

Trabajo Práctico N° 1: Análisis exploratorio de Datos Trocafone

[7506/9558] Organización de Datos
Segundo cuatrimestre de 2018

Nombre	Padron	Mail
Jorge Anca	82399	Ancajorgel@yahoo.com.ar
Federico Elias	96105	fedelias93@hotmail.com
Maria Soledad Escobar	97877	mariasolescobar21@gmail.com
Ignacio Pedro Komarnicki	98872	ignaciokomar@gmail.com

Grupo N° 23

Fecha de entrega: 24 de septiembre de 2018

Índice

1. Introducción	2
1.1. Objetivo del trabajo	2
1.2. Sobre la empresa	2
1.3. Sobre los datos	2
2. Modelo de Dominio	3
3. Análisis Generales	4
3.1. Análisis sobre países y ciudades de donde provienen los usuarios	4
3.2. Análisis sobre usuarios New y Returning	5
3.2.1. Análisis según el canal de acceso	5
3.2.2. Análisis sobre cantidad de visitas por hora	6
3.3. Análisis sobre motores de búsquedas	7
3.4. Análisis sobre la condición de venta de los productos	7
4. Análisis sobre los eventos	8
4.1. Conversiones y usuarios por eventos	8
4.2. Visitas y campañas de marketing	9
5. Análisis sobre modelos de los productos	10
5.1. Checkouts, vistas, conversiones y leads por modelo	10
5.2. Colores	14
5.3. Checkouts, leads, conversion sobre las vistas de cada modelo	15
5.4. Condiciones de los modelos	17
6. Análisis sobre las búsquedas de los usuarios	18
7. Análisis sobre los dispositivos de los usuarios	22
7.1. Resolución de pantalla computadora	22
7.2. Resolución de pantalla celular	23
7.3. Sistemas operativos	24
7.4. Tipo de dispositivos	25
8. Conclusiones	25

1. Introducción

1.1. Objetivo del trabajo

Aplicar los conceptos enseñados en la materia para realizar un análisis exploratorio sobre un conjunto de eventos de web analytics de usuarios que visitaron 'www.trocafone.com', su plataforma de ecommerce de Brasil.

Link GitHub: <https://github.com/IgnacioKomar/OrganizacionDeDatos>

1.2. Sobre la empresa

Trocafone es un side to side Marketplace para la compra y venta de dispositivos electrónicos que se encuentra actualmente operando en Brasil y Argentina. La empresa realiza distintas actividades que van desde la implementación de plataformas de trade-in (conocidos en la Argentina como Plan Canje), logística directa y reversa, reparación y recertificación de dispositivos (refurbishing) y venta de productos recertificados por múltiples canales (ecommerce, marketplace y tiendas físicas).

1.3. Sobre los datos

Para acotar el alcance del trabajo práctico analizaremos un subconjunto de eventos de web analytics que Trocafone tiene disponible para una cantidad representativa de usuarios que visitaron su plataforma de ecommerce. Los mismos se corresponde al período de tiempo desde el 1 de enero al 15 de junio de 2018.

Los datos se encuentran en un archivo CSV que contiene las siguientes columnas:

- 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

Por otro lado, los siguientes tipos de eventos se encuentran disponibles (en el campo event) sobre los cuales se brinda una breve descripción:

- “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 campaña 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.

2. Modelo de Dominio

A continuación hacemos un breve resumen de la empresa cuyos datos vamos a analizar para posibilitar un mejor entendimiento del modelo de dominio ya que es uno de los pilares en los que se basa data Science junto con las ciencias matemáticas y computacionales.

Trocafone implementa un modelo de negocio conocido como ReCommerce que plantea la compra, reacondicionamiento y venta de productos previamente usados. Su principal foco es solucionar la problemática de confianza entre personas al momento de realizar una transacción por un producto electrónico usado en un marketplace tradicional. Trocafone es líder de programas de Trade- Esta modalidad consiste en otorgarle al cliente el beneficio de un descuento para la compra de un nuevo producto a cambio de la entrega de uno previamente usado. Este tipo de programa puede ser aplicado en distintos tipos de negocios, conocido en algunos casos como “plan canje”. Desde el punto de vista tecnológico, Trocafone brinda a sus socios una plataforma que permite evaluar y cotizar un equipo para poder otorgar a los clientes un descuento en la compra de un dispositivo nuevo.

Esto se encuentra respaldado por un conjunto de microservicios que se encargan de:

- Verificar que el dispositivo esté libre de cualquier tipo de bloqueo que imposibilite su adquisición (por ejemplo que no se encuentre declarado en ninguna lista negra).
- Facilitar la interacción del usuario permitiendo la detección automática de su modelo.
- Permitirle a los socios la validación de los descuentos otorgados por el programa de trade-in de forma automática.
- Proceso de re-acondicionamiento de los equipos

Trocafone además de brindar la solución tecnológica, se encarga del soporte logístico para acompañar en estos programas a sus socios. Para ello debe gestionar la logística inversa ya que recibe los celulares de los usuarios desde distintas partes y por diferentes canales (por correo, personalmente en centros habilitados y debe llevarlos a los centros de reacondicionamiento para la evaluación de los mismos y reparación, en caso de ser necesaria.

3. Análisis Generales

3.1. Análisis sobre países y ciudades de donde provienen los usuarios

Analizando el set de datos se puede apreciar que la gran mayoría de los usuarios provienen de Brasil. Esto resulta esperable dado que la empresa es brasilera y el mercado brasileño es considerable

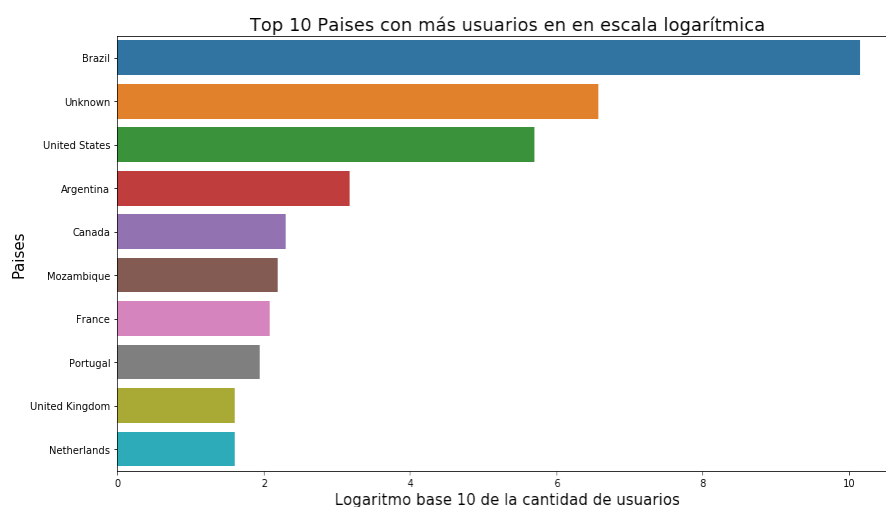


Figura 1: Top 10 países con más usuarios.

Por lo tanto todas las ciudades con mas usuarios son brasileiras.

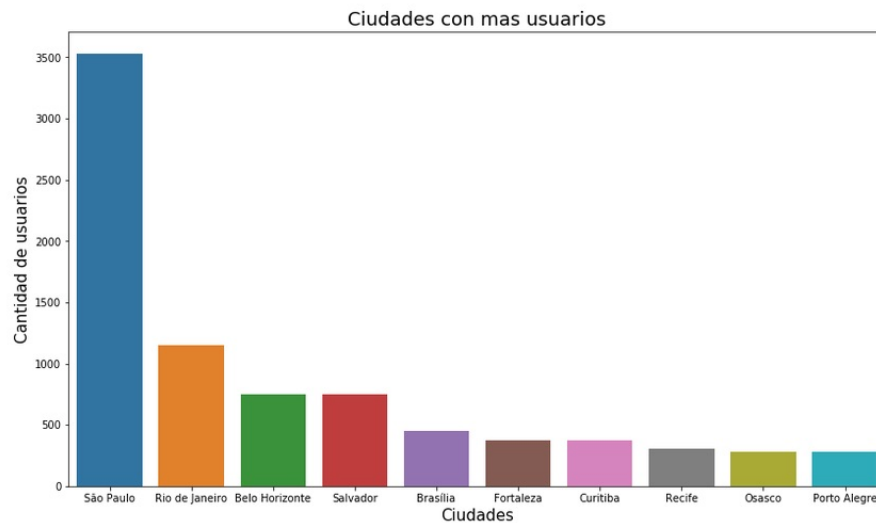


Figura 2: Top 10 ciudades com más usuarios.

Las ciudades con mas visitas coinciden bastante con las de mayor población, una sorpresa es Osasco que es la número 9 con mas usuarios pero figura como número 28 en el ranking de población de Brasil.

3.2. Análisis sobre usuarios New y Returning

Se puede considerar interesante analizar que tipos de usuarios generan los eventos. Consideramos dos tipos de usuarios "Newz returning", para identificar a los usuarios nuevos de los que ya habían visitado el sitio anteriormente. En el set de datos contamos con una columna que nos indica si el usuario es "New." Returning". Para analizar estos datos se decidió no tomar en cuenta a los valores nulos, ya que solo nos interesa ver el comportamiento de los usuarios nuevos y los que retornan.

3.2.1. Análisis según el canal de acceso

Comenzamos que cosas interesantes se pueden encontrar sobre los usuarios new contra los returning según el tipo de canal desde el cuál acceden al sitio. Para esto se hace un conteo para ver la cantidad de usuarios new y returning que hay según el canal desde el cuál acceden al sitio de la empresa y se obtiene la visualización siguiente.

Relación entre cantidad de usuarios new vs returning según el tipo de canal por el que acceden

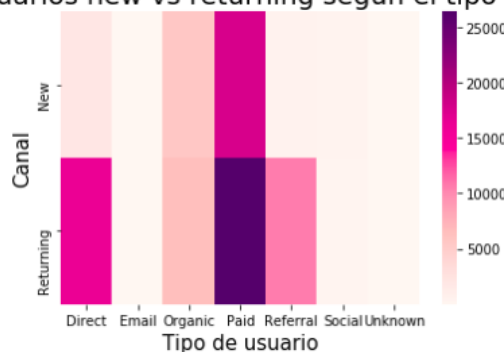


Figura 3: Visitas de usuarios New vs Returning según el canal de acceso.

Se puede observar que la mayoría de los usuarios nuevos visitan el sitio a través de un canal pago. Con los usuarios que ya habían visitado el sitio anteriormente se observan que acceden en gran medida desde un canal pago, pero también hay una gran parte de ellos que lo hace desde un canal directo. Entendiendo que un canal 'Paid' es un canal de publicidad pago, puede notarse la efectividad del mismo, contra los otros canales utilizados, para atraer a nuevos usuarios y a los que ya conocen el sitio a volver a visitar el sitio web. También se puede concluir que los usuarios frecuentes también optan por acceder a través de un canal directo. En tanto para la atracción de

3.2.2. Análisis sobre cantidad de visitas por hora

Ahora podríamos ver cuáles son los horarios en los que los usuarios realizan visitas al sitio. Para representar la información obtenida en ese análisis nos apoyamos en la visualización.

Cantidad de Usuarios New vs Returning que Visitan el Sitio por Hora

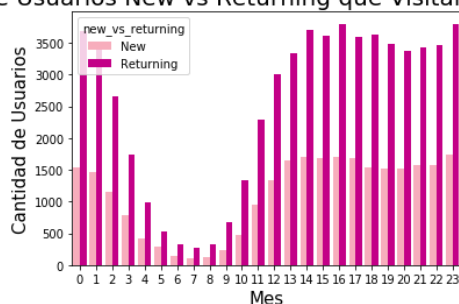


Figura 4: Visitas de usuarios New vs Returning según la hora del día.

Se puede observar que tanto los ingresos nuevos y los de usuarios frecuentes, aumentan a partir del medio día y crecen hasta tarde en la noche donde se alcanza un valor de más de 3500 usuarios. Esto podría deberse a las responsabilidades que tengan los mismos en su vida cotidiana. De esta manera también vemos que el pico mínimo se produce a las 7 de la mañana, con un valor de menos de 300 usuarios. Esta información podría ser útil para determinar en que rango horario es más posible que un potencial usuario o un usuario frecuente pueda interesarse en un aviso publicitario.

3.3. Análisis sobre motores de búsquedas

Nos puede interesar ver si hay algo particular en los motores de búsqueda que utilizan los usuarios. Para realizar este análisis vemos cual es el motor de búsqueda más popular. Para esto utilizamos los datos del evento "search engine hit", que es el único que contiene datos sobre motores de búsquedas, lo cual resulta lógico, pues es el evento en el cual los usuarios realizan búsquedas, por lo que descartamos los otros eventos. Para este análisis obtuvimos la siguiente visualización.

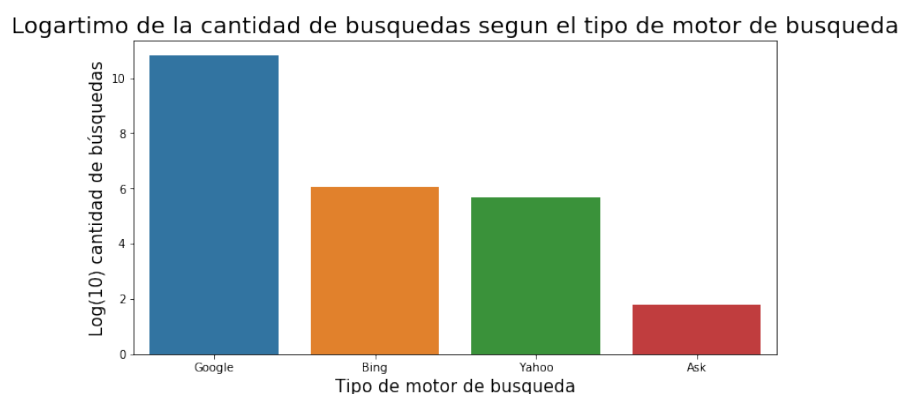


Figura 5: Logartimo de la cantidad de búsquedas para cada motor de búsqueda registrado.

La visualización es muy consistente, Google se lleva la gran parte de las búsquedas, con más de 50.000 búsquedas. Esto era de esperarse, pues es el motor de búsqueda más popular que existe, luego vemos a Bing y Yahoo que no difieren tanto en la cantidad de búsquedas, están entre un 280 y 430 búsquedas aproximadamente. Por ultimo se encuentra ask con menos de 10 busquedas.

3.4. Análisis sobre la condición de venta de los productos

Nos puede interesar en que condición se encuentran los productos más vendidos. En el set de datos se cuenta con cinco clasificaciones diferentes:

- Bom
- Muito Bom
- Excelente
- Bom - Sem Touch ID
- Novo

Analizamos la cantidad de conversiones (compras) registradas para cada una de estas condiciones y obtuvimos la visualización que sigue.

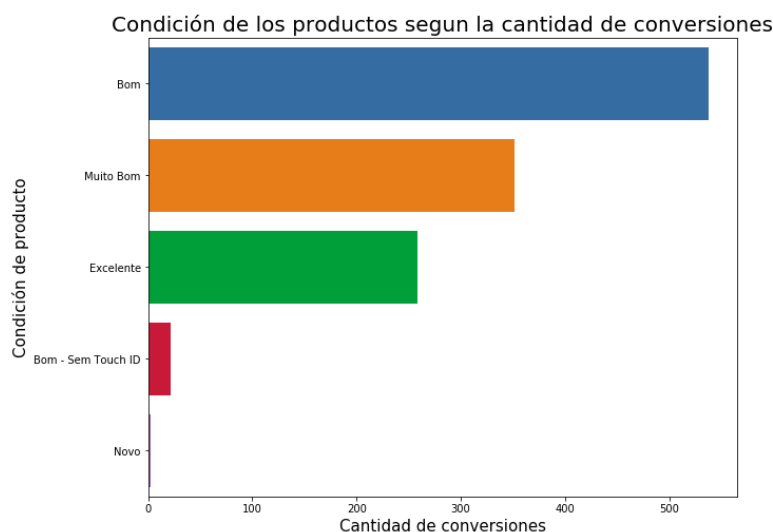


Figura 6: Condición de los productos comprados.

Es curioso ver como la mayoría de las compras se realizan sobre productos en condición "Bom" no "Muito Bom." .Excelente ni hablar de los "Novo" que apenas se llegan a apreciar en la imagen. Podría llegar a pensarse que tal vez esto esté relacionado con el precio del producto en relación con la utilidad y calidad del mismo. Pero no puede concluirse mucho sobre eso pues no contamos con datos de precios. O quizás hay una tendencia a tener los últimos modelos por lo cual hacen cambios frecuentes aun cuando el estado de sus equipos sean buenos.

4. Análisis sobre los eventos

4.1. Conversiones y usuarios por eventos

En esta sección veremos solo uno de los análisis realizados a las conversiones, puesto que es uno de los eventos más interesantes se lo aborda desde diferentes vistas. Comienzo analizando la cantidad de conversiones realizadas a través de los días a lo largo de los meses. El resultado del análisis se puede observar a continuación.

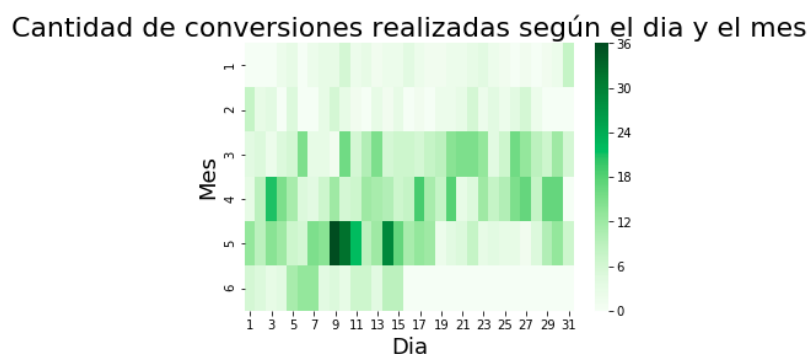


Figura 7: Conversiones realizadas por día de cada mes.

En la visualización se puede observar como las conversiones aumentaron a lo largo de los meses, iniciando muy bajas, entre Enero y Febrero, y alcanzando los valores más altos a mediados del mes

de Mayo. Luego analizamos como se distribuyen la cantidad de usuarios por eventos en general a lo largo de los días y los meses, para ver si hay alguna correlación entre las conversiones y la cantidad de usuarios en los eventos obteniendo la siguiente visualización.

Cantidad promedio de usuarios por Event por cada día según el mes

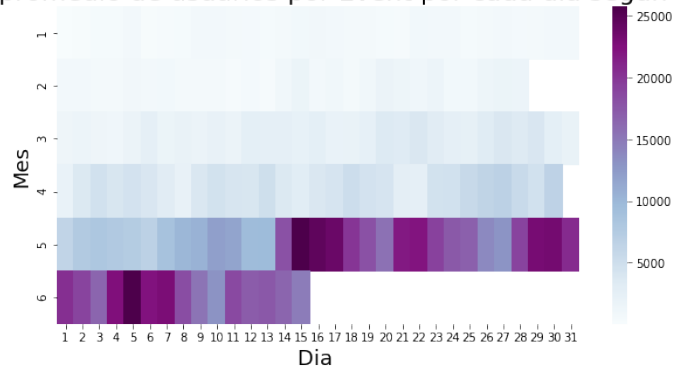


Figura 8: Usuarios por eventos según el día de cada mes.

Viendo esta visualización es claro que los participantes en eventos van en aumento a partir de mediados de Mayo en adelante. Pero se debe concluir que no se observa una relación directa aparente con la cantidad de conversiones, dado que el pico máximo de las conversiones se da entre los días 8 y 10 del mes de Mayo y los usuarios en eventos aumentan a mediados del mes de mayo en adelante.

4.2. Visitas y campañas de marketing

En esta sección nos enfocaremos en las visitas realizadas al sitio web de la empresa. Primero comenzaremos analizando como se distribuye la cantidad de visitas. Para obtener una visualización de mayor comprensión utilizaremos la escala logarítmica para graficar la densidad de las visitas.

Densidad de la cantidad de eventos 'visited site' (log)

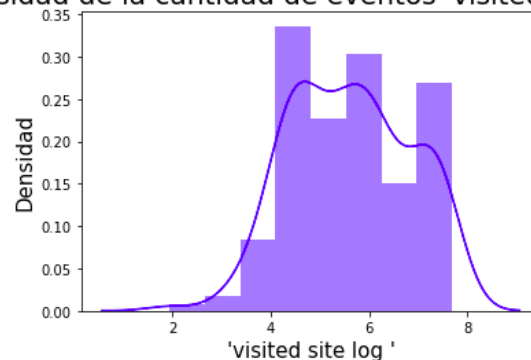


Figura 9: Distribucion de la cantidad de visitas al sitio como "visite site".

Luego visualizamos como se distribuyen las visitas realizadas a través de una campaña de marketing.

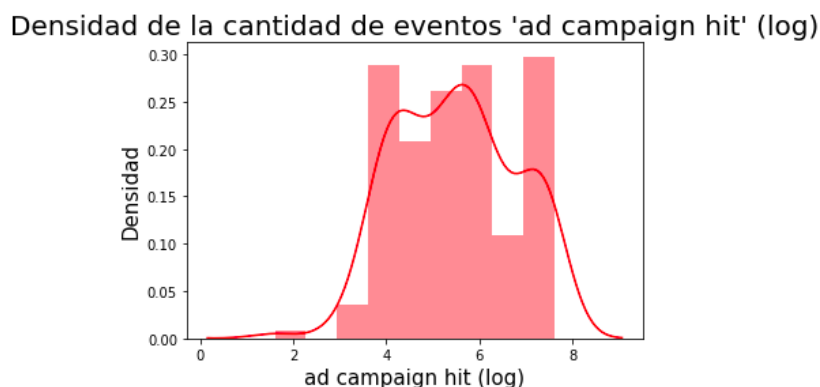


Figura 10: Densidad de visitas a través de campañas de marketing.

A simple vista podemos ver que se ambas se distribuyen de maneras muy parecidas. Pero para apreciar mejor esto recurriremos a un gráfico de densidad superpuesto.

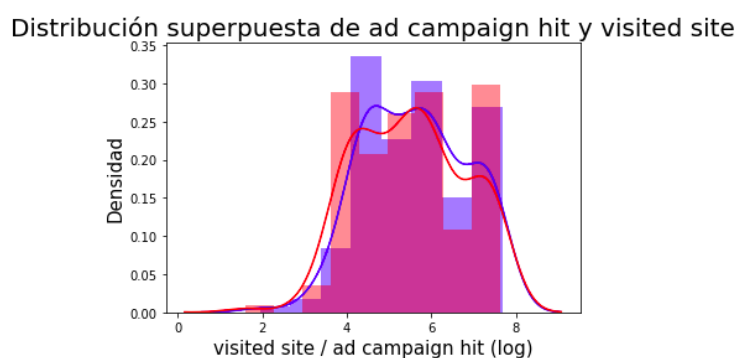


Figura 11: Comparación de densidades entre Visitas y visitas por campaña de marketing.

La visualización representa en rojo a los eventos "visited site" y en azul a los eventos "ad campaign hit". Con esta visualización es claro ver la similitud entre la distribución de ambos eventos. Esto podría explicarse pensando que las visitas "visited site" pudieron realizarse a través de un "ad campaign hit", y esto resulta lógico, pues el usuario ingresa al sitio mediante una campaña de marketing y luego puede hacer un "visited site".

5. Análisis sobre modelos de los productos

5.1. Checkouts, vistas, conversiones y leads por modelo

En esta sección nos interesamos en ver cuáles son los modelos más populares para los eventos 'checkout', 'conversion' y 'leads'. Para realizar este análisis Empezamos agrupando los datos según el tipo de evento para quedarnos con el grupo de checkouts y luego poder realizar un conteo de los modelos para los que se registraron checkouts.

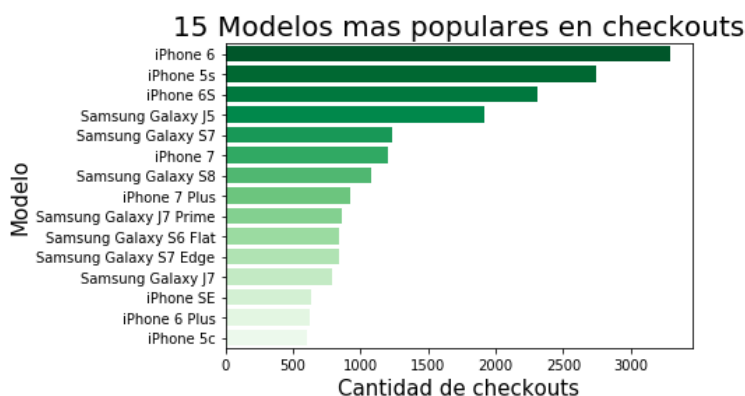


Figura 12: Modelos más populares en checkouts.

En esta visualización se puede observar como el modelo iPhone 6 se lleva la mayor cantidad de checkouts registrados, con más de 3000 checkouts, mientras que el modelo iPhone 5c queda ultimo en el ranking, superando por poco los 500 checkouts. Ahora podríamos ver que sucede con las conversiones por modelos y analizar que sucede entre los checkouts y las conversiones.

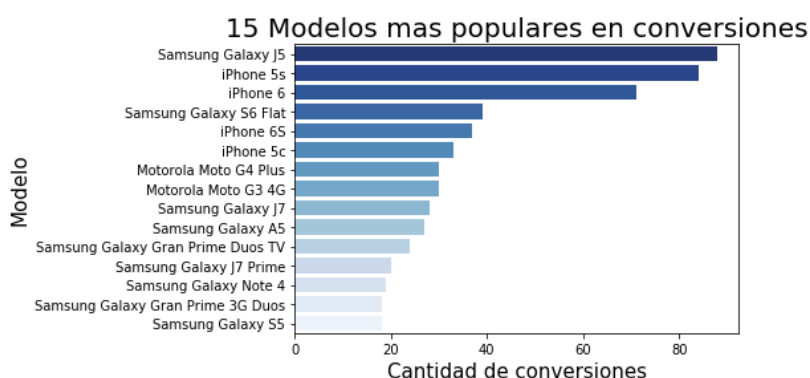


Figura 13: Modelos más populares en conversiones.

Con esta visualización observamos que curiosamente el modelo que lidera las conversiones es el Samsung Galaxy j5. Este modelo no solo no coincide con el de mayor checkouts, sino que se encuentra debajo del iPhone 6 por más de 1000 checkouts. Esta particularidad podría inferir que hay muchos usuarios interesados en el iPhone 6, pero que por alguna razón no llegan a completar en checkout y optan por un modelo diferente. Yendonos un poco hacía la realidad, esto podría ser producto del factor precio de los productos, ya que el iPhone 6 se encuentra a un nivel un tanto superior al Samsung Galaxy j5, tanto en calidad como en nombre y prestigio, o también podría tener que ver con la disponibilidad de stock del mismo.

Para analizar el peso específico de los modelos más vistos hicimos un gráfico de pareto con las cantidad de conversiones por modelo:

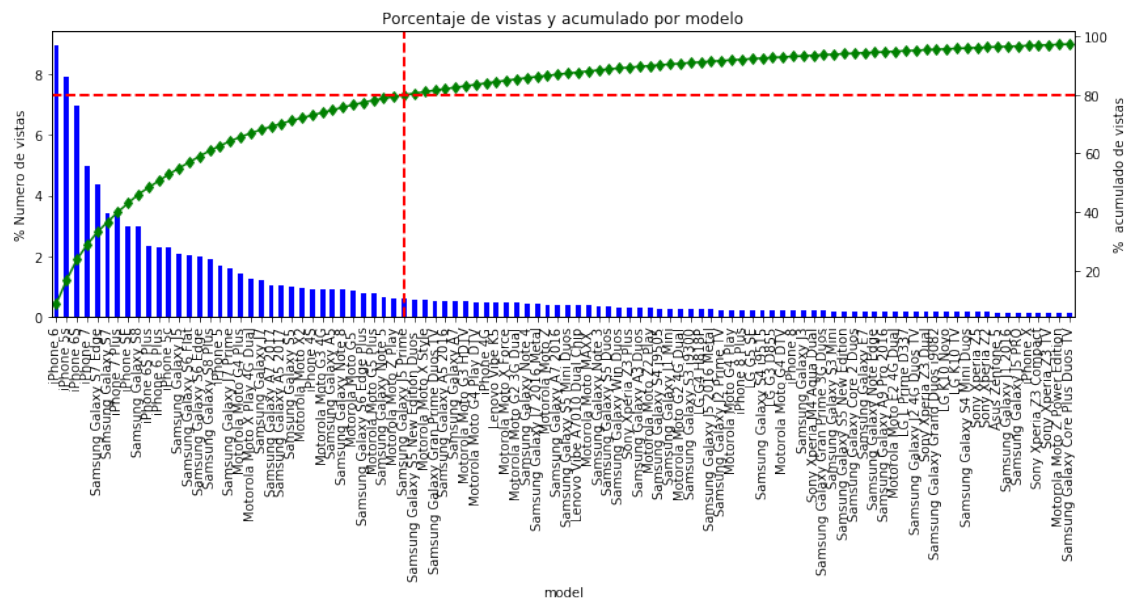


Figura 14: Pareto modelos más vistos.

Del gráfico anterior concluimos que los 33 los modelos de celulares mas vistos (17 por ciento del total) representan el 80 por ciento de las vistas

Como en caso anterior para analizar el peso específico de los modelos hicimos un gráfico de pareto con las cantidad de conversiones por modelo:

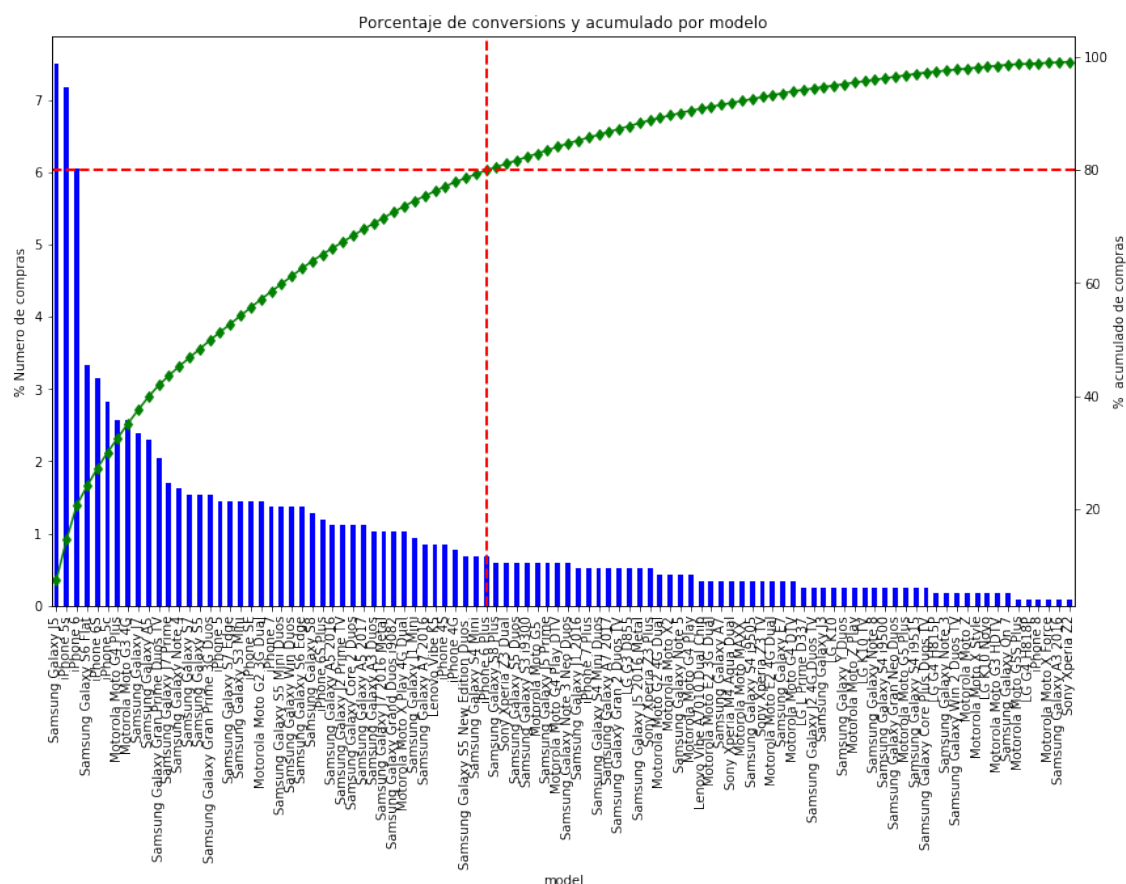


Figura 15: Pareto modelos mas poluarens en conversiones.

En el gráfico anterior se puede apreciar que los 42 modelos de celulares más comprados (21 por ciento del total) representan el 80 por ciento del total de las conversiones. Es interesante señalar que el modelo más buscados (iphone 6) que representa el 9 por ciento de las vistas figura tercero en el ranking de compras con un peso relativo de penas superior al 3 por ciento. Esto puede deberse al precio del mismo y que muchos usuarios optan por otro teléfono.

Para ver que sucede con la disponibilidad de stock por último analizamos que sucede con los Leads. Los mismos hacen referencia al interés del usuario sobre un producto que no tiene stock disponible para la venta en el momento en que el usuario visita el sitio. Volvemos a tomar los modelos más populares y observamos que dato curioso puede haber sobre ellos.

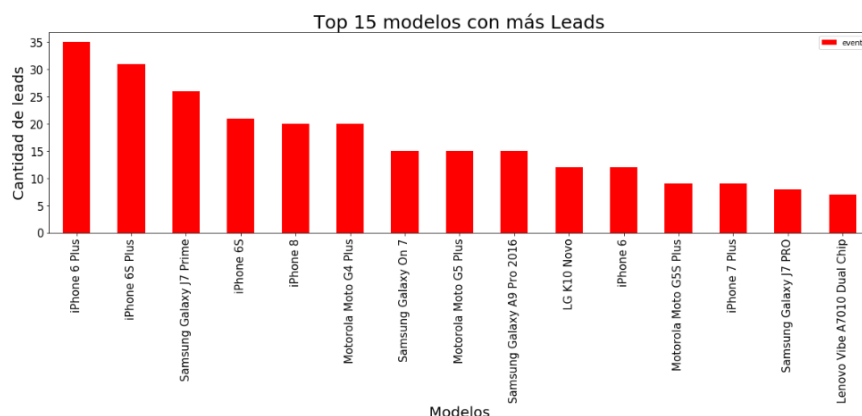


Figura 16: Modelos más populares en leads.

Vemos que los modelos más populares en leads no tienen mucho que ver con los modelos más populares en checkouts y conversiones. Lo cual no resulta descabellado. Pensándolo bien, tiene sentido, observemos que el modelo más comprado, Samsung Galaxy j5, no figura entre los modelos con más leads. Esto nos puede llevar a pensar que su disponibilidad de stock debe tener relación directa con la compra del mismo. También se puede observar que el iPhone 6 se encuentra dentro del ranking de leads y que además este era el modelo con más checkouts, pero que no llegaba a ser primero en las compras. Por lo que podría pensarse que algunos checkouts no pudieron finalizarse debido a la disponibilidad de stock, como habíamos especulado unas líneas más atrás.

5.2. Colores

Es interesante analizar los colores de los modelos más comprados y de los 10 modelos más comprados

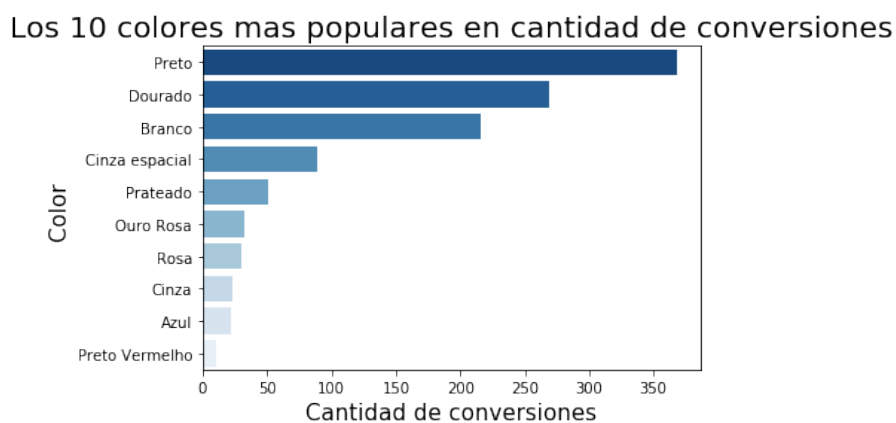


Figura 17: Colores más populares en conversion.

Los 10 colores mas populares en de los top 10 modelos en conversiones

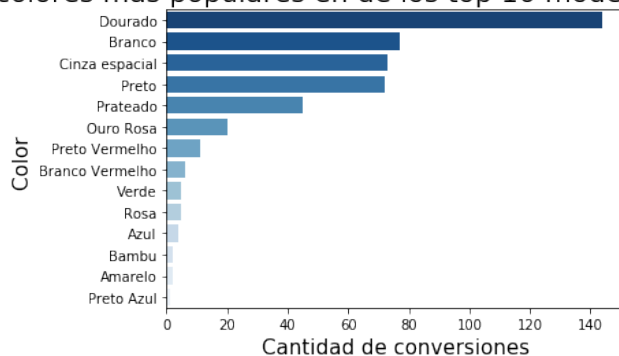


Figura 18: colores más populares del top 10 conversion.

De los gráficos de barras anteriores se puede concluir el color negro es el más popular considerando todos los equipos comprados. Pero si nos centramos en los primeros 10 modelos con más ventas, el dorado es el más popular.

5.3. Checkouts, leads, conversion sobre las vistas de cada modelo

Pensamos que sería interesante analizar eventos en relación a los modelos de los teléfonos y las vistas de los mismos, para ver si había más tendencia a, por ejemplo, comprar un modelo u otro. Usamos el evento 'viewed product' como referencia para determinar qué modelo lleva más tendencia a compra. Por ejemplo si un modelo tiene 10000 vistas y sólo 2000 eventos 'conversion' tendría menos relación de compras que un modelo con 5000 vistas y 2000 'conversion'; el primer modelo sólo tendría un 20 por ciento de sus productos comprados mientras que en el segundo tendría un 40 por ciento.

Para cada visualización, en un principio se intentó visualizar la relación aritmética entre 'checkout' y 'viewed product' y visualizarla junto con la cantidad de 'viewed product' de cada modelo, sin embargo se encontró con una visualización confusa ya que había mucha diferencia de cantidad de eventos 'viewed product' entre los modelos analizados.

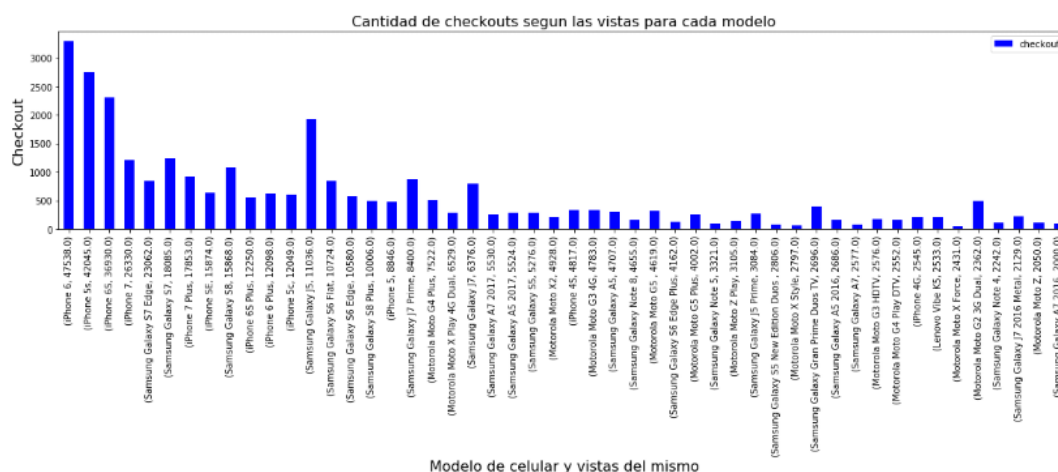


Figura 19: Cantidad de checkouts y viewed product de cada modelo.

La visualización muestra la cantidad de checkouts en cada modelo, el eje horizontal indica

el nombre del modelo y la cantidad de eventos 'viewed product' de cada uno, los modelos en el eje horizontal están ordenados de forma descendente. Se eligió esta configuración debido a que se intentó visualizar con otros plots y no era clara la distribución que quería mostrarse.



Figura 20: Cantidad de leads y viewed product de cada modelo.

Ahora se muestran la cantidad de leads frente a la cantidad de viewed product para cada modelo



Figura 21: Cantidad de conversion y viewed product de cada modelo.

Ahora se muestra la cantidad de 'conversion' frente a la cantidad de 'viewed product' de cada modelo. Se observa que a pesar de que varios modelos cuentan con más miles de vistas, en general los eventos 'conversion' no llegan a los 90



Figura 22: Cantidad de conversion y checkout de cada modelo.

Por último quisimos ver si, a pesar de tener pocas conversiones frente a la cantidad de vistas, los modelos tenían una buena cantidad de conversiones frente a la cantidad de checkouts. Similar a lo que ocurría antes, a pesar de que varios modelos cuentan con más de mil checkouts, en general los eventos 'conversion' no llegan a los 90

5.4. Condiciones de los modelos

Aquí quisimos ver si había alguna correlación entre los modelos y las condiciones de los mismos, para saber cuáles eran los modelos en mejores o peores condiciones

Mapa sobre las condiciones de cada modelo, filtrando los modelos con los que contamos mas de 300 muestras



Por lo visto, no parece que haya una correlación entre los modelos y las condiciones en que estos se encuentran.

Figura 23: Condiciones de cada modelo.

Sólo puede concluirse que el iPhone 5 es el que tiene mayor cantidad de productos en condición 'Bom', y que los iPhone en general tienen una buena cantidad de productos en condición 'Bom', 'Muito Bom' y 'Excelente', pero debido a la falta de datos precisos no pudimos sacar mejores conclusiones.

6. Análisis sobre las búsquedas de los usuarios

Analizamos las búsquedas de los usuarios para ver si se podía encontrar algo interesante.

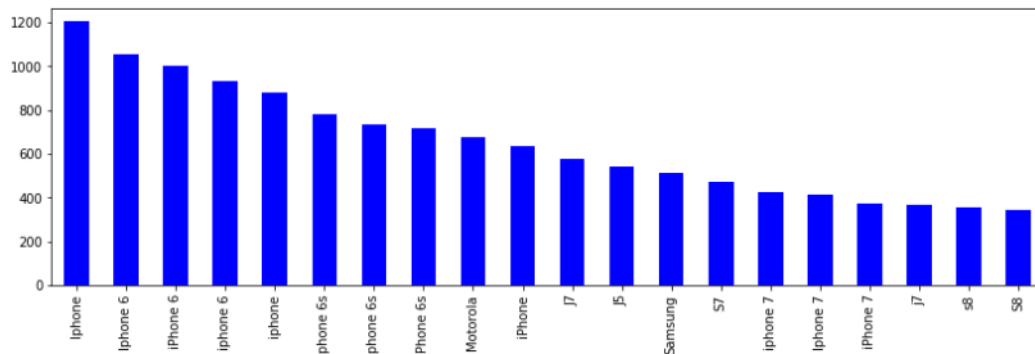


Figura 24: Búsquedas más comunes.

Vimos las búsquedas más comunes de los usuarios y observamos que, como era de esperar, unos escriben diferenciando entre mayúsculas y minúsculas y otros no, por lo que procedimos a convertir todos los caracteres de la lista de búsquedas en minúsculas.

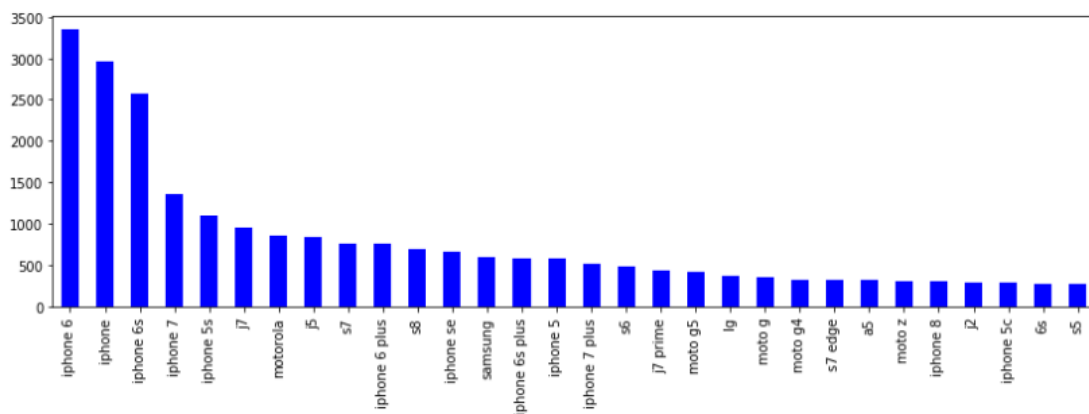


Figura 25: Búsquedas más comunes sin diferenciar entre mayúsculas y minúsculas.

Se observa un cambio sustancial en la cantidad de búsquedas, por ejemplo ahora el iPhone 6 tiene más búsqueda que el iPhone genérico.

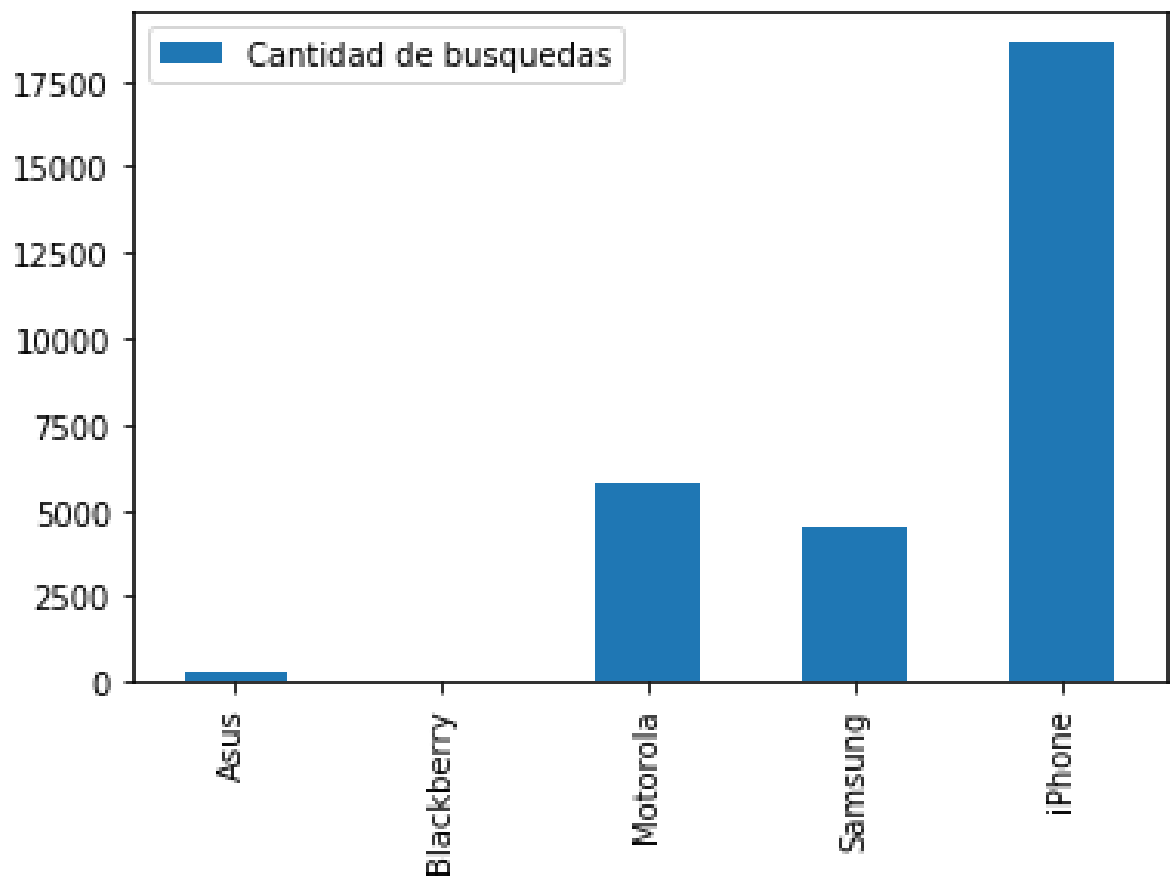


Figura 26: Búsquedas sobre modelo de teléfono.

Otra forma de determinar el modelo más buscado fue por las palabras que escribieron los usuarios, teniendo en cuenta que una persona puede escribir mal o de distintas maneras, por ejemplo iPhone podría haberse escrito como iphone o i phone, y el último no se ve en la visualización anterior. Lo que se encontró fue consistente con el análisis anterior, donde el iPhone es el modelo más buscado, pero este análisis tampoco es muy preciso debido a la cantidad de formas distintas en que podría haberse escrito el nombre de un modelo.

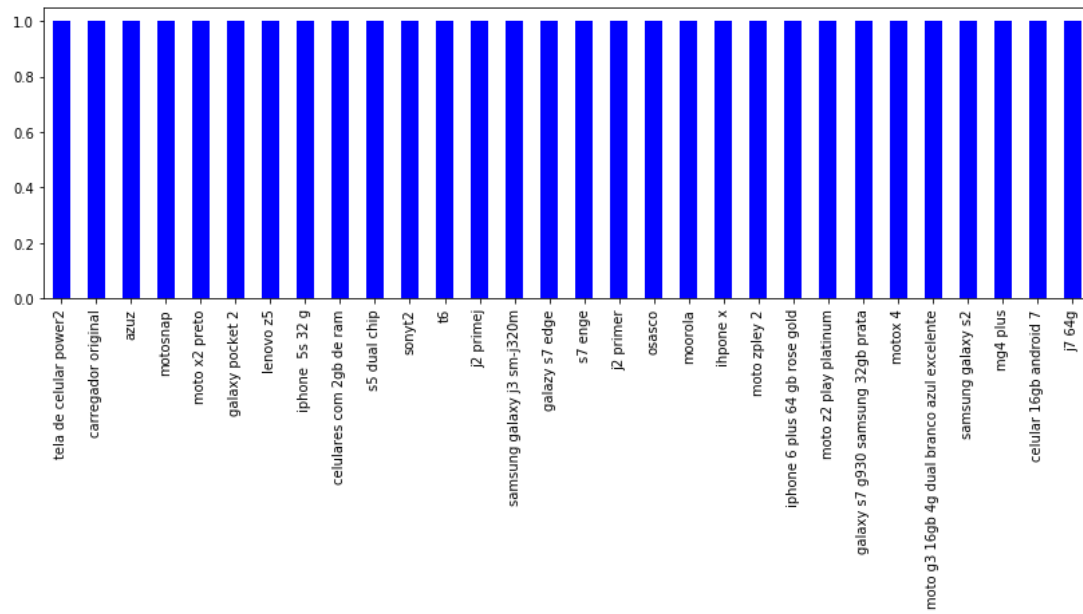


Figura 27: Búsquedas con menor cantidad.

Acá se ve lo que menos buscaron en los usuarios, el problema es que las búsquedas son demasiado específicas como para sacar conclusiones interesantes

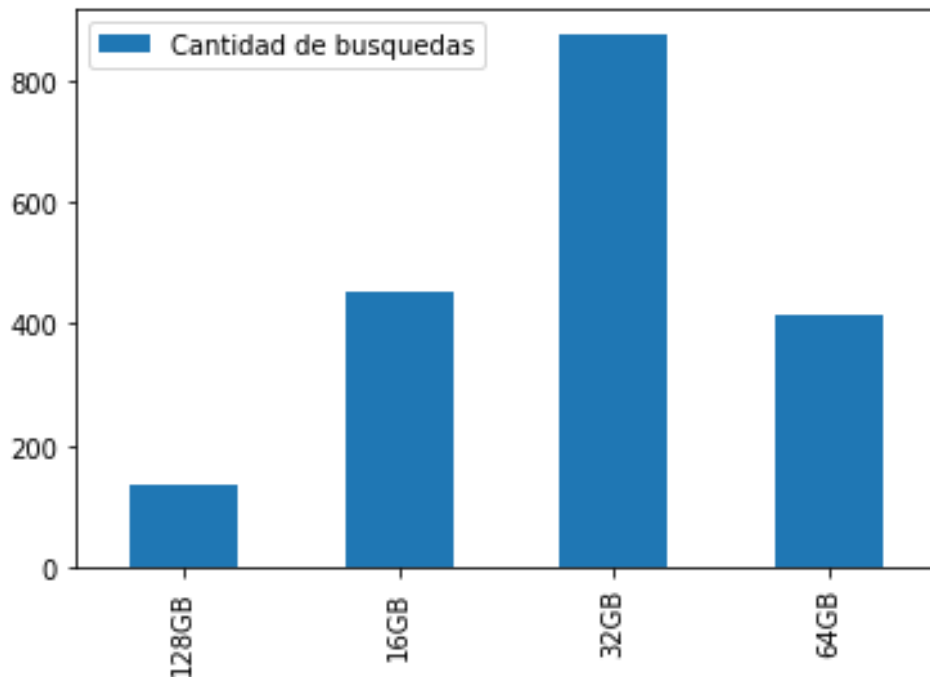


Figura 28: Búsquedas sobre capacidad de memoria.

Tratamos de determinar cuál es la capacidad de memoria más requerida por los usuarios, la desventaja de este análisis es que una persona puede escribir lo mismo de varias maneras, por ejemplo 32GB puede escribirse como 32GB, 32 GB, 32gb, 32 gb, 32g, etc. Por lo que intentamos contemplar todos los casos posibles para cada capacidad de memoria, el problema es que es difícil determinar exactamente de cuántas maneras puede una persona escribir lo mismo. Por lo pronto puede estimarse que la capacidad más buscada es la de 32GB, seguida de 16GB y 64GB, aunque estos últimos dos podrían cambiar de lugares si hay más casos de búsquedas que no se contemplaron.

7. Análisis sobre los dispositivos de los usuarios

7.1. Resolución de pantalla computadora

Vemos que resoluciones utilizan los usuarios de computadora, no se incluyeron a los usuarios que entraron con mas de una computadora. Para hacer las visualizaciones se tomo únicamente el valor mas chico (el de alto).

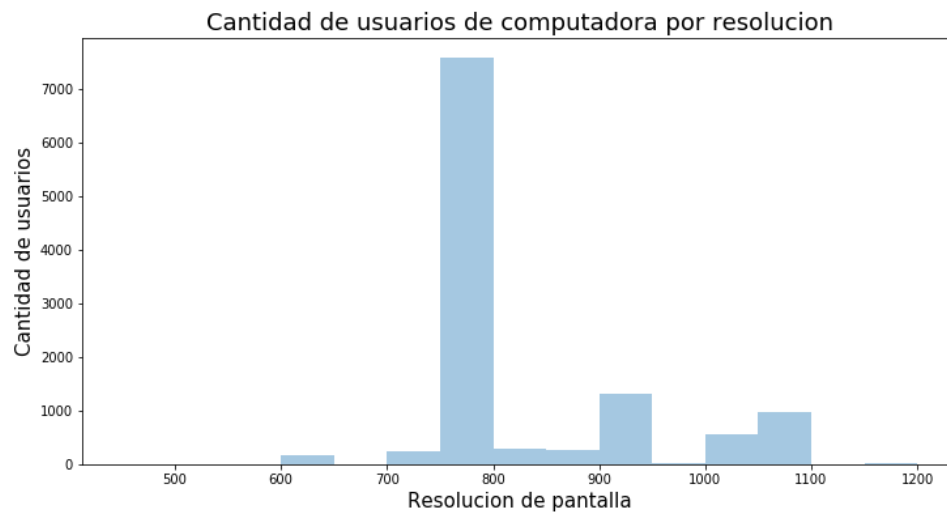


Figura 29: Distribución de las resoluciones de las computadoras de los usuarios.

Se puede observar que la mayoría de los usuarios utiliza resoluciones de entre 700p y 800p. Indicando que sus computadoras no son tan nuevas.

7.2. Resolución de pantalla celular

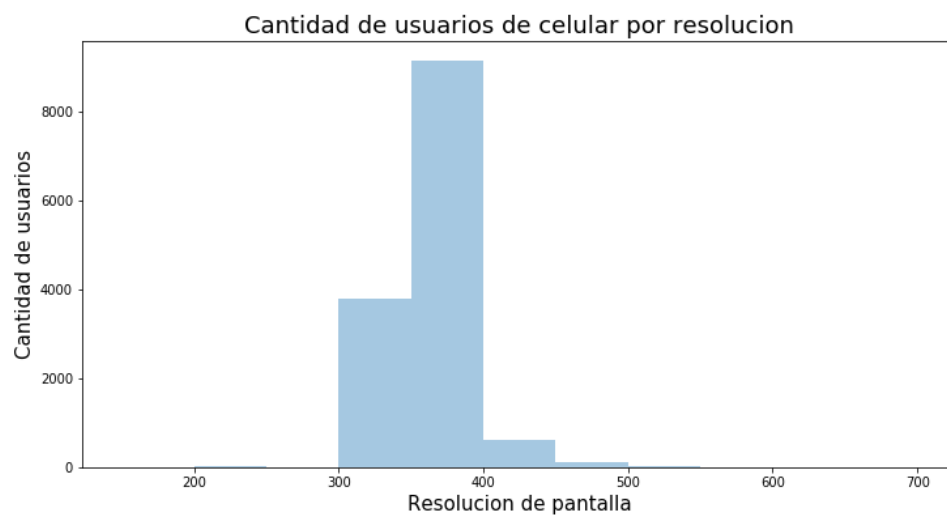


Figura 30: Distribución de las resoluciones de los celulares de los usuarios.

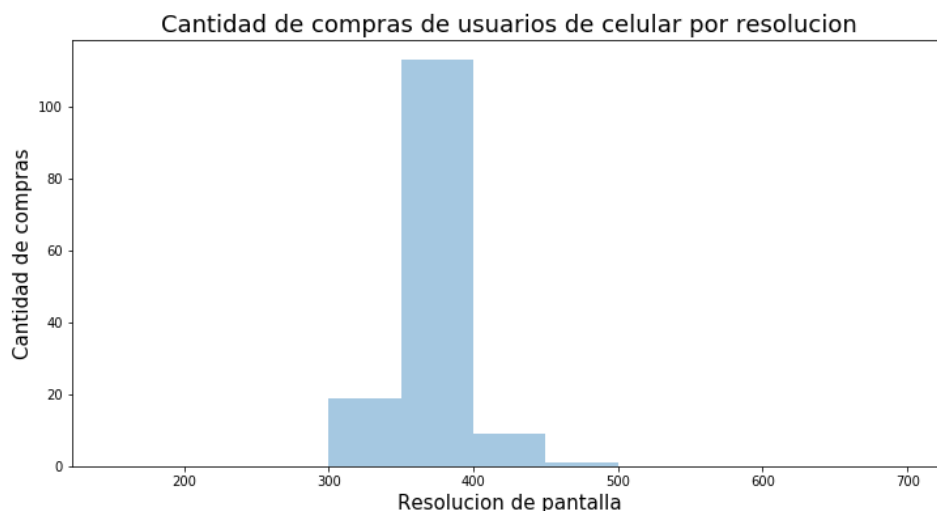


Figura 31: Distribución de las resoluciones de los celulares de los usuarios que compraron.

Comparando el gráfico de usuarios con el de compras se puede ver que la diferencia entre el rango (300p,350p) con el de (350p,400p) aumento.

De los dos gráficos anteriores se puede concluir que la resolución de pantalla de los dispositivos con los cuales los usuarios hacen sus consultas y compras en general es baja. Esto es importante para ajustar la tamaño de las imágenes del sitio web de manera tal no hacer más lenta la navegación a los mismos de lo necesario. Se podría diferenciar entre las resoluciones del navegador de escritorio y móvil para que la experiencia del usuario sea óptima de la siguiente manera:

- web para navegador de escritorio: entre 700 y 800 pixels
- web para navegador móvil: entre 300 y 400 pixels

7.3. Sistemas operativos

Se indica con multi los usuarios que ingresaron con mas de un sistema operativo. Dado que la gran mayoría accede mediante sistemas Android y se hizo un gráfico con escala logarítmica.

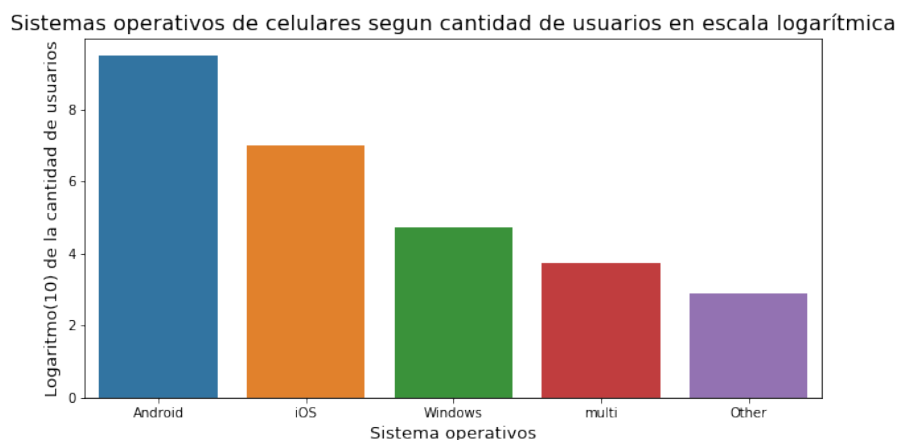


Figura 32: Distribución de los sistemas operativos utilizados por los usuarios.

Como se puede observar el sistema operativo más utilizado es Andorid, seguido de lejos por apple IOS. Esto confirma la tendencia mundial de preponderancia de las navegaciones por medio de teléfonos móviles por sobre computadoras y del sistema Android sobre el resto.

7.4. Tipo de dispositivos

Veamos que tipos de dispositivos utilizan los usuarios o si utilizan mas de un tipo.

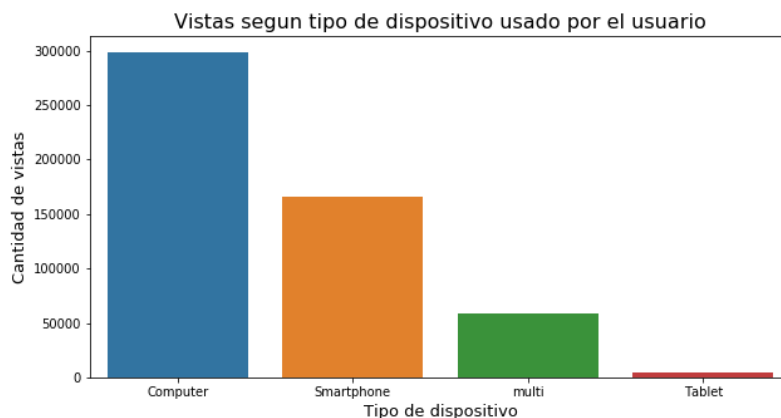


Figura 33: Distribución de las vistas segun los dispositivos utilizados por los usuarios.

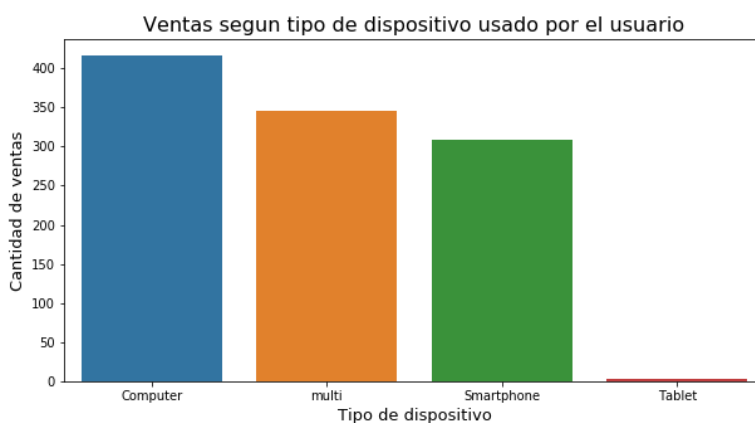


Figura 34: Distribución de las ventas según los dispositivos utilizados por los usuarios.

De los gráficos anteriores se puede concluir que la mayor parte de las vistas y compras de productos se realiza por medio de computadoras. Los celulares inteligentes figuran segundos en cuanto a dispositivos para búsquedas y terceros en el uso para efectuar compras. Esto se puede deber a que los usuarios prefieren efectuar las transacciones mediante una computadora. Sin embargo seria interesante contar con más datos de años anteriores para observar tendencias.

8. Conclusiones

En primer lugar es pudimos interiorizarnos un poco sobre el modelo de negocios de Trocafone algo desconocido por todos los integrantes del equipo. Asimismo pudimos interiorizaron sobre

el uso de herramientas de Big Data para depurar datos y analizarlos con el objeto de obtener conclusiones que permitan maximizar la experiencia del usuario.

De los análisis realizados podemos sacar las siguientes conclusiones generales:

- los usuarios están altamente concentrados en la ciudad de San Pablo, Brasil
- El horario pico de visitas se concentra entre las 12 horas del mediodía a 2 am del otro día. Por lo cual, las publicidades deben ser orientadas a dichos horarios. A su vez cualquier mantenimiento del sitio web debe realizarse entre las 5 am y las 8 am
- La mayor parte de las consultas se realizan a través de Google, por lo que mayor esfuerzo publicitario debe concentrarse en dicho buscador
- Los usuarios cambian de teléfono por motivos no relacionados al estado de los mismos, Posiblemente porque los usuarios quieren tener los últimos modelos disponibles. Esto es importante porque evidencia una necesidad reducida de personal dedicado a la restauración de los mismos. Es más, se podría evaluar eliminar el servicio de restauración de equipos y vender los equipos por lote asumiendo el buen estado de los mismos, considerando el bajo porcentaje de equipos en mal estado
- A partir del mes de Mayo se observa un gran crecimiento de las visitas al sitio web. Como solamente tenemos datos del primer semestre del 2018, no podemos concluir si es un fenómeno estacional o se debe a otros motivos. En caso de que la tendencia se haya repetido en el 2017 se podría concluir que los primeros meses del año se reduce la actividad fuertemente debido al periodo de vacaciones donde la gente no quiere o no puede cambiar sus teléfonos. Si esta tendencia pudiera ser confirmada, resultaría muy importante para ajustar la campaña de marketing a dicha realidad, reduciéndola a un mínimo de enero a marzo. Por otra parte, en caso de que Trocafone contase con una estructura de costos laboral fija alta, podría reducir la planta permanente, tercerizando servicios durante los meses de demanda pico. De esta manera se podría obtener una estructura óptima y maximizar los beneficios
- El color negro es el color más elegido por los usuarios que compran, seguido por el dorado y el blanco. Por otra parte, el dorado es el favorito de los modelos que más se venden seguido por el blanco y el gris espacial. De esta manera podemos concluir que debe haber un mayor stock de estos 4 colores que del resto
- La mayoría de las búsquedas se orientan a celulares de Apple seguidos por Motorola y Samsung. Esto es importante para diseñar campañas de marketing y políticas de stock para los equipos
- Los usuarios buscan en su mayoría celulares con 32 gb de memoria seguidos por 16Gb y 64Gb. Esto coincide con la capacidad de los modelos más buscados
- Es importante tener en cuenta las bajas resoluciones de los equipos con los cuales los usuarios realizan sus búsquedas en la web para reducir el tamaño de las imágenes a la medida justa de manera de optimizar la experiencia de usuario
- Los usuarios utilizan equipos con sistemas operativos Android en su mayoría, seguidos por iOS. Por lo cual es importante adaptar la web móvil para dichos sistemas
- La computadora es el dispositivo para búsquedas y compras por excelencia. Sin embargo, habría que analizar datos de otros años para ver las tendencias. Es esperable que los teléfonos inteligentes superen pronto a las computadoras tanto para búsquedas como para efectuar transacciones