

TRABAJO PRÁCTICO 1

Análisis Exploratorio

Organización de Datos

75.06

GRUPO 18 - Cuadernos 2x1 2.0

Nombre	Padrón	Mail
Belosevich Mugerli, Victor	97757	victor.belosevich94@gmail.com
Lencina, César Emanuel	96078	cel_1990@live.com
Pereira, Martin	95793	pereira.martin21@gmail.com
Petrucci, Maximiliano	95872	maximilianoangel.petrucci@gmail.com

- 1. Introducción general**
- 2. Procesamiento y Análisis de los Datos**
 - 2.1. Información general sobre el análisis realizado**
 - 2.1.1. Datos utilizados**
 - 2.1.2. Pre-procesamiento de los datos**
 - 2.1.3. Lenguaje y librerías utilizadas**
 - 2.1.4. Repositorio de GitHub**
 - 2.2. Eficacia de los Ads de publicidad**
 - 2.2.1. Introducción**
 - 2.2.2. Comparación de formas de ingresar a la página**
 - 2.2.3. Ad con mayor cantidad de visitas**
 - 2.2.4. Campaña publicitaria más exitosa**
 - 2.3. Ventas de celulares por mes**
 - 2.3.1. Introducción**
 - 2.3.2. Cantidad de celulares vendidos por mes**
 - 2.3.3. Ratio de ventas por mes**
 - 2.3.4. Relación checkouts/ventas por marca**
 - 2.4. Ventas de celulares según el día de la semana**
 - 2.4.1. Introducción**
 - 2.4.2. Día con mayor cantidad de checkouts.**
 - 2.4.3. Día con mayor cantidad de ventas**
 - 2.4.4. Comparativa entre ambos resultados**
 - 2.4.5. Mapa de calor de las compras por mes y día de la semana**
 - 2.4.6. Mapa de calor de las compras por hora y día de la semana**
 - 2.5. Actividad de usuarios en el sitio**
 - 2.5.1. Introducción**
 - 2.5.2. Cantidad de eventos realizados por hora del día**
 - 2.5.3. Evolución semanal de nuevos ingresos**
 - 2.5.4. Evolución semanal de los ingresos desde sitios externos**
 - 2.6. Tipo de dispositivo utilizado para navegar por la página**
 - 2.6.1. Introducción**
 - 2.6.2. Comparación de tipos de dispositivos utilizados**
 - 2.7. Usuarios que ingresaron con celular con un SO y terminaron comprando un celular con otro SO**
 - 2.7.1. Introducción**
 - 2.7.2. Usuarios ingresaron con iOS y compraron un celular con Android**
 - 2.7.3. Usuarios ingresaron con Android y compraron un celular con iOS**
 - 2.8. Calidad y color de los celulares vendidos**

- 2.8.1. **Introducción**
 - 2.8.2. **Stock según su condición**
 - 2.8.3. **Cantidad de celulares vendidos según su condición**
 - 2.8.4. **Top 10 productos más comprados y clasificados por su condición**
 - 2.8.5. **Cantidad de celulares vendidos según su color**
- 2.9. **Búsquedas de dispositivos**
 - 2.9.1. **Introducción**
 - 2.9.2. **Dispositivos más visitados**
 - 2.9.3. **Dispositivos con más suscripciones al servicio de disponibilidad de stock**
 - 2.9.4. **Top 10 términos más buscados**
 - 2.9.5. **WordCloud de términos más buscados**
- 2.10. **Ventas de celulares por región**
 - 2.10.1. **Introducción**
 - 2.10.2. **Regiones con mayor cantidad de ventas**
 - 2.10.3. **Distribución de las ventas en las regiones**
 - 2.10.4. **Porcentaje de Checkout vs Conversión**
 - 2.10.5. **Dispositivos desde los que se realizan las ventas**
- 3. **Conclusiones**

1. Introducción general

Este informe se encargará de analizar los datos de los eventos que realizan los usuarios en la página web www.trocafone.com para obtener información acerca de cuáles son los eventos que realizan los usuarios a la hora de comprar un celular.

El objetivo de este informe es analizar los datos brindados por *Trocafone* para observar si se encuentra información que pueda aportarles datos útiles para mejorar su servicio y proporcionarles estadísticas para mostrarles cuál es el estado de su empresa en estos últimos 6 meses.

Este informe también puede serle útil al lector para entender un poco más el negocio de Trocafone y mejorar su experiencia a la hora de comprar un celular en el sitio web.

2. Procesamiento y Análisis de los Datos

2.1 Información general sobre el análisis realizado

2.1.1 Datos utilizados

Para el análisis propuesto utilizaremos los datos de conjunto de eventos de web analytics de usuarios que visitaron www.trocafone.com, su plataforma de ecommerce de Brasil. Dichos datos se encontraban en un CSV que gentilmente la empresa brindó.

2.1.2 Pre-Procesamiento de los datos

El dataset proporcionado para el análisis se encontraba, a simple vista, lleno de campos sin datos. Si bien estos datos estaban incompletos porque cada evento es distinto y tiene sus propios campos necesarios (esto es, un evento tipo 'visited_site' no requiere tener el campo 'color' o 'model' con valores ya que no corresponde), se decidió completar algunos campos (como 'country', 'city', 'region', 'browser_version', 'operating_system_version', 'device_type') con los valores provenientes de otros eventos, del mismo usuario, que sí tenían esos datos.

El criterio que tomamos para rellenar los campos faltantes, consistió en agrupar los eventos por persona y fecha. Luego utilizamos los datos del evento 'visited_site' para completar los campos necesarios en los demás eventos.

También al adoptar este criterio, eliminamos los registros pertenecientes a los usuarios que no registraban el evento 'visited_site', ya que la información que proporcionaban no era concluyente.

2.1.3 Lenguaje y librerías utilizadas

Para el procesamiento de datos se utilizó Python 3, junto a la librería de Pandas.

Para el tema de visualización y gráficos se utilizaron las librerías Matplotlib, Seaborn y WordCloud.

2.1.4 Repositorio de GitHub

Se utilizó un repositorio para la integración del trabajo. En él se encuentran los notebooks realizados por cada integrante y este mismo informe.

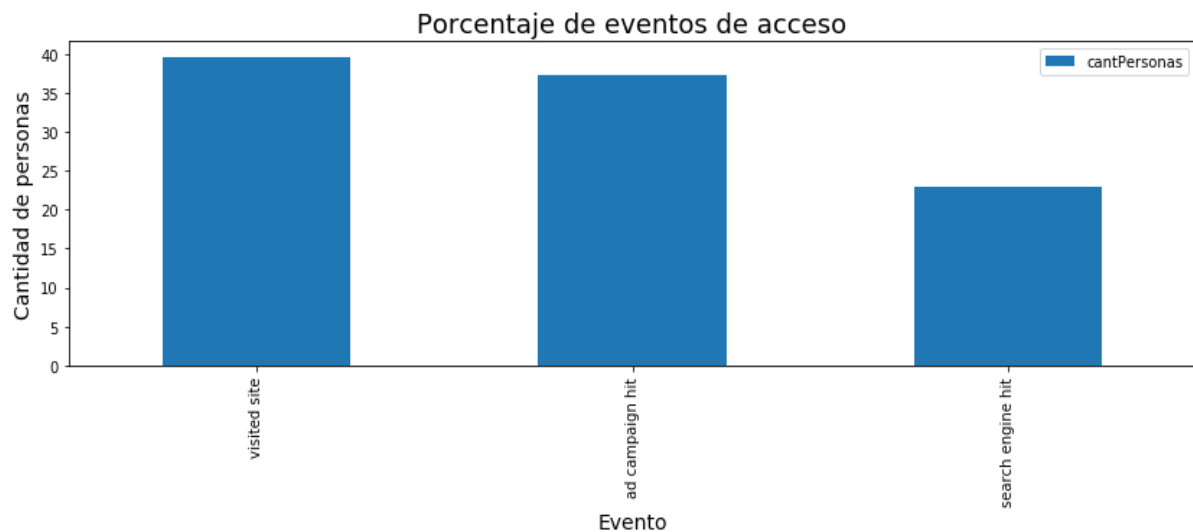
Link: <https://github.com/Belovic94/OrgaDatos-TP-2018>

2.2 Eficacia de los Ads de publicidad

2.2.1 Introducción

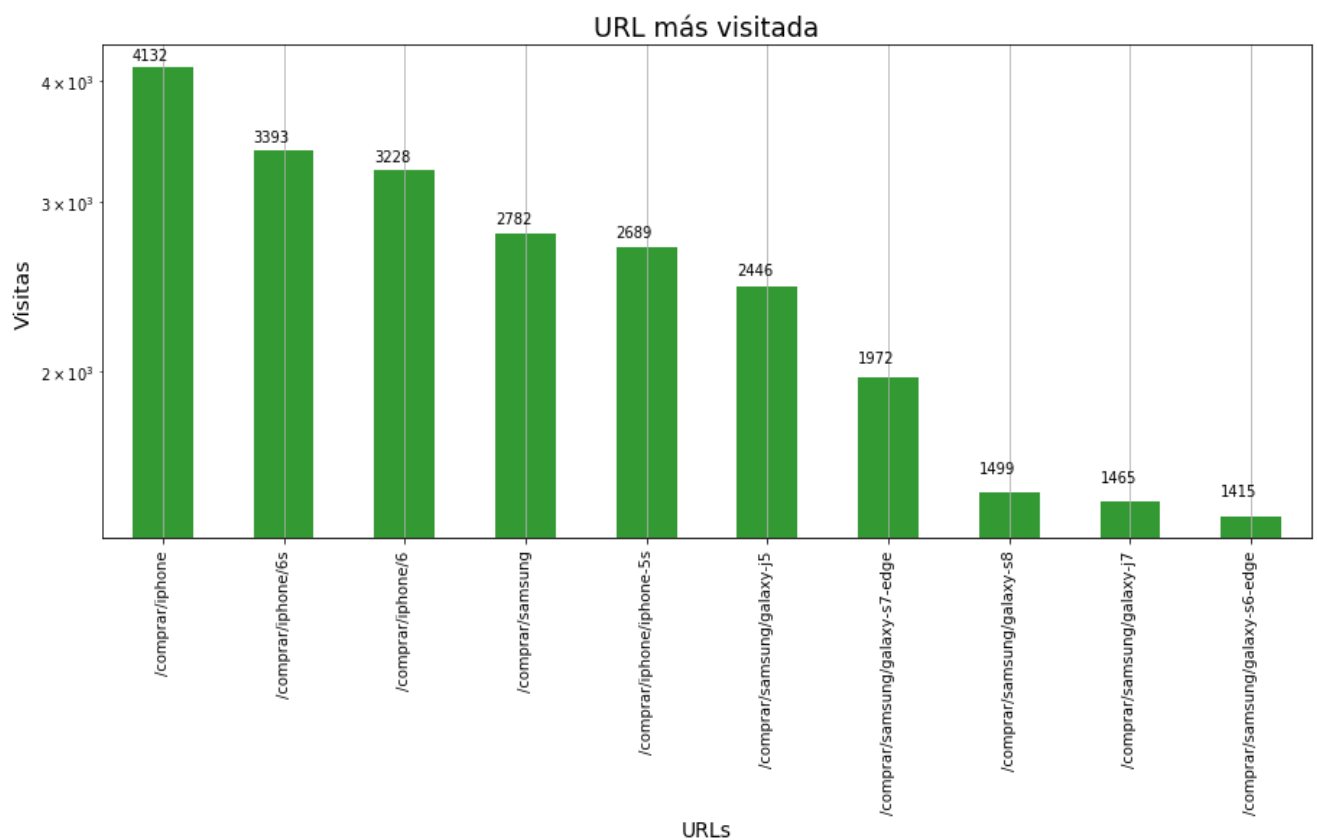
Queríamos ver las formas de ingresar a la página de *Trocafone*, si la forma de ingresar por un ad era significativa o no. De los eventos del set de datos, los 3 que utilizamos para el análisis son: “Visited site”, “Ad campaign hit” y “Search engine hit”. Además realizamos un análisis sobre las campañas de publicidad que el sitio había lanzado y cuán eficaces son.

2.2.2 Comparación de formas de ingresar a la página



Como se puede ver la cantidad de personas que ingresan por un Ad es muy parecida a la gente que ingresa directamente a una URL del sitio. Las campañas de marketing utilizadas por la empresa son eficaces y aumentan la cantidad de personas que visitan el sitio.

2.2.3 Ad con mayor cantidad de visitas



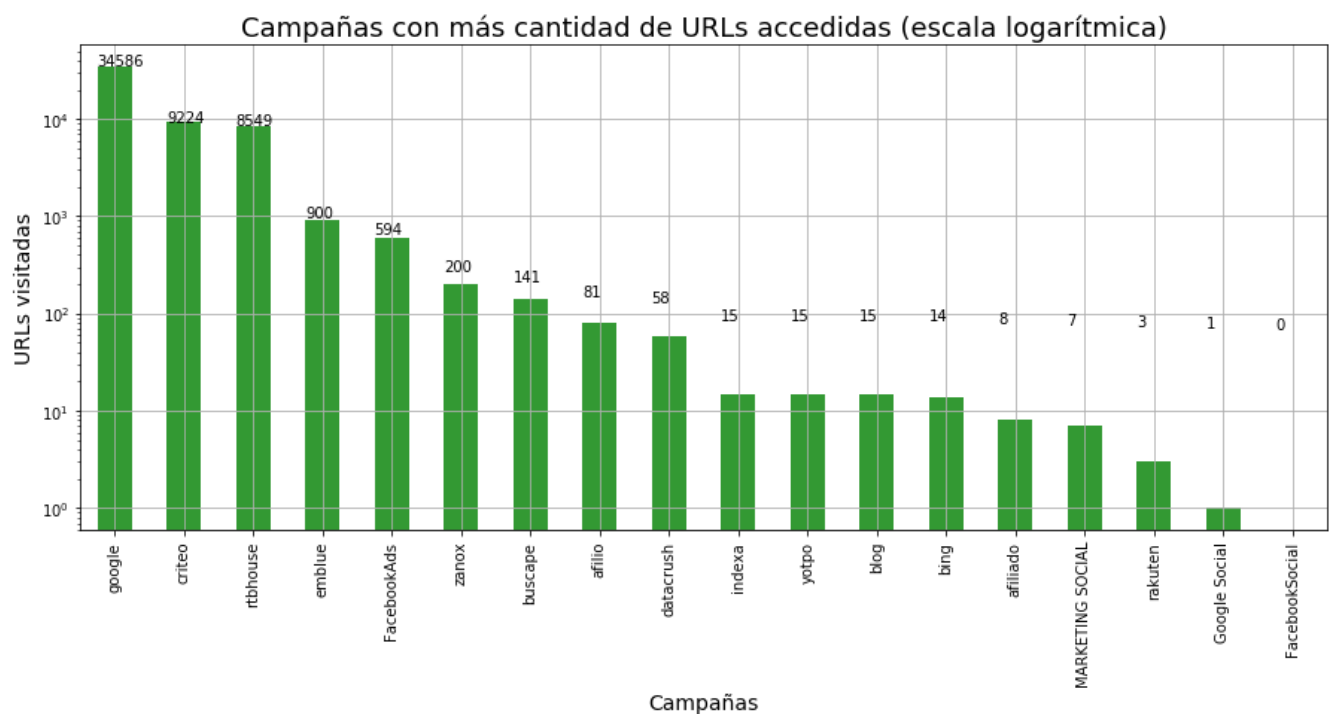
Como se puede apreciar el Ad más visitado fue el de *iPhone* lo que parecería indicar que a la gente le interesa solamente el producto sin importarle la versión del mismo.

También se puede notar que los siguientes dos ads más visitados también hacen referencia a dispositivos *iPhone*. En cuarto lugar recién aparece otro dispositivo, el ad de *Samsung* genérico.

Investigando en la red se encontró que los modelos de *iPhone* más clickeados son de productos con más de tres años en el mercado de marca *iPhone* en detrimento de celulares más recientes pero de marca *Samsung*.

Como **conclusión** Trocafone debería enfocarse en comercializar preferentemente productos *iPhone* con una antigüedad de más de tres años.

2.2.4 Campaña publicitaria más exitosa



Es evidente que la campaña publicitaria más exitosa es la de *Google*, seguramente debido a su popularidad. En segundo lugar se encuentra *Criteo* una empresa especializada en publicidad al igual que *RTB House* que ocupa el tercer lugar.

Como el set de datos abarca seis meses, queda en evidencia que las campañas con menos de una visita por día se las podría considerar como un fracaso.

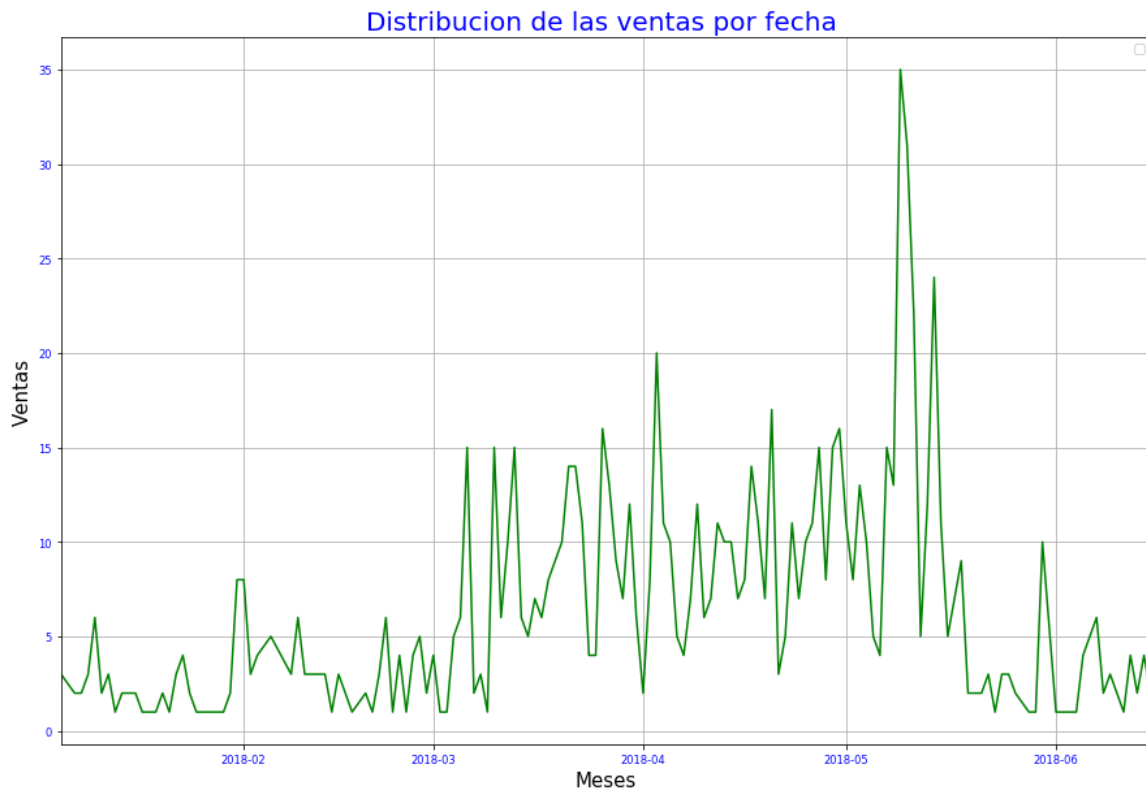
En **conclusión** para optimizar su margen de ganancias *Trocafone* debería dejar de gastar dinero en utilizar los servicios de las empresas de ads con menos visitas aportadas, o sea desde *Buscape* en adelante.

2.3 Ventas de celulares por mes

2.3.1 Introducción

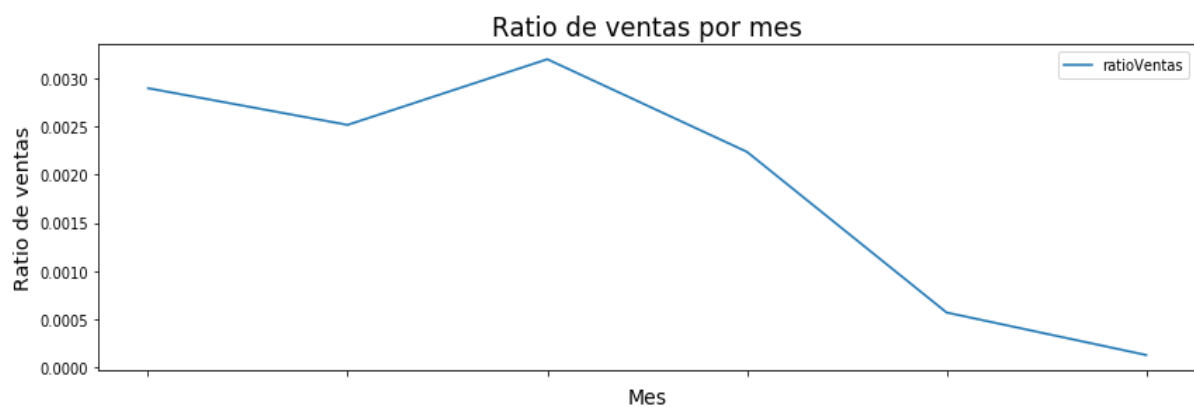
Queríamos ver cómo le estaba yendo a la empresa *Trocafone* en estos últimos 6 meses y si la pérdida de valor del real vs el dólar tuvo un impacto negativo en las ventas. Para realizar este análisis, evaluamos la cantidad de “conversiones” que se realizaron por mes.

2.3.2 Cantidad de celulares vendidos por mes



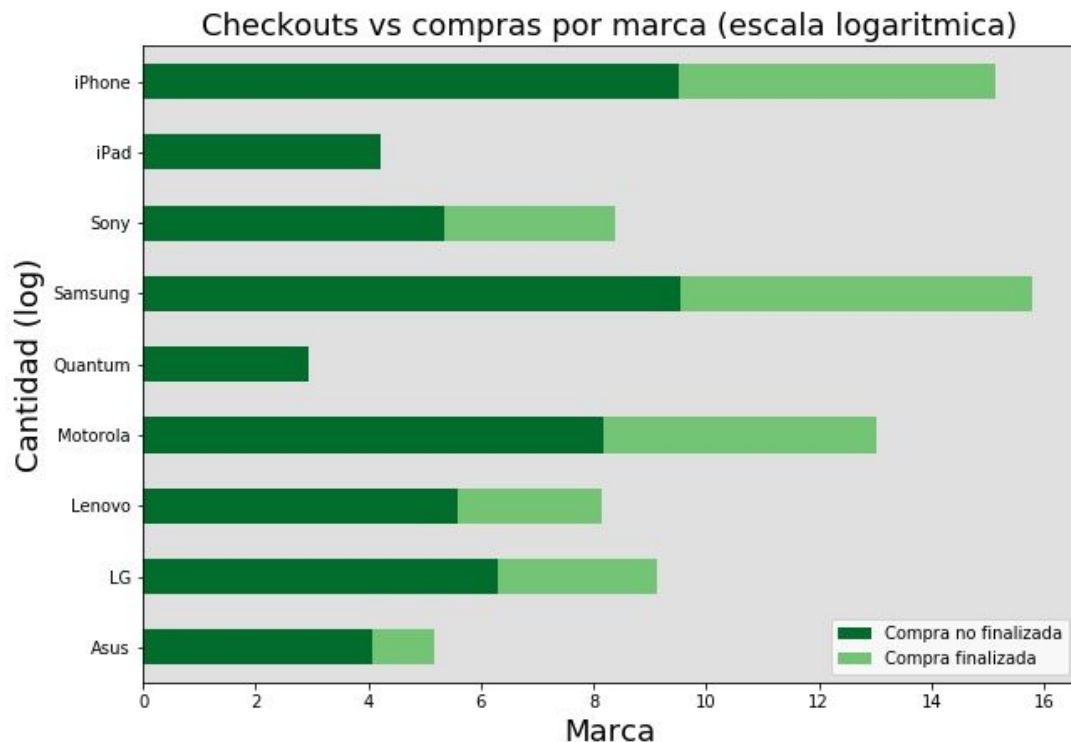
Como se puede ver, la cantidad de celulares vendidos fue en aumento desde Enero hasta Abril en donde se ve una pequeña reducción de ventas hasta el mes de Mayo. A partir del mes de Mayo las ventas cayeron fuertemente. Esto puede ser debido a la caída de la moneda Brasileña respecto al dólar que comenzó por esa época.

2.3.3 Ratio de ventas por mes



En este caso analizamos la cantidad de conversiones que se hicieron en el mes respecto de la cantidad de eventos que se realizaron ese mes para poder conseguir un ratio de ventas. Como se puede ver a partir de Marzo, las ventas sufrieron una fuerte caída de la cual no se pudieron recuperar. Lamentablemente la devaluación de la moneda hizo fuerte estragos en las ventas de la empresa *Trocafone* y esperamos que pronto comience su recuperación.

2.3.4 Relación checkouts/ventas por marca



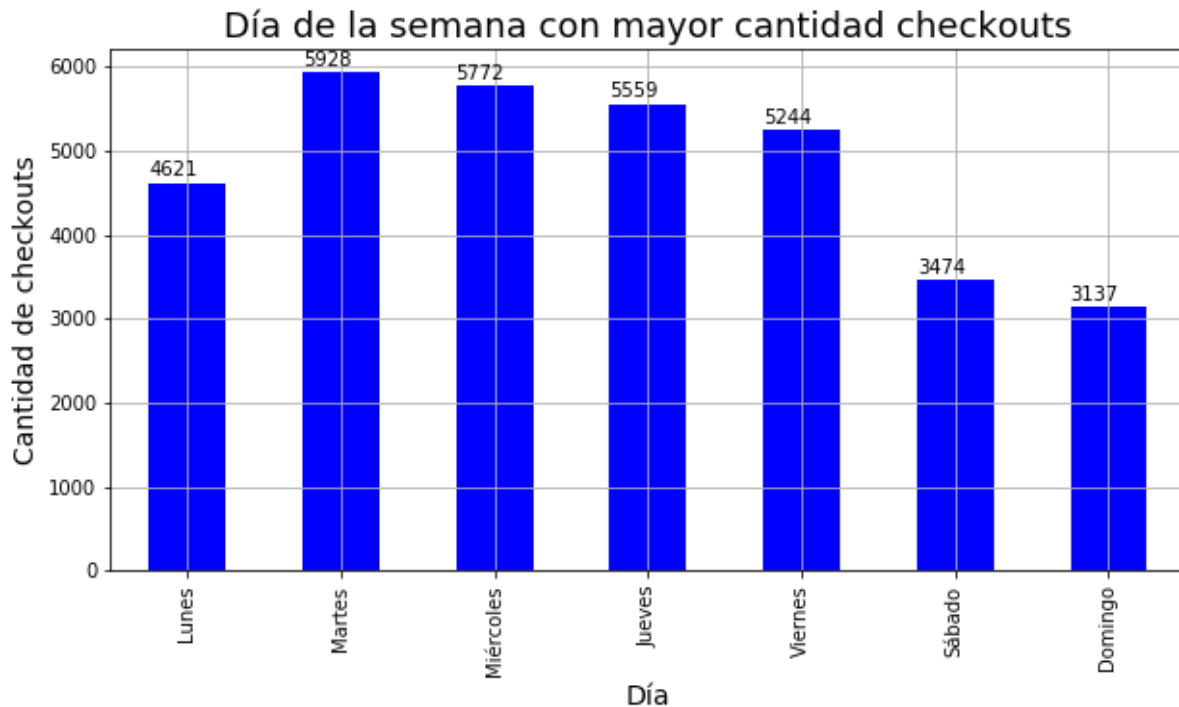
En este gráfico (**a escala logarítmica**), se puede observar que en la plataforma de *Trocafone* **no se concretan la mayoría de los checkouts**. Al parecer, el sitio no brinda a los usuarios la seguridad necesaria como para finalizar la compra del producto, independientemente de la marca en cuestión. Como se verá en el Top de los productos más vistos en la plataforma, **quienes tienen más checkouts y compras son Samsung e iPhone**.

2.4 Ventas de celulares según el día de la semana.

2.4.1 Introducción

Se revisó que día de la semana era el cual se realizaba la mayor cantidad de ventas y checkouts con el objetivo de ver si alguna destacaba en especial. Se utilizaron las columnas “timestamp” y “event”, este último con los valores “conversión” y “checkout”.

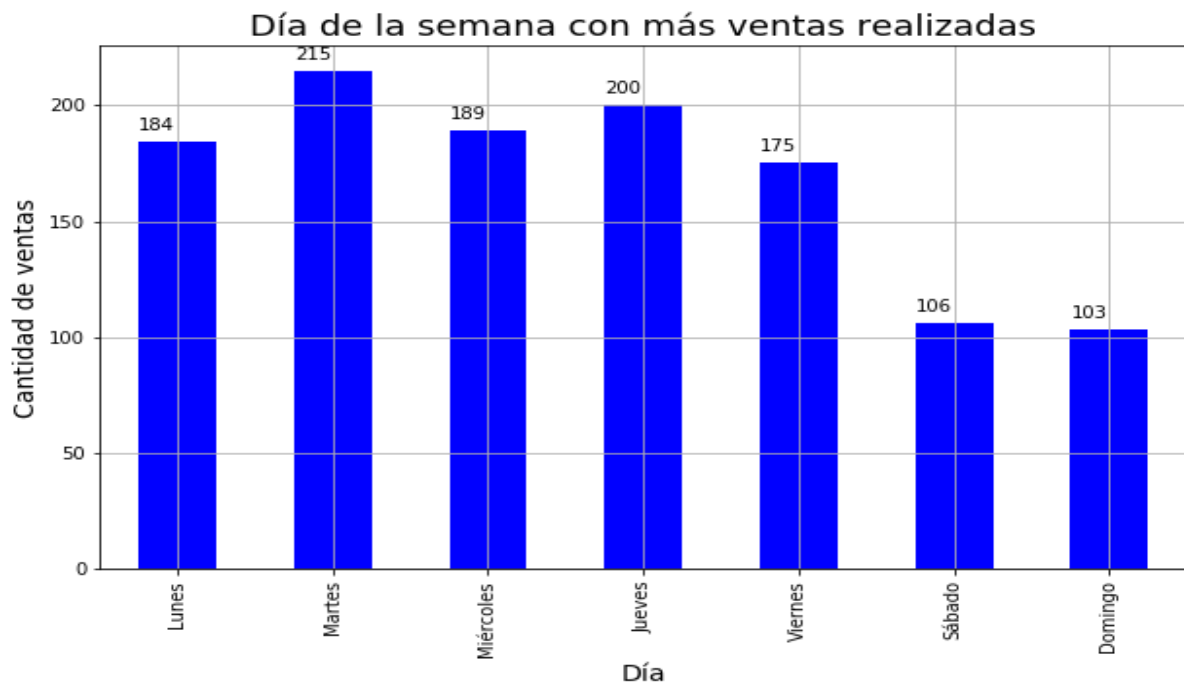
2.4.2 Día con mayor cantidad de checkouts



Se puede ver que el día martes es aquel con mayor cantidad de checkouts (usuarios que llegan hasta el carrito de compras online sin comprar) pero bastante cerca se encuentran en orden decreciente los días miércoles, jueves y viernes. El fin de semana y el lunes muestran una merma de la actividad respecto al resto de la semana.

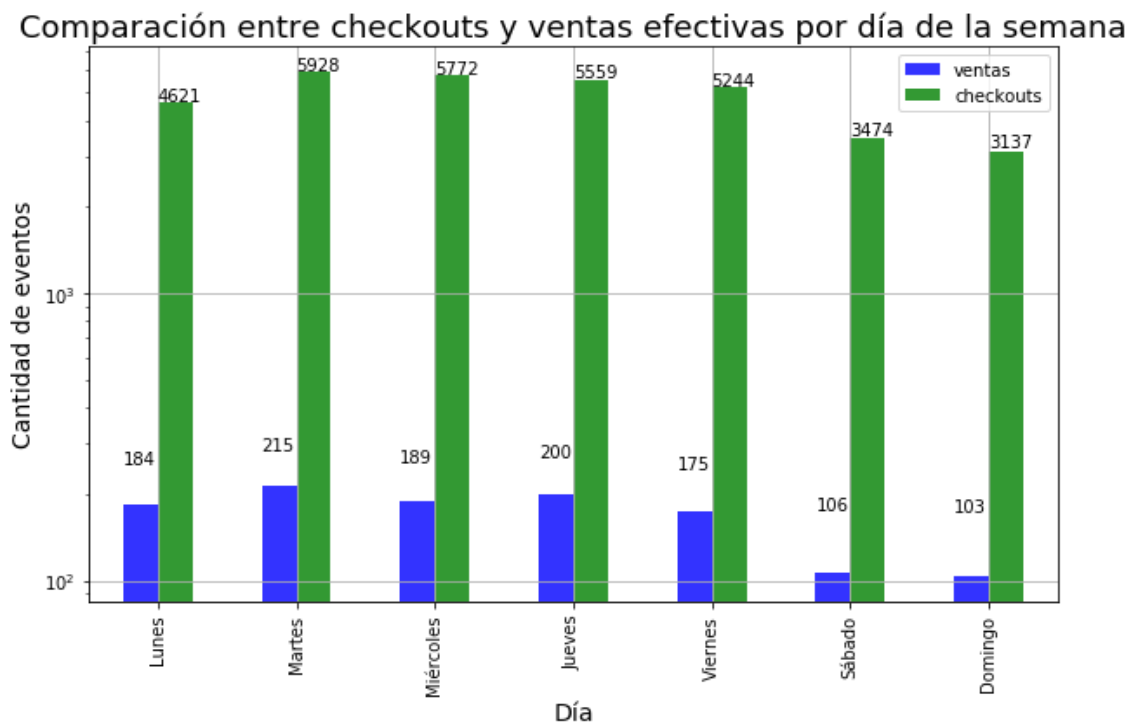
Se podría teorizar que esta merma se debe a que los fines de semana la gente se encuentra en actividades de ocio y que los lunes los atacan las obligaciones de la vida cotidiana. No es muy concluyente por qué ocurre esto.

2.4.3 Día con mayor cantidad de ventas



Se puede ver que la distribución de datos es muy similar a la de los checkouts salvo por el día miércoles.

2.4.4 Comparativa entre ambos resultados

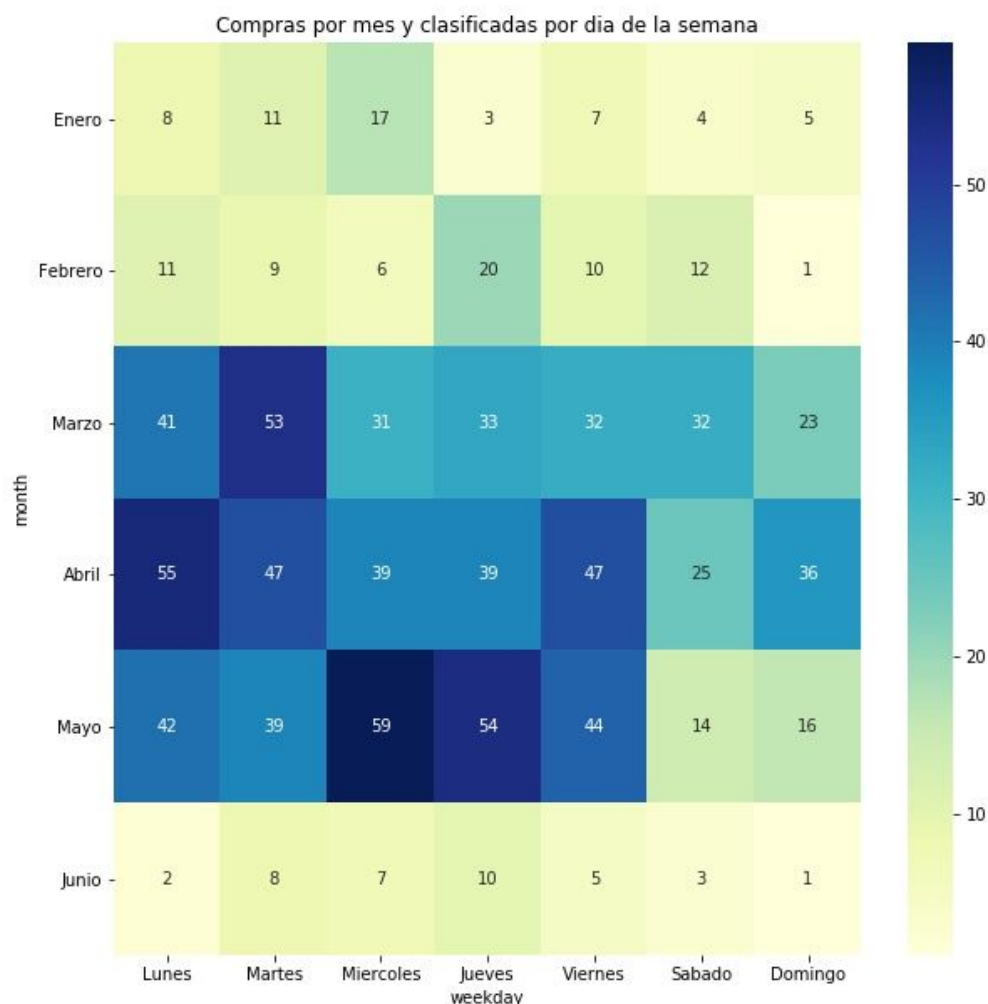


Se puede notar una amplia diferencia entre la cantidad de checkouts y la cantidad de ventas efectivas. La gran mayoría de la gente no compra el producto que se encuentra en el carrito.

Como **conclusión**, para disminuir este margen *Trocafone* podría incluir ofertas de descuento en el caso de que la compra se realice dentro de las primeras 24 horas de estar el producto en el carrito o regalar descuentos en la próxima compra.

2.4.5 Mapa de calor de las compras por mes y día de la semana

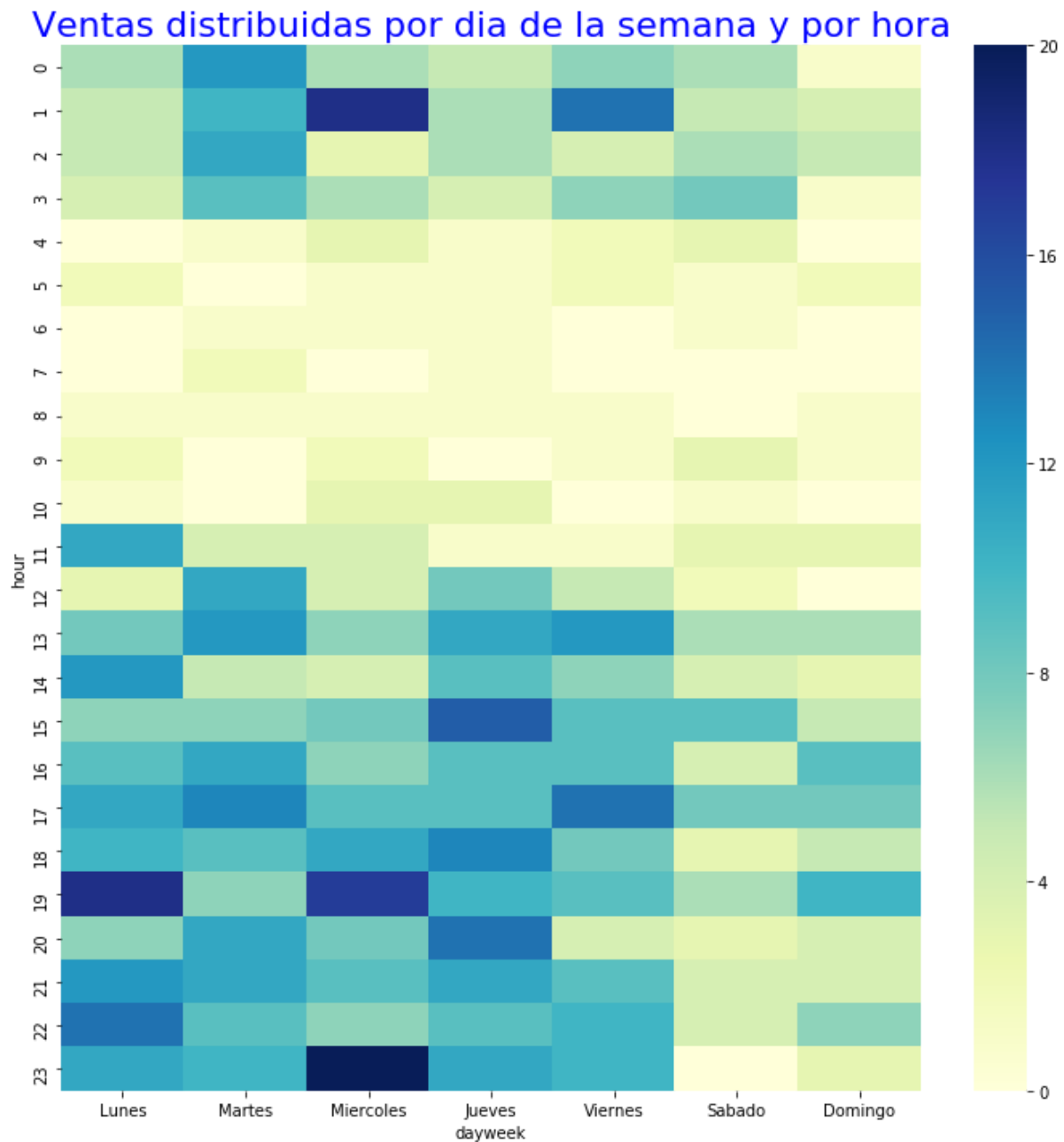
Para cerrar estos dos apartados (2.3 y 2.4), podemos visualizar en este mapa de calor las compras en la plataforma a lo largo de los seis meses y clasificados por día de la semana:



Se puede apreciar que durante todo Abril y todo Mayo hubo ingresos a la plataforma. **Esto puede deberse**, en parte, **al día de la madre celebrado en Brasil** que fue el Domingo 13 de Mayo, por lo cual la gente ingresó a la plataforma a buscar un presente. De este mapa de calor, se evidencia que hubo más actividad

(de compras) en Trocafone durante Marzo, Abril y Mayo que durante Enero, Febrero y Junio.

2.4.6 Mapa de calor de las compras por hora y día de la semana



En este gráfico se puede apreciar que la mayoría de las ventas son realizadas principalmente durante los días de semana a partir de las 14 hs hasta las primeras horas de la madrugada. También se nota una concentración de ventas los días miércoles alrededor de la medianoche.

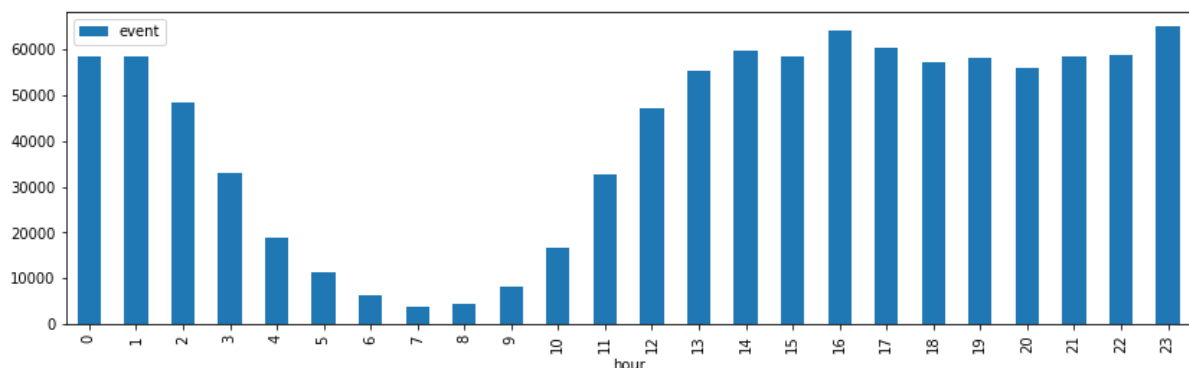
Cabe destacar que durante la mañana y los fines de semana, las ventas disminuyen considerablemente. Una forma de evitar estas recaídas sería ofrecer descuentos y promociones durante los fines de semana.

2.5 Actividad de usuarios en el sitio

2.5.1 Introducción

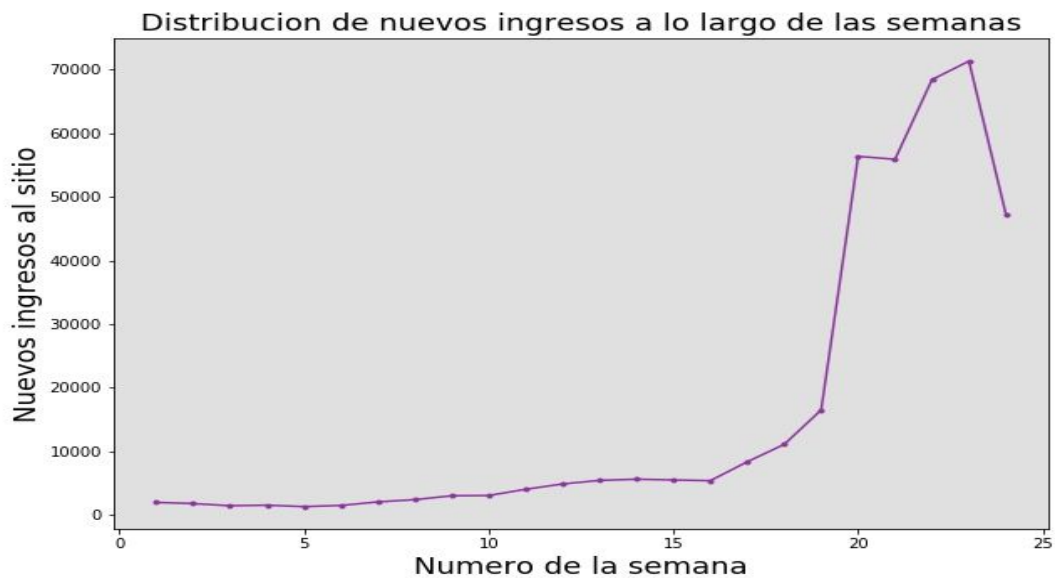
En esta sección del análisis exploratorio, se mostrarán datos sobre la actividad de los usuarios en la plataforma.

2.5.2 Cantidad de eventos realizados por hora del día



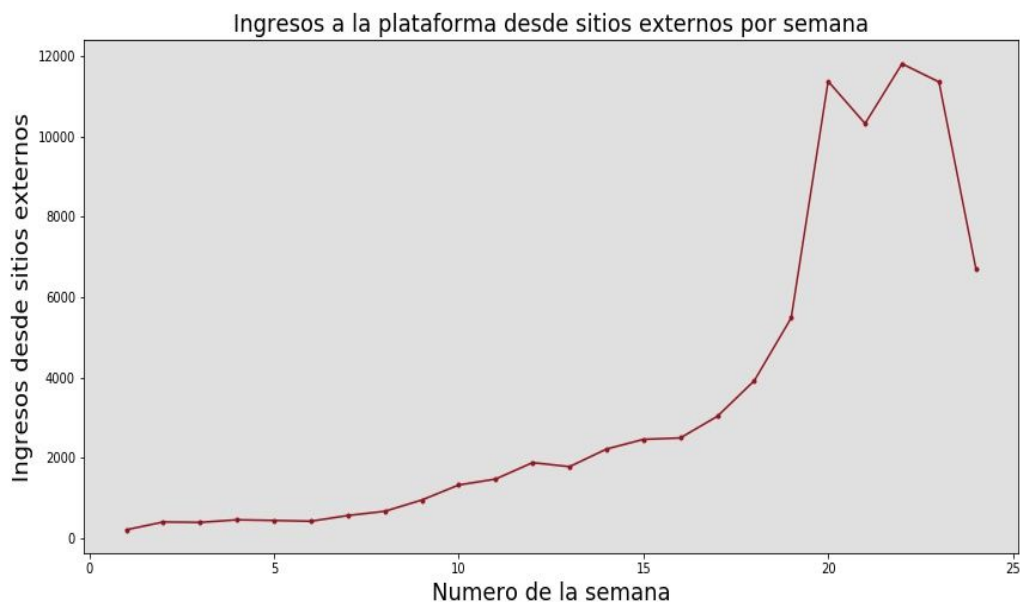
Como se puede ver, la cantidad de eventos por hora se mantiene constante a partir de las 2pm hasta la 1am en donde se ve una caída en los eventos hasta las 7am. Con el comienzo del día laboral de los trabajadores, la cantidad de eventos comienza nuevamente a aumentar.

2.5.3 Evolución semanal de nuevos ingresos



Como podemos ver, **la cantidad de ingresos de nuevos usuarios** (según los datos de Google Analytics) **se mantuvo *planchada*** (y relativamente constante) durante las primeras 16 semanas. Luego tuvo un leve incremento hasta la semana 18 y **obtuvo su gran pico en la semana 19, la cual corresponde al día de la madre** anteriormente mencionado (semana del 13 de Mayo). Podríamos aventurarnos a decir que **este tipo de eventos favorece enormemente a la plataforma.**

2.5.4 Evolución semanal de los ingresos desde sitios externos



Analizando la evolución semanal de los ingresos desde sitios externos (esto es, publicidades), podemos observar una gran similitud con el gráfico anteriormente

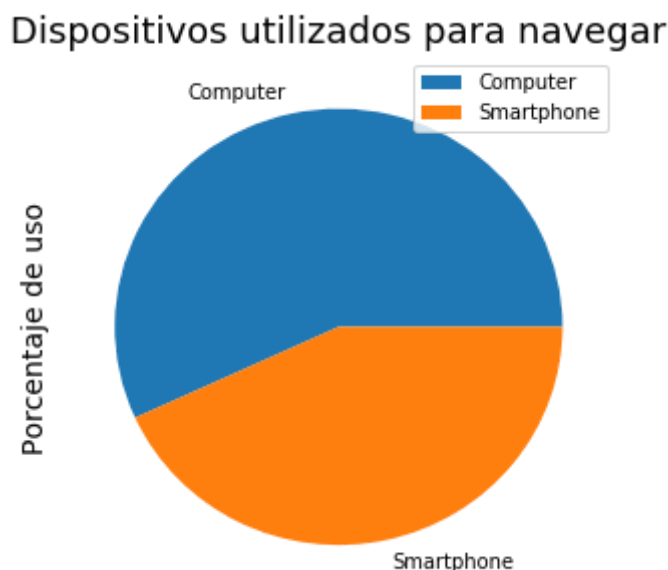
visto. De esta manera, **se evidencia que el sitio obtiene sus clientes gracias**, en gran parte, **a la publicidad que tiene en otros sitios**.

2.6 Tipos de dispositivos utilizados para navegar por la página

2.6.1 Introducción

Queríamos analizar qué tipos de dispositivos utiliza la gente para navegar por el sitio. Para este análisis dejamos de lado las “Tablets” y “Otros” ya que solamente representaban únicamente el 1% de los dispositivos utilizados.

2.6.2 Comparación de tipos de dispositivos utilizados



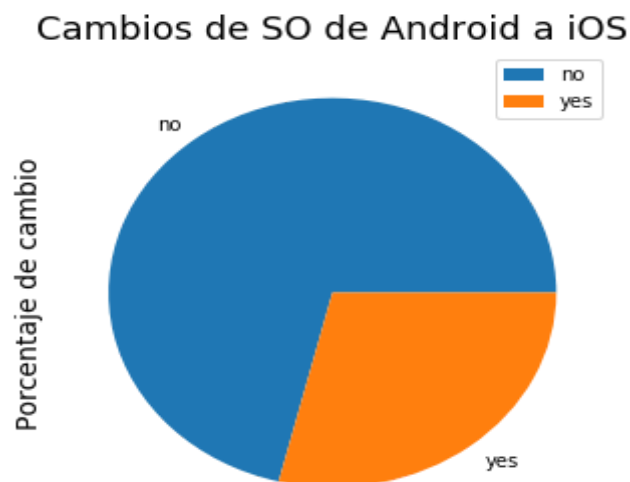
Como se puede ver en el gráfico, hay una leve tendencia al ingreso desde una computadora aunque hay una cantidad considerable de gente que ingresa desde el celular. Teniendo esto en cuenta, sería una buena idea que *Trocafone* implemente una aplicación para celular para facilitarle las cosas a los usuarios de Smartphones.

2.7 Usuarios que ingresaron con celular con un SO y terminaron comprando un celular con otro SO

2.7.1 Introducción

Queríamos ver si los usuarios que ingresaron con un determinado sistema Operativo tenían la tendencia de comprar un celular que tuviera otro sistema operativo o se mantenían con el mismo. Esto lo hicimos para ver cuán leales son los usuarios a su sistema operativo.

2.7.2 Usuarios ingresaron con iOS y compraron un celular con Android



Por lo visto en el gráfico, la gente que tiene un celular con Android termina en su mayoría comprando un celular también con Android. Se podría decir que los usuarios de Android están a gusto con el sistema operativo y lo vuelven a elegir.

2.7.3 Usuarios ingresaron con Android y compraron un celular con iOS



Por lo visto en el gráfico, la gente que tiene un celular con iOS está repartida entre comprar un celular con iOS o con Android. Se podría decir que los usuarios de iOS

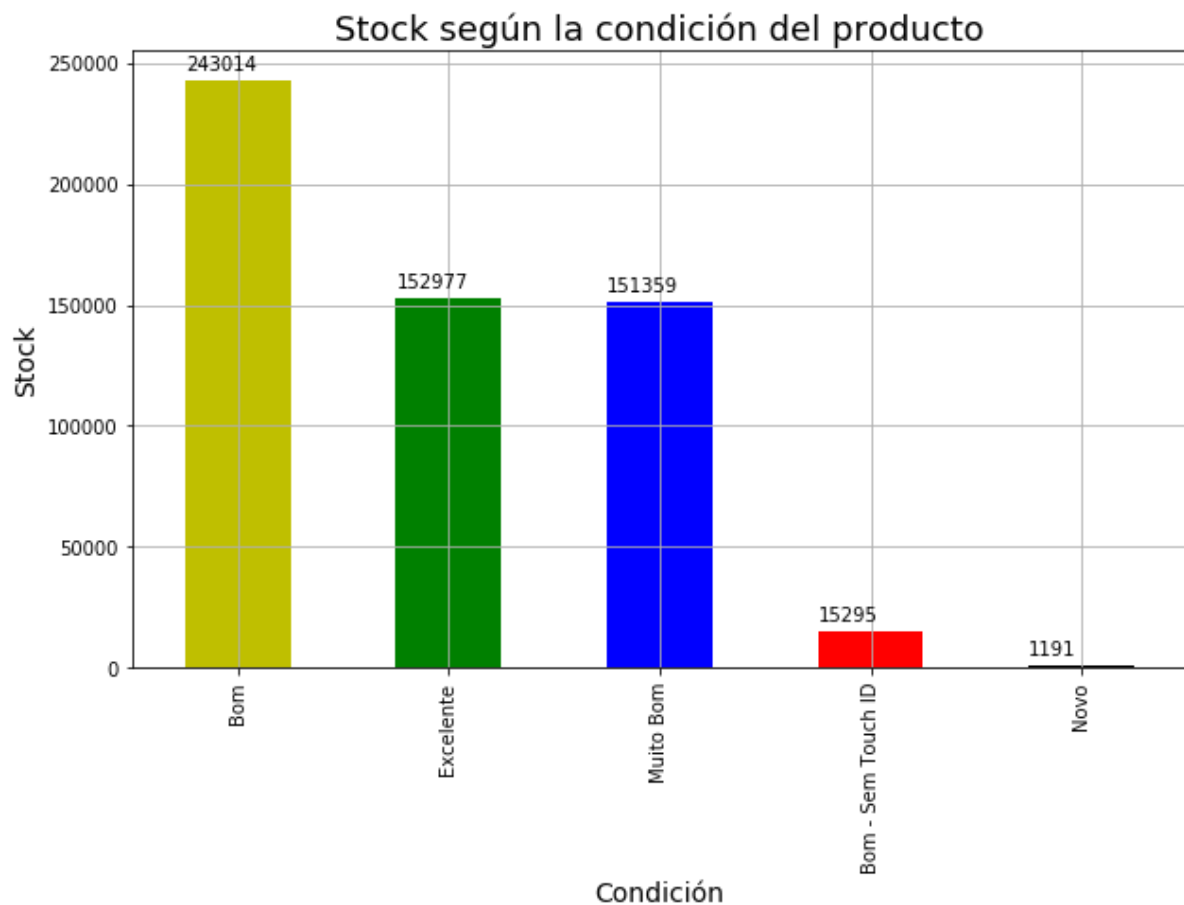
no son muy leales a su sistema operativo y Android le estaría sacando usuarios a iOS. Cabe aclarar que la cantidad de celulares que utilizan Android es mucho mayor que los que usan iOS.

2.8 Calidad y color de los celulares vendidos

2.8.1 Introducción

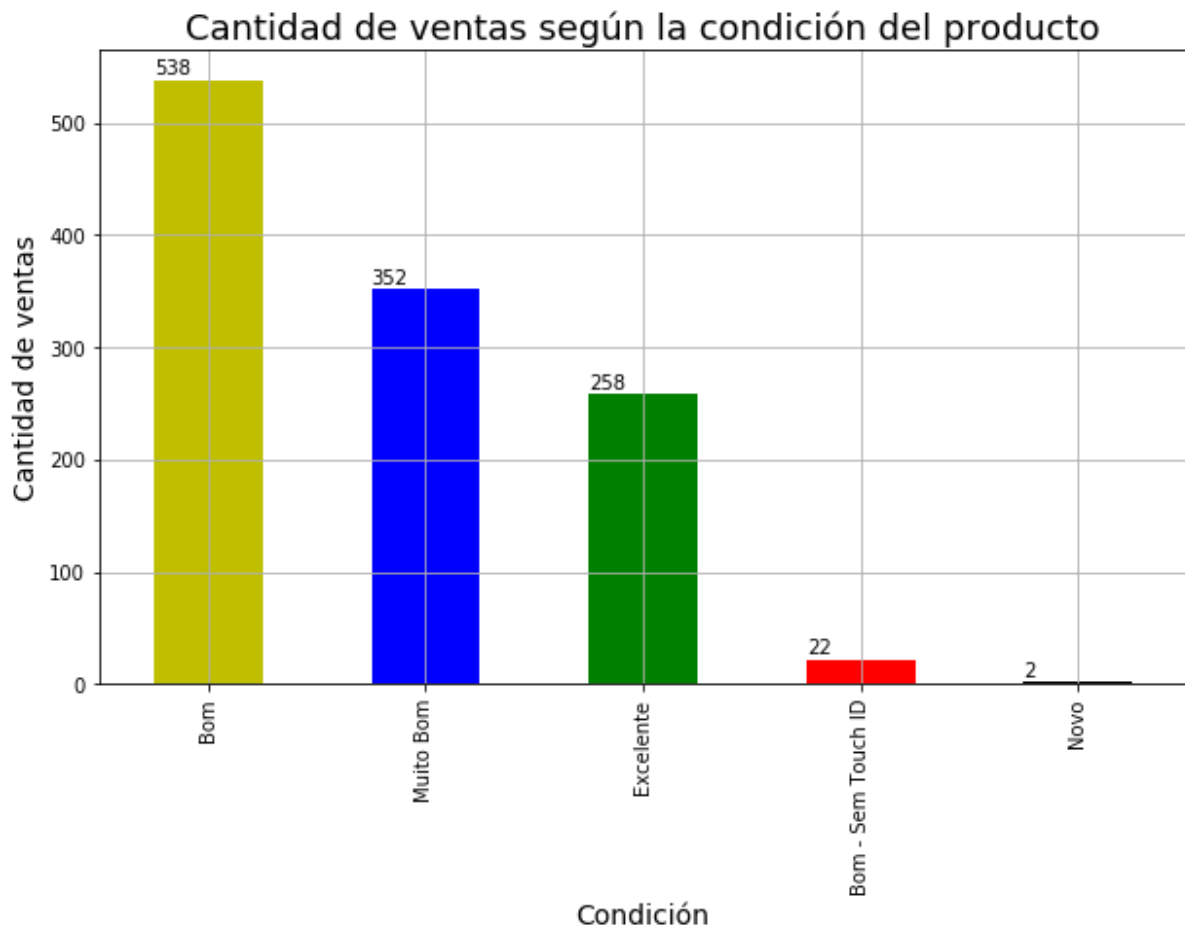
Se busca mostrar las preferencias de los usuarios respecto a la calidad y el color de los productos comprados. Se utilizó del set de datos las columnas “condition” ,“color” y “event” con el valor de “conversion” (venta).

2.8.2 Stock de celulares según su condición



A modo ilustrativo se puede apreciar que la condición de la mayoría de los productos de *Trocafone* es de buena a excelente.

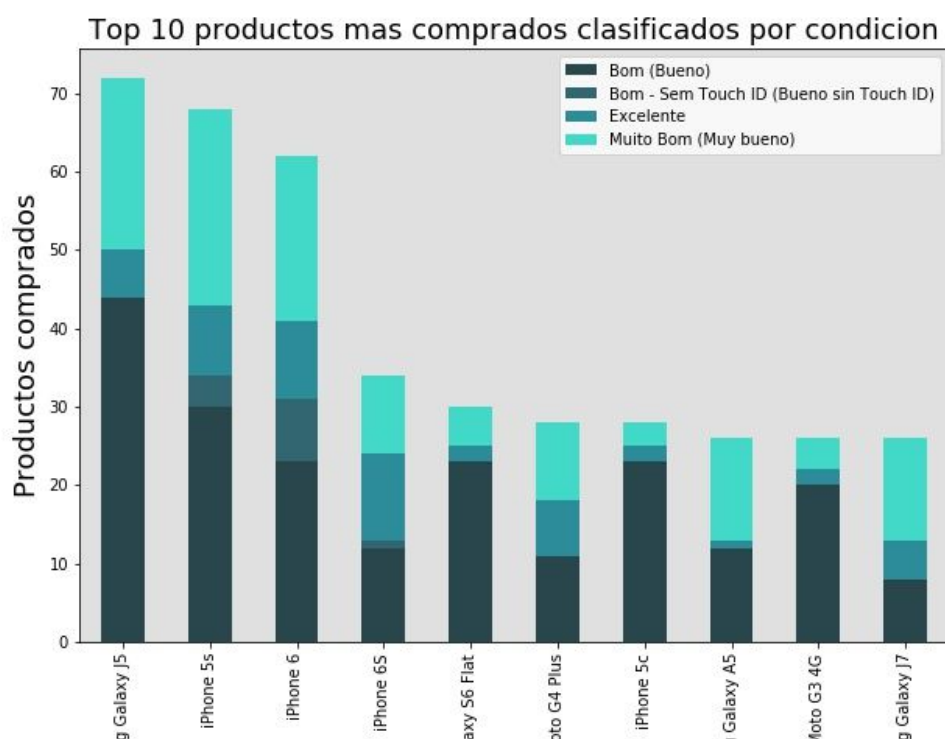
2.8.3 Cantidad de celulares vendidos según su condición



Se puede ver que los usuarios prefieren celulares que estén en buen estado y en menor medida en condiciones mejores. Sin embargo quedan muy atrás los celulares totalmente nuevos y aquellos que están rotos (sin Touch ID).

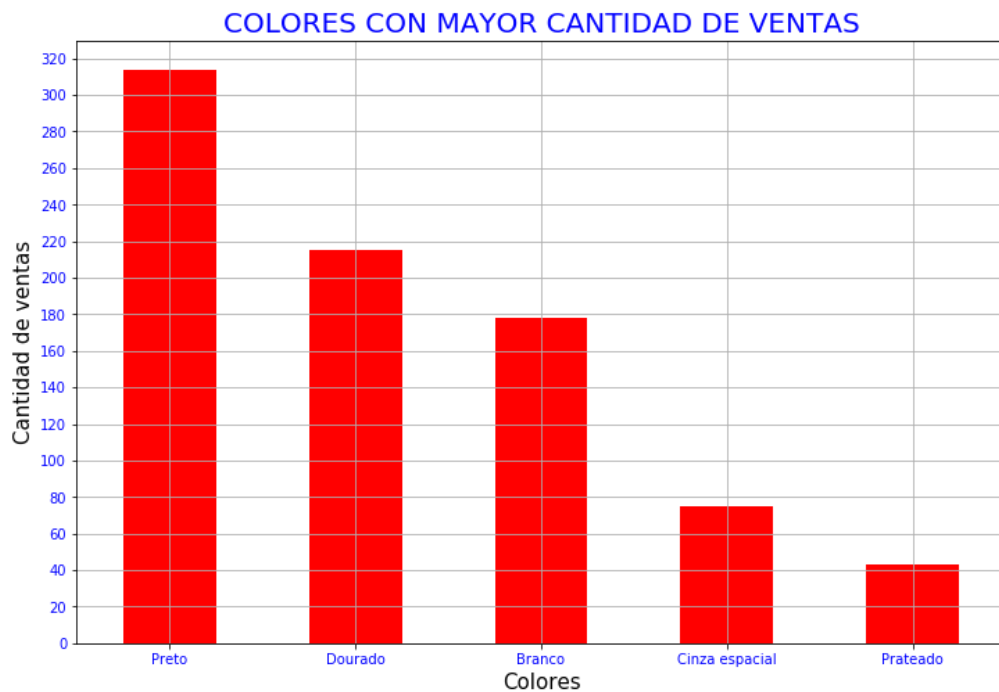
Como **conclusión** Trocafone debería enfocarse en comercializar celulares de las categorías dominantes dejando de lado aquellos productos nuevos o fallados.

2.8.4 Top 10 productos más comprados y clasificados por su condición



En esta visualización se puede apreciar lo anteriormente dicho: **los usuarios de Trocafone tienden a comprar dispositivos en estado “Muy bueno” o “Bueno”** (con Touch ID). Además, podemos ver que, si bien el teléfono **más comprado es el Samsung Galaxy J5**, hay una clara supremacía de los iPhone.

2.8.5 Cantidad de celulares vendidos según su color



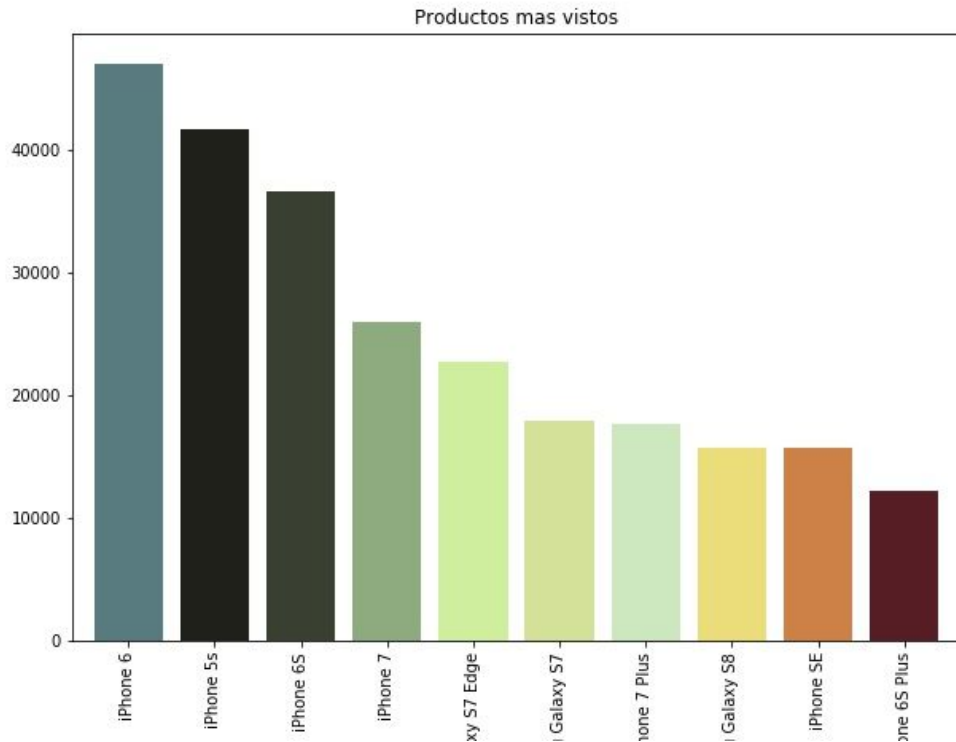
Este gráfico nos permite observar la preferencia de los usuarios por el color **Preto**(Negro) sobre los demás. Un poco más abajo se llegan a ver los colores **Dourado**(Dorado) y **Branco**(Blanco) y con mucha diferencia los colores **Cinza Espacial**(Gris Espacial) y **Prateado**(Plateado)

2.9 Búsquedas de productos

2.9.1 Introducción

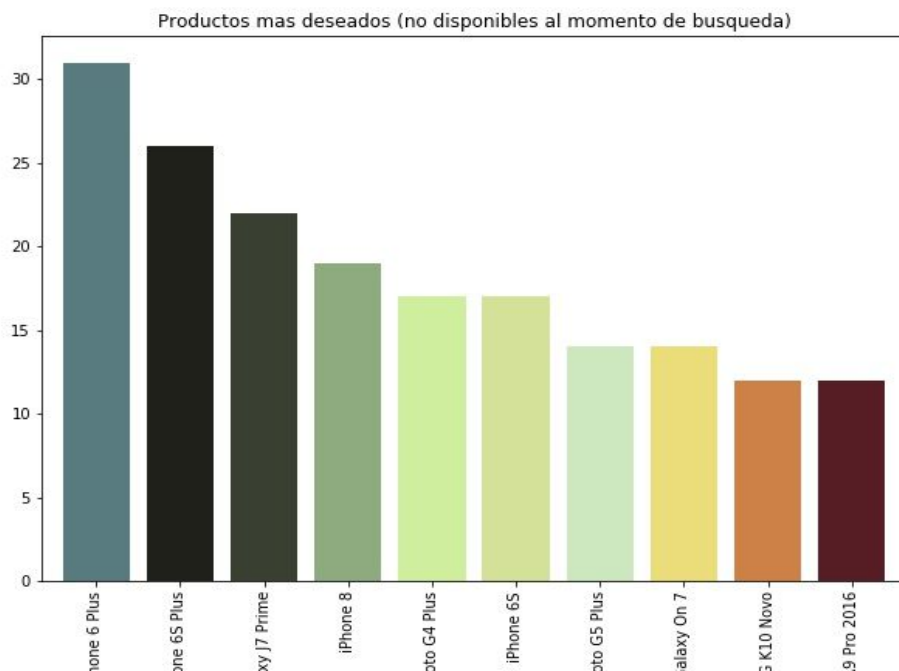
En este apartado se mostrarán gráficos y conclusiones sobre las búsquedas de productos realizada por los usuarios.

2.9.2 Dispositivos más visitados



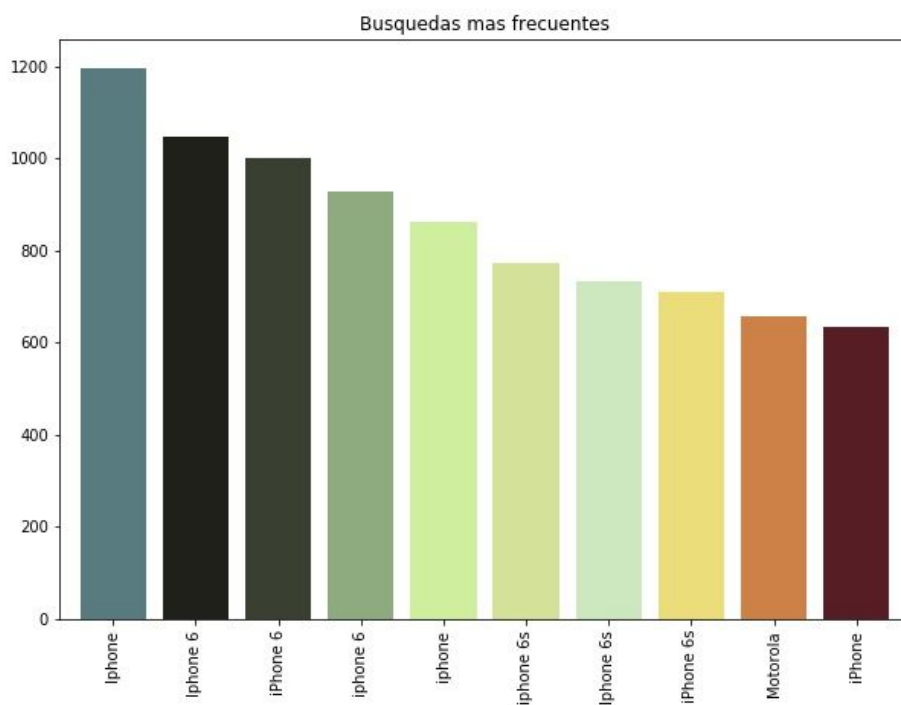
Estos fueron los productos más visitados en Trocafone. Como se pudo ver en los checkouts y las compras, **los productos más visitados son los de Apple (iPhone) seguidos por la línea Galaxy de Samsung.**

2.9.3 Dispositivos con más suscripciones al servicio de disponibilidad de stock

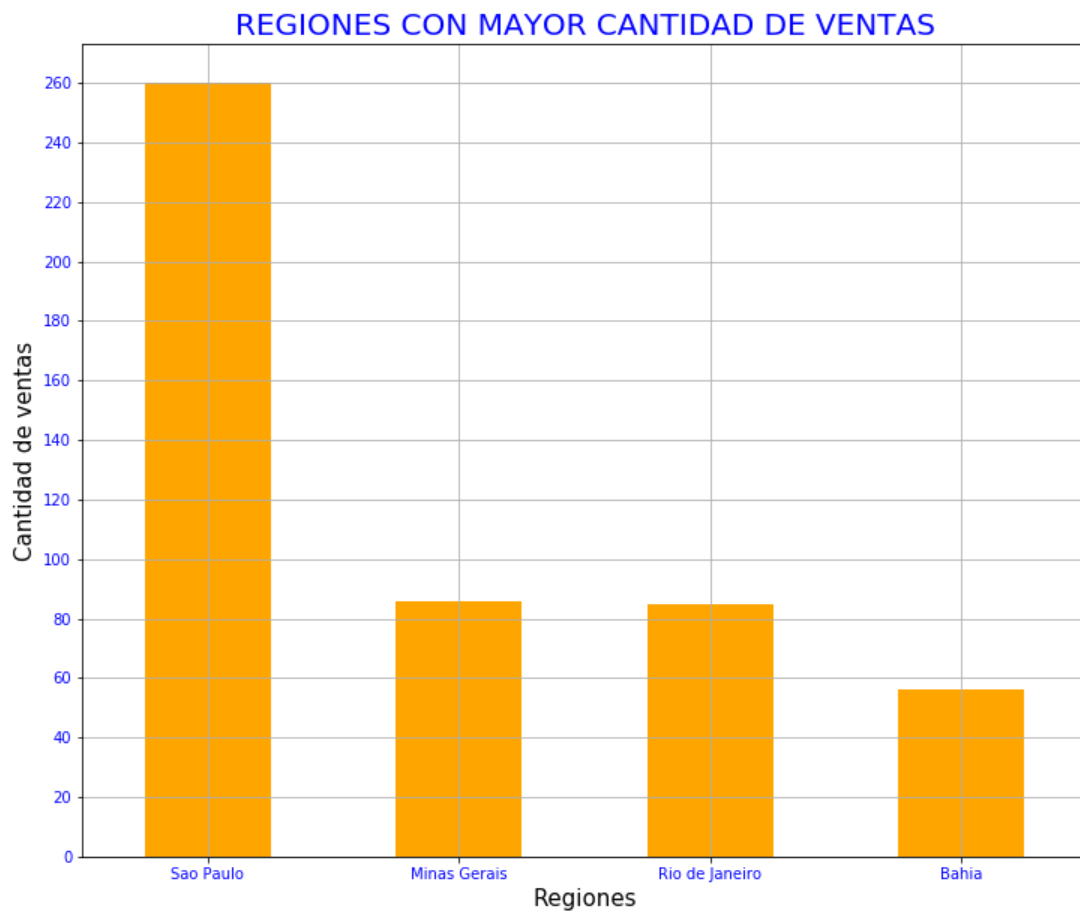


Analizando ahora la suscripción de usuarios al servicio “informarme cuando este producto esté disponible”, vemos que lo más solicitado son los iPhone y que no están disponibles. **iPhone posee mucha demanda y no la suficiente oferta.**

2.9.4 Top 10 términos más buscados



2.10.1 Regiones con mayor cantidad de ventas



En esta visualización se pueden observar las cuatro regiones con mayor cantidad de ventas. Las mismas se corresponden con las regiones más pobladas de Brasil por lo tanto es lógico que se de esta distribución ya que a mayor cantidad de gente, mayor posibilidades de compra.

2.10.3 Distribución de las ventas en las regiones

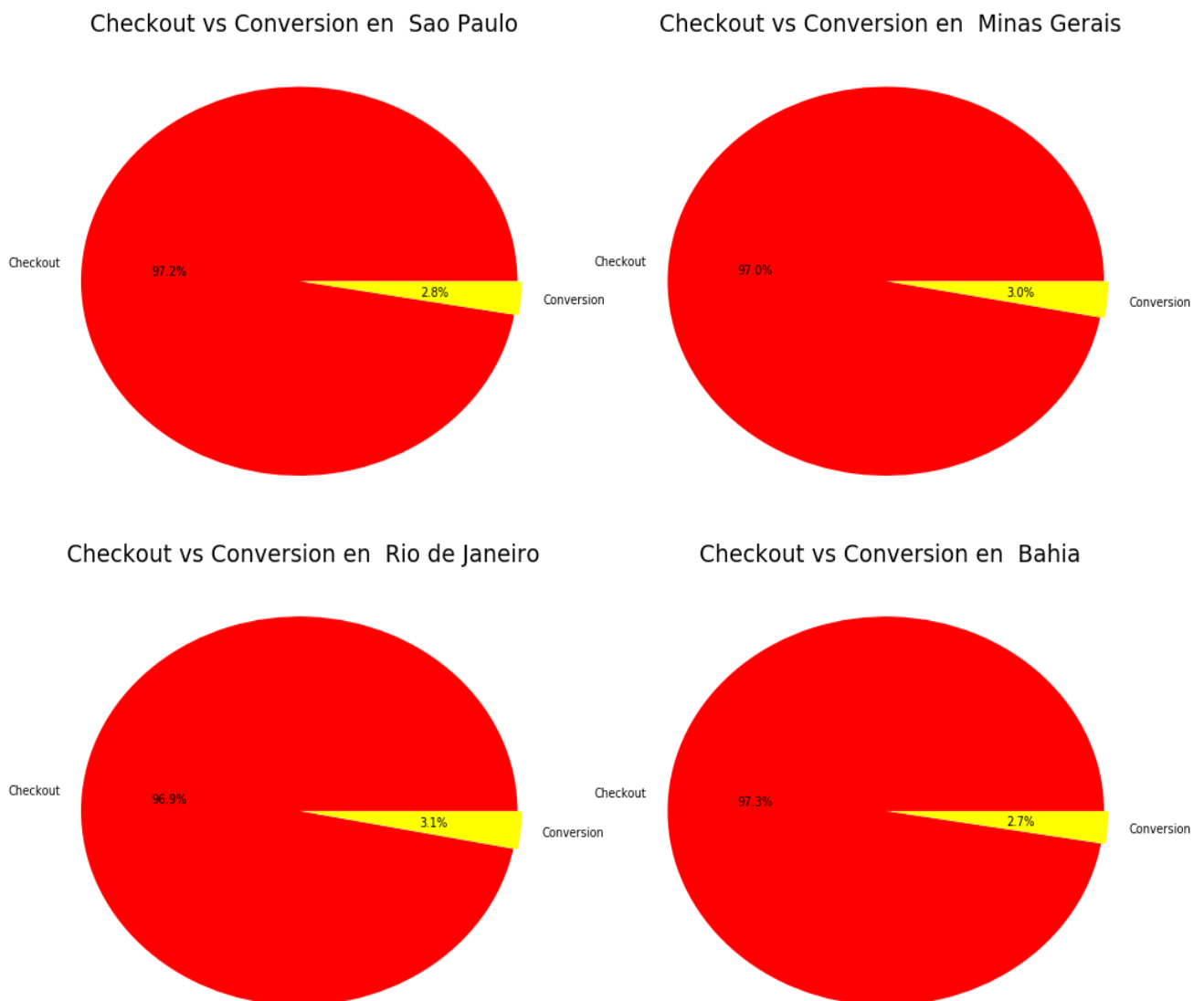


En este gráfico se puede ver que tanto en la región de Sao Paulo como la de Minas Gerais, se registra una mayor concentración de ventas en los meses de Mayo y Abril, la misma tendencia vista en la sección **2.32**.

Mientras que en las otras 2 regiones, se observan una cantidad de ventas más uniforme a lo largo de los últimos 6 meses.

Como conclusión, podemos ver que en todas las regiones se observa una disminución de las ventas hacia el mes de Junio y en los meses de Febrero y Marzo.

2.10.4 Porcentaje de Checkout vs Conversión



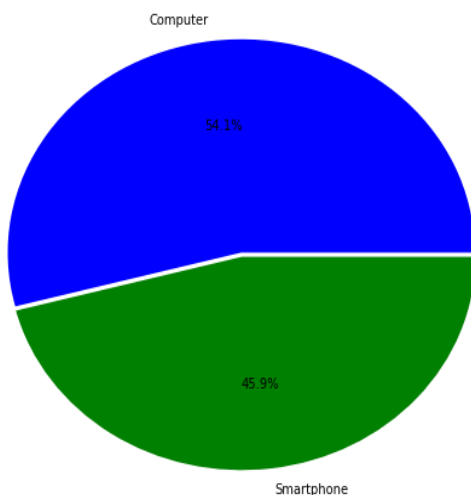
En la sección **2.4.4** se observó una gran diferencia entre los checkout y las conversiones realizados por día de la semana. En esta visualización se quiere

mostrar si la misma relación se mantiene en las regiones con mayor cantidad de ventas.

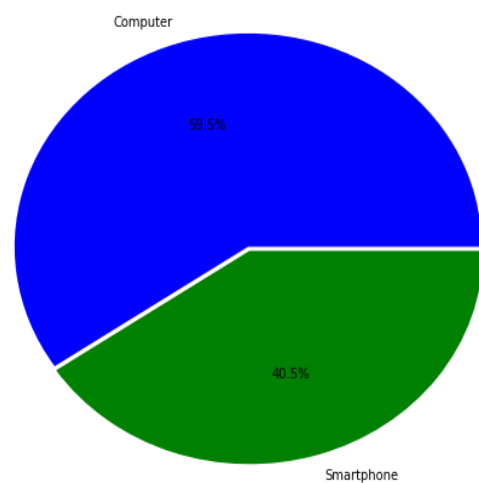
Como bien se ve el porcentaje de checkout es muy superior al de conversiones, determinando que hay un grave problema a la hora de concretar una venta. Esto puede ser causado por el dispositivo por el cual el usuario se conecta a la página. Por esa razón vamos a analizar desde qué dispositivo se realizan la mayor cantidad de ventas.

2.10.5 Dispositivos desde los que se realizan las ventas

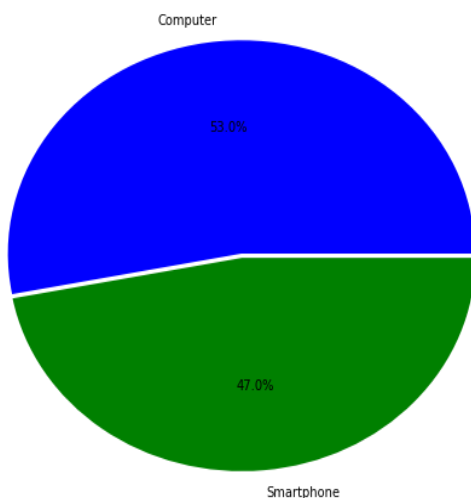
Computer vs Smartphone en Sao Paulo



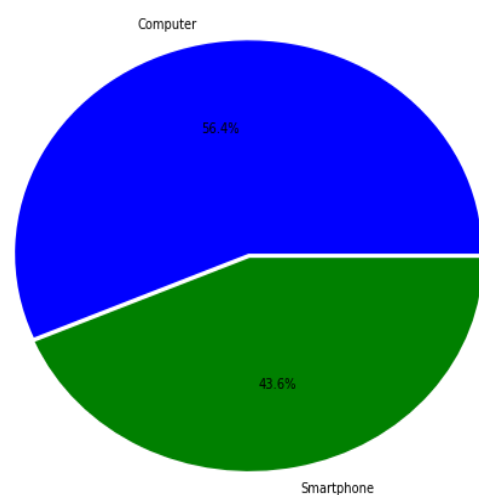
Computer vs Smartphone en Minas Gerais



Computer vs Smartphone en Rio de Janeiro



Computer vs Smartphone en Bahia



Se observa que en todas las regiones donde se registraron la mayor cantidad de ventas, no hay un dispositivo que concentre la mayoría de las mismas. Es más se visualiza una gran paridad entre ambos dispositivos, teniendo las computadoras una cierta ventaja.

Por lo tanto no podemos determinar que la causa de la gran cantidad de checkouts incompletos se deba al dispositivo por el cual el usuario se conecta a la página.

3. Conclusiones

Luego de realizar el análisis de los datos provistos por Trocafone, podemos concluir lo siguiente:

1. La cantidad de eventos de checkouts es muy superior a la cantidad de conversiones, esta situación se da hasta en las regiones más grandes y no se ve afectada por el dispositivo con el cual se conecta el usuario. Tratar este tema es de suma importancia.
2. Con respecto a los dispositivos, las ventas están muy igualadas tanto para los usuarios que se conectan mediante una computadora como los que lo hacen a través de sus Smartphone. Por lo tanto consideramos que sería un gran avance si Trocafone hiciera una app que le facilitara el movimiento por la página a los usuarios que ingresan mediante el celular.
3. Las marcas que más ventas registran son iPhone y Samsung. En el caso particular de los iPhone, sus celulares son más demandados pero Trocafone no tiene la suficiente oferta para satisfacer esos pedidos. El solucionar este hecho, puede verse en un gran incremento de sus ventas.
4. El sitio experimenta muy poca actividad durante los fines de semana, una posible mejora sería ofrecer descuentos y promociones para aumentar el número de usuarios.
5. Se nota un incremento en las ventas en fechas especiales, en este caso en el día de la madre. Aprovechar estos eventos, sería de gran utilidad para fomentar las ventas y la cantidad de usuarios.
6. La mayoría de los usuarios que entran por primera vez a la página lo hacen a través de la publicidad que Trocafone tiene en otros sitios, por lo tanto es importante invertir en las campañas que tuvieron más éxito (Google, Criteo y RTB House)
7. Los usuarios tienen una preferencia por los artículos con una buena o muy buena condición y el color que más ventas tiene es el negro. Se debería dar importancia a estos datos a la hora de incorporar nuevos productos.
8. Los usuarios que ya poseen un celular con el sistema operativo Android son más proclives a comprar un teléfono con el mismo sistema. Por lo tanto si se

implementara un sistema de recomendación dentro de la pagina, esto debería tomarse en cuenta.