conclusiones</b>	<b>29</b>

# 1 - Introducción

Este trabajo está enfocado en hacer un primer análisis de los datos ofrecidos por Trocafone, de manera que encontremos particularidades que puedan ser de interés para dicha entidad.

Siendo este el caso, lo primero que vamos a hacer es interiorizarnos de uno de los pilares que tiene la ciencia de datos, el negocio.

Se hará un análisis general para entender un poco más el negocio y luego un análisis más fino sobre los datos para contestar algunas preguntas que podrían resultar curiosas.

## 1.1 - ¿De qué se encarga Trocafone?

Trocafone es esencialmente una plataforma de comercio online para comprar y vender celulares/tablets usados de forma segura, con la garantía de que Trocafone se encarga de validar la procedencia de los aparatos y de garantizar el correcto funcionamiento de los mismos.

En este caso, Trocafone nos facilitó un set de datos que contiene distintos tipos de eventos que se registran en su sitio web, ya sea visitas, compras u otros tipos de eventos que se detallarán luego.

## 2 - Estructura y manejo de los datos

A fin de mejorar la performance de ejecución de las operaciones realizadas, y por otro lado facilitar las tareas de análisis, se ha hecho un tratamiento previo sobre el set de datos proporcionado por la cátedra.

Se puede consultar el tratamiento completo, así como también el código utilizado para generar el análisis y los gráficos, en el notebook disponible en la carpeta **'REENTREGA'** del repositorio de *Git* presentado en la carátula de este informe.

Como primer medida pasamos a 'category' todas las columnas que nos resultan acordes para ser categorizadas, lo que nos permite darle otro tipo de manejo como así también mejorar el rendimiento de las diferentes operaciones que involucren estas columnas.

Por otro lado, también hacemos una diferenciación sobre los eventos, como para poder tener una visión sobre la importancia que cada uno de estos nos puede proveer, puesto que aunque haya eventos que se relacionan entre sí o pueden significar cosas similares, vamos a considerar que aquellos eventos con mayor cantidad son los que aportan más riqueza al informe.

En forma similar, diferenciamos aquellos eventos que corresponden a las visitas propiamente dichas con aquellas que corresponden al acceso a través de una campaña publicitaria, lo que nos podrá permitir observar el cambio en el tiempo de ambas y compararlas, para estos es necesario agruparlas por mes y diferenciar unas de otras.

Continuando con nuestro análisis, realizamos un ordenamiento por persona y fecha del evento, lo que nos permite dar un primer paso para luego continuar con una agrupación sobre algunos tipos de eventos particulares (uno con poca cantidad como es conversion y otro con mucha como es campaña) para poder de esta manera hacer un análisis sobre en que momentos del día los usuarios hacen tal o cual acción.

Por otro lado, armamos algunas columnas auxiliares que nos permitan separar el tiempo en día, mes y hora para luego poder aplicar esta información como parámetros de búsqueda y organización en los diferentes gráficos que planteamos, como así también por ejemplo plantear el día de la semana con un formato personalizado que nos permita hacer más legible futuros gráficos.

Independientemente del tipo de evento o incluso de la cantidad de los mismos, buscamos armar un set de datos que nos muestre las preferencias de los usuarios sobre la condición del aparato, entendiendo que a mayor cantidad, mayor es el número de aparatos con esa condición. Dado que este dato es algo que obtenemos en otro idioma y podemos aplicarle un pequeño trabajo para pasarlo a nuestro idioma y hacerlo más legible en el informe, como así también tomar la salvedad de que hay dos tipos de condición que hacen referencia al estado 'Bueno', lo que decidimos agruparlo en un solo estado, puesto que no es relevante en cantidad ni tampoco brinda valor como conceptos separados.

También decidimos separar los sistemas operativos de los diferentes equipos que conforman este set de datos, pero no vamos a entrar en el detalle de diferenciar las versiones puntuales de cada sistema operativo (al menos en una primera instancia no nos interesa), sino más bien que vamos a agruparlos de manera de poder conocer las familias de sistemas operativos y poder plantear u observar cuales son aquellas que predominan en el mercado.

De todas maneras esto no quiere decir que no vayamos a diferenciar entre los sistemas que son móviles o para computadoras, aunque la empresa madre sea la misma pretendemos diferenciar los sistemas operativos a gran escala pero por ejemplo tendremos Windows y Windows Phone, lo que requerirá en algunos casos un tratamiento especial.

También vamos a diferenciar las marcas con mayor importancia en el mercado y ver cuáles son los equipos o modelos en particular que éstas tienen, considerando que salen modelos nuevos de manera bastante continua, pero al menos nos sirve como para tener una foto del momento actual en el que se está haciendo el informe.

Nuestra intención radica también en averiguar que navegador usa la mayoría de la gente, sin tratar de diferenciar mucho cuáles son navegadores mobile y cuáles no, sino simplemente agruparlos y analizar el resultado que esto nos puede traer.

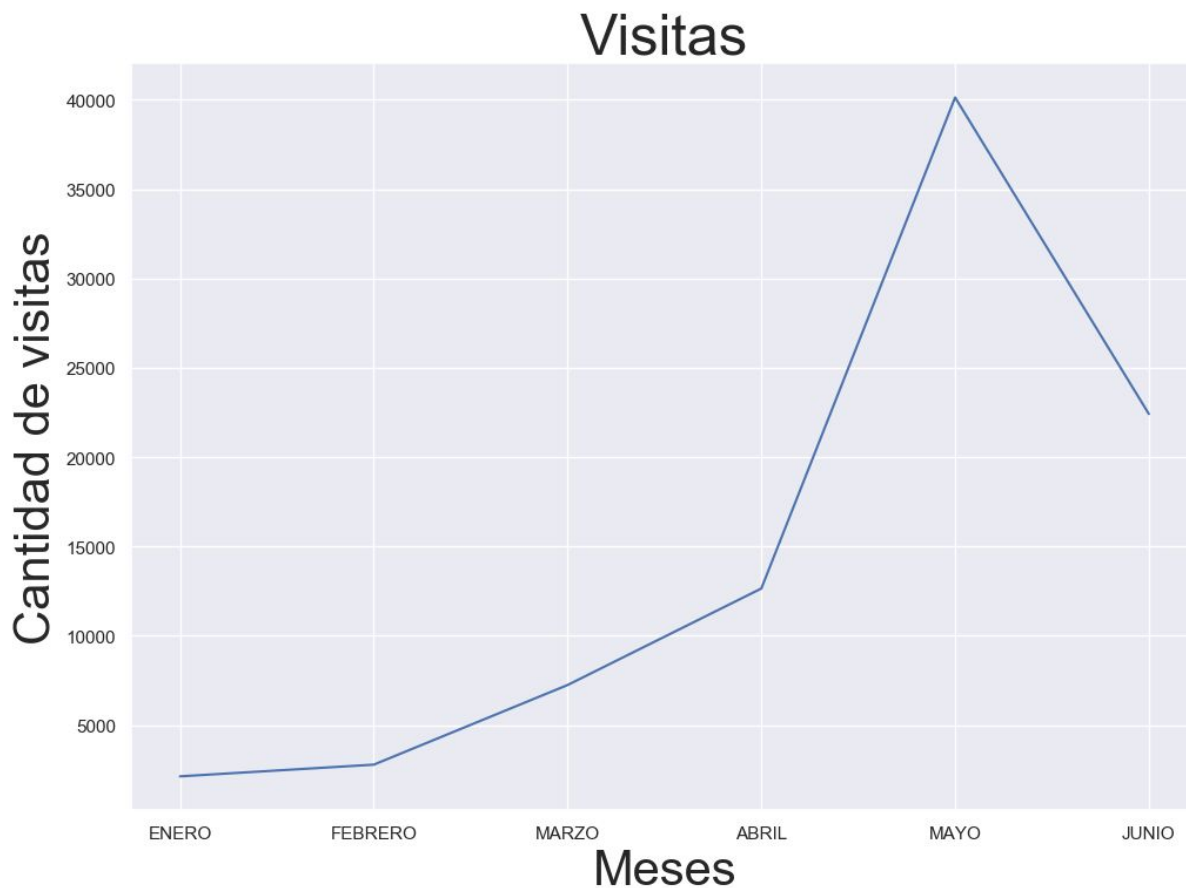
Al plantear el análisis geográfico respecto a la cantidad de visitas, efectuamos un agrupamiento de los datos para lograr llegar a los 5 países más populares, categorizando como 'Otros' al resto de los valores presentes en el set de datos. Los valores 'Unknown' fueron ignorados. Asimismo, para sacar la proporción entre Brasil y Argentina, utilizamos columnas adicionales para identificar fácilmente si un evento era de alguno de dichos países.

Por último vale agregar que en algunos casos nos hemos tomado la libertad de considerar algunos valores como despreciables en cuanto a relación de cantidad sobre el total, como así también hacer un análisis sobre los eventos que más cantidad de registros nos brindan lo que nos permite entender la situación con un volumen de datos mayor lo cual nos acerca un poco más a la realidad.

### 3 - Análisis sobre visitas y campañas publicitarias

#### 3.1 - ¿Cómo evolucionó la cantidad de visitas en el año 2018?

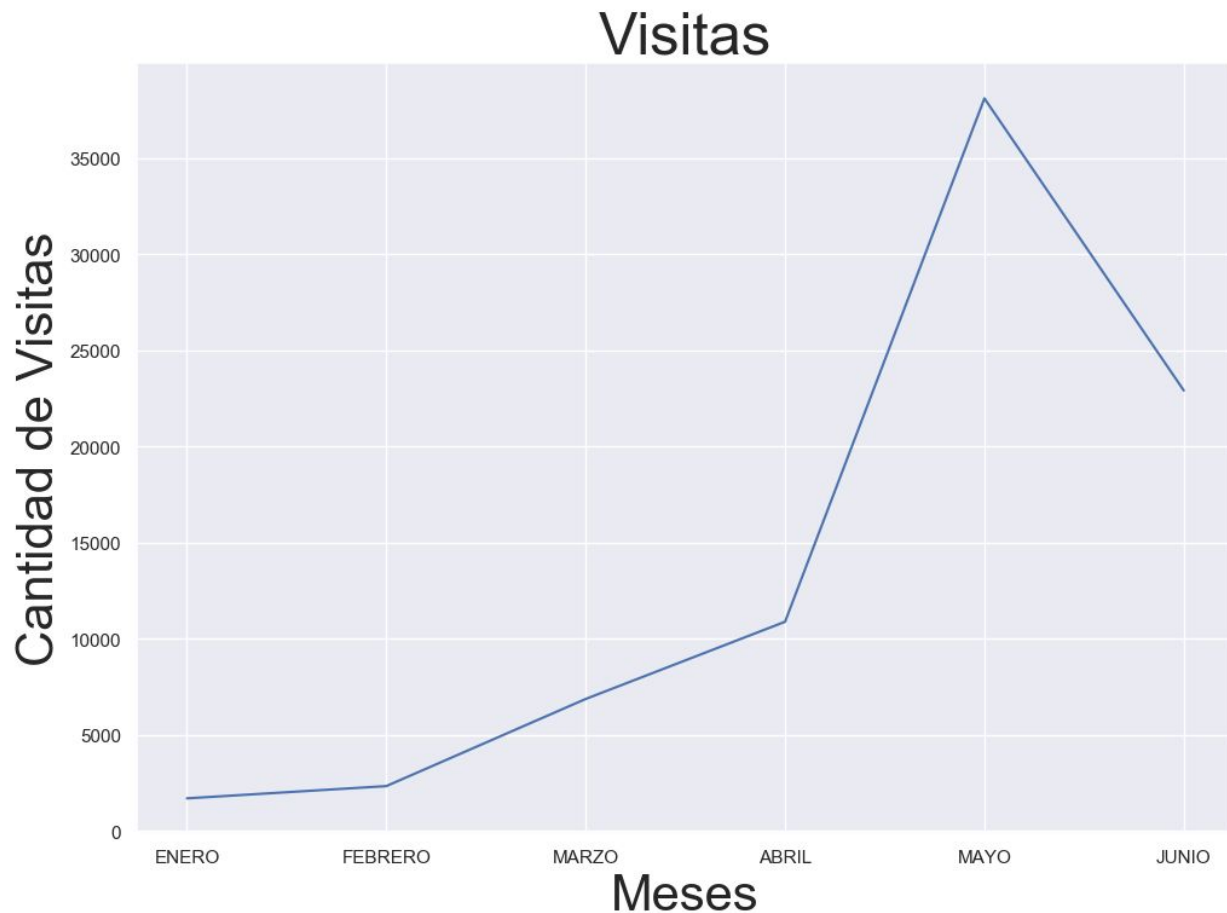
Sabiendo que el set de datos sólo proporciona datos sobre el año 2018, lo que se plantea es ver cómo ha ido cambiando la cantidad de accesos/visitas al sitio a través de ese año, al menos de los meses sobre los cuales hay datos. Como se puede ver, los primeros meses es un número bajo y luego va incrementándose progresivamente, mientras que al final posee una baja sensible.



Veremos si estos números son respaldados por ingresos debido a una campaña publicitaria.

### 3.2 - ¿Se realizó alguna campaña publicitaria fuerte en los meses de crecimiento?

Aunque similar, podemos ver graficadas las visitas al sitio a través de campañas publicitarias, lo que nos indica el valor que estas campañas pueden tener a la hora de aplicarlas o no en la empresa.

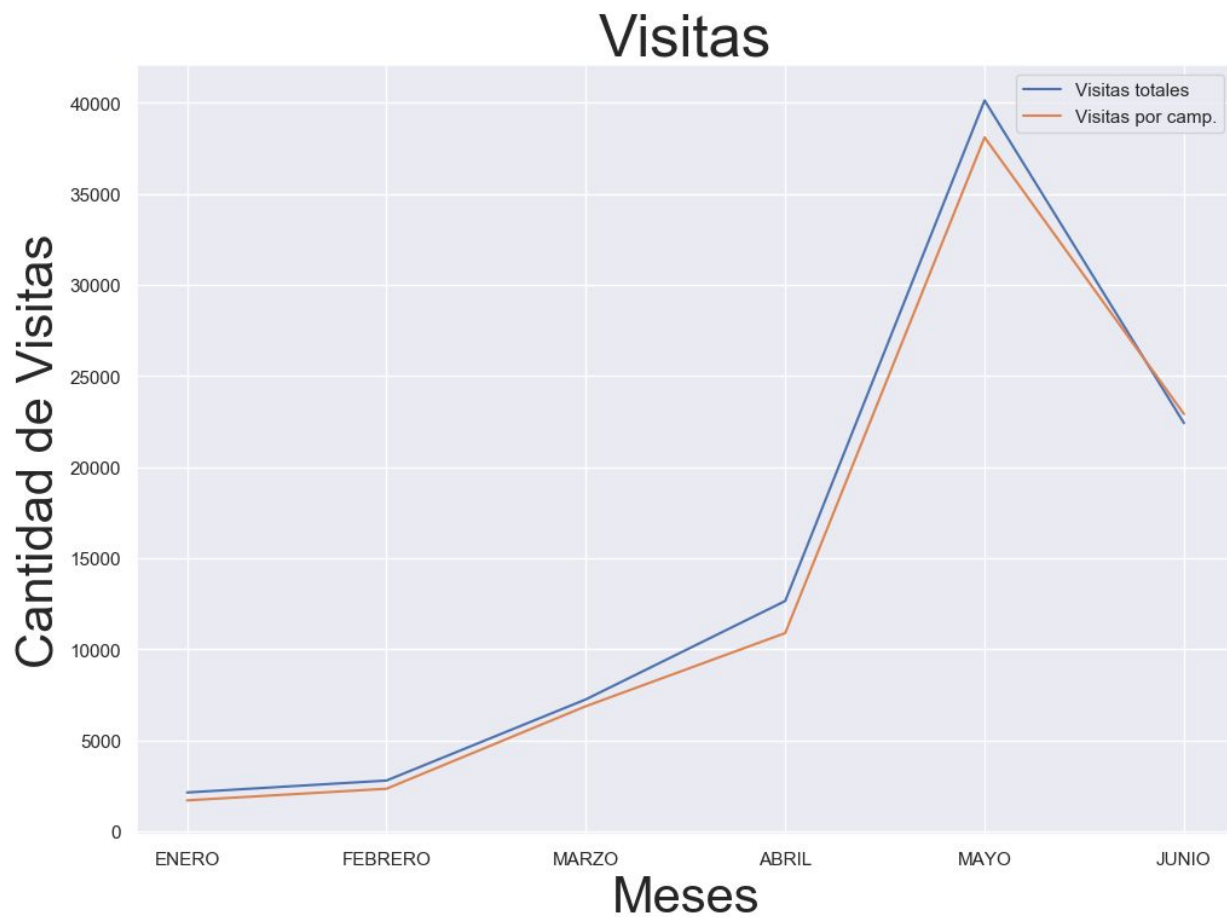


Queda claro que es muy similar al de visitas. Es decir la cantidad de entradas al sitio por campañas sigue la misma forma que las visitas al sitio en general.



### 3.3 - ¿Habrá correlación entre las campañas publicitarias y las visitas?

Se grafican los datos frente a frente, notar como las visitas por campaña son la principal fuente de visitas al sitio.

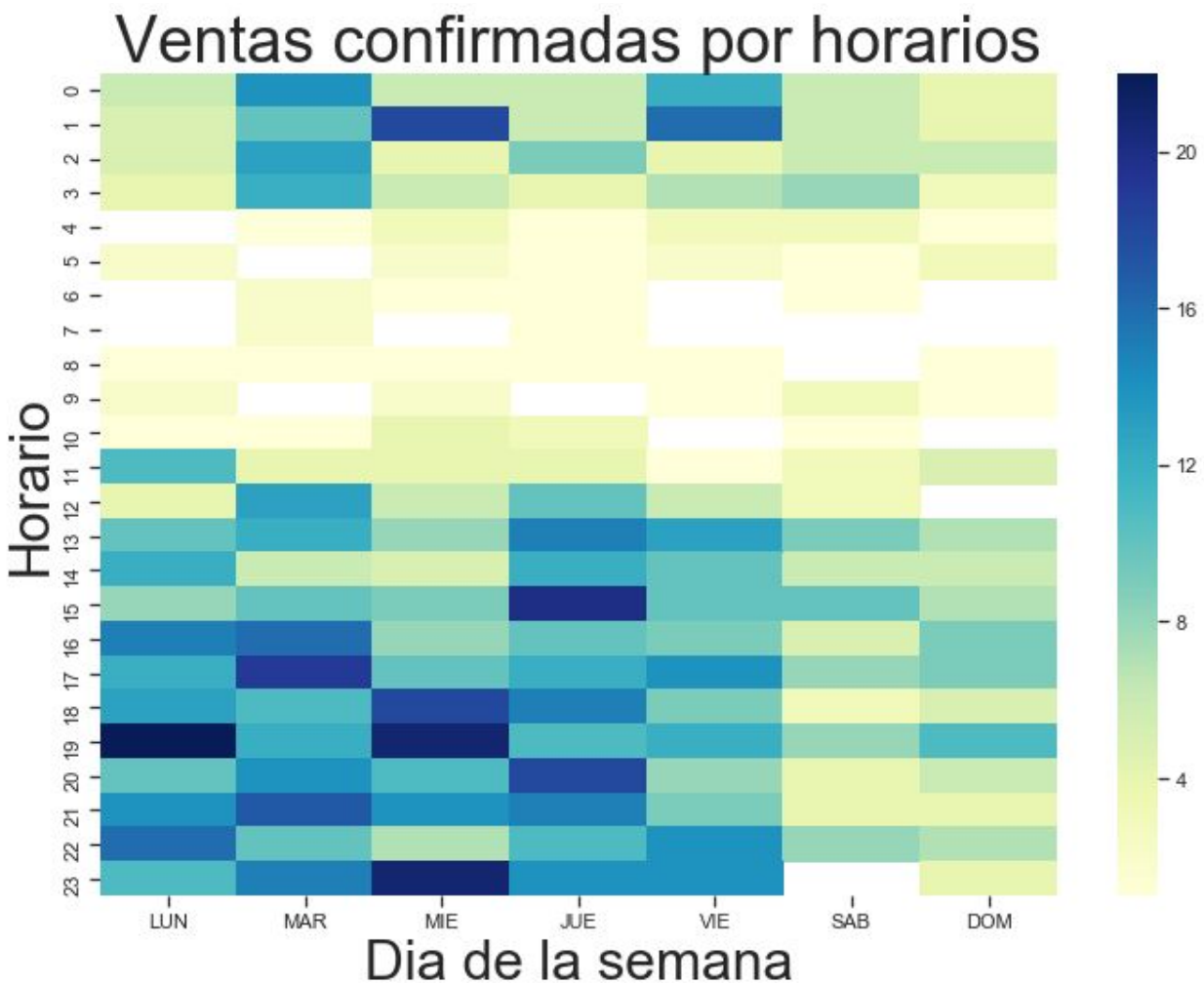


Esto nos indica que estas campañas son bastante provechosas para la empresa y marcan una dependencia bastante crítica con la cantidad de visitas. Observar que si en el último mes graficado bajaron las campañas publicitarias, las visitas siguieron esta tendencia, lo que demostraría una peligrosa dependencia.

## 4 - Análisis sobre los horarios y la actividad:

### 4.1 - ¿Cuáles son los horarios donde se efectivizan mayor cantidad de ventas?

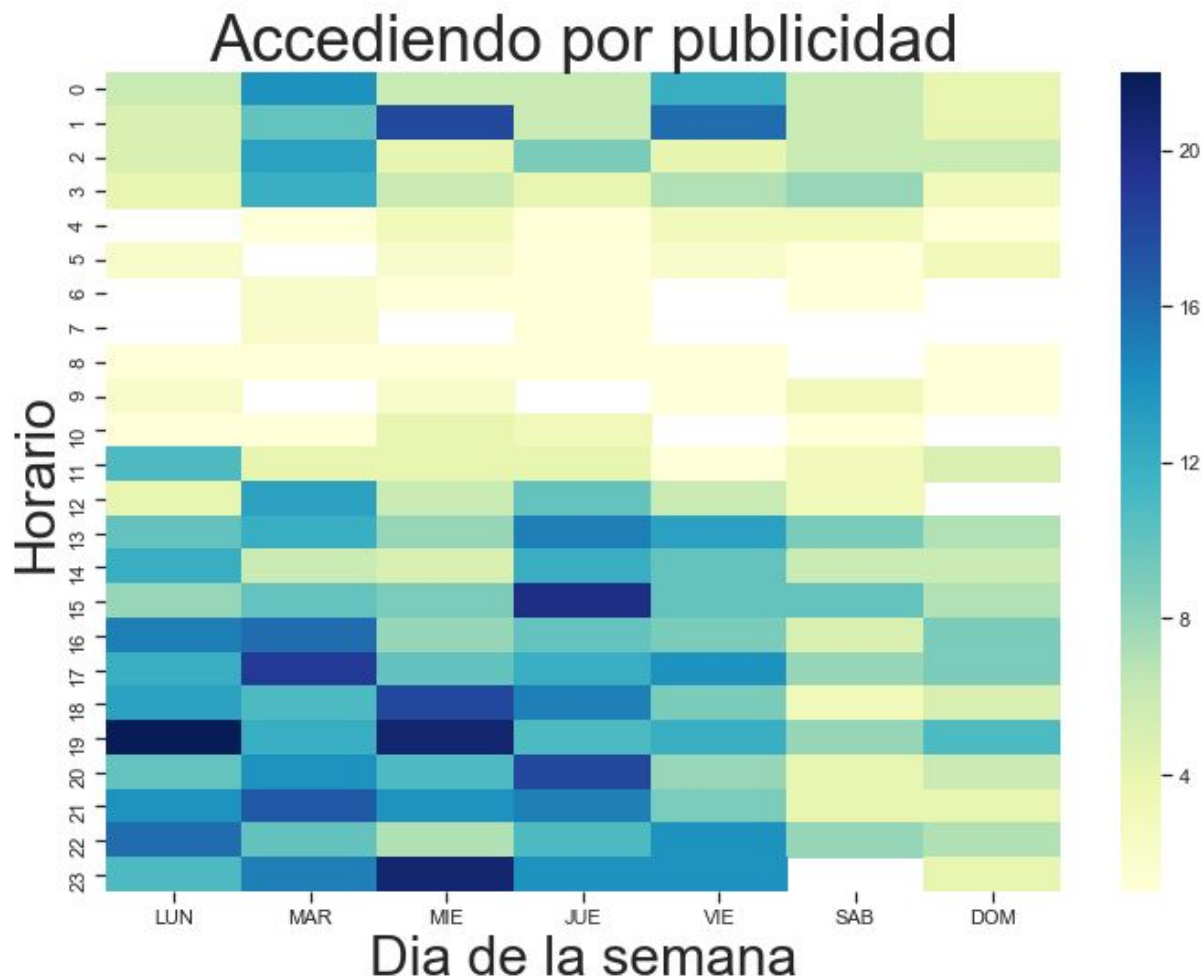
Lo que se observa a continuación es un mapa que relaciona las horas con los días de la semana y la cantidad de ventas que se han confirmado, siendo éste el evento conversión que nos asegura que se realizó la venta. Si bien su cantidad en relación a otros eventos no es tan representativa, de todas maneras se pueden ver las franjas horarias más escogidas por la gente para las ventas.



En conclusión, los días de semana en horario post laboral y hasta la hora de dormir es el período de tiempo donde más se llevan a cabo estos eventos.

#### 4.2 - ¿Hay correlación con los horarios donde se efectúan las campañas publicitarias?

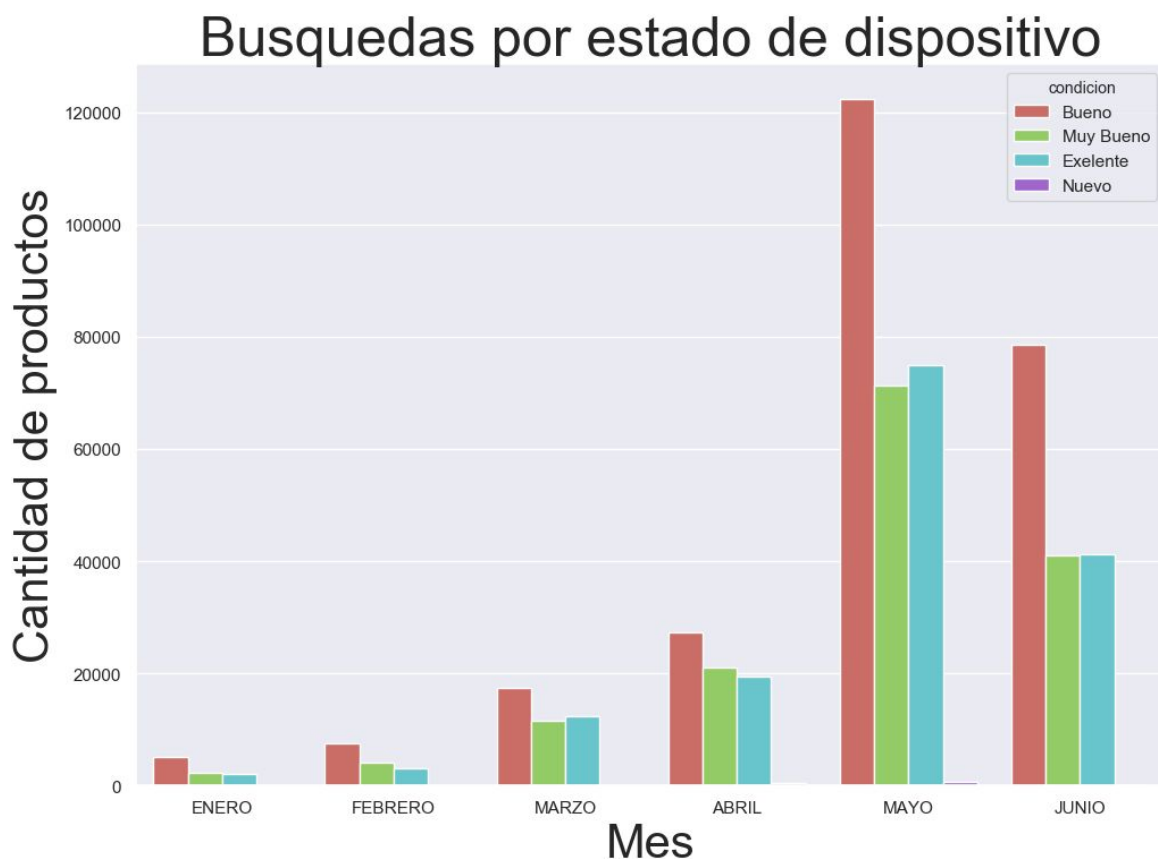
Como se sabe que los eventos de conversión son mucho menos que los demás, se buscó otro tipo de evento, por ejemplo, los accesos a través de las campañas publicitarias, para corroborar si el gráfico anterior refleja verdaderamente la relación de horarios y accesos al sitio.



Esto demuestra que efectivamente el horario post laboral y en los días de semana son los horarios donde se da mayor tráfico y accesos al sitio.

### 4.3 - ¿Cómo varía la condición de los elementos buscados en los últimos meses?

A continuación se busca mostrar, según el estado de los productos, cómo termina escogiendo la gente y cómo esta evolución se da en los sucesivos meses donde tenemos información para analizar.



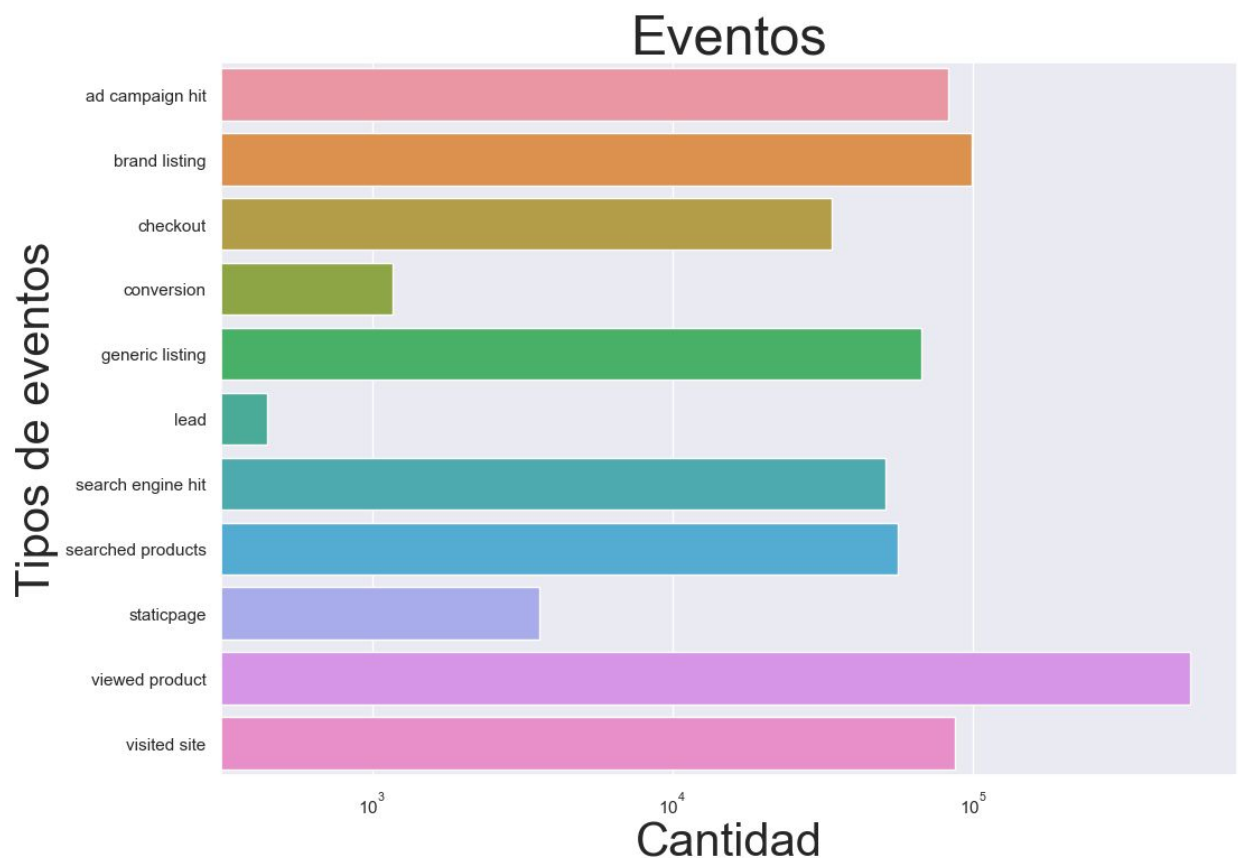
Como podemos ver, esto se condice con los gráficos que mostraban la evolución temporal, puesto que en los primeros meses el valor numérico es muy inferior y luego se ve un incremento. Con respecto a lo que se analizó, los productos en un estado 'Bueno' son los que dominan, y son los que han tenido un incremento todavía mayor en los últimos meses, llegando casi a la suma del segundo y tercero en disputa.

Por otro lado vemos que nuevos casi no hay, cosa lógica puesto que la empresa trabaja con equipos usados.

## 5 - Análisis sobre cantidades en general

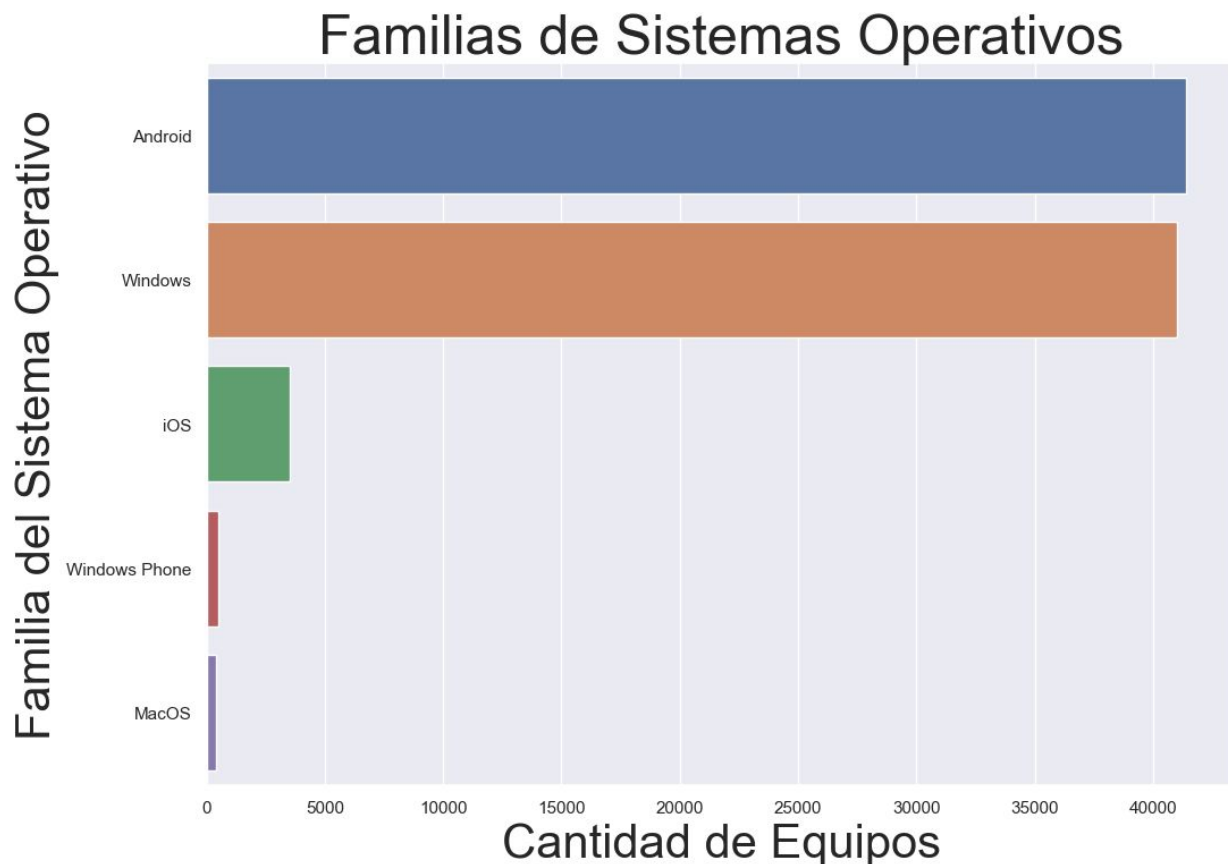
### 5.1 - ¿Cuáles son los eventos más predominantes?

Este set de datos está basado en los eventos que se van registrando en la web, a continuación se enumeran según su frecuencia, lo que permite sacar una conclusión sobre cuáles son los eventos que más nos interesa prestarle atención, puesto que a mayor cantidad mejor se podrá analizar su variación.



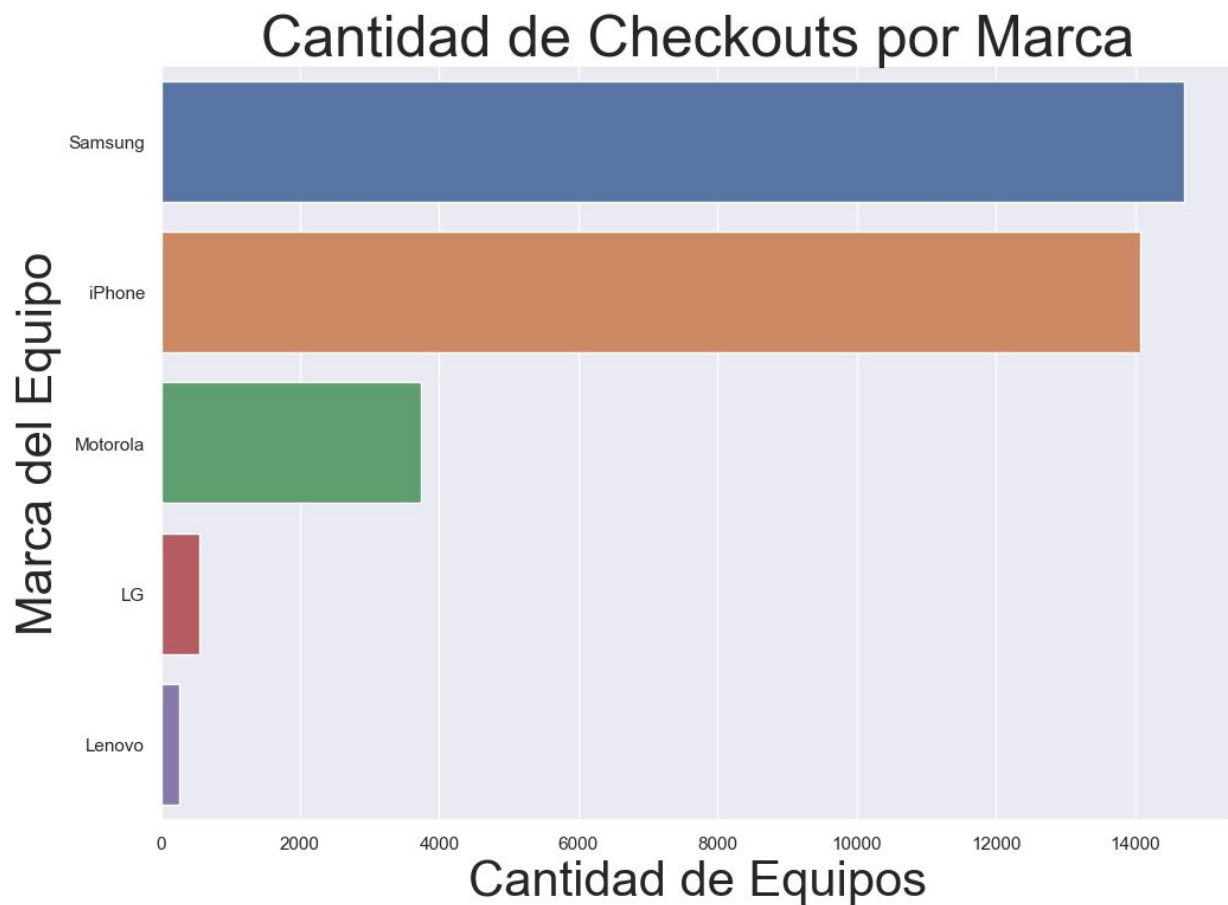
## 5.2 - ¿Desde qué sistemas operativos se utilizó más el sitio?

Se verán los diferentes sistemas operativos y su predominio en el mercado, vale aclarar que no se hace una distinción entre el tipo de equipo, pero como vemos en el gráfico Android y Windows son los que predominan, siendo el primero para smartphones y el segundo para PC, con lo cual tenemos de manera muy sencilla una idea de que sistema domina en cada tipo de aplicación, siendo despreciable las tablets por su pequeña cantidad en el set de datos.



### 5.3 - ¿Cuáles son las marcas con más checkouts hechos?

Se tomó el evento de checkout como aquel más cercano a una compra. Si bien es cierto que el evento de conversión representa fehacientemente una compra, éste contiene muy pocos registros en comparación a los checkouts realizados, por lo cual optamos por tomar los checkouts para determinar cuáles son las marcas que más dominio tienen sobre el mercado.

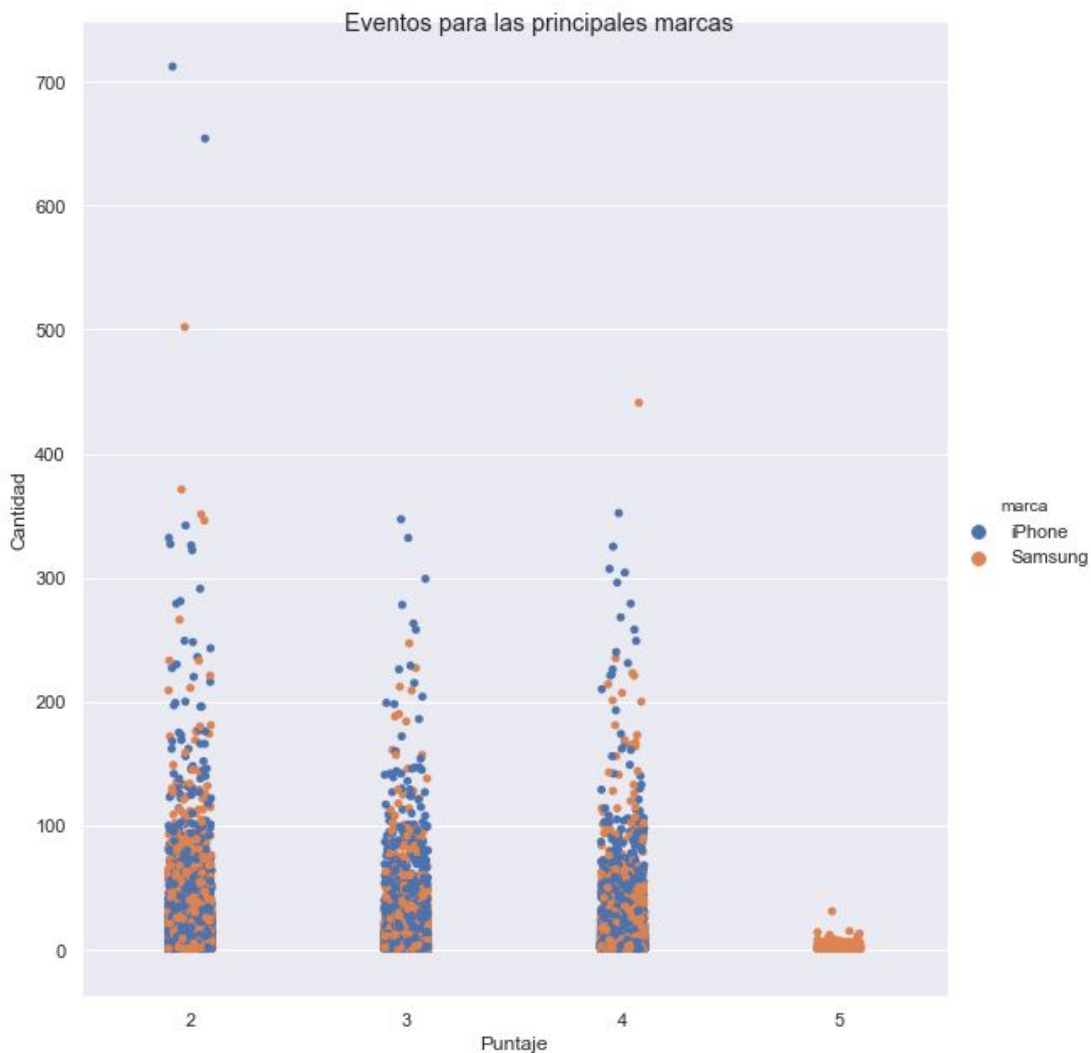


Como se puede ver Samsung e iPhone son las marcas que más predominan, quedando en tercer lugar Motorola, aunque la diferencia entre los dos primeros y el resto es verdaderamente notable.



## 5.4 - ¿Cuántos eventos hacen los usuarios en relación a las marcas principales?

Teniendo en cuenta que los usuarios generan muchos tipos de eventos, vamos a obviar por un momento los eventos de checkout y conversión, que son los que nos indican la idea de 'compra', y así trabajar sobre el resto de eventos, analizando la relación existente con respecto al estado/puntaje de los equipos (bueno, muy bueno, excelente, nuevos).

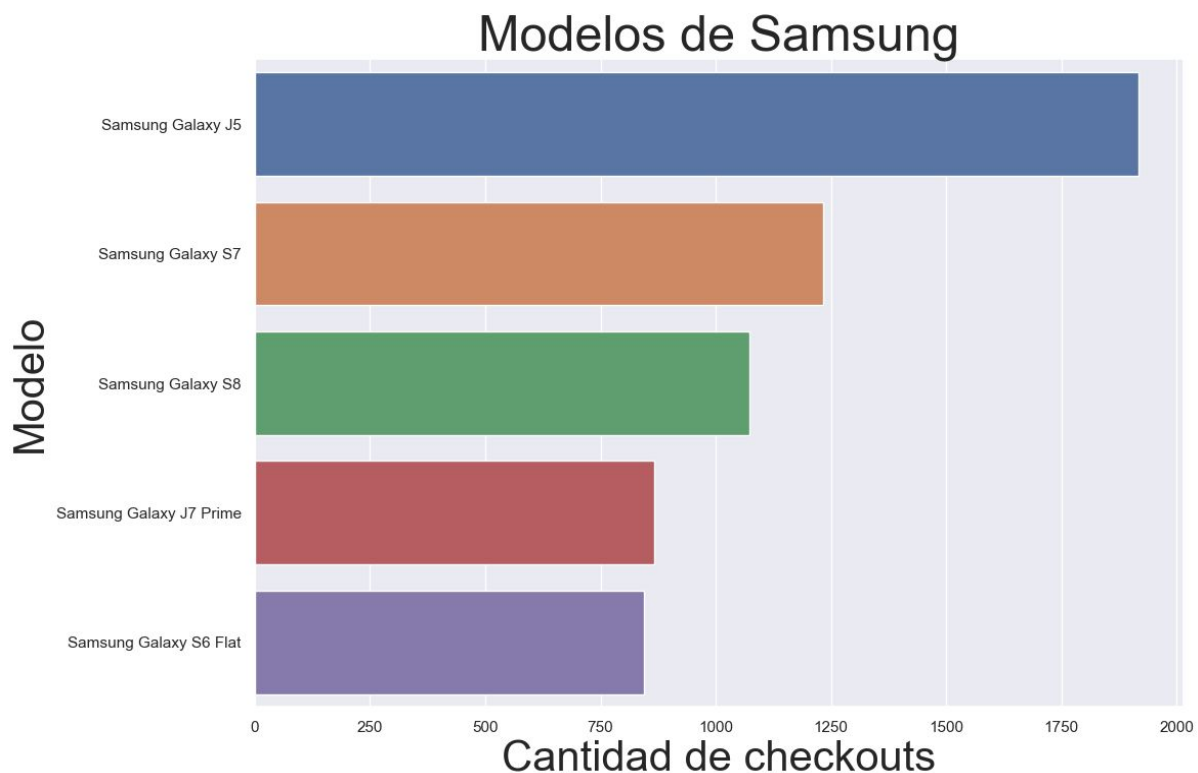


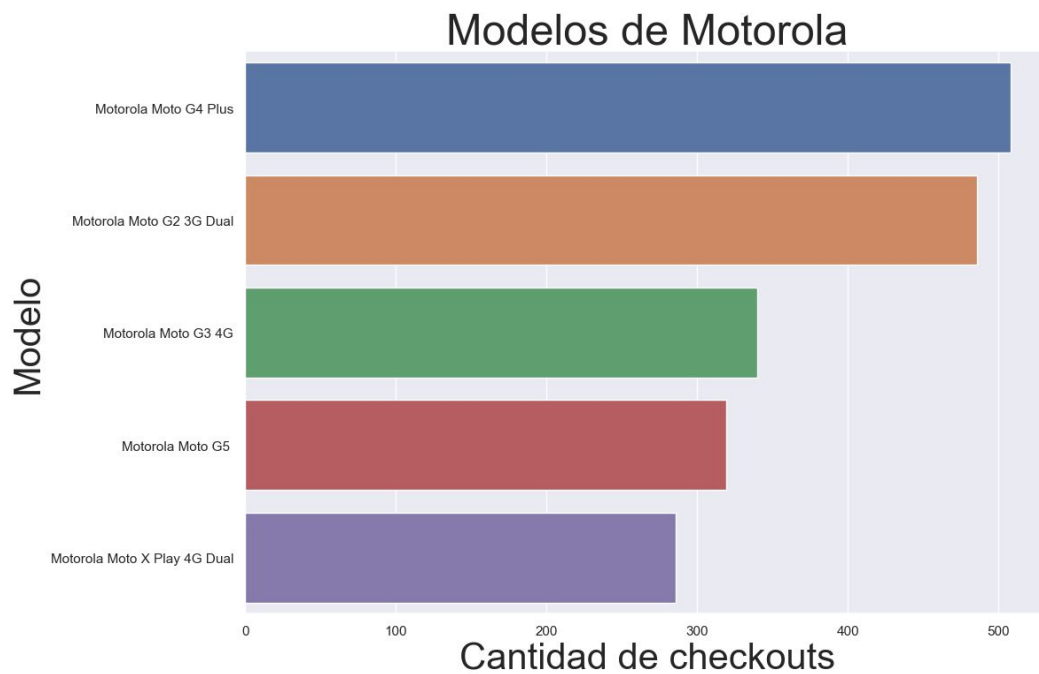
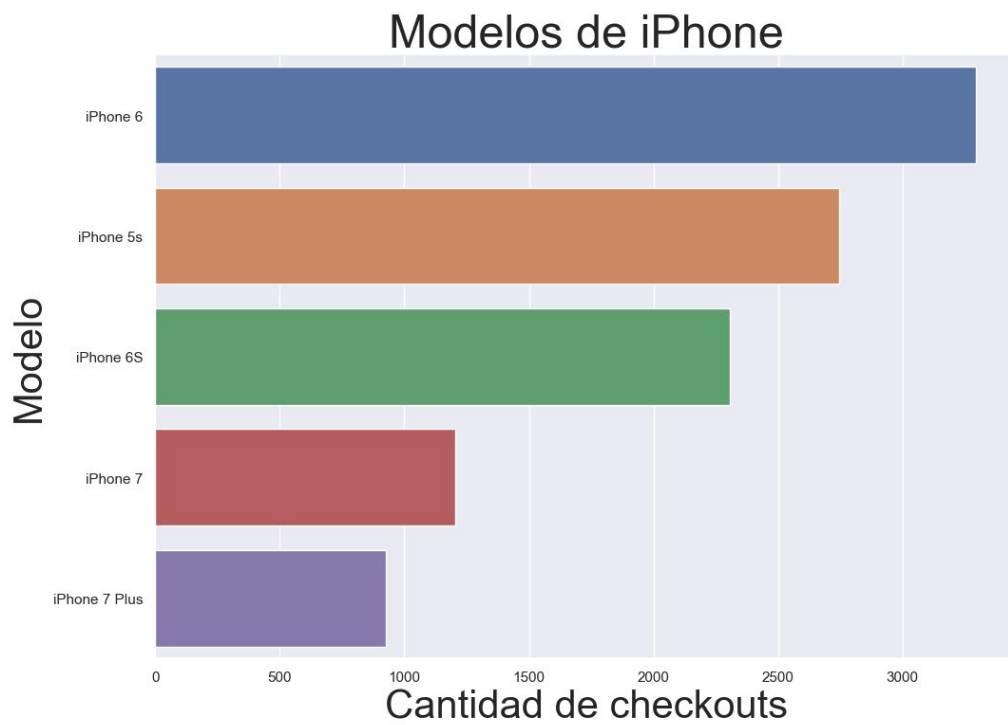
Podemos observar que la mayoría de los eventos están repartidos equitativamente entre ambas marcas, y que hay una franja lógica entre 0 y 100 eventos por usuario, luego a mayor cantidad serían los casos más atípicos de usuarios.

Por otro lado, podemos observar que sólo para la marca Samsung encontramos eventos relacionados con un puntaje que corresponde a 'nuevo', sin tener siquiera participación los iPhone, lo que nos da el indicio que por algún motivo la gente que entra al sitio en busca de iPhone no piensa en los productos nuevos, mientras que para Samsung al menos hacen una búsqueda.

### 5.5 - ¿Qué podemos observar con respecto a los modelos escogidos de las primeras marcas?

Los siguientes tres gráficos, aunque son sencillos, permiten visualizar cuáles son los 5 modelos más buscados para las marcas que dominan el mercado. En Samsung podemos ver que el Galaxy J5 supera al resto de manera más marcada, mientras que en las otras dos marcas están más parejos los resultados.



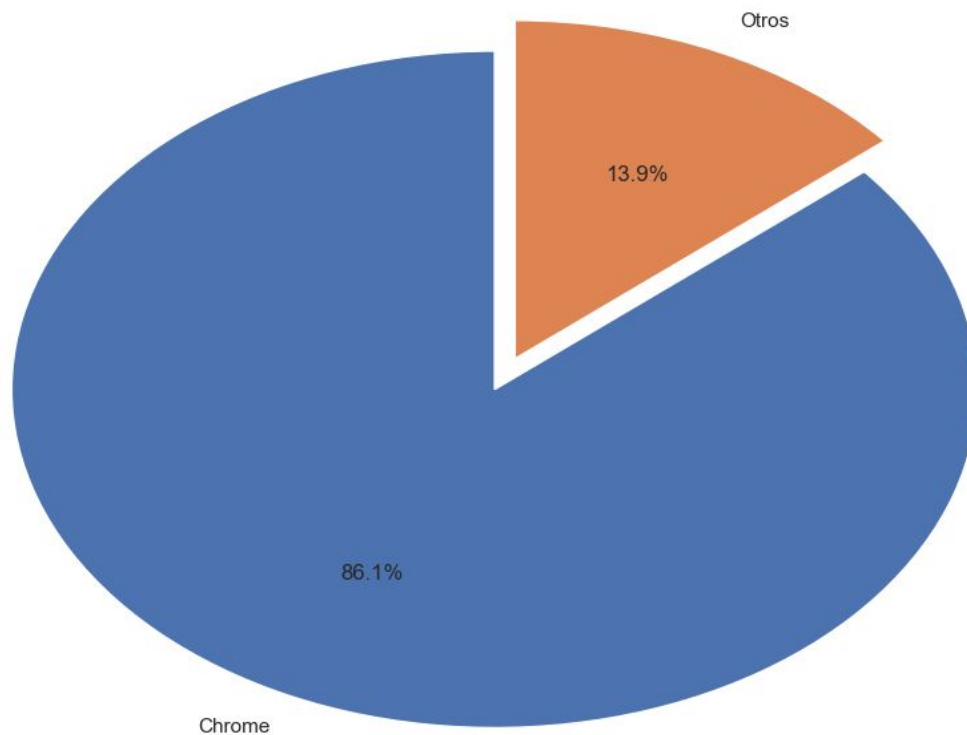


## 5.6 - ¿Con qué navegador la gente accede al sitio?

Lo que nos proponemos a continuación es analizar cuál es la preferencia de los usuarios con respecto a la elección del navegador a la hora de acceder al sitio, aunque esto no tenga relación directa sobre los aparatos que proporciona la empresa.

Entendemos que este análisis sirve para conocer a las preferencias de los usuarios del sitio.

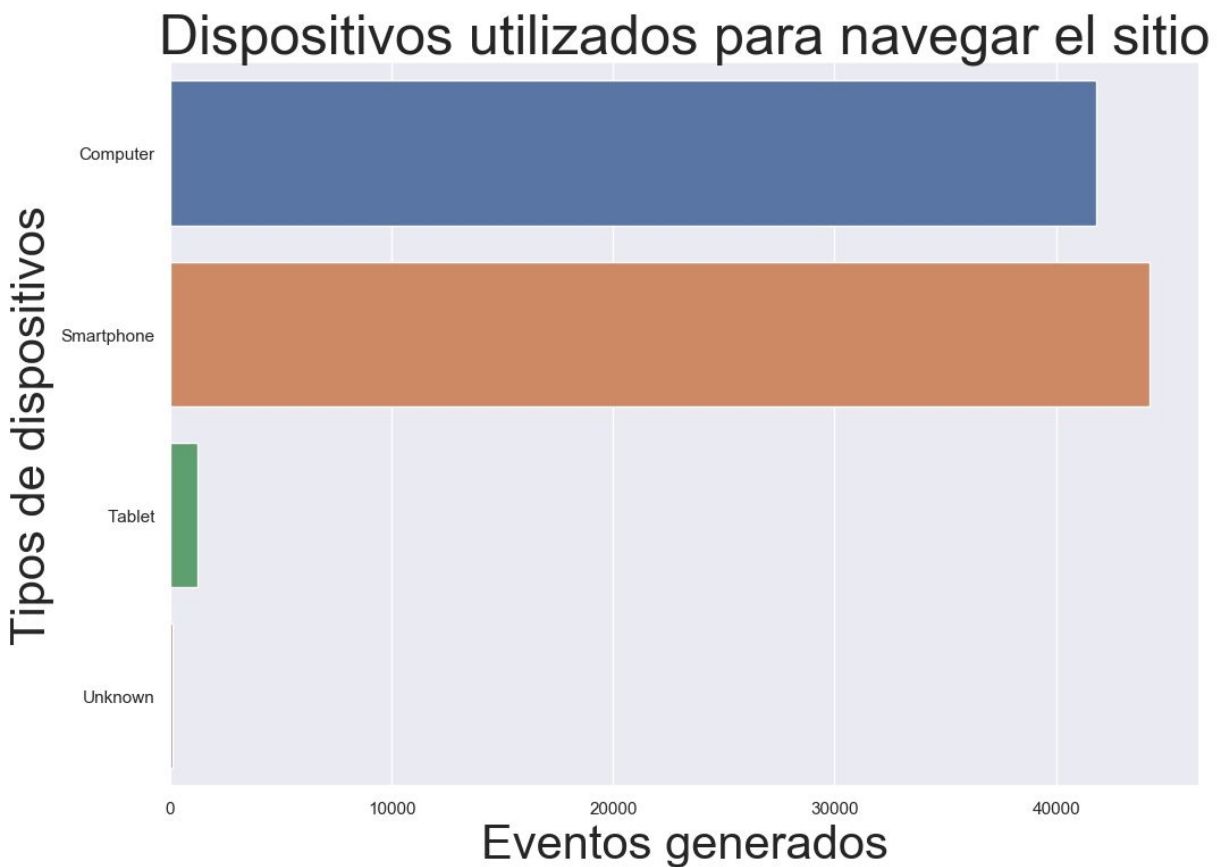
### Navegadores usados



Viendo la abrumadora la diferencia que tiene el Chrome sobre el resto, esta información puede llevar a enfocar el trabajo de mejoras en el sitio con respecto a este navegador. Aunque entendemos que hoy esté mucho más estandarizado el desarrollo web puede ser de utilidad saber cual es la preferencia de los usuarios y enfocar las mejoras a este navegador o estar siempre en consonancia con las actualizaciones que este lleva a cabo.

## 6 - Análisis sobre los dispositivos utilizados por los usuarios del sitio

### 6.1 - ¿Cuáles son los dispositivos que se usan para ingresar al sitio?



Por lo que se observa en el gráfico, hay una cantidad muy similar de eventos generados desde smartphones y computadoras, quedando relegadas las tablets por un amplio rango. Esto puede indicarnos que tal vez no es tan necesario dedicar desarrollo de visualizaciones para dispositivos de tamaño mediano como una tablet, y poner más énfasis en el desarrollo para mejorar la interfaz en la web y en dispositivos pequeños como los smartphones.

Un punto a destacar es que hay una gran cantidad de eventos que se generaron de una fuente desconocida:

**Cantidad total de eventos generados: 1011288.**

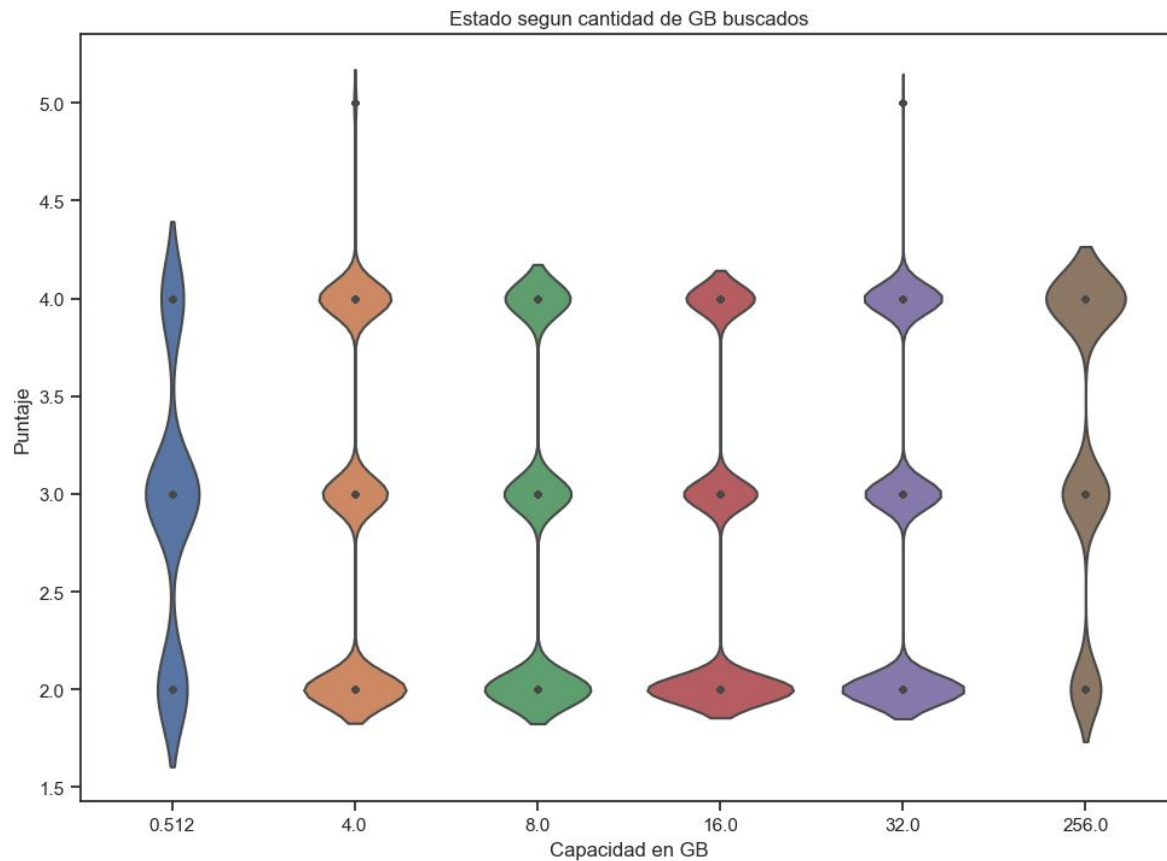
**Cantidad total de eventos generados desde dispositivos desconocidos: 923910.**

Por lo tanto desconocemos el dispositivo que originó evento para el: 91.36% de los eventos. Pero aun así, los pocos datos que se tienen alcanzan para determinar una tendencia.

## **6.2 - ¿Cómo se relaciona la capacidad de almacenamiento de los dispositivos según su condición de uso?**

En este caso interesa ver una relación entre un puntaje dado sobre la condición de uso (de menor a mayor en modo numérico) y su capacidad de almacenamiento, lo que podría brindar información sobre por ejemplo si los equipos de mayor capacidad son buscados en mejores condiciones (muy bueno o excelente), como también saber si a menor capacidad de almacenamiento interesa poco su condición.

Estas son hipótesis en un análisis previo, el gráfico en cambio nos muestra una realidad donde son muy pocos los equipos de alta capacidad y la mayoría son buscados en una condición básica o normal, lo que denominamos 'Buenos'.

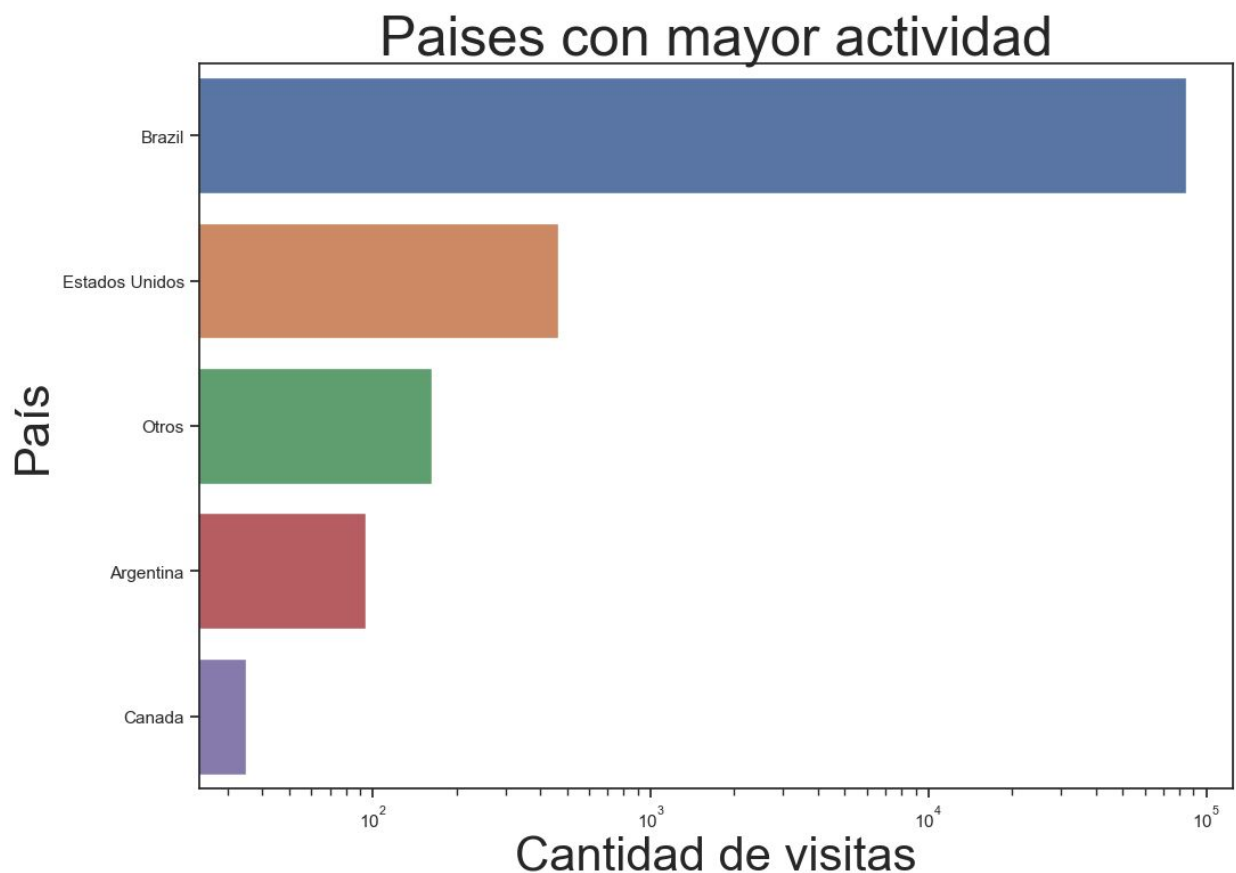


Se puede interpretar el comportamiento de las personas de la siguiente manera, vemos que las que buscan un smartphone con poca capacidad es más variable con relación al estado del producto, por otro lado aquellas personas que buscan una franja de capacidad un poco más amplia o estándar al día de hoy notamos que terminan cayendo mucho más en los que son de estado 'bueno' y que a la inversa aquellos que buscan una capacidad bien alta tratan de caer en aquellos que tienen un mejor estado a ser 'Excelente'.

## 7 - Análisis sobre la zona geográfica de los usuarios del sitio

### 7.1 - ¿Cuáles son los principales países desde los que se ingresa al sitio?

Se explorará a continuación los principales países desde donde se produce la mayor cantidad de accesos del sitio. El único evento con información de regiones es 'visited site', por lo que nos centraremos en esos registros. En primer lugar, nos quedamos con los 5 países más populares y el resto los agrupamos en 'Otros'. También ignoramos los registros con valor Unknown:

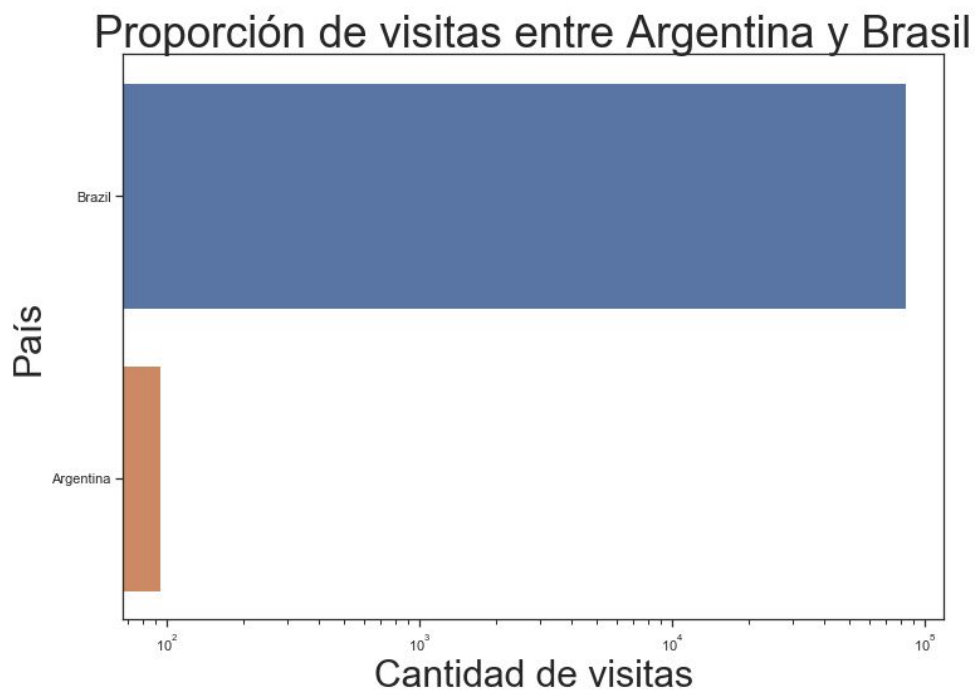




Principalmente se está accediendo al sitio desde Brasil, seguido de Estados Unidos un orden de magnitud por debajo, Argentina, el otro país donde opera Trocafone, aparece en cuarto lugar, dos órdenes de magnitud por debajo de Brasil.

## 7.2 - ¿Cuál es la relación de visitas entre Brasil y Argentina?

Dado que la empresa opera en estos dos países, rehacemos el gráfico anterior pero solo con esos datos:

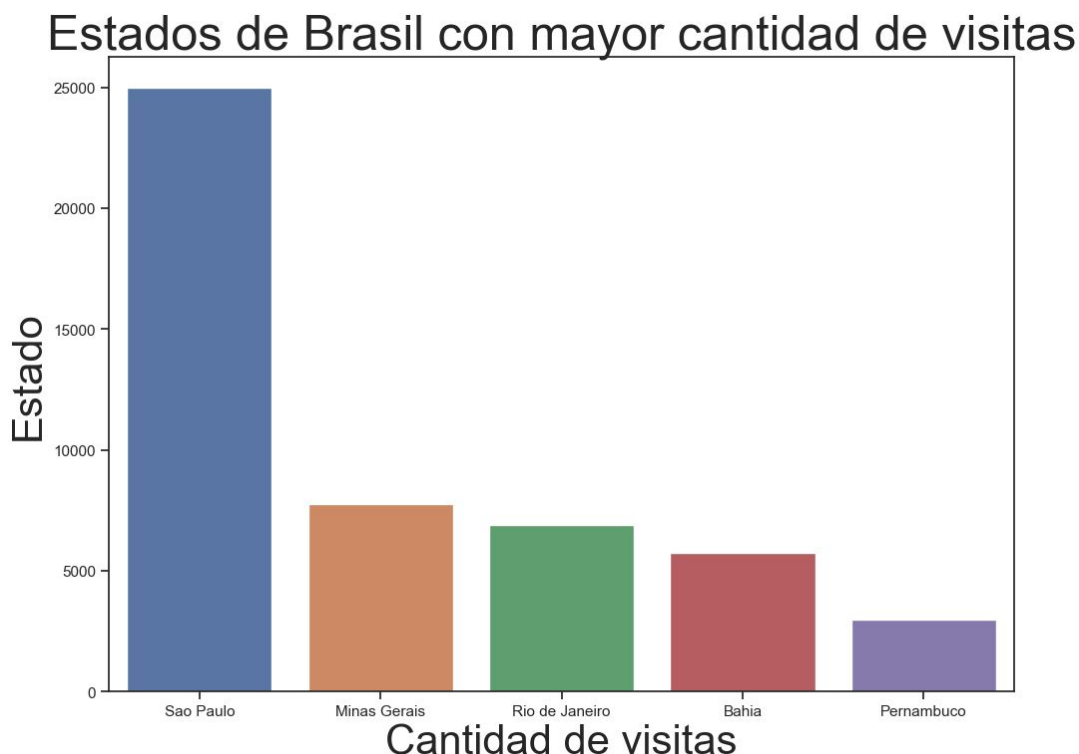


Obtenemos como resultado que Brasil representa el 99,9% del total de tráfico y Argentina tan solo el 0,1%.

Se debe tener en cuenta, a la hora de analizar estos resultados, el tamaño en cuanto a cantidad de población de los países: sólo el estado de San Pablo cuenta con una población de 41 millones de habitantes (dato censo IBGE 2010), que es aproximadamente 1 millón más que el total de habitantes de Argentina según el censo del INDEC del año 2010. En total, la IBGE estima que en el 2014 Brasil contaba con unos 200 millones de habitantes, mientras que en Argentina el INDEC estimó para el mismo año unos 42 millones.

### 7.3 - ¿Para Brasil y Argentina, cuáles son las provincias o estados más populares?

Con el objetivo de entender cómo se reparte geográficamente la información presentada en el punto anterior, para cada uno de estos países estudiaremos los estados (en el caso de Brasil) y las provincias (en el caso de Argentina) desde donde se originan las visitas al sitio:



En el caso de Brasil, el estado de San Pablo es el que reúne la mayor cantidad de visitas, con una cantidad aproximada de 25.000 eventos. Por debajo aparecen Bahia y Rio

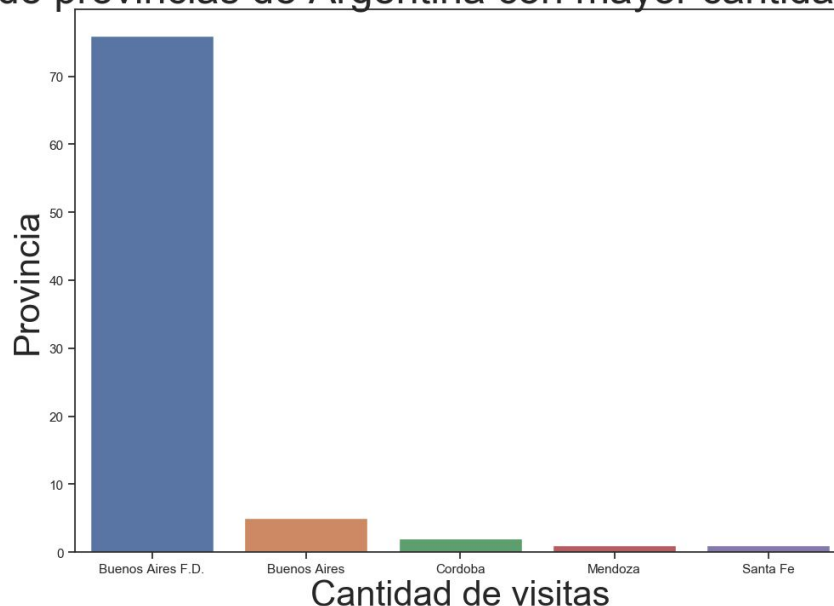
de Janeiro, ya con menos de 10.000. Si se analizan estos resultados, y se comparan por ejemplo con la cantidad de habitantes de estas regiones, la lista prácticamente coincide con dicho ranking (datos censo IBGE 2010):

- San Pablo (22% de la población total)
- Minas Gerais (10% de la población total)
- Rio de Janeiro (8% de la población total)
- Bahia (7% de la población total)
- Pernambuco (5% de la población total)

Asimismo, entre Minas Gerais, Rio de Janeiro y Bahía está concentrado el 25% de la población, apenas por encima de la cantidad del estado de San Pablo. Los tres primeros de la lista, además, son los que tienen mayor producto bruto (y el 52% de todo el país). Estos datos fueron tomados del Reporte de cuentas regional, publicado por el IBGE en el año 2015:

- San Pablo (32% del total)
- Rio de Janeiro (11% del total)
- Minas Gerais (9% del total)

### Top 5 de provincias de Argentina con mayor cantidad de visitas



Para el caso de Argentina, se puede apreciar que la gran mayoría de las visitas (aproximadamente el 80%) proviene de la Ciudad Autónoma de Buenos Aires, seguida por la provincia de Buenos Aires y Córdoba, las tres áreas con mayor población y concentración de riqueza del país.

Si bien la provincia de Buenos Aires es la que concentra la mayor cantidad de habitantes, hay que considerar que la zona de mayor población dentro de la provincia es el Gran Buenos Aires, que junto con la Ciudad Autónoma de Buenos Aires concentran aproximadamente el 46% de la población del país (datos del censo del INDEC de 2010).

Si repetimos el estudio utilizando las ciudades de ambos países, en vez de las regiones, obtenemos resultados análogos.

## 8 - Conclusiones

A lo largo de este trabajo, hemos estudiado diversos aspectos del set de datos proporcionado por Trocafone. Se han analizado las visitas al sitio y su relación con las campañas publicitarias realizadas por la empresa, los horarios en que los usuarios son más activos, las marcas de dispositivos más populares, los modelos más populares, los dispositivos utilizados para visitar la web de la empresa (y sus sistemas operativos) y por último la distribución geográfica de los usuarios de la página.

En el período analizado, de Enero a Junio de 2018, vemos que en los primeros meses la cantidad de visitas es baja, registrándose un aumento significativo a partir de Abril. Estudiando las visitas por campaña encontramos una relación entre la cantidad de visitas y la realización de campañas publicitarias. No solo las visitas por campaña representan la principal fuente de visitas al sitio, sino que la cantidad de visitas están fuertemente atadas a la realización de dichas campañas.

Respecto a los horarios en los que se efectivizan la mayor cantidad de ventas, la conclusión surgida del análisis es que ocurren en los días de semana, por las noches, luego del horario laboral. Sobre el estado de los aparatos que fueron comprados, hallamos que los que se encuentran en estado 'bueno' son los más populares. Con esto podemos proponer que haya mayor soporte dentro de estos horarios que son los de mayor flujo, como así también entendemos que las personas que ingresan al sitio se enfocan en dispositivos buenos y no necesariamente nuevos o un estado excelente.

Respecto a los equipos más utilizados para consultar la plataforma de Trocafone, encontramos que los sistemas Android y Windows son los más utilizados. Muy por debajo, se encuentran iOS y MacOS respectivamente. El navegador más utilizado para navegar el sitio, es por amplio margen, Google Chrome. Por otro lado tanto las PC como los smartphones son utilizados en prácticamente igual proporción para visitar el sitio. Las tablets quedan relegadas muy por detrás, en tercer lugar.

Las marcas más populares de productos comprados son Samsung, Apple y Motorola respectivamente. Encontramos asimismo que para las dos marcas más populares, los

usuarios buscan equipos en estado bueno, muy bueno y excelente, mientras que para el caso de Samsung también hay una proporción que busca aparatos nuevos. Los usuarios que buscan equipos con una capacidad de almacenamiento alta tratan de caer en aquellos que tienen un mejor estado, por ejemplo 'Excelente'. Para capacidades más bajas, encontramos que el resultado es más variable.

Sobre la distribución geográfica de las visitas al sitio, observamos que principalmente se está accediendo desde Brasil. Argentina, el otro país donde opera Trocafone, aparece en cuarto lugar, dos órdenes de magnitud por debajo de Brasil. De ambos países de interés, obtuvimos que el 99,99% de tráfico corresponde a Brasil. Un análisis de regiones y ciudades arroja en ambos casos que los lugares desde donde se registra la mayor cantidad de casos corresponde a las zonas de mayor población e importancia económica. Por otro lado entendemos que de momento en Argentina el flujo es poco representativo y su estudio al menos con estos datos no arroja información de relevancia.

A modo de cierre, podemos decir que el análisis realizado nos ha permitido entender el comportamiento de los usuarios del portal de Trocafone, sus gustos y preferencias. De esta manera, el resultado de un análisis de este tipo puede utilizarse para mejorar la oferta de productos, las prestaciones o características del sitio web, o bien detectar zonas donde la compañía podría expandir su operación, entre otras aplicaciones.

Asimismo, nos gustaría agregar que el estudio de un set de datos de estas características, que puede obtenerse en forma muy sencilla realizando *logs* de actividad en un sitio web, resulta una herramienta poderosísima a la hora de tomar decisiones estratégicas para la compañía, en una amplia variedad de campos, a fin de administrar de la mejor forma posible los recursos disponibles.