



**UNIVERSIDAD DE BUENOS AIRES**

**FACULTAD DE INGENIERÍA**

**Organización de Datos (75.06)**

**TRABAJO PRÁCTICO Nro. 1**

**ANÁLISIS EXPLORATORIO**

<b>GRUPO 11</b>	
Ituarte, Lautaro	93639
Simizu, Nicolás	94556
Aguinaco, Carlos	80969
Castellanos, César	81404

Segundo Cuatrimestre 2018



## Índice

<b>Resumen</b>	<b>4</b>
<b>2. Análisis preliminar del set de datos</b>	<b>5</b>
2.1 Análisis del rango de fechas	5
<b>3. Análisis del tráfico del sitio</b>	<b>6</b>
3.1 Cantidad de visitas	6
3.1.1 Cantidad de visitas por mes	6
3.1.2 Cantidad de visitas por día y hora	8
3.2 Origen del tráfico	8
3.2.1 Países que acceden al sitio	9
3.2.2 Regiones con más actividad	10
3.2.3 Personas más activas	11
3.3 Información de Navegación	11
3.3.1 Sistemas operativos más utilizados	12
3.3.2 Resoluciones de pantalla más utilizadas	13
<b>4. Análisis de Popularidad</b>	<b>14</b>
4.1 Popularidad de marcas en eventos	14
4.2 Análisis de palabras	15
4.2.2 Modelos más listados en base a palabras buscadas	15
4.2.3 Marcas más listadas en base a búsquedas por marca (eventos de brand listing)	17
4.3 Modelos con mayor cantidad de solicitudes de stock	18
4.4 Productos más deseados	18
<b>5. Análisis de ventas</b>	<b>20</b>
5.1 Calidad de productos vendidos	20
5.2 Arrepentidos	21

5.3 Análisis de ventas y checkouts	23
5.3.1 Checkouts por usuario	23
5.3.2 Marcas con más checkout	24
5.3.3 Capacidades(Storages) más vendidas	25
5.3.4 Checkouts y Ventas por día por marca	26
5.3.4 Ventas por marca y modelo	28
5.3.5 Ventas vs Checkouts	29
<b>6. Análisis de campañas</b>	<b>30</b>
6.1 Nuevos usuarios por campañas	30
<b>Conclusiones</b>	<b>31</b>

### **Resumen**

El presente trabajo muestra el Análisis Exploratorio del set de datos propuesto por la cátedra perteneciente a la empresa Trocafone para el cual se analizaron con distintas herramientas los datos de las eventos que ofrece Analytics relacionadas con usuarios que accedieron a la plataforma.

El trabajo se puede consultar en el sitio <https://github.com/lituarte/Datos-TP1>

## **2. Análisis preliminar del set de datos**

En esta sección vamos a incluir todos los análisis que consideramos debemos realizar antes que cualquier otro, ya que de no ser tenidos en cuenta pueden causar análisis con resultados contaminados o erróneos.

### **2.1 Análisis del rango de fechas**

¿Entre qué fechas son válidos nuestros análisis?

Encontramos que el set de datos se encuentra comprendido entre el 1 de enero de 2018 y el 15 de junio de 2018, lo que significa que los datos que tenemos de Junio son parciales y esto lo debemos tener especialmente en cuenta cuando hagamos análisis por meses.

### **3. Análisis del tráfico del sitio**

Los objetivos de esta sección son varios, primero que nada queremos darnos una idea de la magnitud de usuarios que utilizan el sitio. Además esperamos que estos análisis nos cuenten un poco la historia del sitio, cómo va cambiando a medida pasa el tiempo y de dónde provienen sus clientes.

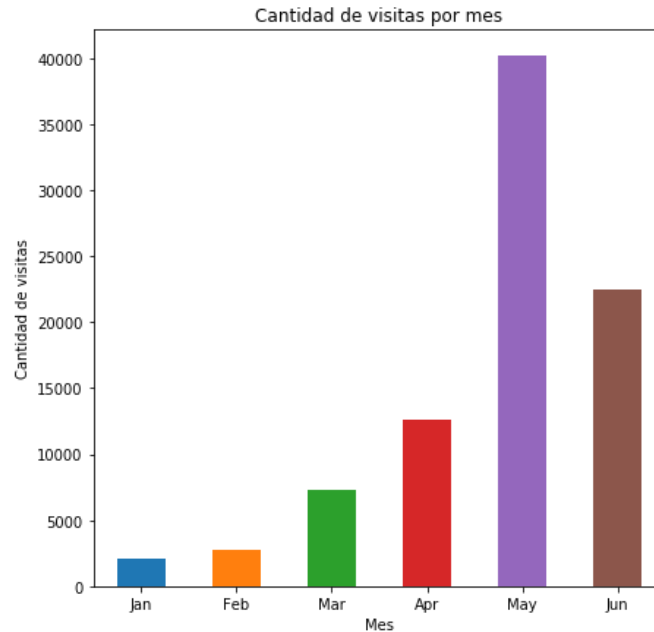
#### **3.1 Cantidad de visitas**

##### **3.1.1 Cantidad de visitas por mes**

Este primer análisis busca resolver dos simples preguntas: ¿Con qué volumen de usuarios estamos tratando? y ¿Cómo varía el volumen de usuarios a lo largo del tiempo?

Es muy diferente tratar con diez usuarios a tratar con mil o cien mil, no es lo mismo a nivel negocio y no es lo mismo a nivel infraestructura tampoco. Estas cantidades también pueden servir a Trocafone para saber dónde están parados en relación a dónde quieren estar.

También es importante saber cómo varían estos números con el tiempo, que disminuyan puede significar un indicio de fracaso para Trocafone mientras que si aumentan demasiado rápido puede encender las alarmas si no tiene una estructura que soporte el crecimiento.



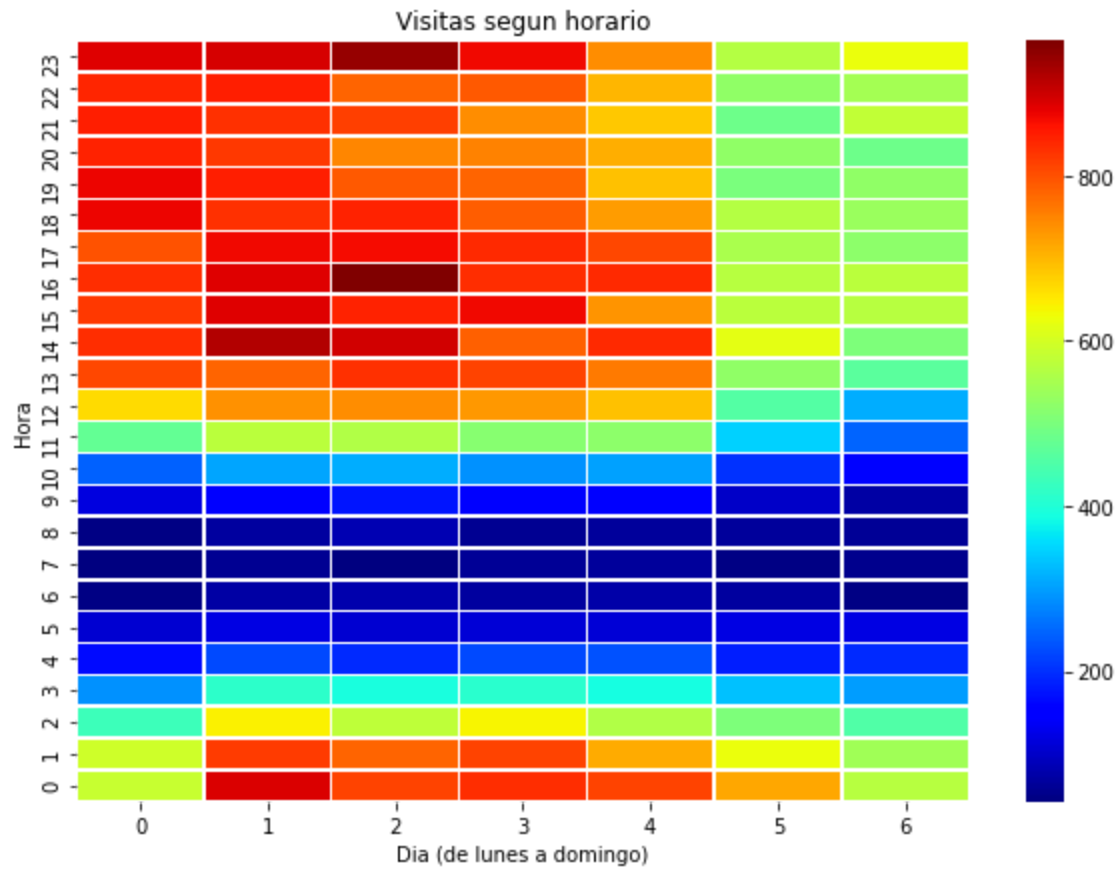
***Aclaración importante: los datos de junio son parciales (Ver 2.1)***

Podemos ver en el gráfico que Trocafone ha tenido un crecimiento sostenido de visitas al sitio en estos últimos meses, si bien los datos de junio son parciales (solo tenemos hasta mitad de mes) de continuar la tendencia tendríamos un mes similar a mayo en cantidad de visitas.

Cómo se mencionó anteriormente, si Trocafone no estaba preparado para el aumento de usuarios entonces puede haber significado malas noticias para el sitio pero en caso contrario pueden estar muy felices con el crecimiento.



### 3.1.2 Cantidad de visitas por día y hora

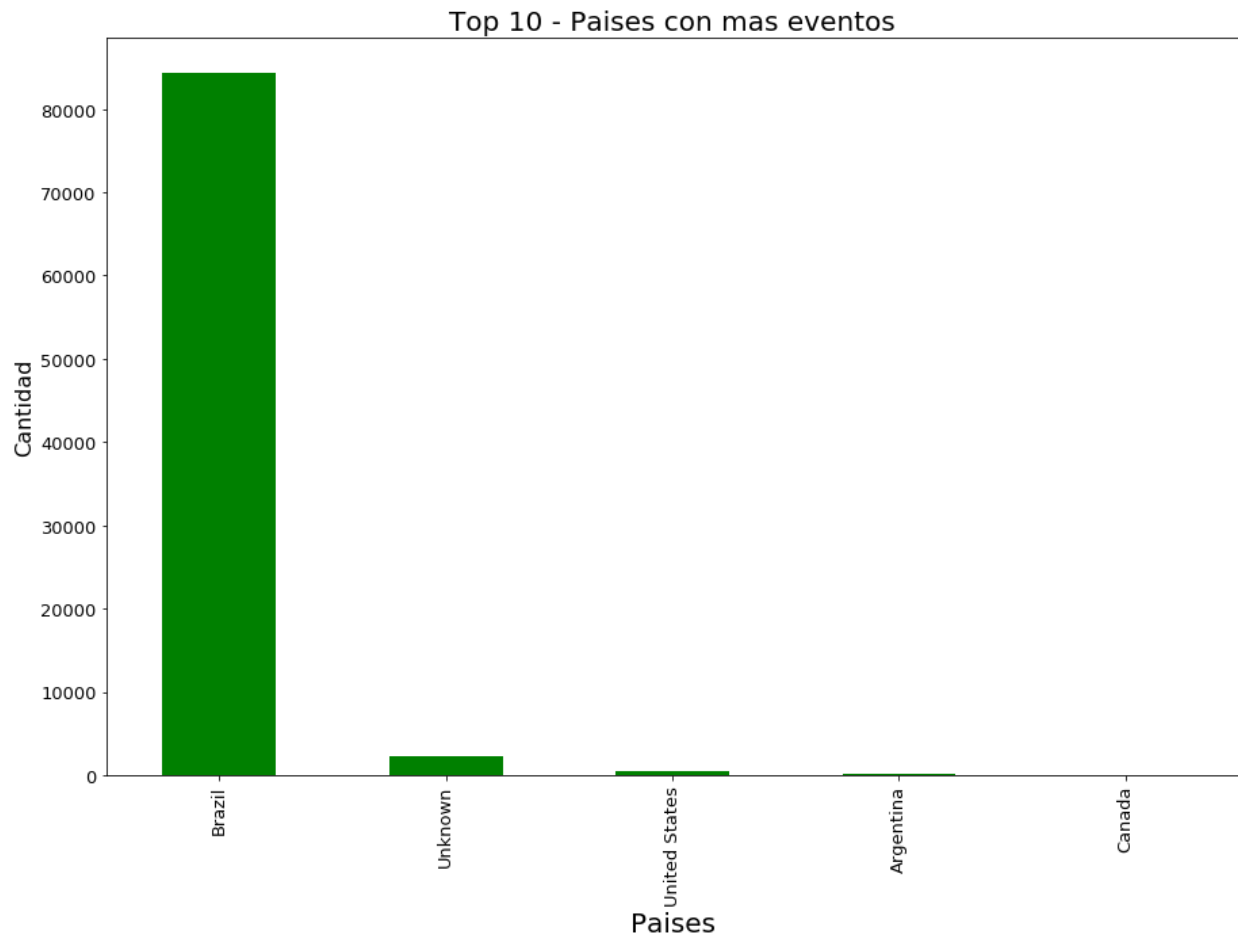


Con este gráfico podemos apreciar que la mayor cantidad de visitas que recibe el sitio se da entre el mediodía y la medianoche de los días de semana. También podemos observar un decremento bastante importante los fines de semana. Trocafone podría, por ejemplo, poner algún tipo de promoción el fin de semana para incrementar las visitas o tener menos gente trabajando esos días.

### 3.2 Origen del tráfico

¿De dónde provienen los usuarios? Para cualquier negocio es importante conocer sus clientes y de dónde provienen en este caso nos ayuda a conocerlos mejor a ellos y a nuestro mercado.

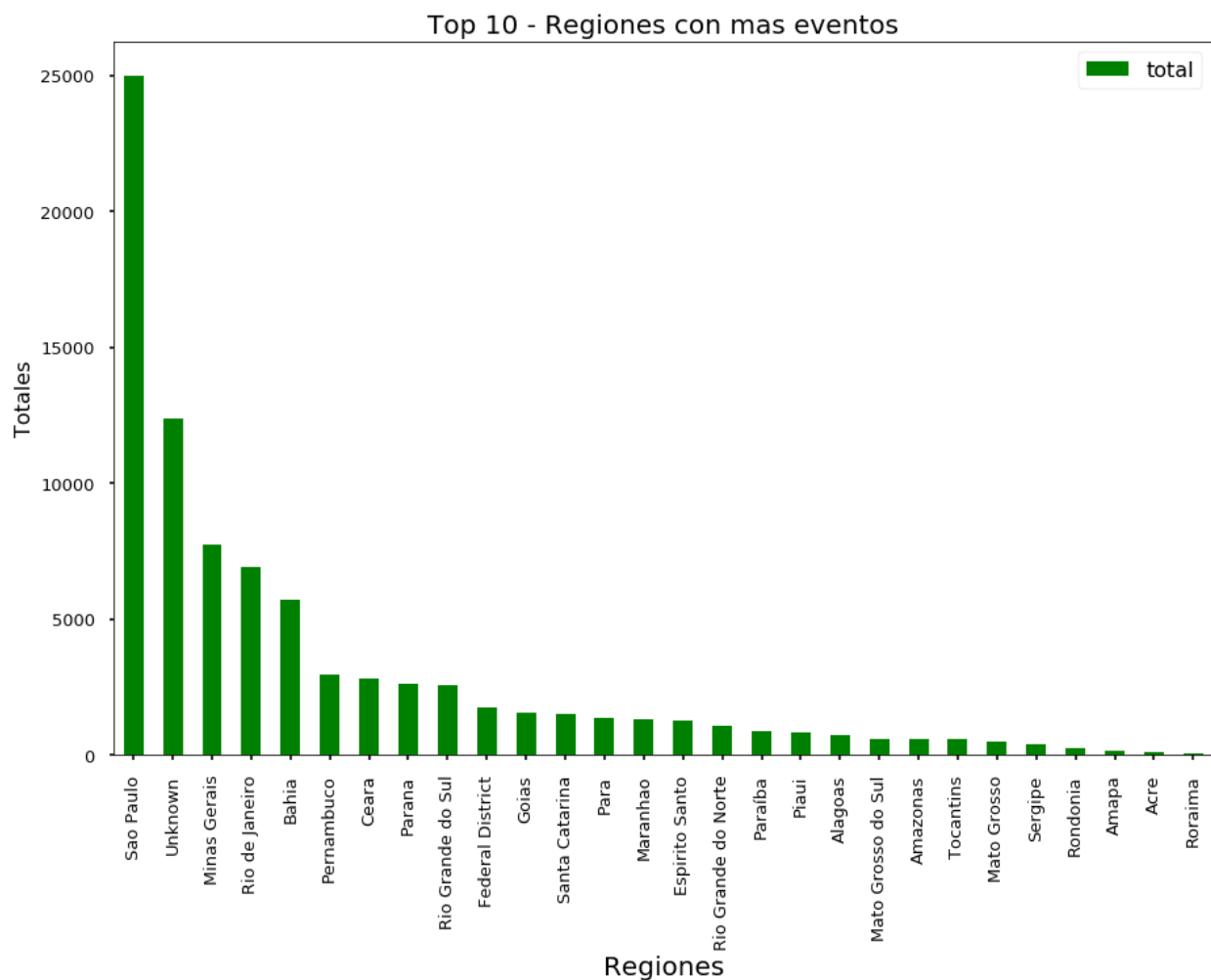
### 3.2.1 Países que acceden al sitio



Vemos que la fuente de datos es principalmente Brasil, de donde surge la principal operatoria de la empresa. Hay también un porcentaje pequeño que no se conoce el origen. Es muy amplia la diferencia a favor de Brasil por lo tanto no se elimina ya que es un análisis inicial. Luego siguen (y muy por debajo) Estados Unidos, Argentina y Canadá.

Estos resultados son muy importantes para la empresa, puede ayudarlos a tomar decisiones de alto nivel. Sabiendo que Brasil es un porcentaje tan alto de su mercado probablemente nunca tomen una decisión que perjudique dicho sector, mientras que la baja cantidad de visitas desde Argentina puede significar la necesidad de alguna campaña publicitaria por ejemplo.

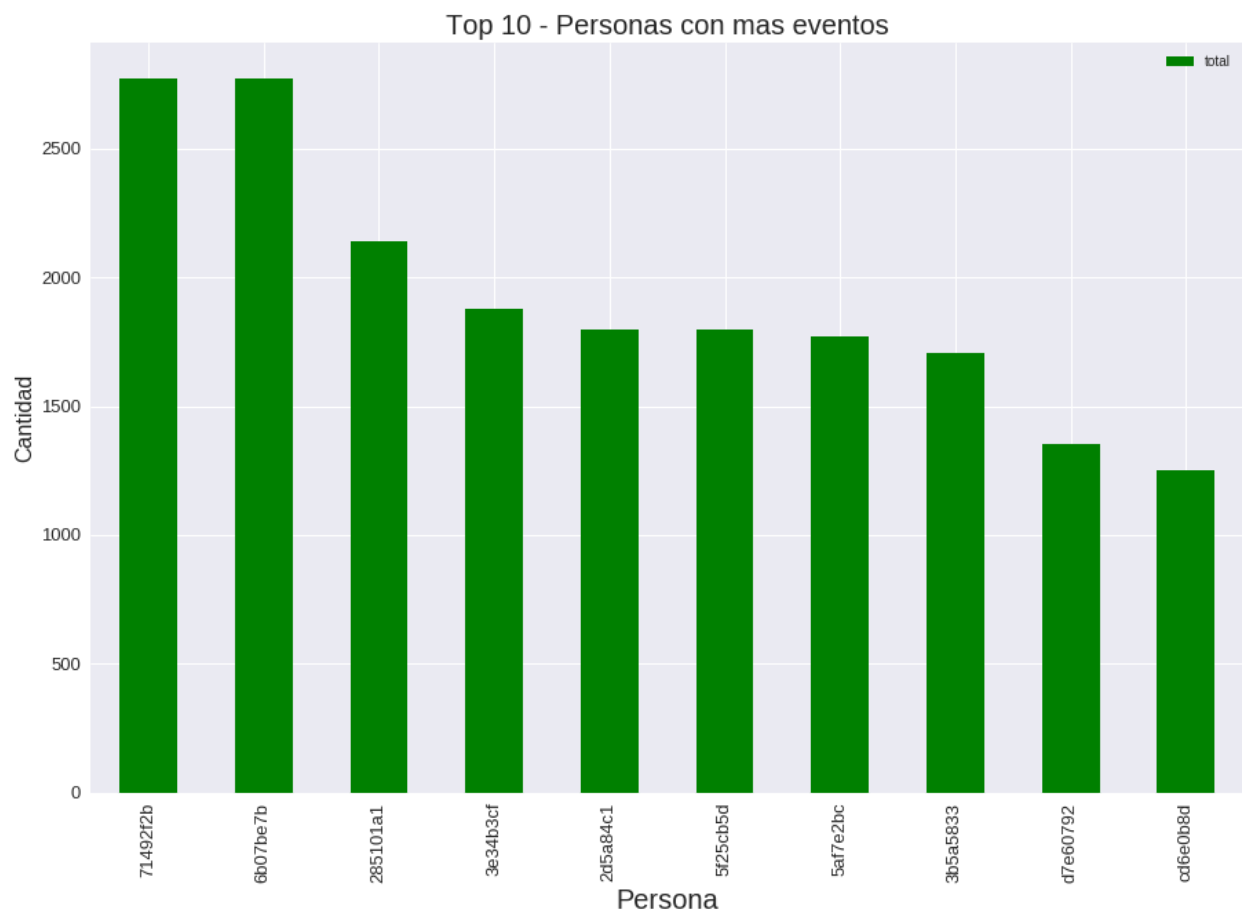
### 3.2.2 Regiones con más actividad



Profundizamos el análisis de los países con las gráfico de regiones con más eventos de Brasil, el país más activo por amplio margen.

### 3.2.3 Personas más activas

Otra de las métricas que nos resultó interesante obtener es la de los usuarios con más eventos en el sitio, números demasiado altos podrían derivar en una investigación más detallada de dicho usuario.



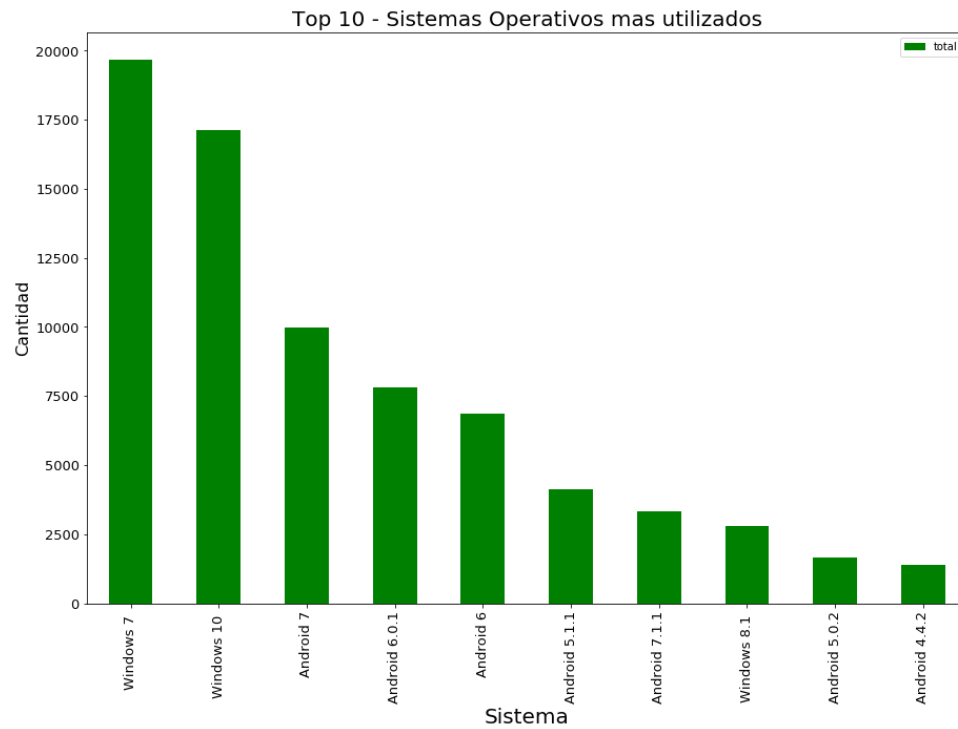
Aquí se puede ver los usuarios con mayores actividad.

### 3.3 Información de Navegación

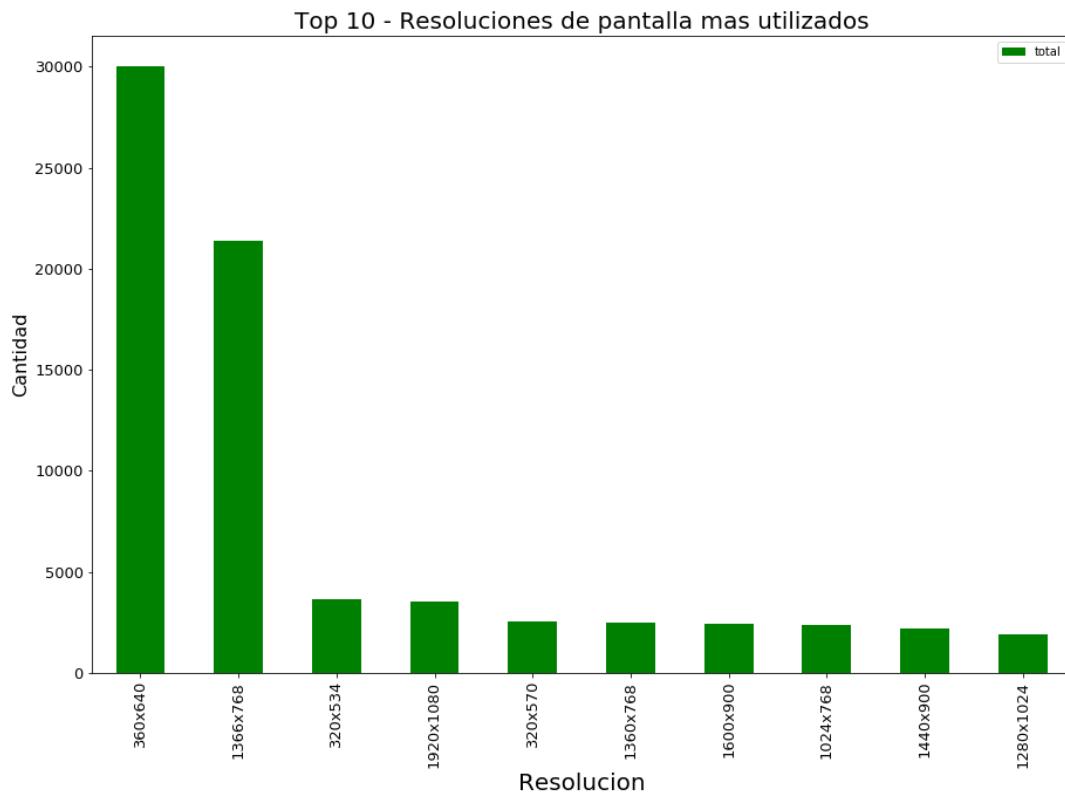
Vamos a analizar algunos datos de la navegación de las personas que visitan el sitio, en particular qué Sistemas Operativos y qué resoluciones son más utilizados, de esta forma se puede dedicar

más esfuerzo a que se revise periódicamente el correcto funcionamiento del sitio en estos dispositivos.

### 3.3.1 Sistemas operativos más utilizados



### 3.3.2 Resoluciones de pantalla más utilizadas

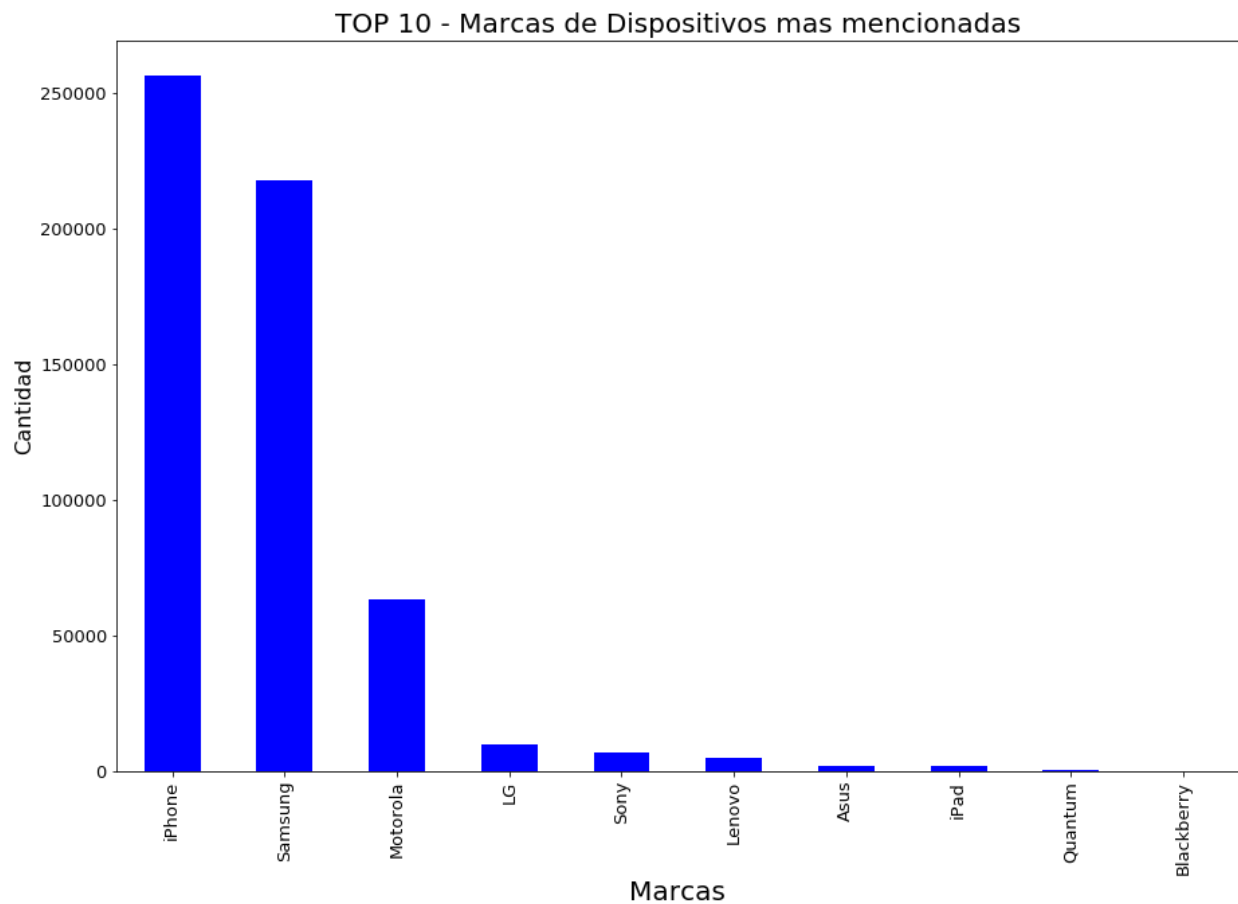


De ambos gráficos podemos ver que si bien Windows 7 y Windows 10 son los Sistemas Operativos más frecuentes, en el global, hay una gran cantidad de personas que van a acceder al sitio a través de un dispositivo móvil.

#### 4. Análisis de Popularidad

En esta sección analizamos la popularidad de distintos productos, modelos y marcas según distintos parámetros

##### 4.1 Popularidad de marcas en eventos



Es clara la importancia tanto de Apple como Samsung como líderes del sector por la cantidad de datos, seguido por Motorola. Las restantes marcas tienen una presencia menos importante.

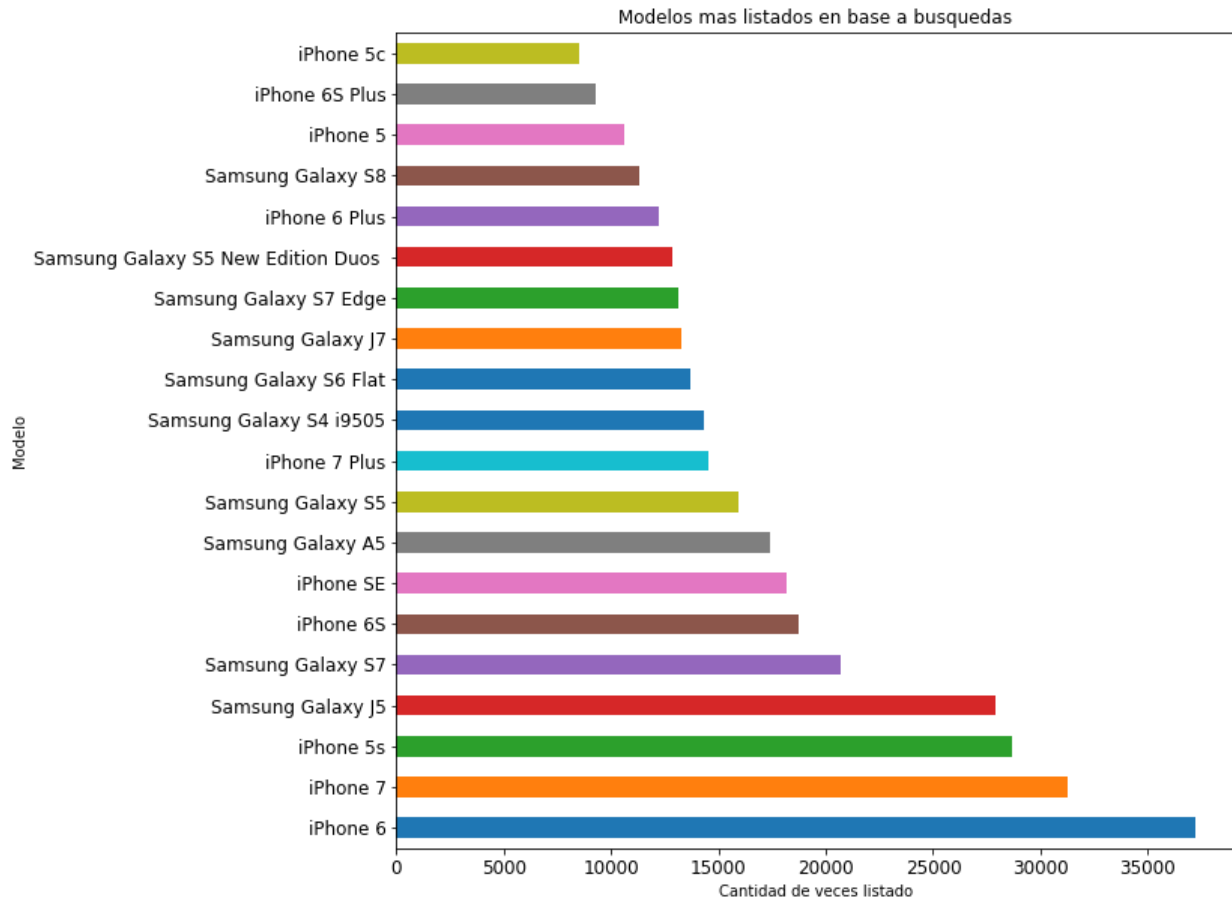
## 4.2 Análisis de palabras



Claramente se ve la popularidad de las palabras “iPhone” y “Samsung” como más resonantes dentro del conjunto, lo cual confirma la gráfica anterior.

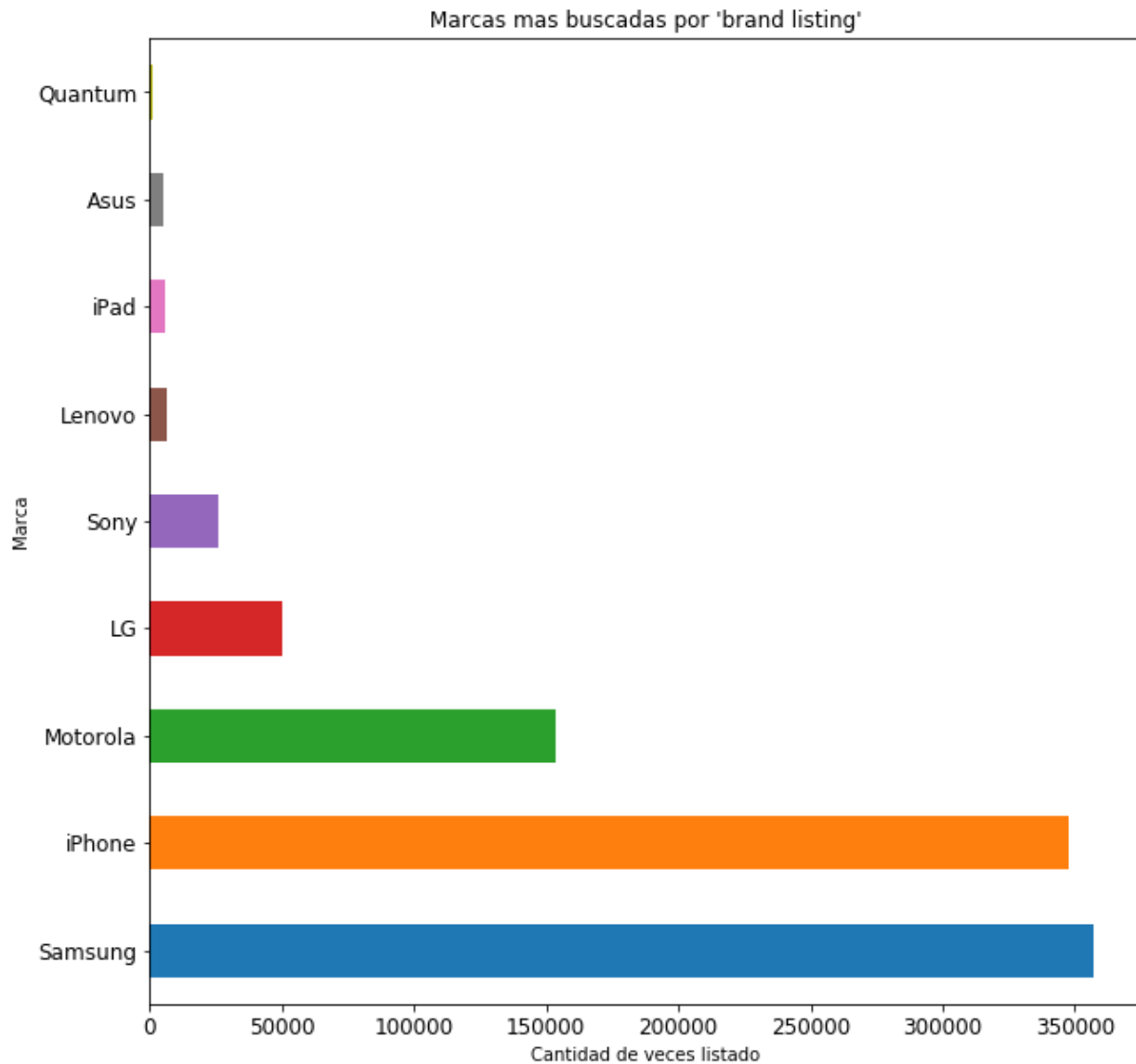
### 4.2.2 Modelos más listados en base a palabras buscadas





Se puede ver que los dispositivos más listados según las búsquedas que realizan los usuarios son los iPhone en sus modelos 6, 7 y 5S y siguen los modelos S7 y J7 de Samsung.

#### 4.2.3 Marcas más listadas en base a búsquedas por marca (eventos de brand listing)



Se puede ver que los productos Samsung son los más listados, si bien en la imagen previa se veía que los Iphone eran los más listados de forma individual. Esto según entendemos es porque Samsung posee más alternativas de modelos que Apple.

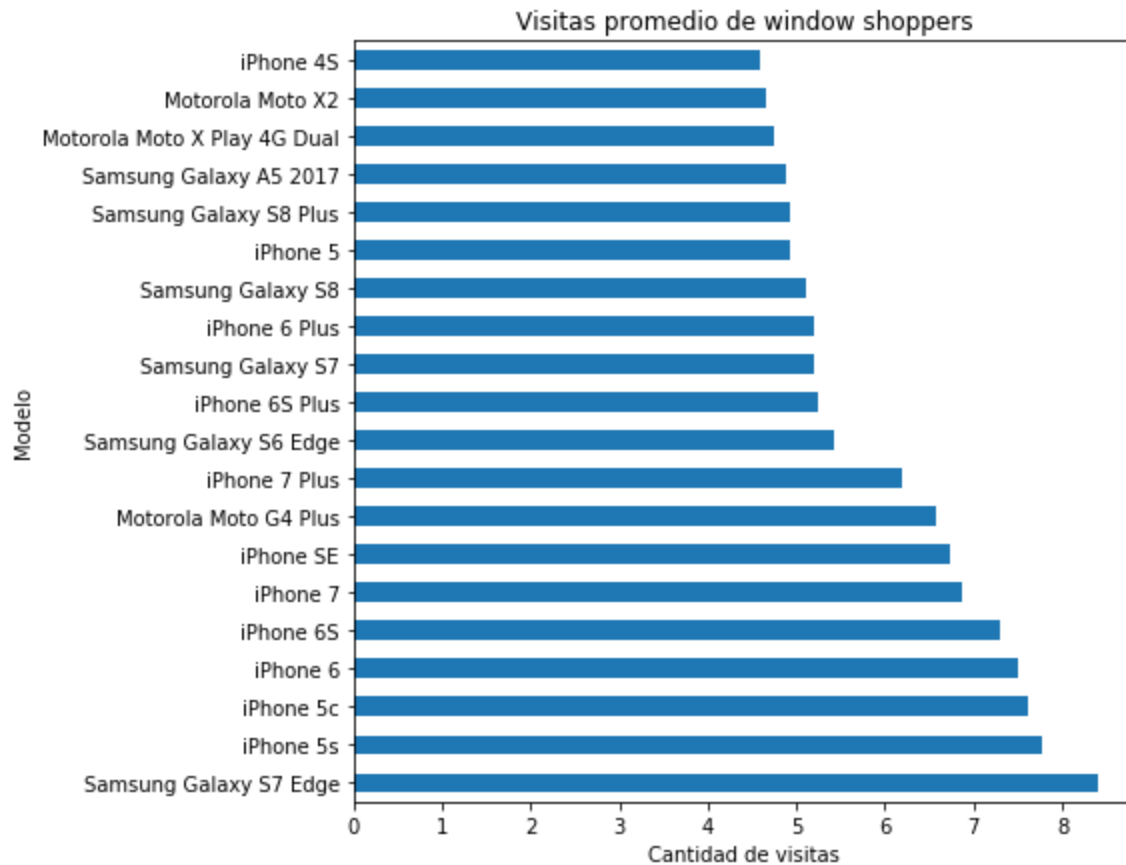
### **4.3 Modelos con mayor cantidad de solicitudes de stock**

Hicimos un breve análisis de qué modelos tuvieron más solicitudes de stock por mes, en particular nos interesaba saber el más solicitado en junio (último mes cargado), el cual tiene 9 pedidos de stock para el “I phone X”. Este dato puede servir a Trocafone para restablecer el stock lo más rápidamente posible.

### **4.4 Productos más deseados**

¿Qué modelos de celular son los que más desea la gente? Esa es la pregunta que intentamos responder en esta sección, para eso definimos el “desear” como la cantidad de veces que un usuario entra a ver el producto sin comprarlo.

Para empezar buscamos los usuarios que más veces hayan entrado a ver un modelo de celular sin comprarlo (en primer lugar el usuario “c501ff3a” entró 812 veces a ver productos con el modelo “iPhone 5c”), este listado puede servirle a Trocafone para, por ejemplo, otorgarle un descuento especial a esos usuarios y “empujarlos” a que por fin compren el producto que tanto quieren. Luego sacamos las vistas promedios de lo que denominamos “window shoppers”:



Lo que nos dice esta tabla es la cantidad de veces, en promedio, que entra a ver un producto un usuario que no lo compró, en inglés a estas personas se las llama "window shoppers". Es muy probable que esta gente esté a la espera del precio adecuado y Trocafone sea uno de los tantos lugares que visita para encontrar ese precio.

## 5. Análisis de ventas

### 5.1 Calidad de productos vendidos

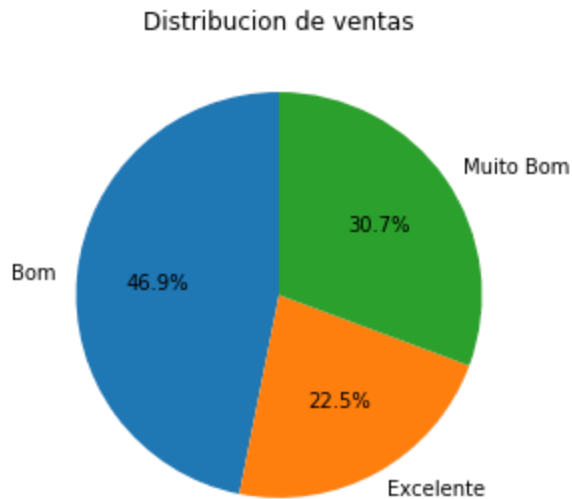
Se analizó la calidad de los productos vendidos utilizando la plataforma, los cuales primero se obtienen buscando en el set de datos a los eventos del tipo “conversion” que identifican una venta.

Dentro del conjunto de dispositivos vendidos se han encontrado 5 valores posibles: “Novo”, “Excelente”, “Muito bom”, “Bom” y “Bom - Sem Touch ID” de los cuales los valores “Novo” y “Bom -Sem Touch ID” presentaron muy pocas ventas, 2 y 22 respectivamente. A raíz de esta situación se plantearon 3 causas posibles:

1. El sitio ofrece poca cantidad de productos en esas condiciones.
2. Los productos en esas condiciones tienen muy poco éxito
3. Los productos en esas condiciones fueron cargados erróneamente.

Dado que el set de datos provisto es reducido, es posible que el conjunto solo tomara una pequeña parte de las ventas en esas categorías, descartamos las causas 1 y 2 y consideramos que los productos fueron cargados erróneamente. Por lo tanto, descartamos aquellos productos en las categorías “Novo” y “Bom - Sem touch ID” del análisis.

Para las categorías restantes, obtiene la siguiente distribución.

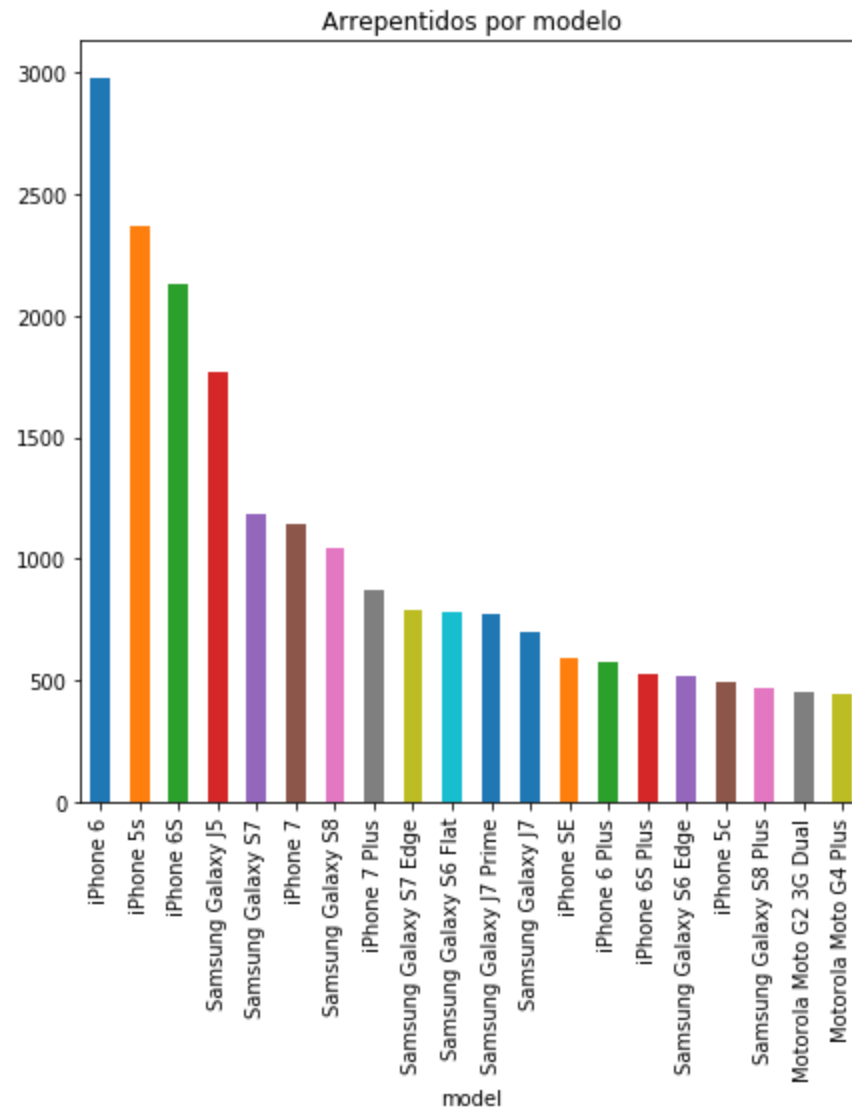


Del gráfico se ve como a mejor calidad se reducen las ventas. Una posible causa de este fenómeno es que los usuarios prefieran comprar un teléfono nuevo a comprar uno usado de calidad “Excelente”. Otra posible causa es que, al tratarse de un sitio que comercializa dispositivos usados, los usuarios estén dispuestos a buscar calidades más bajas a cambio de un mejor precio.

## 5.2 Arrepentidos

Se considera Arrepentido a los eventos del tipo ‘checkout’, que representan que el usuario llegó a ver el carro de compras del sitio, pero no concretó la compra.

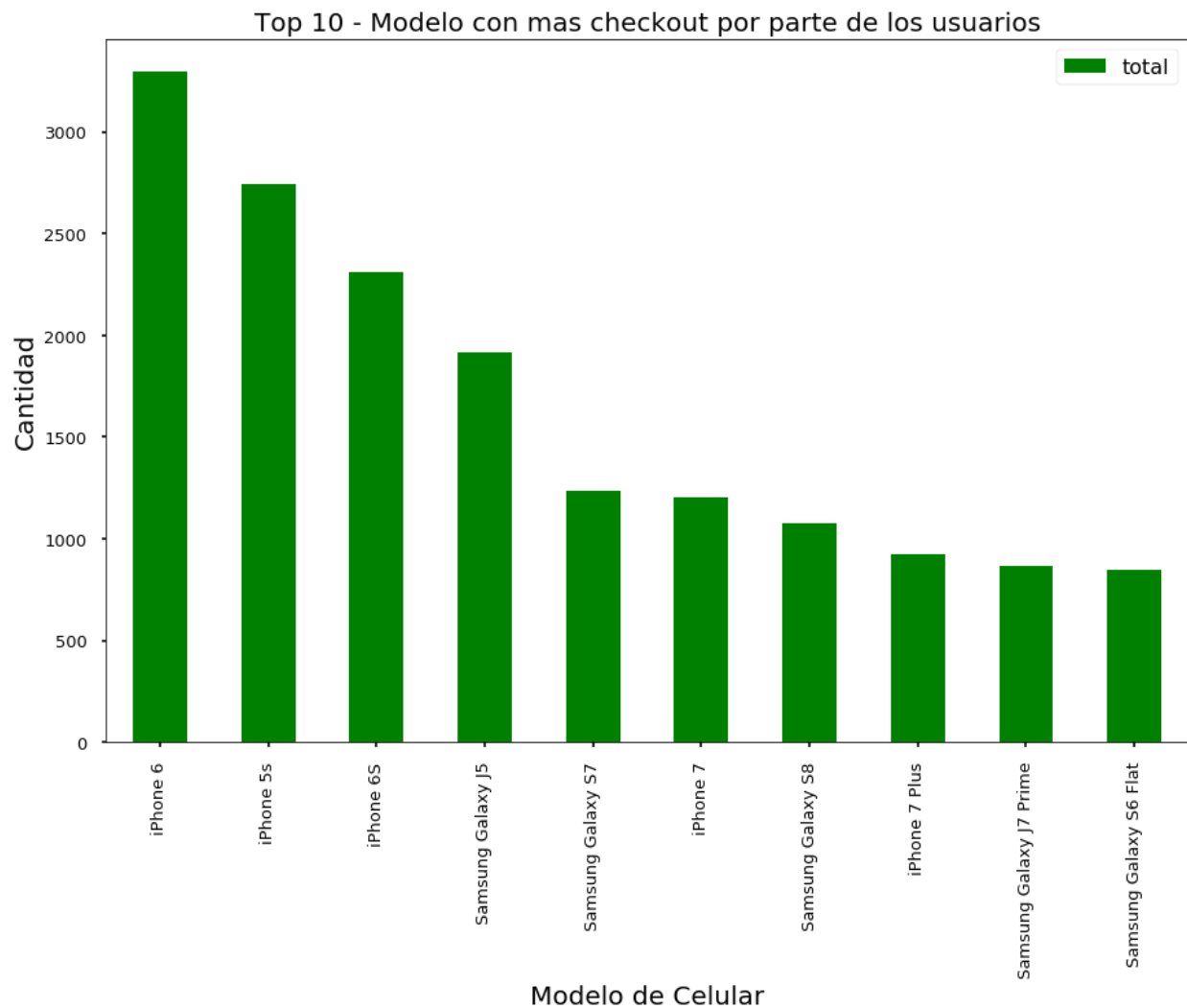
Para esto, necesitamos obtener los eventos del tipo ‘checkout’ por un lado y del tipo ‘conversion’ por otro, con el fin de poder eliminar los eventos que sí concretaron la venta del conjunto de checkout. Una vez obtenidos aquellos arrepentidos, podemos agrupar los eventos por modelo



Del gráfico podemos ver que los productos con más arrepentidos son iPhone y Samsung Galaxy.

### 5.3 Análisis de ventas y checkouts

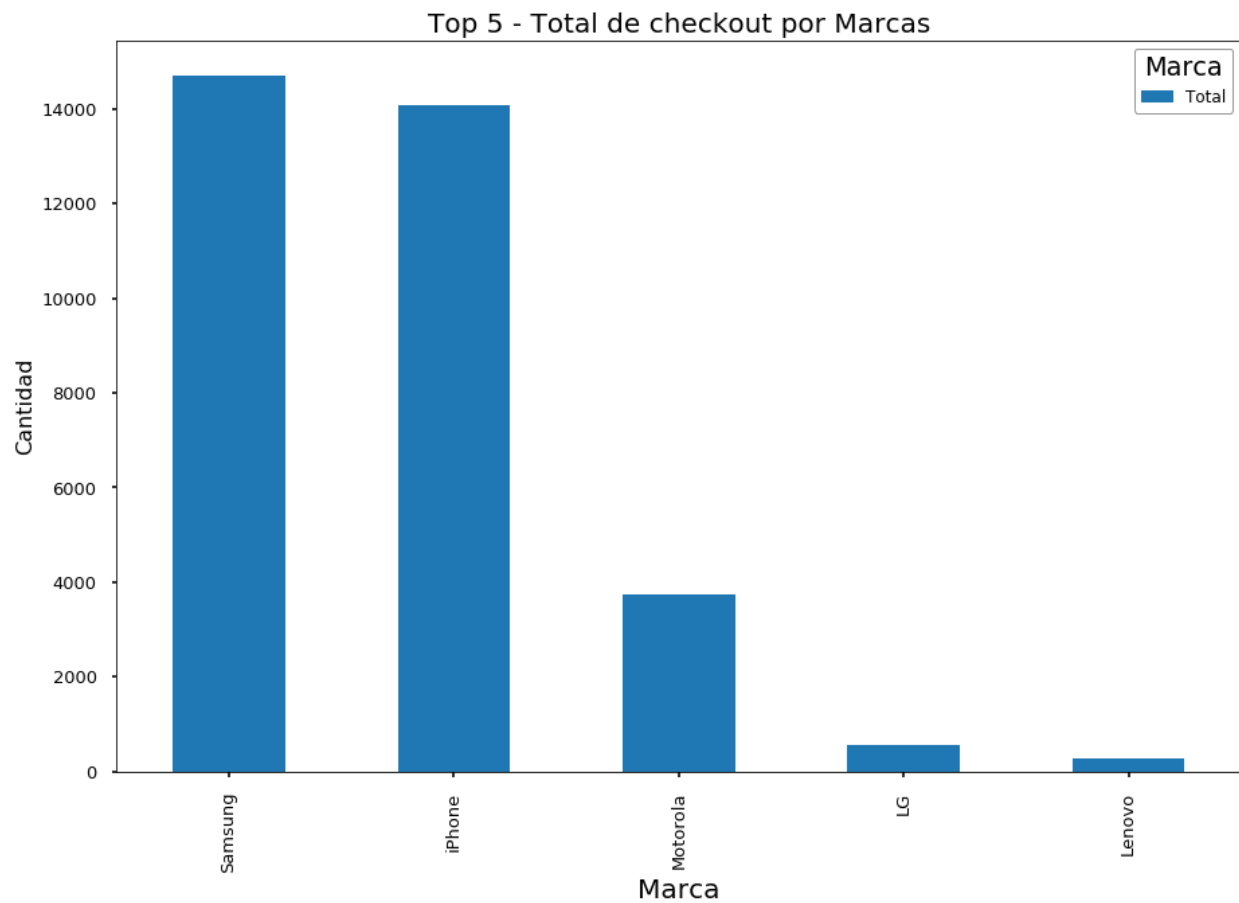
#### 5.3.1 Checkouts por usuario



Vemos que predomina en los primeros 10 tanto Apple como Samsung como marcas favoritas al momento de elegir un producto.

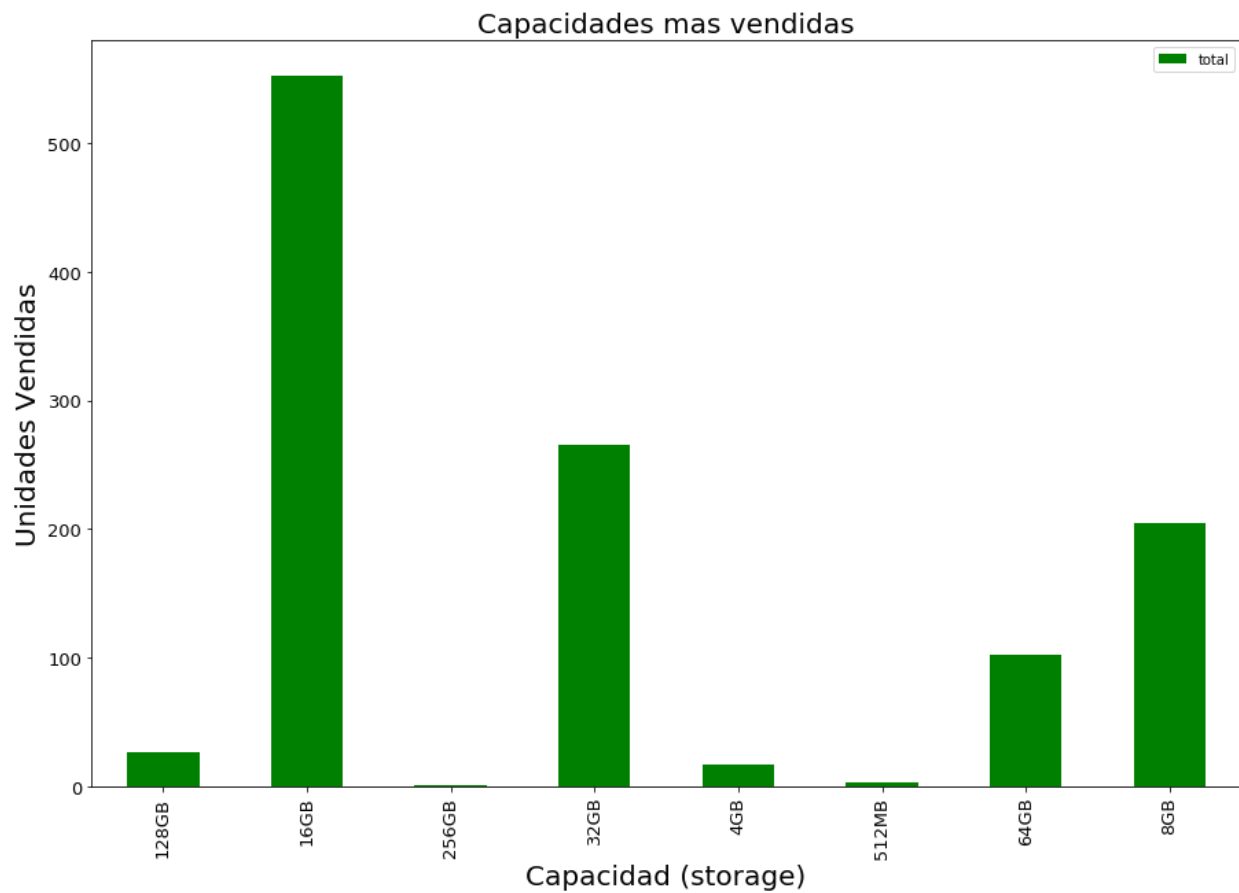


### 5.3.2 Marcas con más checkout



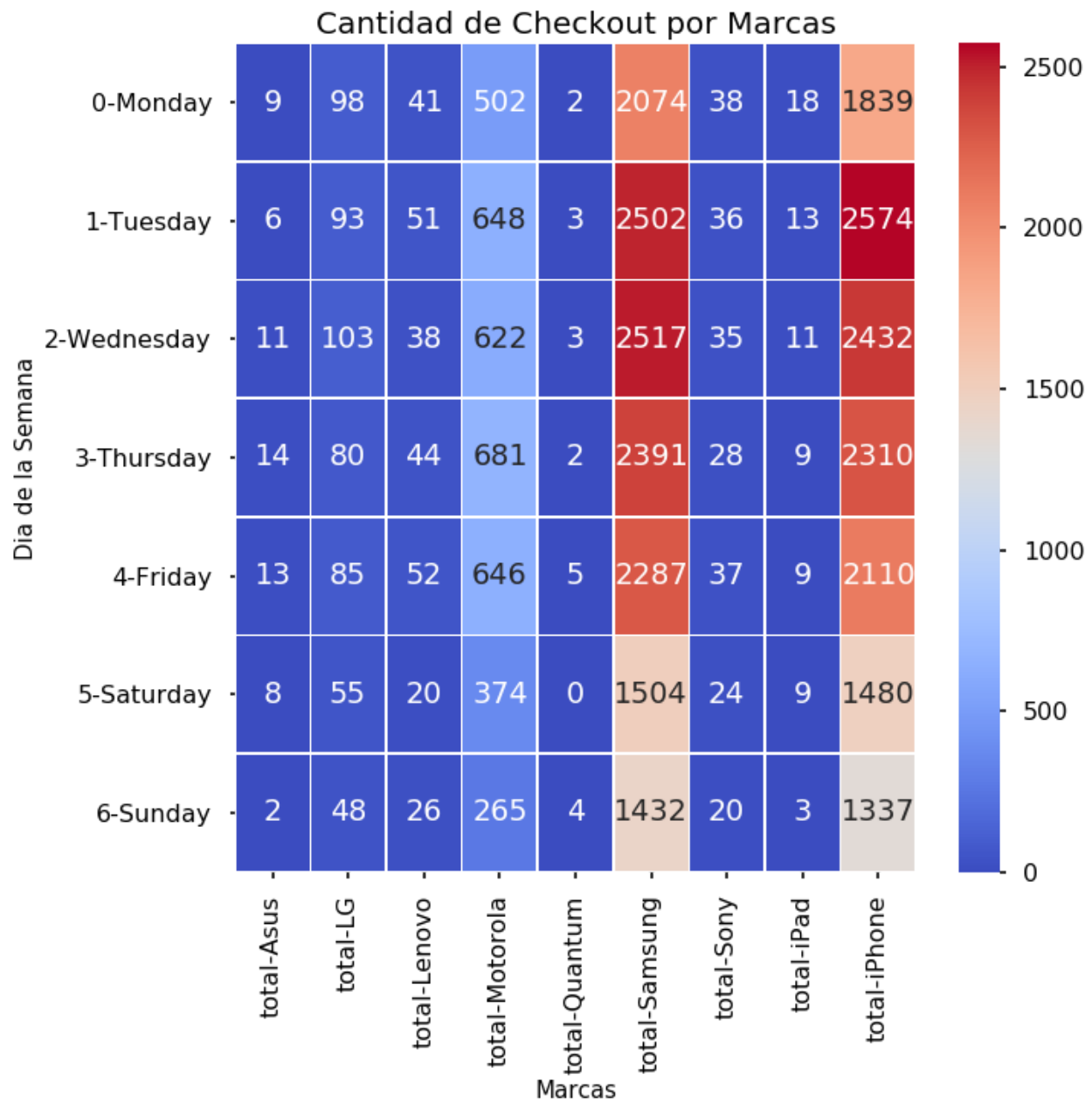
Aquí encontramos como primer insight que, a pesar de que los 3 modelos con más checkouts son de Apple, Samsung lo supera. Una de las causas puede ser que Samsung tenga mayor variedad de modelos que Apple.

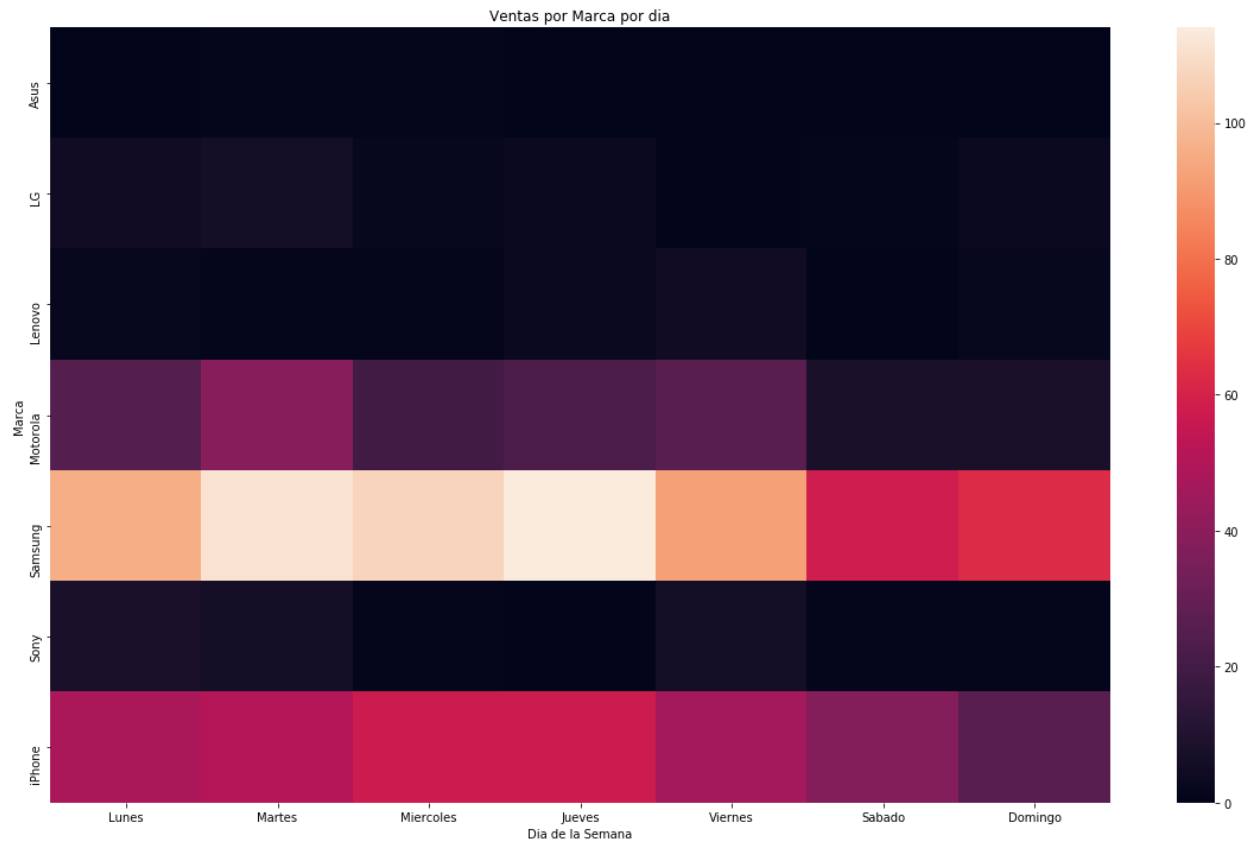
### 5.3.3 Capacidades(Storages) más vendidas



En este caso, se ve que la mayoría de los usuarios buscan una capacidad de almacenamiento promedio. Esta decisión puede estar influida la relación costo/beneficio en el momento de elegir.

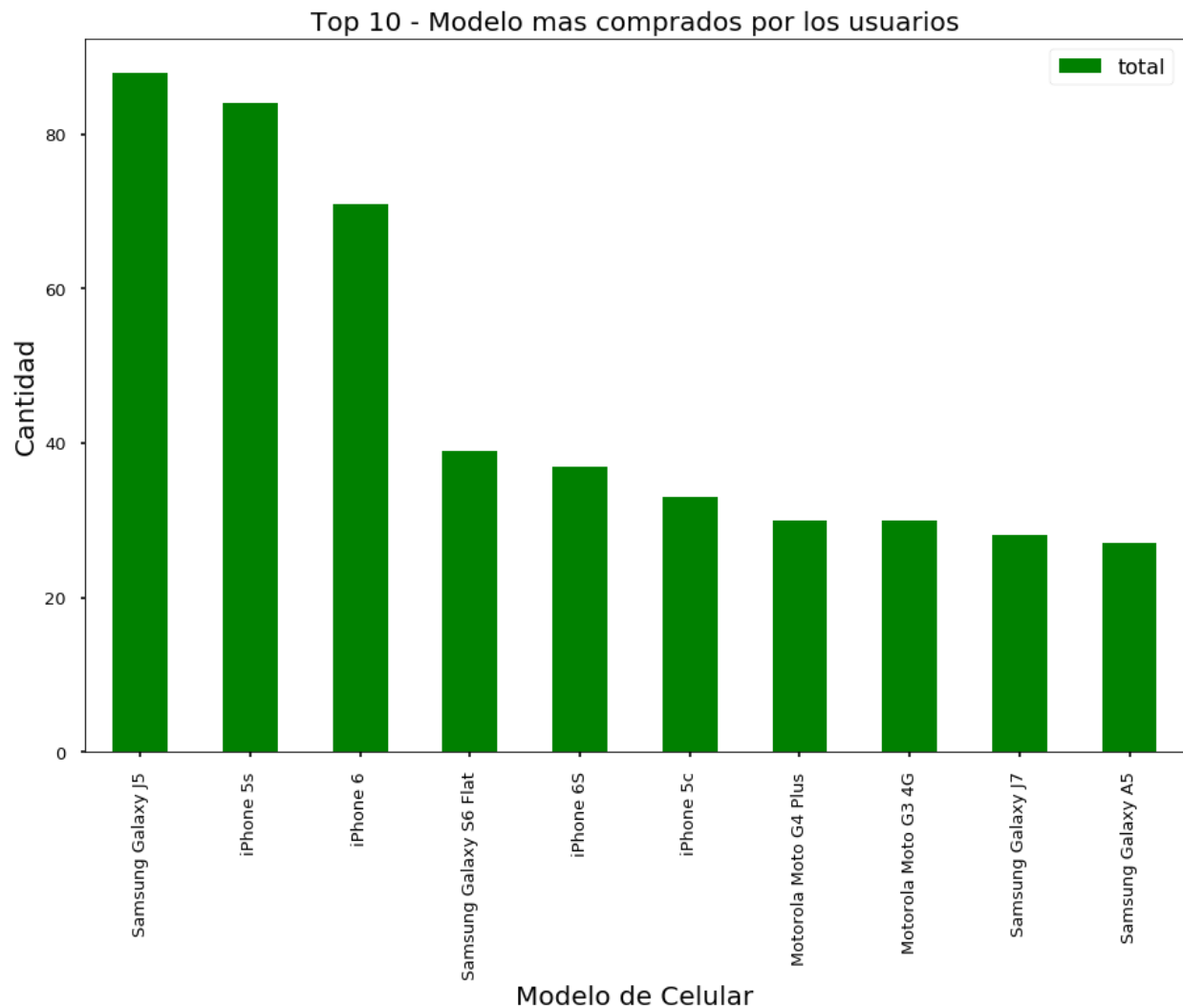
## 5.3.4 Checkouts y Ventas por día por marca





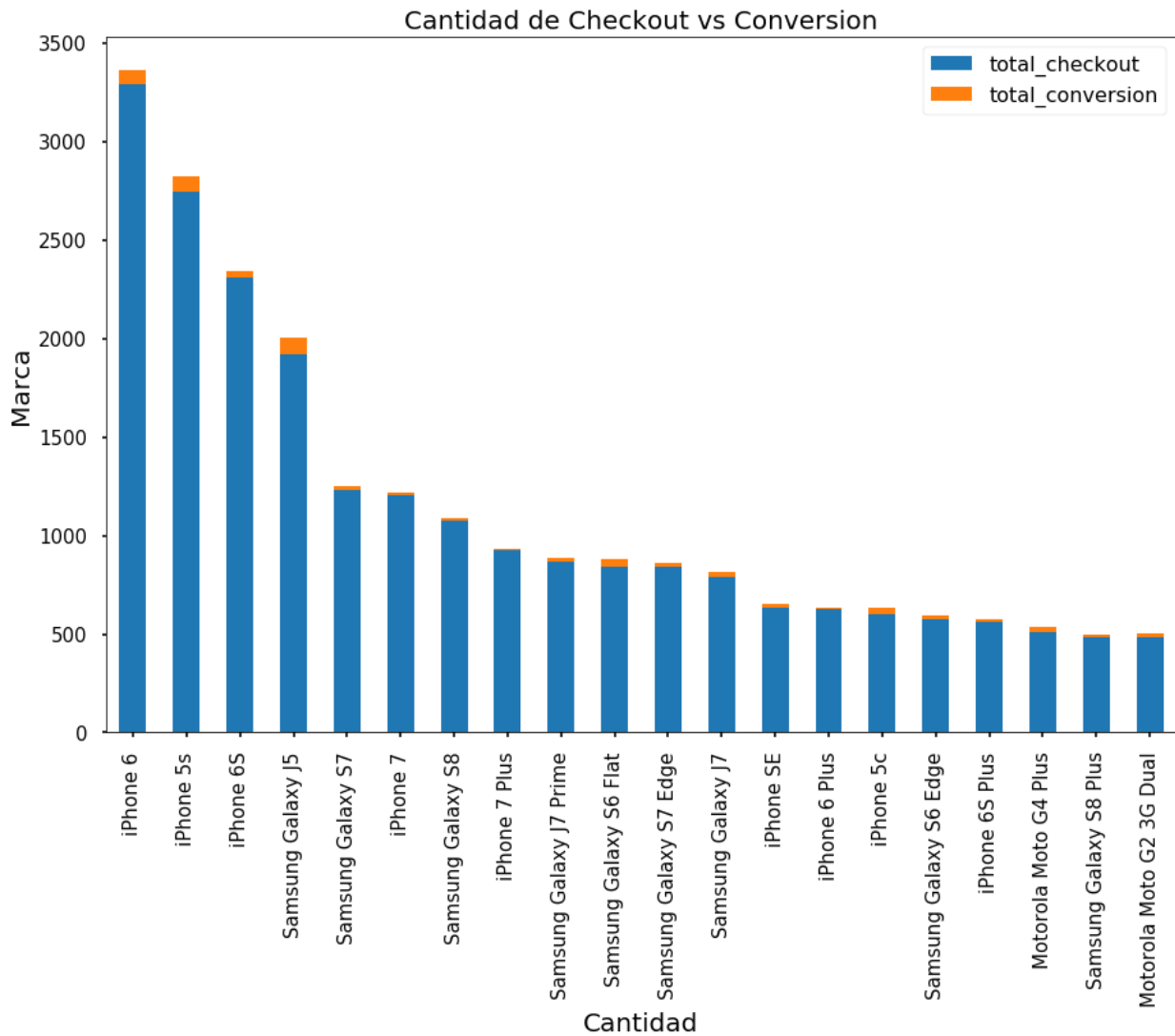
En los gráficos se observa que la mayoría de las ventas de las 3 marcas más vendidas (Samsung, Apple y Motorola) se concentran entre los días hábiles, Lunes a Viernes. Una posible causa es que las ventas registradas en el set de datos provengan de negocios, mientras que las ventas registradas durante sábado y domingo sean realizadas por particulares.

### 5.3.4 Ventas por marca y modelo



En este gráfico puede verse cuales son los modelos más vendidos por la plataforma, que está repartido entre 3 marcas: hay 4 modelos de Apple, 4 de Samsung y 2 de Motorola. Sin embargo, se destaca la diferencia entre 2 grupos: los primeros 3 modelos y los últimos 7 ya que hay un salto de aproximadamente 20 ventas.

### 5.3.5 Ventas vs Checkouts

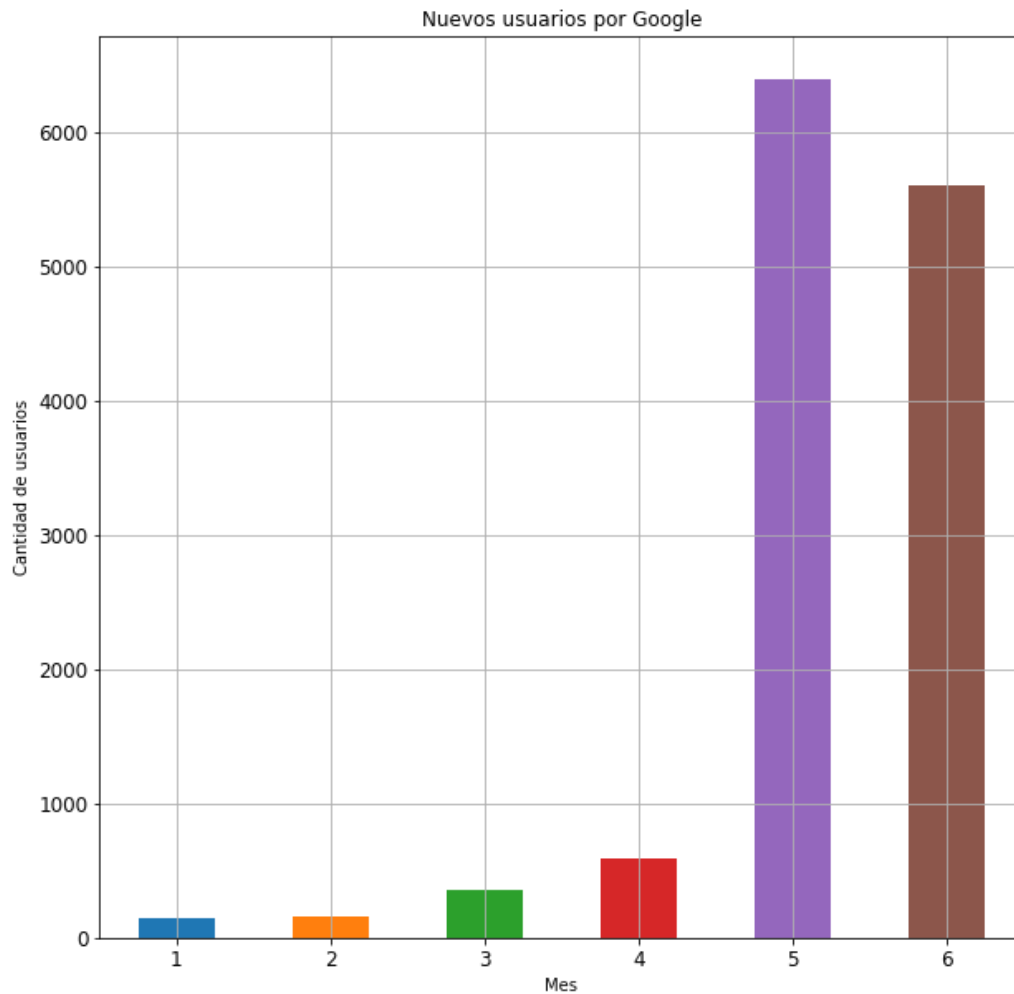


A partir del gráfico, se obtiene que los usuarios acceden muchas más veces al checkout que a la compra del producto, lo cual muestra que analizan varias veces antes de efectuar la compra.

## 6. Análisis de campañas

En esta sección analizamos el efecto de las campañas de marketing detectadas en el set de datos.

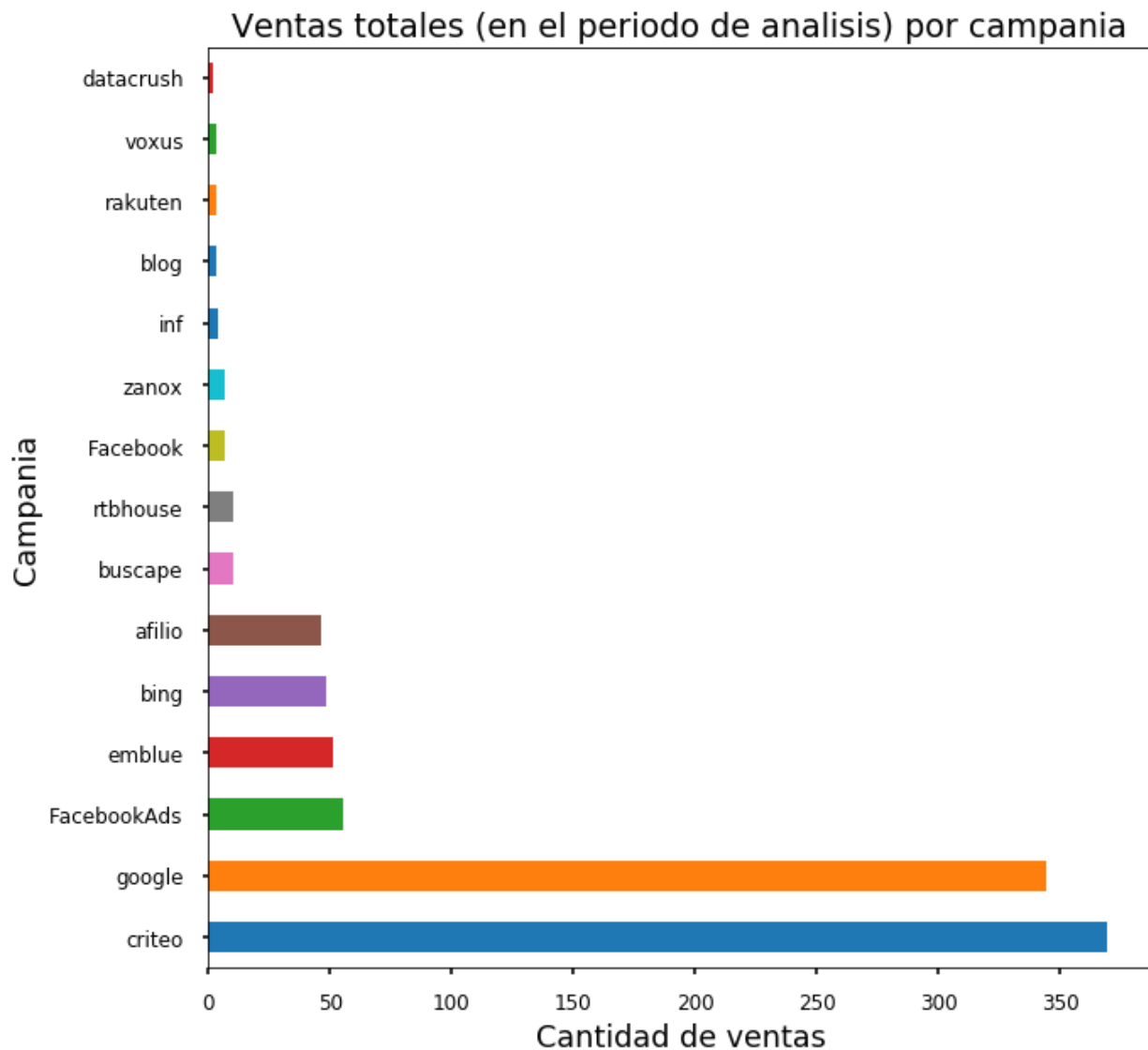
### 6.1 Nuevos usuarios por campañas



Podemos observar que en mayo la cantidad de nuevos usuarios que trajo Google aumentó muchísimo, para este gráfico también hay que tener en cuenta que los datos de junio son parciales (sólo tenemos la mitad de junio cargada)

Otro gráfico que nos resulta interesante es ver el aporte de cada campaña en el mes de mayo. Elegimos el mes de mayo ya que es el último mes que fue cargado completamente.

## 6.2 Ventas por campaña



Si tomamos en cuenta nuevamente el gráfico de nuevos usuarios por campaña, podemos ver que, si bien Google ha atraído una gran cantidad de nuevos usuarios, la mayoría de éstos no compran en el sitio, mientras que en comparación los llegan a través del servicio de publicidad de Amazon, "Criteo", lo hacen en un mayor porcentaje.



A partir de estos resultados puede definirse con más claridad a qué campañas destinar más recursos: ¿Queremos que nos visite más gente?, ¿Queremos que la gente que nos visite compre?

## 7. Conclusiones

A través del análisis de tráfico pudimos determinar que Trocafone está creciendo en popularidad, en particular tuvo un aumento notable en visitas durante el mes de mayo y es importante que la empresa esté preparada tanto comercial como estructuralmente para dicho aumento. También es importante notar que Brasil es el país que, por lejos, más eventos tiene registrados, por lo que debería enfocarse en el mismo.

El set de datos posee una gran cantidad de checkouts en comparación a la cantidad de ventas, este fenómeno debería evaluarse con mayor detenimiento y datos ya que de poder concretar un mayor porcentaje de checkouts la empresa se vería muy beneficiada por las ventas.

En cuanto a popularidad iPhone, Samsung y Motorola siempre destacan por lo que es importante trabajar siempre con estas marcas. En cuanto al estado de los productos la mayoría de las personas optan por comprarlos en “Buen” estado, lo cual puede afectar la filosofía de la empresa al comprar productos para hacer el refurbishing.

Según los resultados de las campañas de marketing, podría planear en dejar de utilizar aquellas que fueron menos exitosas en el plazo y concentrarse en aquellas que sí lo fueron.

A partir de los datos obtenidos de analizar la popularidad de los productos en las búsquedas y cuáles fueron los más vendidos, la empresa podría empezar a predecir la demanda de los productos más populares.

