



FACULTAD DE INGENIERIA

Universidad de Buenos Aires

Universidad de Buenos Aires
Facultad de Ingeniería

75.06 Organizacion de Datos
2do Cuatrimestre 2018
Grupo 3

Gaido, Nicolás	100856
Sallis, Nicolás	87236
Tusca, Bautista	92028

Link de github: <https://github.com/btusca/7506/tree/master/TP2018>

Introducción

Datos de entrada

El análisis se realizó sobre un conjunto de eventos de web analytics de usuarios que visitaron www.trocafone.com, la plataforma de ecommerce de Brasil de la empresa Trocafone. Trocafone es una empresa que se dedica a la compra y venta de dispositivos electrónicos, y se encuentra operando actualmente en Brasil y Argentina, con una propuesta interesante de compra y venta de dispositivos móviles usados, fomentando un consumo consciente y preocupado por el medio ambiente.

Como parte del análisis, realizamos búsquedas relacionadas con la empresa para entender su contexto de negocio y así poder apuntar mejor qué analizar.

Enfoque

La visión del análisis fue dirigida a mejorar las ventas y conocer más al cliente de Trocafone.

Herramientas

Se utilizó la Suite de Anaconda con Jupyter Notebook bajo lenguaje Python en conjunción con las bibliotecas especializadas en el análisis y manejo de datos Pandas y Numpy, y Seaborn y Matplotlib para realizar visualizaciones.

A su vez, se trabajó con aplicaciones de paradigma colaborativo en línea para la documentación (Google Docs) y versionado (GitHub: <https://github.com/btusca/7506>).

Preparando el dataset

El dataset utilizado cuenta con los siguientes atributos:

- **timestamp:** Fecha y hora cuando ocurrió el evento. (considerar BRT/ART).
- **event:** Tipo de evento
- **person:** Identificador de cliente que realizó el evento.
- **url:** Url visitada por el usuario.
- **sku:** Identificador de producto relacionado al evento.
- **model:** Nombre descriptivo del producto incluyendo marca y modelo.
- **condition:** Condición de venta del producto
- **storage:** Cantidad de almacenamiento del producto.
- **color:** Color del producto
- **skus:** Identificadores de productos visualizados en el evento.
- **search_term:** Términos de búsqueda utilizados en el evento.
- **staticpage:** Identificador de página estática visitada
- **campaign_source:** Origen de campaña, si el tráfico se originó de una campaña de marketing
- **search_engine:** Motor de búsqueda desde donde se originó el evento, si aplica.
- **channel:** Tipo de canal desde donde se originó el evento.
- **new_vs_returning:** Indicador de si el evento fue generado por un usuario nuevo (New) o por un usuario que previamente había visitado el sitio (Returning) según el motor de analytics.
- **city:** Ciudad desde donde se originó el evento
- **region:** Región desde donde se originó el evento.
- **country:** País desde donde se originó el evento.
- **device_type:** Tipo de dispositivo desde donde se generó el evento.
- **screen_resolution:** Resolución de pantalla que se está utilizando en el dispositivo desde donde se generó el evento.
- **operating_system_version:** Versión de sistema operativo desde donde se originó el evento.
- **browser_version:** Versión del browser utilizado en el evento

Para realizar un mejor análisis, realizamos una limpieza de datos:

- *timestamp* fue formateado a un tipo de dato de datetime.
- *model* fue separado en dos columnas para identificar la marca de los móviles.
- *weekday* fue agregado como atributo que se desprende del timestamp para apuntar directamente a los días de la semana. *hour* y *month* fueron agregados de la misma manera.

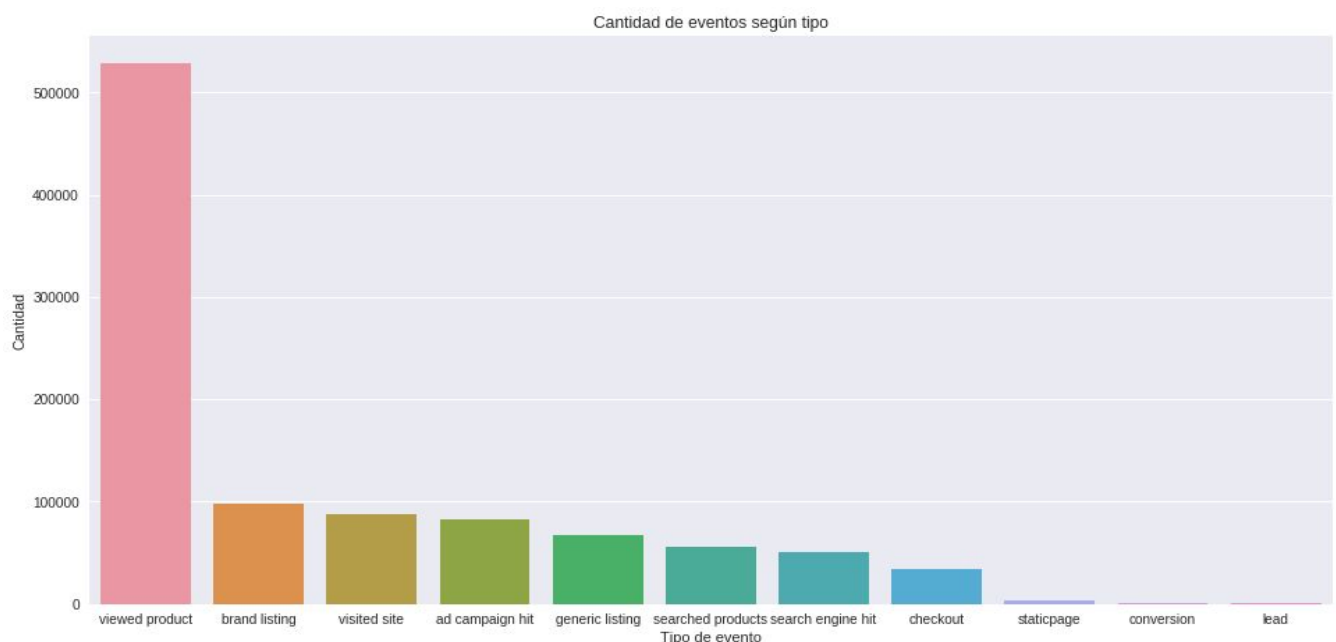
Analizamos los datos nulos y los categorizables. Estudiamos el contenido de los registros distribuidos a lo largo del tiempo: el rango temporal contenido en los registros ronda de: 2018-01-01 07:32:26 como fecha mínima y 2018-06-15 23:59:31 como fecha máxima. Con este dominio de datos, iniciamos nuestro análisis.

Tipos de eventos

El set de datos cuenta con 11 tipos de eventos:

- **“viewed product”**: El usuario visita una página de producto.
- **“brand listing”**: El usuario visita un listado específico de una marca viendo un conjunto de productos.
- **“visited site”**: El usuario ingresa al sitio a una determinada url.
- **“ad campaign hit”**: El usuario ingresa al sitio mediante una campana de marketing online.
- **“generic listing”**: El usuario visita la homepage.
- **“searched products”**: El usuario realiza una búsqueda de productos en la interfaz de búsqueda del site.
- **“search engine hit”**: El usuario ingresa al sitio mediante un motor de búsqueda web.
- **“checkout”**: El usuario ingresa al checkout de compra de un producto.
- **“staticpage”**: El usuario visita una página
- **“conversion”**: El usuario realiza una conversión, comprando un producto.
- **“lead”**: El usuario se registra para recibir una notificación de disponibilidad de stock, para un producto que no se encontraba disponible en ese momento.

Estos eventos se distribuyen a lo largo del dataset de la siguiente manera:



Hay un amplio predominio del tipo de evento Viewed product, con el 52% de los registros.

A continuación la proporción que ocupa cada evento sobre el total:

viewed product	0.523027
brand listing	0.097534
visited site	0.086403
ad campaign hit	0.081902

generic listing	0.066780
searched products	0.055447
search engine hit	0.050388
checkout	0.033358
staticpage	0.003558
conversion	0.001159
lead	0.000443

Cantidad de usuarios y usuarios que hacen checkout

Analizando el set de datos, obtuvimos que el total de usuarios distintos es de 27.624.

Después, revisamos la cantidad de usuarios distintos que realizaron checkout y concluimos que todos llegaron a la instancia de checkout.

Cantidad de compras (conversiones) en relación con checkouts

Haciendo un análisis más es profundidad, identificamos que del total de casos de checkouts, sólo en 1172 se efectivizó la compra.

Es decir que sólo 3.47% de los usuarios que llegó al checkout, realizó la compra.

Otro dato llamativos es la proporción de los iPhones “checkouteados” respecto a los “convertidos”, que es de aproximadamente la mitad que en otras marcas:

Modelo	Cant. de checkouts	Cant. de conversiones	Tasa de conversión por checkout
Samsung	14707	641	4,36%
iPhone	14082	323	2,29%
Motorola	3738	149	3,99%

Notificaciones de teléfonos fuera de stock

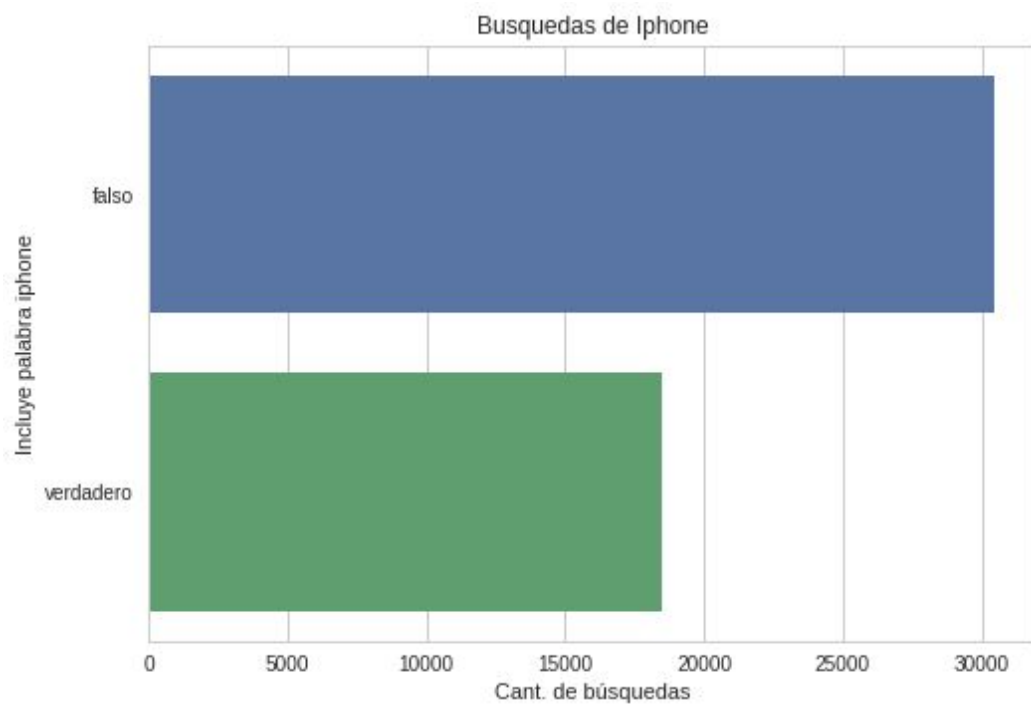
Uno de los eventos que menos ocurre es el de registrarse para recibir novedades de algún dispositivo que no se encuentra en Stock. De todos modos, analizamos estos casos, y vimos que 291 personas (de 27.624) se registraron. De ellas, ¡hubo una que se registró para recibir notificaciones 16 veces!.

Los teléfonos más buscados fueron iPhones, particularmente el iPhone 6:

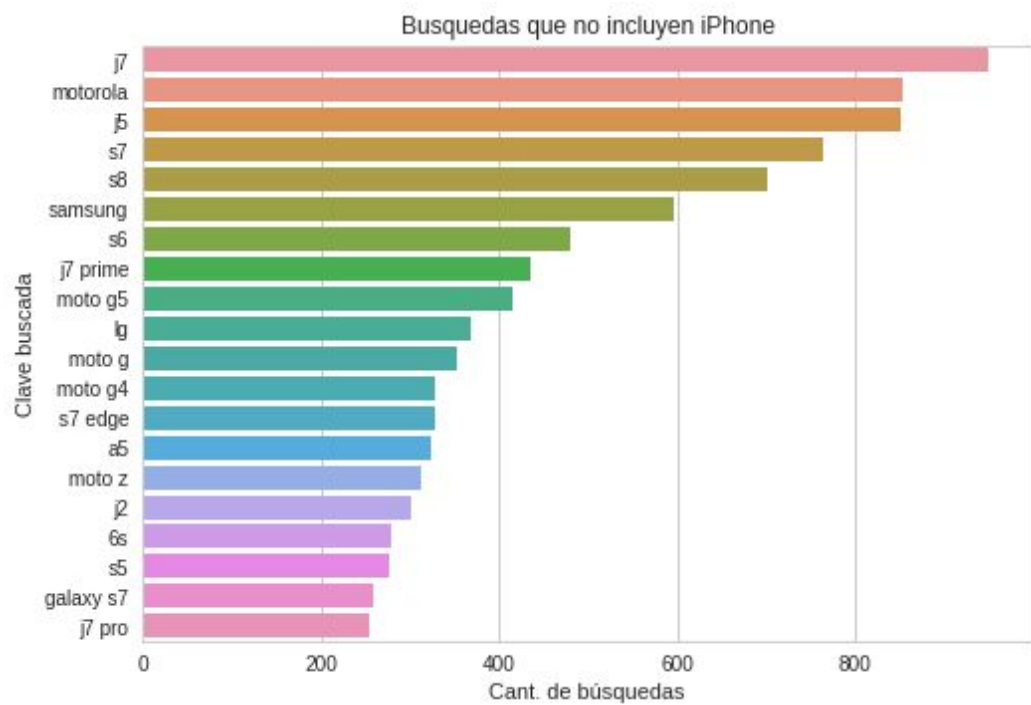
iPhone 6 Plus	35
iPhone 6S Plus	31
Samsung Galaxy J7 Prime	26
iPhone 6S	21
iPhone 8	20

Búsquedas realizadas

Analizando los patrones de búsqueda, observamos que de 48.967 búsquedas, más del 37% incluía la palabra iPhone (18.508 búsquedas):



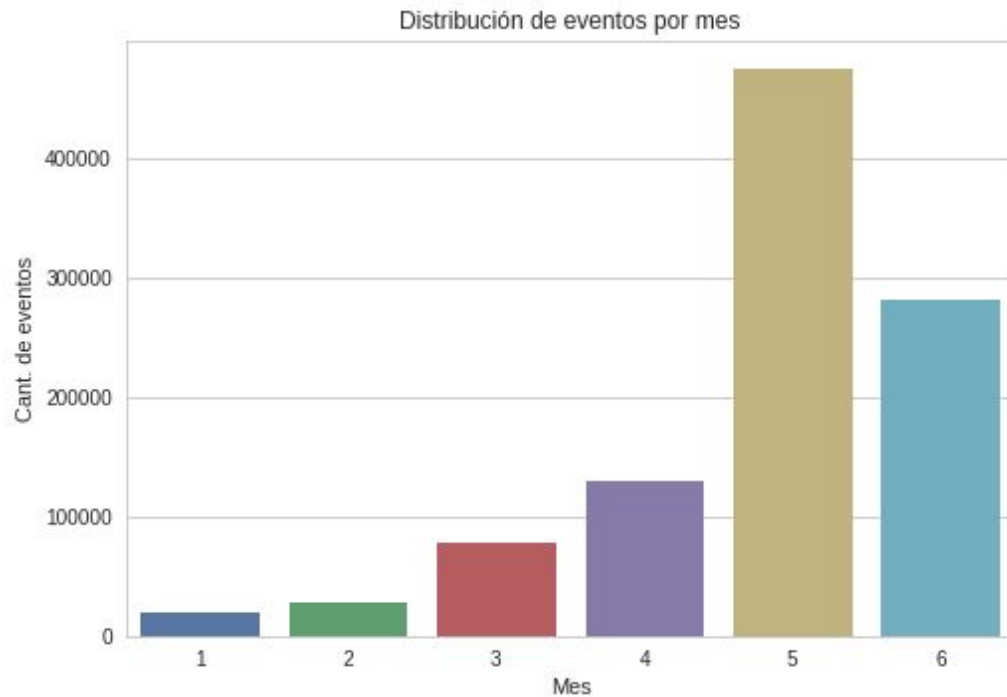
De las búsquedas que no incluían la palabra iPhone, quedaron las siguientes como claves más buscadas:



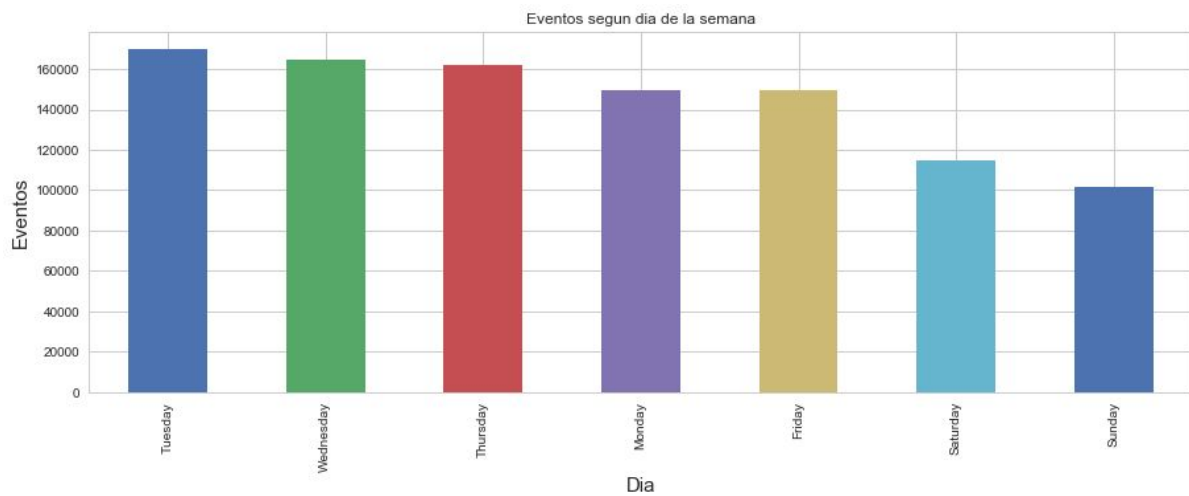
Análisis temporal

El rango temporal contenido en los registros ronda de: 2018-01-01 07:32:26 como fecha mínima, y 2018-06-15 23:59:31 como fecha máxima.

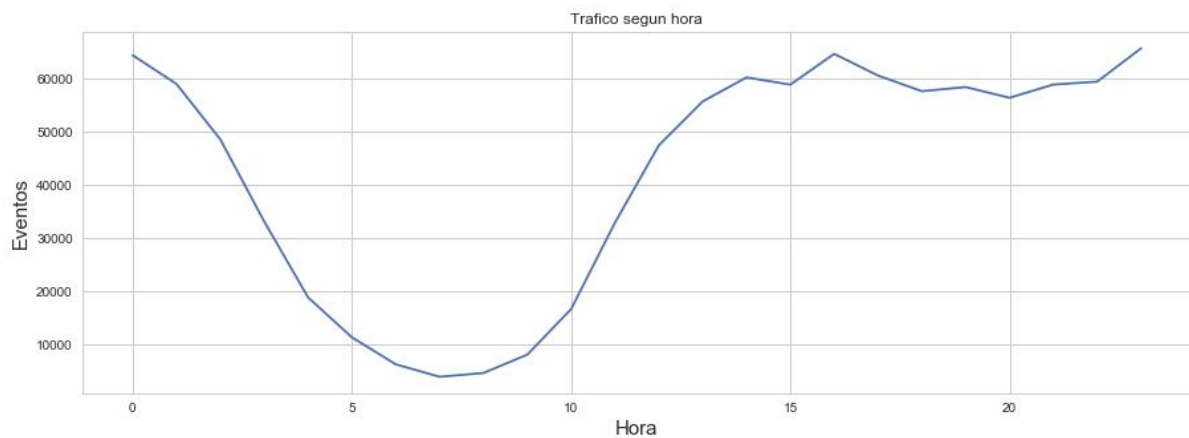
De aquí extraemos la siguiente información:



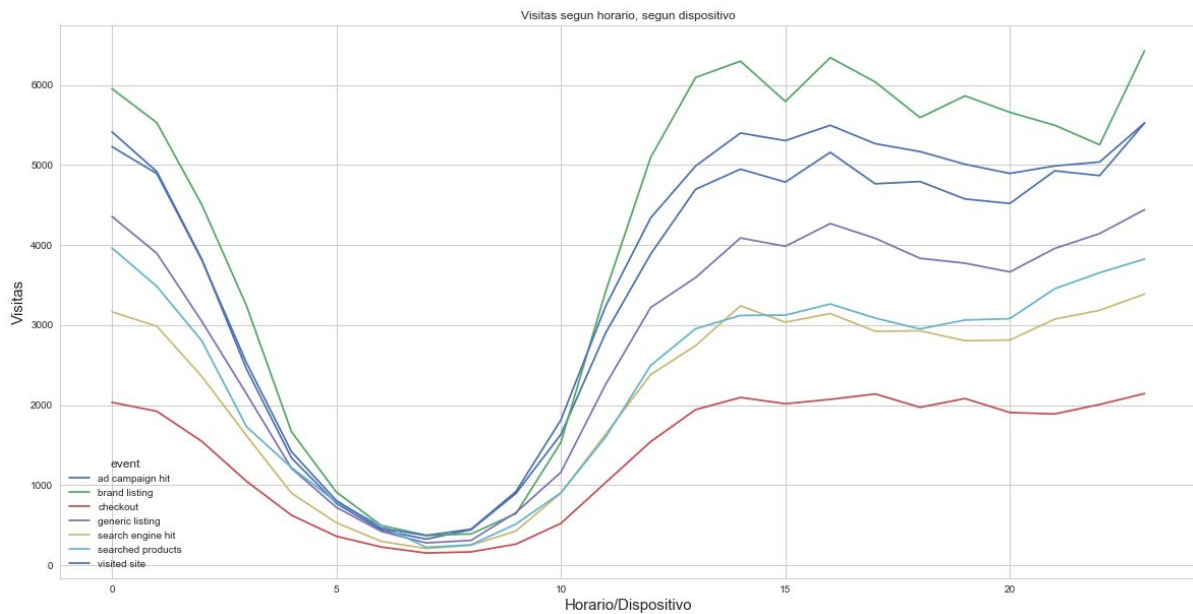
Observamos que el volumen de tráfico en la página web de Trocafone pareciera ir incrementando con el tiempo. También observamos que el mes con el mayor tráfico resulta ser Mayo, pero teniendo en cuenta que se tienen datos hasta mitad de Junio, nos arriesgamos a suponer que de tener los datos del mes completo, se mantendrá la tendencia de alza.



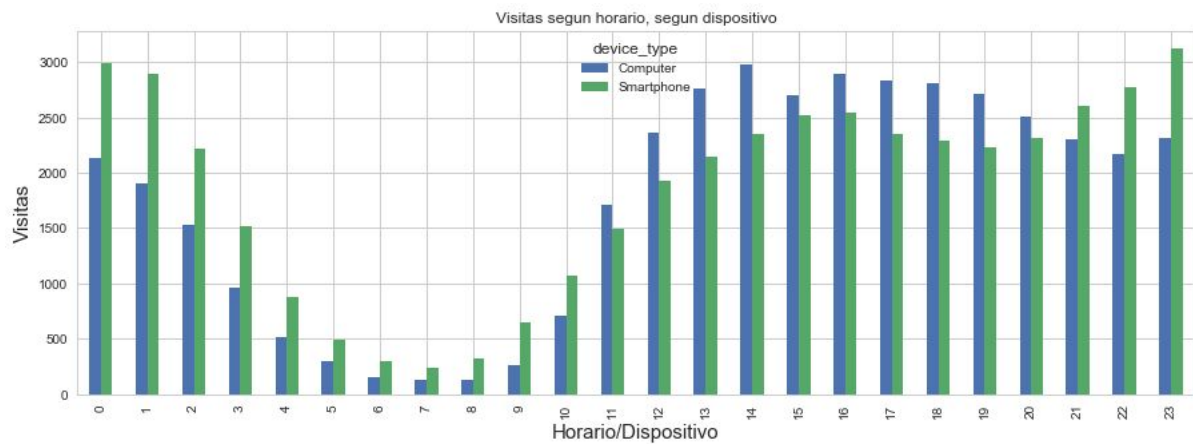
Observamos que la página posee un menor tráfico durante los fines de semana.



El análisis realizado cumple con lo esperado, que es un decremento en el tráfico a partir de las 00:00 hs, y hasta las 07:00 hs, debido a que es el horario de descanso.



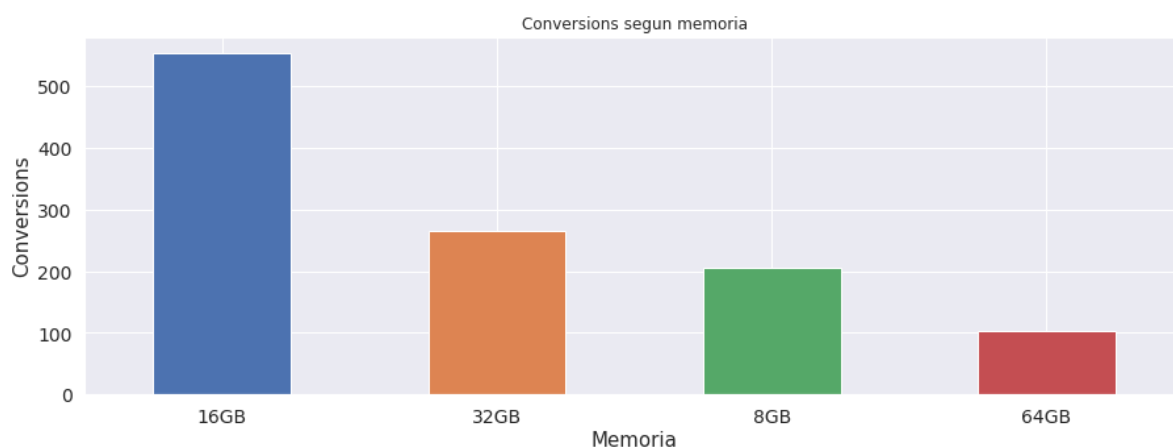
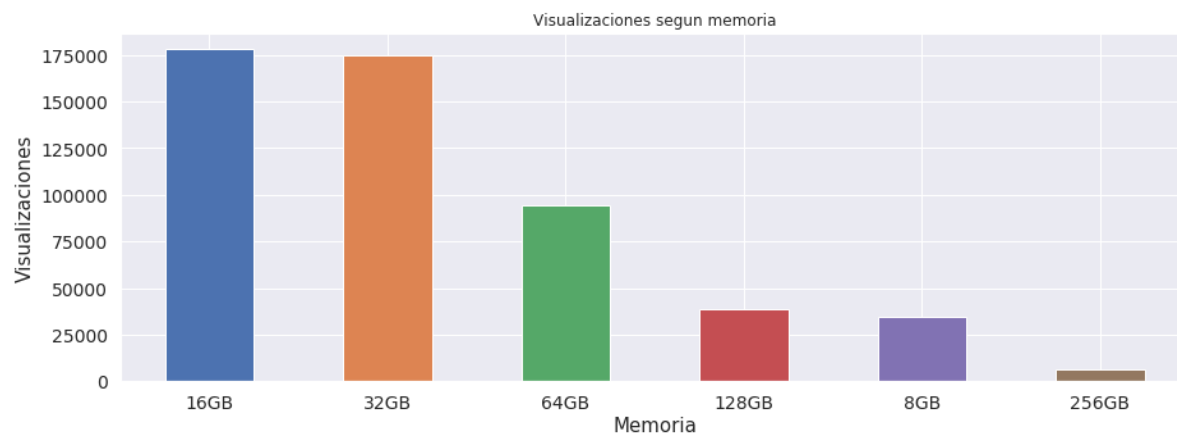
La tendencia se mantiene para todos los distintos tipos de eventos.



De aquí observamos que el tráfico según dispositivo varía, nuestra suposición es que en la franja horaria donde el tráfico de computadoras es mayor (11:00-20:00 hs) los usuarios se encuentran en sus lugares de trabajo, donde es normal el uso de computadoras, mientras que en el resto del día, calificable como "tiempo de ocio" los usuarios prefieren acceder desde la comodidad de sus dispositivos móviles.

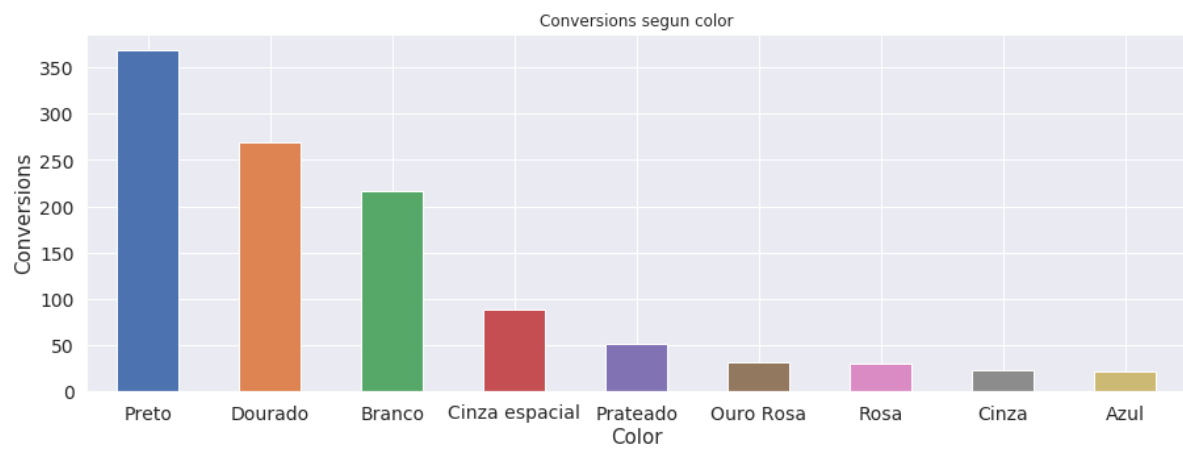
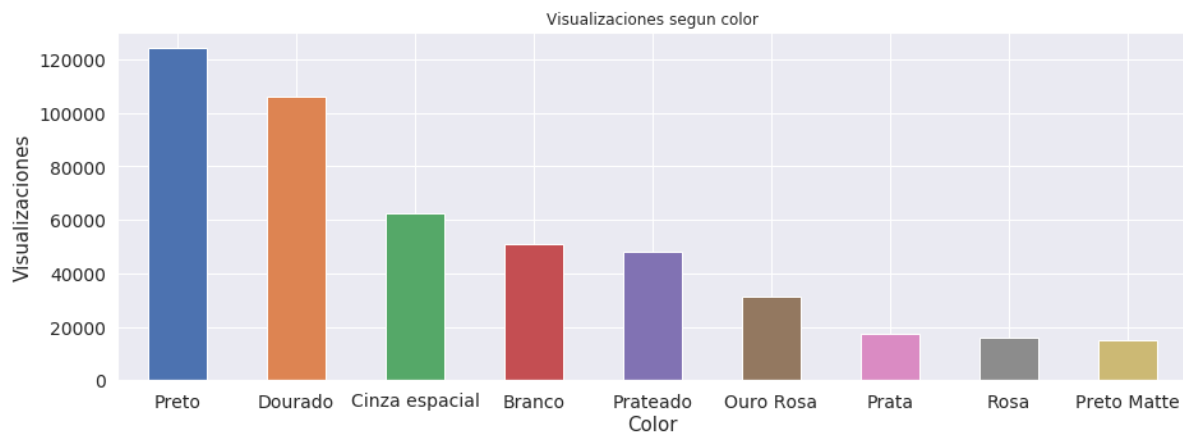
Analisis segun características de producto

Memoria



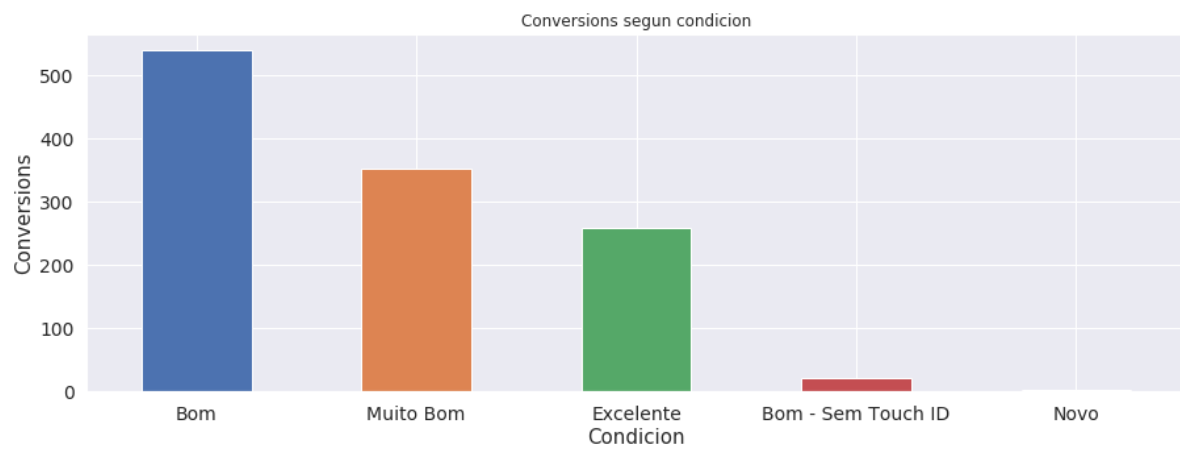
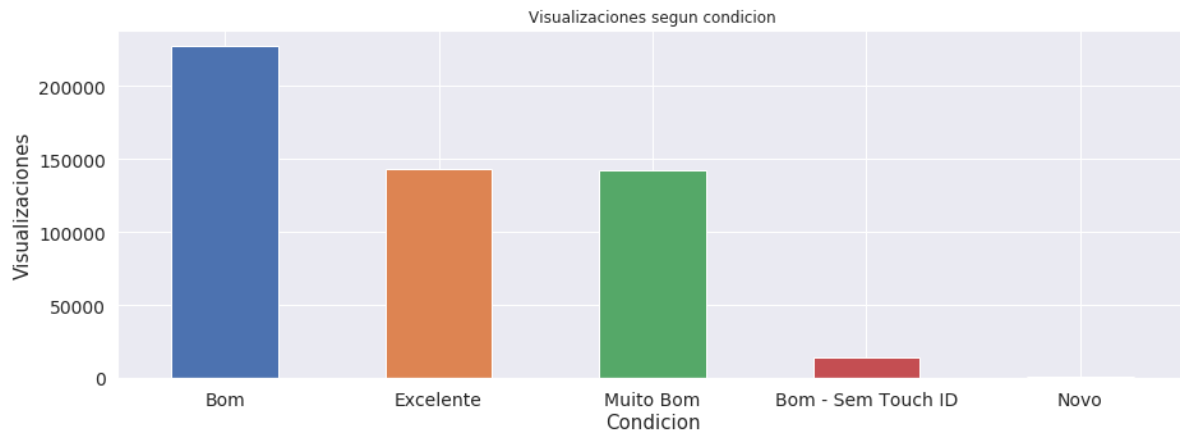
Observamos que es similar la cantidad de visualizaciones a productos con 16, o 32 GB de memoria, pero a la hora de realizar la compra, la gente opta por los productos de 16 GB, muy probablemente debido a que estos poseen, en general, un precio menor.

Color



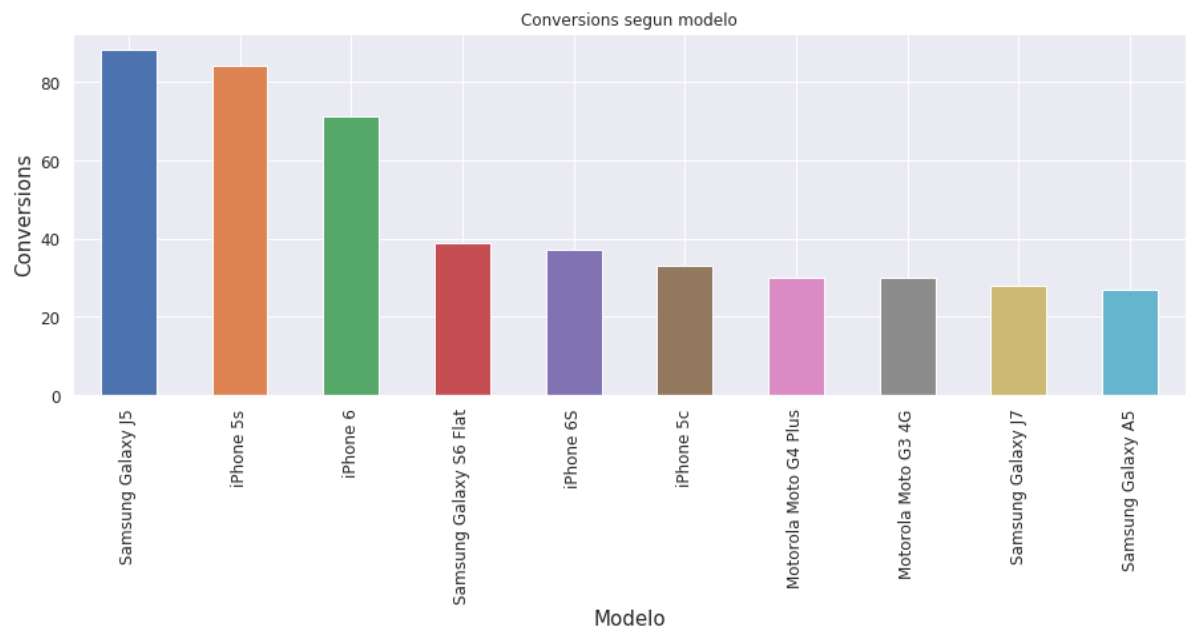
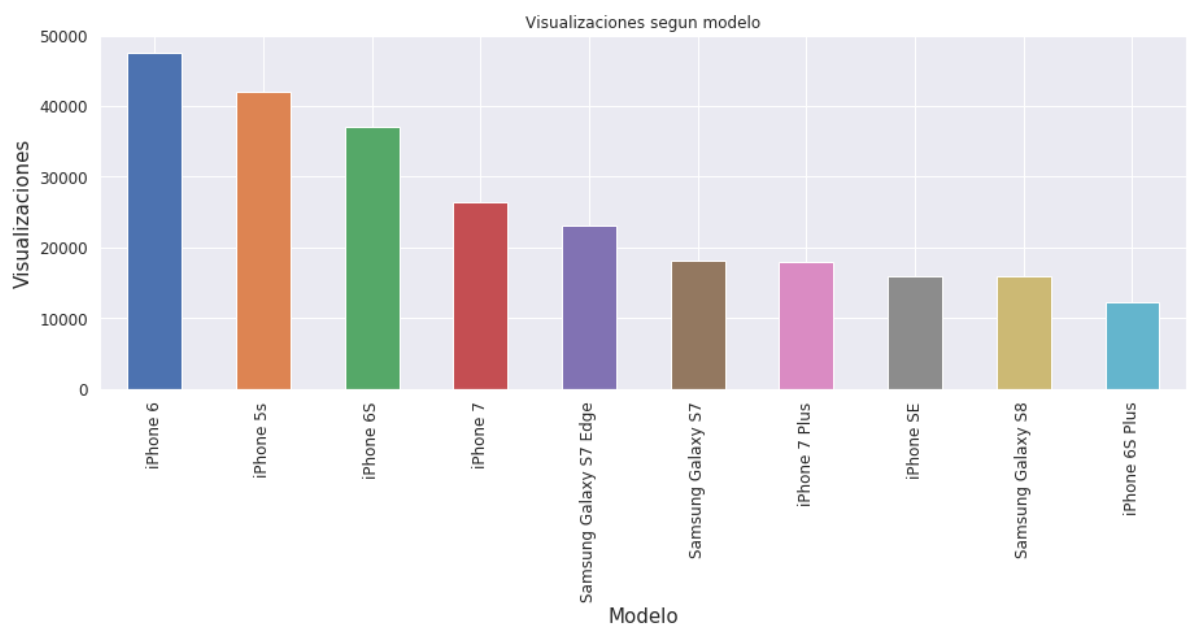
Observamos que el interés es mayor para los celulares color preto(negro). Nuestra teoría es que según la concepción de la sociedad, el negro trae asociado la imagen de sobriedad, y seriedad, por lo tanto, los usuarios prefieren este tipo de productos para dar una mejor imagen.

Condición

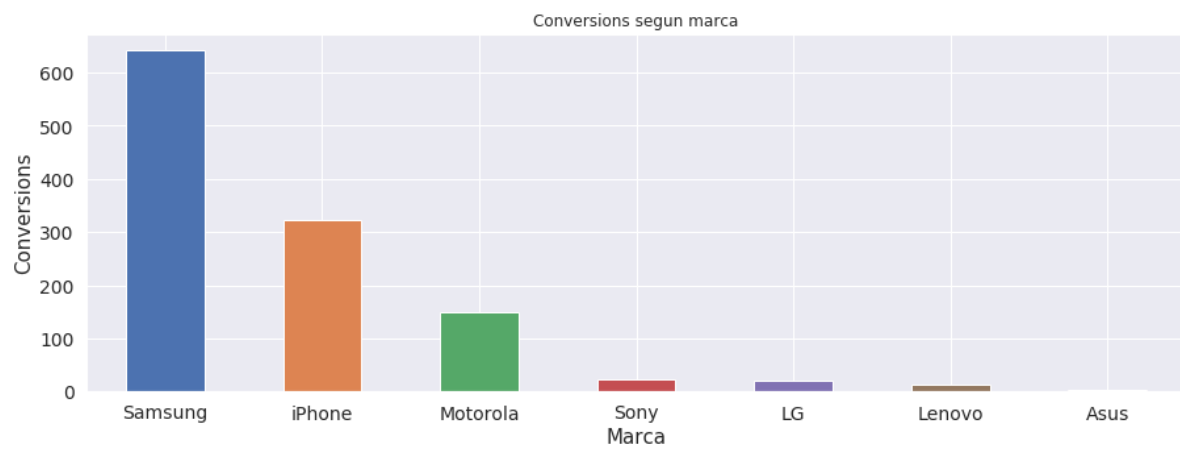
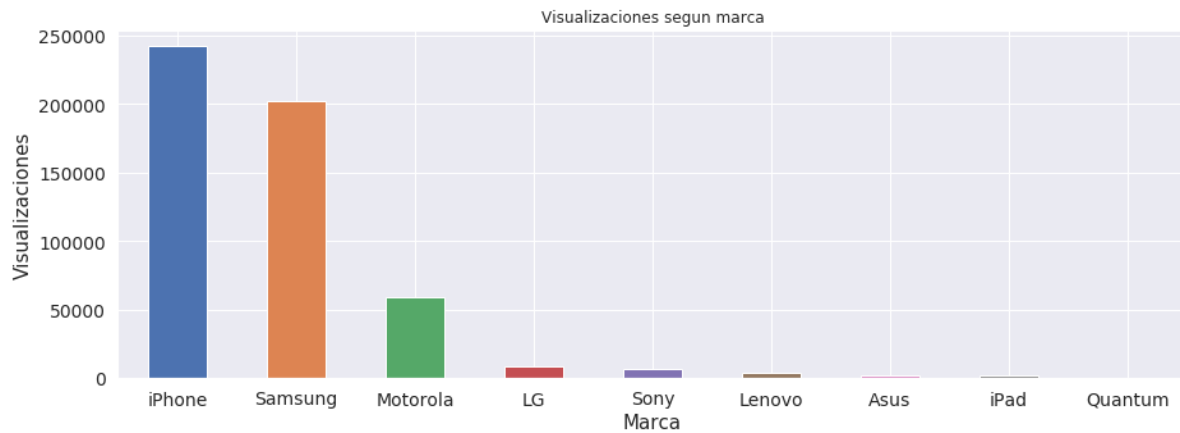


Creemos que la condición del producto está firmemente relacionada con el valor económico de cada uno de los productos y como vimos en el caso de la memoria, creemos que los usuarios de Trocafone deciden optar por productos con un costo económico menor, lo que lleva a resultados como el que observamos.

Modelo

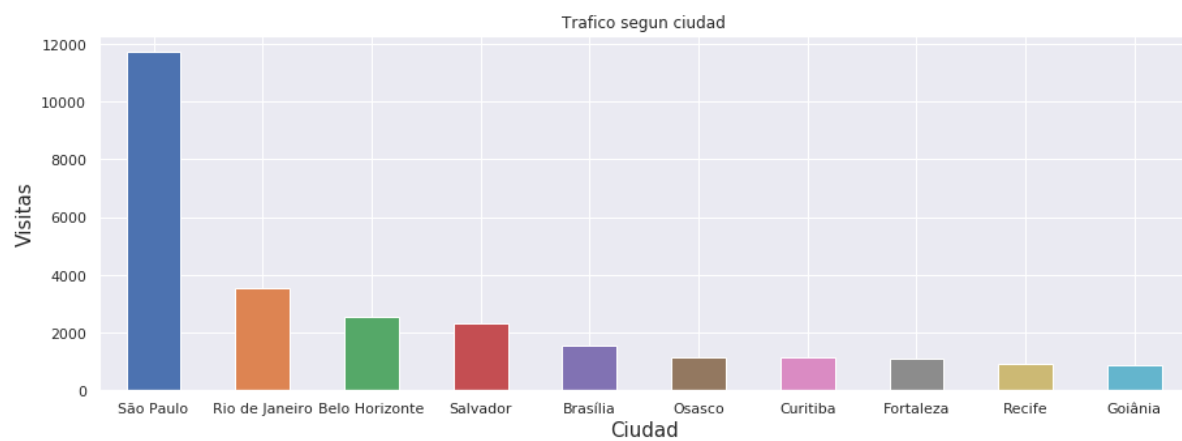
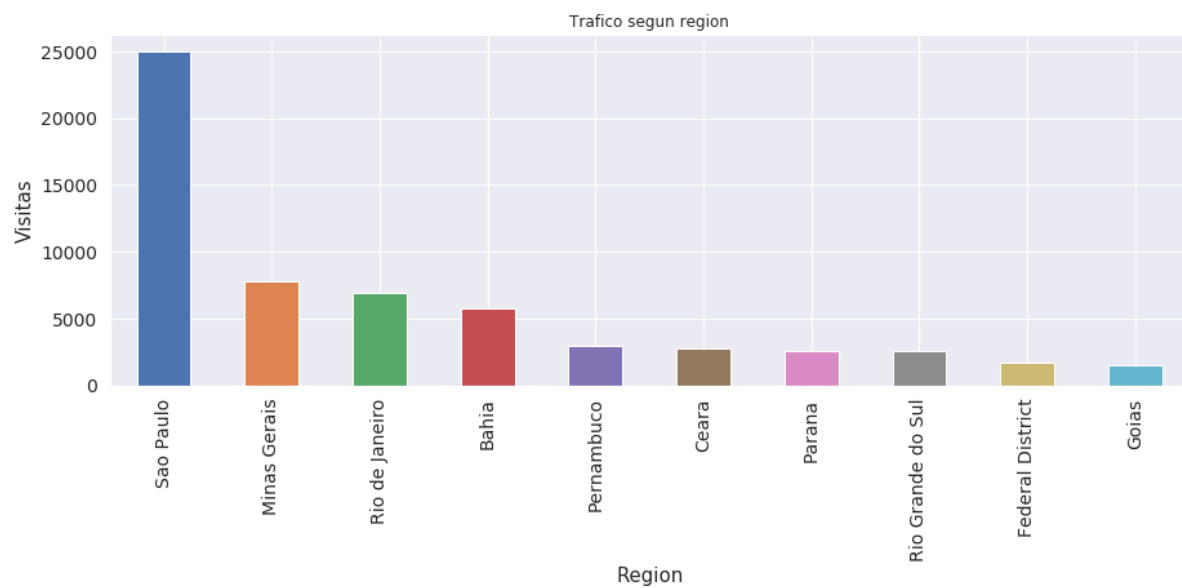


Marca



Observamos que se contraponen la cantidad de visualizaciones con la cantidad de compras, creemos que el público en general posee un mayor interés en los celulares del modelo Iphone, pero por algún motivo (podría ser su elevado precio en el mercado) terminan optando por celulares marca Samsung.

Análisis geográfico



Observamos que es inmensa la diferencia de tráfico desde Brasil en comparación a otros países. Véase la proporción de visitas:

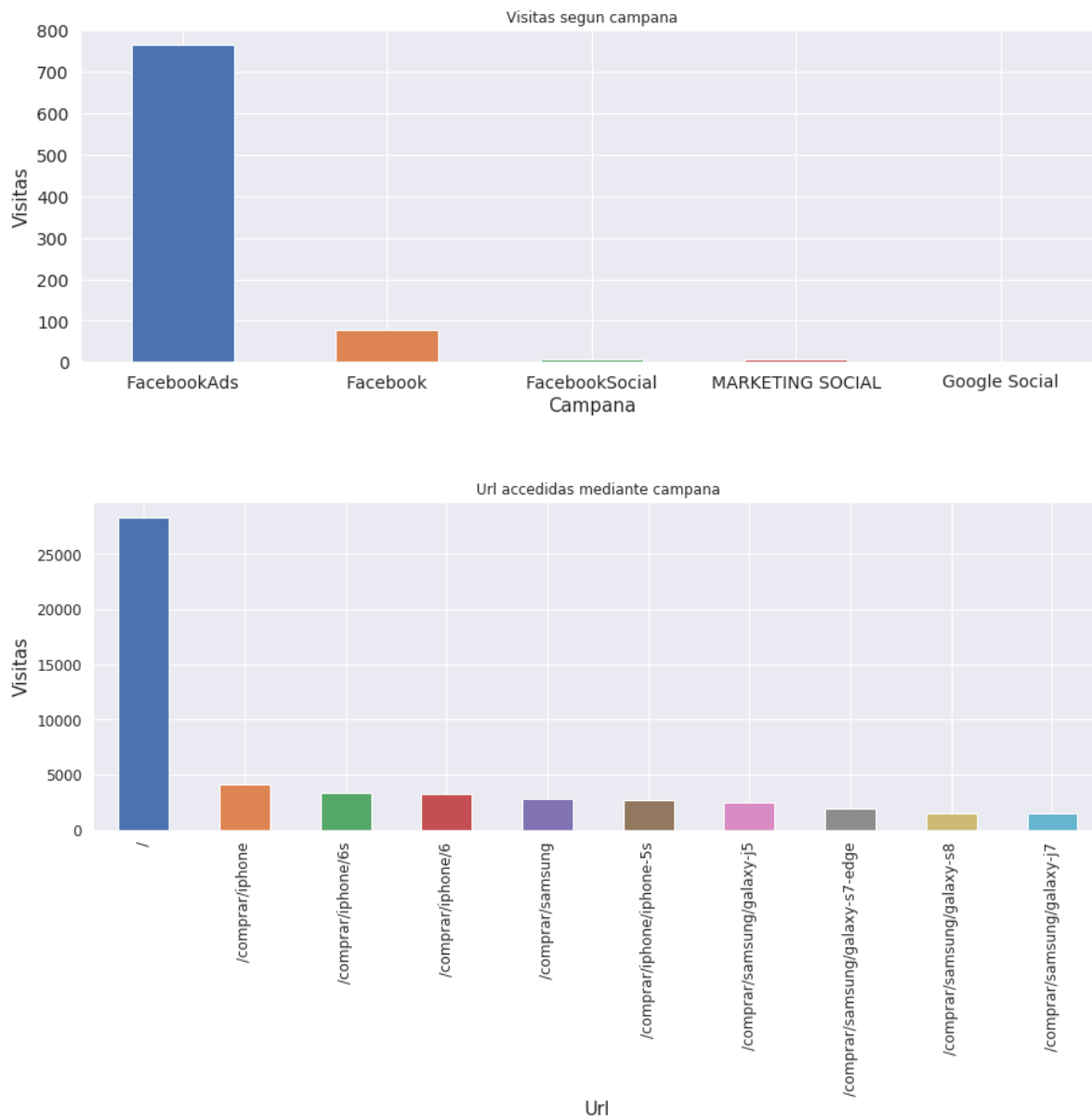
Brazil	0.991101
United States	0.005466
Argentina	0.001105
Canada	0.000411
Netherlands	0.000153
France	0.000153
United Kingdom	0.000141
Mozambique	0.000141
Japan	0.000129
Poland	0.000106

Naturalmente es mayor el tráfico de Brasil por ser el país de origen de Trocafone.

Además hemos omitido algunos países ya que al ser un volumen tan pequeño de visitas los consideramos despreciables.

Observando los datos, asumimos que Trocafone realiza una campaña mucho más exhaustiva en la región de Sao Paulo, en particular, en la ciudad de Sao Paulo.

Análisis de campaña



Podemos observar que el método de campaña más exitoso es, claramente Facebook Ads, seguido a su vez por otros dos canales pertenecientes a la misma compañía. De estos resultados creemos que en términos económicos sería óptimo usar solamente este método, que tiene mucho mejores resultados que los demás.

Además podemos ver que la gran mayoría de los usuarios que acceden mediante campaña son redireccionados a la página principal del sitio. A grandes rasgos, también podemos inferir que los usuarios de Facebook presentan un mayor interés en celulares modelo Iphone.

Conclusiones

Finalmente, creemos que a partir del análisis se desprenden las siguientes conclusiones a ser analizadas:

- El 100% de los usuarios en el set de datos suministrado (27.624) llegaron a la etapa de checkout. De estos casos, sólo el 3.47% realiza la compra (1.172).
Se debe evaluar la posibilidad de que la gran deserción en el interés en el producto se produce en esta etapa al conocer el precio del producto.
- El celular más buscado es el de la marca iPhone, habiendo sido buscada 18.508 veces. (37% del total de búsquedas) y tiene a los usuarios más ansiosos por recibir notificaciones por stock.
- El tráfico en el portal está en crecimiento.
- El tráfico a través de computadoras es mayor en el intervalo de 11:00-20:00 hs mientras que fuera de ese horario prevalece el uso del móvil.
- Los celulares más comprados son aquellos que tienen una memoria de 16gb, negros y en buen estado.
- El día con menos ventas en el Domingo.
- La región de Brasil con más éxito de inserción es en Sao Pablo, con casi 12.000 visitas, mientras que en las regiones de Osasco, Curitiba, Fortalea, Recife y Goiania tiene un tráfico casi 10 veces menor.
- El portal no solo interesa al público brasileño, si no tambien, en menor medida, al:
 - Estadounidense.
 - Argentino.
 - Canadiense.
 - Holandés.
 - Francés.

