



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE INGENIERÍA

2^{do} Cuatrimestre - 2018

ORGANIZACIÓN DE DATOS (75.06)

INFORME TRABAJO PRÁCTICO 1

Análisis Exploratorio de Datos

GRUPO 12: **Zulma LoDato**

INTEGRANTES:

Nombre y Apellido	Padrón	Correo Electrónico
Donato, Juan Pablo	100839	juan.pablo.donato1996@gmail.com
Pistillo, Carolina Rocío	99177	caropistillo@gmail.com
Velasco, Ignacio	99796	iignaciovelasco@gmail.com
Viere, Constanza Irene	98692	constanza.viere@gmail.com

FECHA: 24 de septiembre de 2018

Link de **GitHub**: https://github.com/IgVelasco/Tp1_Datos

Link de **Kaggle**: <https://www.kaggle.com/jpdonato/analisis-trabajo-practico-1>

Índice

1. Introducción general	1
2. Análisis temporal	1
2.1. Rating temporal	1
2.1.1. Cantidad de visitas a lo largo del tiempo	1
2.1.2. Cantidad de conversiones a lo largo del tiempo	1
2.2. Comportamiento de un usuario según frecuencia de visitas	3
2.2.1. Frecuencia diaria	3
2.2.2. Frecuencia mensual	5
2.3. Tiempo de permanencia de un usuario en la página	6
2.3.1. Tasa de rebote	8
2.4. Usuarios por mes dependiendo el sistema operativo	8
2.5. Cantidad de ventas por día de semana.	9
2.6. Análisis de Campañas en el tiempo	10
2.6.1. Visitas	10
2.6.2. Compras	10
3. Análisis Demográfico	12
3.1. Cantidad de visitas al sitio en todo el mundo	12
3.2. Análisis de visitas al sitio en Brasil	13
3.3. Análisis de visitas al sitio en Estados Unidos	13
3.4. Tipo de acceso al sitio	14
3.5. Teléfonos más buscados por país	16
3.5.1. Brasil	16
3.5.2. Estados Unidos	17
3.5.3. Argentina	18
4. Análisis de sistemas operativos	18
4.1. Usuarios de Android	19
4.1.1. Relación de modelos	19
4.2. Usuarios de Windows	21
4.2.1. Relación de modelos	21
4.3. Usuarios de iOS	22
4.3.1. Relación de modelos	22
5. Análisis sobre los productos	23
5.1. Productos más vendidos	23
5.2. Condición de los teléfonos mas vendidos	24
5.3. Análisis del feature lead	25
5.4. Análisis de ventas	27
5.4.1. Ventas por modelo	27
5.4.2. Año de lanzamiento	30
5.4.3. Pauta de ventas	30
6. Conclusiones finales	31

1. Introducción general

A lo largo de este informe se analizarán los datos de eventos de *web analytics* de usuarios que visitaron la página www.trocafone.com, su plataforma e-commerce de Brasil. Trocafone es una empresa de tecnología con operaciones en Brasil y Argentina. Implementan un modelo de negocio conocido como *ReCommerce* que plantea la compra, reacondicionamiento y venta de celulares previamente usados.

Los datos correspondientes a los eventos de *web analytics* fueron proporcionados en el archivo “*eventos.csv*”. Este set representa un subconjunto de la totalidad de eventos y dispone de una cantidad representativa de usuarios que visitaron la plataforma.

A partir del análisis exploratorio se pretende interpretar comportamiento. Es decir, observar que características o tendencias presentan los datos, detectar posibles irregularidades, generar algún tipo de estadística, realizar visualizaciones rápidas, entre otras cosas. En nuestro caso, este análisis puede resultar útil para la investigación comercial y de mercado, y para evaluar o mejorar la efectividad del sitio web. También puede ayudar a la empresa a medir los resultados de las campañas publicitarias, estimando cómo cambia el tráfico de un sitio web después del lanzamiento de una publicidad. En adición, puede servir para medir las tendencias de popularidad que son útiles para la investigación de mercado.

2. Análisis temporal

Ver notebook: */Notebooks/Analisis_temporal.ipynb*

Dentro del set de datos resulta interesante analizar la columna *timestamp*, la cual indica la fecha y hora (BRT/ART) de cuándo ocurrió el evento. Como se mencionó previamente el set representa un subconjunto de la totalidad de eventos, por lo que se espera que las fechas estén acotadas en un cierto rango. En un primer acercamiento pudimos apreciar que dichos valores se correspondían a eventos ocurridos entre los meses de enero y junio del año actual (2018). Siendo más rigurosos, el rango obtenido fue:

[2018-01-01 07:32:26, 2018-06-15 23:59:31]

A partir de los datos dados en esta columna podemos realizar distintos tipos de análisis en función de la relación existente de los mismos con los presentes en el resto de las columnas.

2.1. Rating temporal

2.1.1. Cantidad de visitas a lo largo del tiempo

Si analizamos la cantidad de visitas al sitio que realizan los usuarios por día, podemos darle una idea a la compañía de cómo varía su *rating* a lo largo del tiempo. Esto puede resultar útil para establecer en qué épocas es más conveniente realizar campañas publicitarias para atraer clientes a la página.

Ahora bien, para realizar este análisis debemos tener una forma de computar las visitas. Para ello tendremos en cuenta el evento “*visited site*”, que nos indica que el usuario ingresa al sitio a una determinada url. Este evento está presente en la amplia mayoría de las sesiones de cada usuario. Sin embargo se presentan excepciones, en las que el usuario solo ingresa a la página para efectuar un evento como un “*checkout*” o “*conversion*”. Estos casos también serán tenidos en cuenta en el criterio que utilizaremos: “Se computarán todos los eventos “*visited site*” por usuario y por día, de forma tal que si un día el usuario no presenta el evento de visita pero se registró algún otro evento, este caso también se compute como una visita.”

Aprovechando que contamos con una variable temporal, decidimos realizar una visualización mediante un gráfico de líneas (ver Figura 1). Podemos observar que la mayor cantidad de visitas ocurre durante los meses de mayo y junio. A mediados de mayo el número de visitas aumenta repentinamente y se mantiene dentro del rango de las 1000-2200 visitas. Estos valores contrastan en gran magnitud con la cantidad dada para los meses de enero y febrero, donde el número de visitas está por debajo de los 250.

2.1.2. Cantidad de conversiones a lo largo del tiempo

Procederemos de forma similar que en la sección anterior, pero esta vez analizando la cantidad de conversiones a lo largo del tiempo y relacionando esta cantidad con la cantidad de visitas. Es decir, tendremos información no solo de la cantidad de conversiones a lo largo del tiempo si no también de qué porcentaje representa esta cantidad respecto al número de visitas al sitio. En primer lugar realizamos una visualización superpuesta de ambas

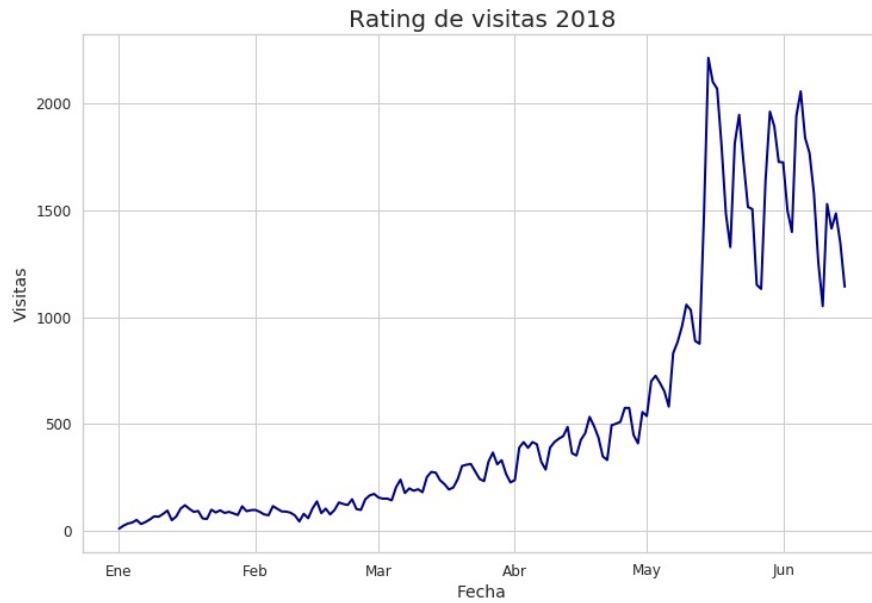


Figura 1: Cantidad de visitas al sitio a lo largo del tiempo

cantidades a lo largo del tiempo utilizando nuevamente gráficos de línea como puede observarse en la Figura 2. De esta visualización notamos que el número de conversiones por día a lo largo del tiempo es insignificante en relación a la cantidad de visitas efectuadas. Podemos observar que la proporción es pequeña pero resulta muy difícil sacar conclusiones respecto a si más visitas implican más conversiones. Para ello necesitamos saber en qué época hubo más conversiones. Por lo tanto, se decidió graficar de forma separada la cantidad de conversiones a lo largo del tiempo, tal como se muestra en la Figura 3.

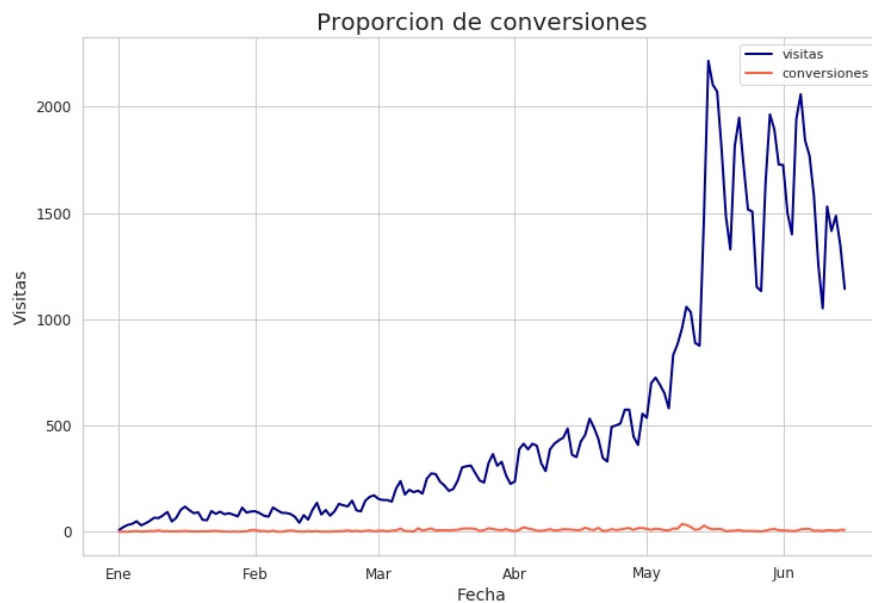


Figura 2: Cantidad de visitas al sitio y conversiones a lo largo del tiempo

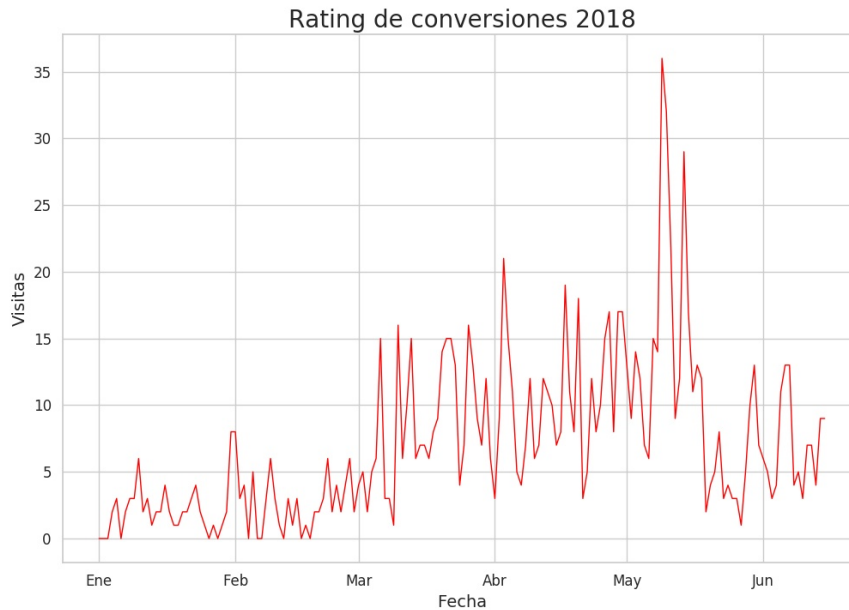


Figura 3: Cantidad de conversiones a lo largo del tiempo

Del gráfico de conversiones a lo largo del tiempo, comienza a notarse una cantidad significativa de conversiones a partir del mes de marzo. Es interesante notar que es el mismo mes en el que el número de visitas comienza a incrementarse en la Figura 1. Los resultados se vuelven notablemente mayores para el mes de mayo, coincidente con el mayor número de visitas. Esto nos da una idea de la relación presente entre la cantidad de visitas y la de conversiones efectuadas. Hay una tendencia que al haber más visitas, hay más conversiones. Sin embargo, la Figura 2 nos dice que estas conversiones son de muy baja proporción respecto al número de visitas.

Debemos tener en cuenta que un usuario puede realizar más de una visita por día, por lo que resultaría interesante analizar como varía nuestro panorama si clasificamos a los usuarios según cantidad de visitas para cada día.

2.2. Comportamiento de un usuario según frecuencia de visitas

2.2.1. Frecuencia diaria

Resulta interesante analizar como varía el comportamiento de un usuario según cuántas visitas hizo en el día. Previamente para el cálculo de las visitas por día realizamos un cálculo de las visitas por día para cada usuario. Esto fue necesario para que no se nos pasen por alto aquellos que habían visitado el sitio un día pero no habían efectuado un evento *“visited site”*. Si realizamos un conteo de aquellos datos obtenidos obtenemos la tabla de la Figura 4a.

Aquí se ven representadas las cantidades totales de visitas según la cantidad de veces que fueron efectuadas por un usuario en el día. Podemos notar que la mayoría de las visitas por usuario fueron únicas en el día y una gran mayoría se dio dos veces por día. Sin embargo hay usuarios que visitan aún más veces por día el sitio y es interesante analizar como influye este aspecto respecto a las operaciones que realizaron en el día.

Para analizar subgrupos que tengan una cantidad significativa, se decidió unir las columnas que presentan visitas mayores a 5 en una nueva columna *“+5”* (ver Figura 4b). Para cada subgrupo (dado por el número de visitas por día), se computaron los eventos ocurridos. Dado que la cantidad de visitas que fueron realizadas de forma única en el día es preponderante respecto al resto de los subgrupos, se esperará encontrar una cantidad mucho mayor de eventos para este caso. Para hacer comparables las cantidades obtenidas, se dividieron los valores de cada subgrupo por el total de visitas dado por la Figura 4b. A su vez, como solo nos interesa la relación entre subgrupos y no la cantidad, se decidió normalizar aquellos valores para cada evento. Los resultados obtenidos pueden contemplarse en la Figura 5.

En general se nota una tendencia a que cuántas más visitas realizó un usuario en el día, más eventos generó. Las relaciones que resultan más equitativas son las dadas en los eventos *“ad campaign hit”*, *“checkout”* y *“search”*.

	1	2	3	4	5	6	7	8	9	10	11	12	14	13	42	35
visitas	52767	9763	2822	1003	364	159	71	44	24	21	6	4	3	3	1	1

(a)

	1	2	3	4	+5
visitas	52767	9763	2822	1003	701

(b)

Figura 4: Cantidad de visitas de un usuario por día

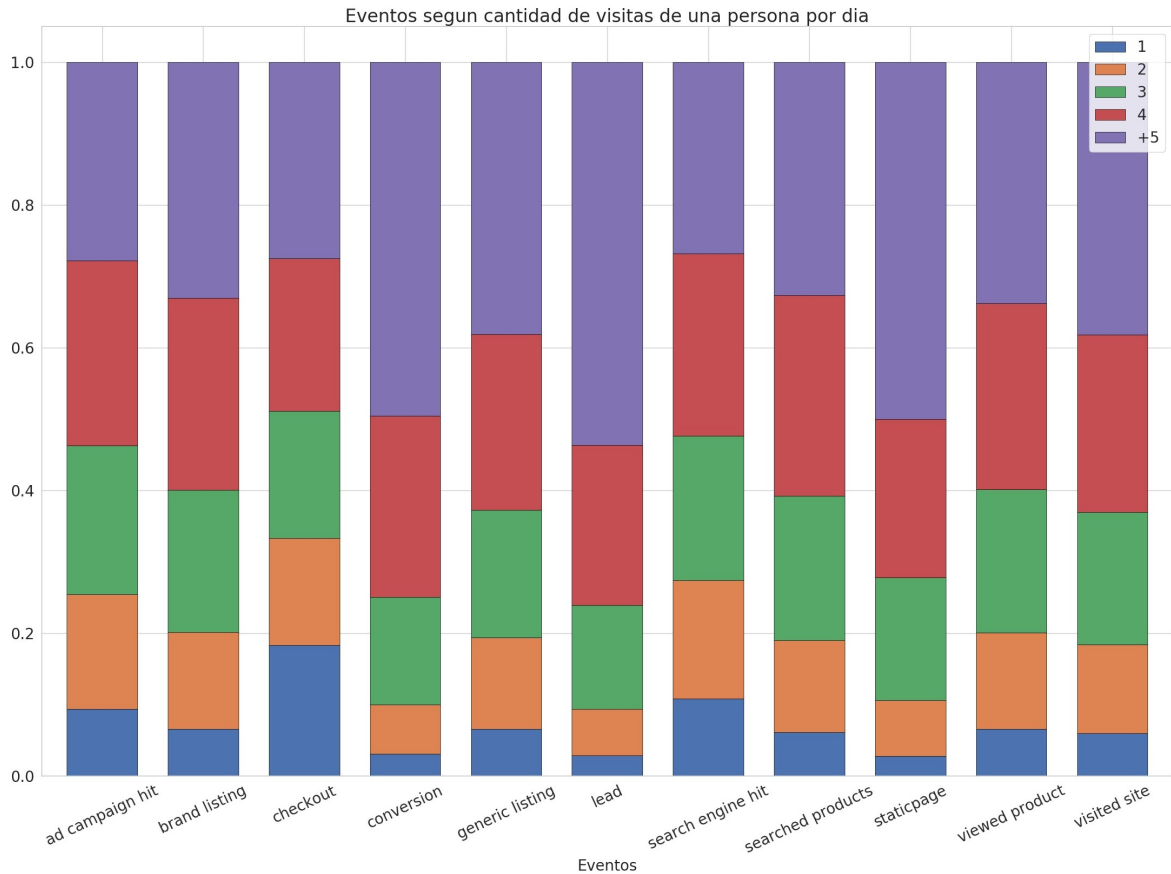


Figura 5: Proporción de eventos según cantidad de visitas de una persona por día

engine hit". Esto nos da una idea que dichos eventos no dependen de la cantidad de veces que el usuario visita el sitio en el día.

El primero puede resultar lógico si pensamos en que una persona que ya accedió a la página por una campaña, no volverá a acceder por hacer clic en una publicidad sino que ya habrá explorado el sitio y vuelto por otro medio.

El evento *"checkout"* por su parte, tampoco parece depender de la variable en cuestión. Esto puede deberse a que las personas acostumbren a realizar la operación independientemente de la cantidad de veces diaria en que entraron a la página o que sean personas que ya hayan visitado la página días anteriores.

Las proporciones del evento *"search engine hit"* por su parte nos sugieren que una persona que en el día ya entró a la página por un motor de búsqueda luego no vuelve a ingresar por el mismo medio. Si observamos los valores correspondientes a *"lead"* y *"static page"* podemos notar que las proporciones correspondientes a visitas mayores a 5 son mucho mayores que la de visitas de menor cantidad. Si relacionamos los eventos, podemos suponer que el usuario no vuelve a ingresar por motor de búsqueda sino como consecuencia de haberse registrado

para recibir información de disponibilidad de stock o bien ha guardado la página en algún tipo de acceso rápido.

Por otro lado resulta interesante fijarse en el evento de conversiones. Los usuarios que realizaron más de una visita por día suelen realizar más compras que los que ingresan solo una vez.

Podríamos seguir analizando las proporciones pero no resultan de nuestro interés. Para los eventos como “*searched products*” o “*viewed product*” nos interesa más saber si el producto buscado o visto fue siempre el mismo o no y en esta sección todavía no procesamos los datos al respecto.

Para continuar con el análisis temporal, procederemos de forma similar pero teniendo en cuenta las visitas por mes.

2.2.2. Frecuencia mensual

Como se mencionó en la sección 2.2, hay parámetros que resultan insuficientes o incompletos de analizar si tenemos en cuenta solo la cantidad de veces que entró un usuario por día. Pueden, por ejemplo, haber casos de usuarios que entren solo una vez por día todos los días del mes que no estemos contemplando.

Procediendo de forma similar que en la sección anterior, debemos contabilizar las cantidades totales de visitas según la cantidad de veces que fueron efectuadas por mes. Esta vez también agrupamos en subgrupos que tengan una cantidad significativa, tal como se observa en la Figura 6.

Podemos notar que la mayoría de las visitas fueron únicas en mes. Si comparamos con el cuadro de la Figura 4a podemos establecer que de las 52767 visitas que fueron únicas en el día, tenemos la certeza de que 21795 de ellas (casi la mitad) no volvieron a efectuarse otro día en el mismo mes. Esto nos indica que las proporciones que vayamos a calcular para este caso de visitas únicas en el mes deben ser por lo menos la mitad que las calculadas para visitas únicas en el día.

Hay una cantidad importante de visitas que se dieron 2 veces al mes o entre 3 y 15 veces. Los números comienzan a ser mucho menores para las visitas con frecuencia mayor a 16, pero consideramos de gran importancia analizar qué realizan en la página aquellos que entraron tantas veces.

	1	2	3-15	16-30	31-60	+61
visitas	21795	5554	7764	426	102	21

Figura 6: Cantidad de visitas de un usuario por mes

Realizando un análisis porcentual y normalizando como antes, obtenemos los resultados dados en la Figura 7. Se decidió utilizar el mismo tipo de gráfico para poder hacer un análisis comparativo.

Del gráfico de barras apiladas podemos observar que nuestra predicción se cumple para las visitas con frecuencia única en el mes. Efectivamente las proporciones son la mitad y en algunos casos menos que las dadas para las visitas ocurridas una vez por día en el gráfico anterior.

La mayor proporción de eventos está dada por las visitas que fueron efectuadas más de 16 veces al mes. Ahora bien, estas visitas pudieron ser efectuadas en casos extremos todas el mismo día o todas en días distintos. Es decir, pueden corresponderse al subgrupo de 1 visita por día del gráfico de la Figura 5 o bien a otro de los subgrupos. Sin embargo, podemos afirmar con seguridad que las visitas entre 31-60 deben repartirse al menos en alguno de los subgrupos con visitas diarias mayores a 1 y que las mayores a 60 por mes tienen que corresponderse a visitas ocurridas al menos 2 veces por día.

Una de las columnas con proporciones dentro de todo “equitativas” es nuevamente la de “*checkout*”. Esto de alguna forma confirma nuestra hipótesis de que un usuario realiza un checkout independientemente de la cantidad de veces que haya entrado en el día o en el mes. Sin embargo, parece ser una costumbre entrar a la opción de checkout y no realizar la compra. Si observamos la columna “*conversion*” notamos que la proporción de usuarios que compraron es insignificante para aquellos que entraron pocas veces en el mes.

Por otro lado resulta interesante analizar la apilación para el evento “*ad campaign hit*”. En este caso las proporciones ya no son tan equitativas como en el gráfico anterior. Anteriormente supusimos que una persona que ingreso por campaña no volvería a hacerlo en el día, pero eso no quiere decir que no vuelva a hacerlo en el mes. Al contrario, este hecho nos da un parámetro de que las campañas han logrado que un usuario entre varias veces al mes.

Las diferencias en las proporciones de los eventos “*lead*” y “*static page*” para las visitas con menos frecuencia mensual resultan aún más marcadas que en el gráfico de proporciones de visitas diarias. Esto es lógico ya que ahora tenemos la certeza de que la visita fue única en el mes. Antes sabíamos que la visita había sido única en el

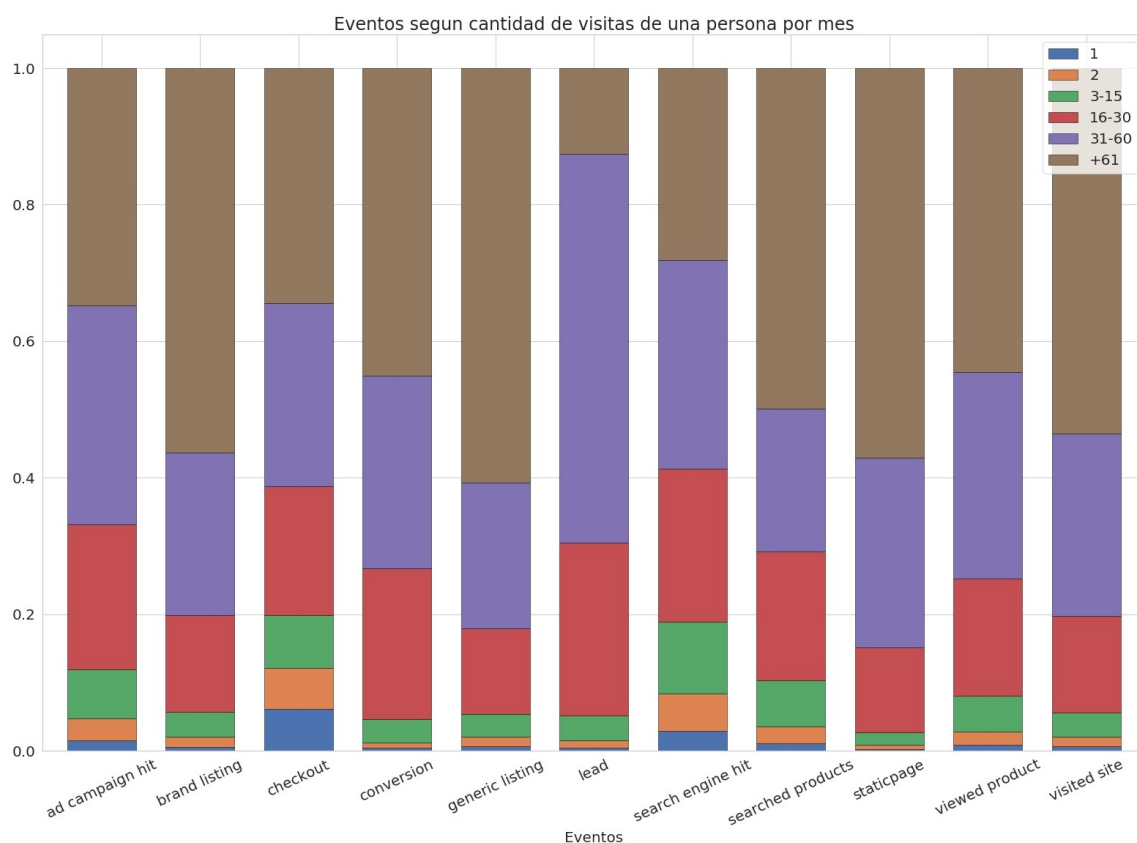


Figura 7: Proporción de eventos según cantidad de visitas de una persona por mes

día pero probablemente estábamos computando también a personas que ya habían ingresado con anterioridad y se habían suscripto o guardado la página para acceder directamente.

Para concluir este análisis de proporciones podemos decir que si bien los usuarios presentan ciertas tendencias según la cantidad de visitas por día, las diferencias de proporciones entre eventos resultan aún más marcadas cuando analizamos la frecuencia mensual de las visitas. Esto se debe a que en la frecuencia diaria no tenemos en cuenta los eventos que pudo haber realizado días anteriores o posteriores. De todos modos, ambos análisis sirven para tener una idea de la importancia de la cantidad de veces que un usuario ingresa tanto por día como por mes.

2.3. Tiempo de permanencia de un usuario en la página

Resulta de gran importancia analizar el tiempo de permanencia de un usuario en la página. Aunque la compañía pudo haber puesto todo su esfuerzo en crear el contenido de una página web, la calidad de la misma está en el ojo del que la visita. Tiempos bajos de permanencia en el sitio pueden indicar que no se está captando el interés del usuario. Además es importante conocer esta medida para tener una idea de cuanto tiempo se dispone para conseguir su atención.

Es importante tener en cuenta que la medición solo podrá realizarse entre dos interacciones en la web. Esto implica que si una página es la última que se visita durante la sesión, no habrá forma de calcular el tiempo de permanencia de un usuario en ella ya que no se ha visitado ninguna otra página del sitio posteriormente. Por lo tanto, si un evento es el último que un usuario realiza durante su sesión, el tiempo que permaneció luego no se tomará en cuenta.

Para el cálculo en cuestión procedimos como se detalla a continuación. En primer lugar, se calcularon las diferencias de tiempo entre los distintos eventos temporales *por usuario*. Si nos quedamos solo con este cálculo estaríamos omitiendo los lapsos de tiempo existentes entre sesiones de un usuario que ingresó más de una vez al sitio. Por ello es que seguidamente se calcularon las diferencias de tiempo entre eventos, sin importar el usuario. Si fijamos un umbral máximo de sesión podremos contemplar que cuando entre dos eventos haya una diferencia

mayor a ese umbral, se tratará de una nueva sesión del mismo usuario. Si también aplicamos el umbral a las diferencias de tiempo entre eventos *por usuario* estaríamos detectando nuevas sesiones de nuevos usuarios que se nos hayan pasado por alto en el primer paso.

Elección del umbral

La elección del umbral no fue azarosa. Se realizaron varias pruebas con distintos umbrales para ver cuál se adaptaba mejor a nuestros datos. Para ello se tuvo en cuenta el número total de visitas calculado previamente en la sección 2.2. Si observamos el cuadro de la Figura 4a y realizamos los cálculos necesarios, obtenemos un total de visitas de 89092. Se esperaría obtener entonces un número de sesiones similar, ya que puesto en otras palabras una sesión vendría a ser lo mismo que una visita. En el cuadro de la Figura 8 se observan los resultados obtenidos para el número de sesiones variando el umbral.

Umbral [horas]	Número de sesiones
0.50	88743
0.40	90752
0.45	89655
0.46	89461
0.47	89273
0.48	89071

Figura 8: Número de sesiones para distintos umbrales de tiempo de sesión

Resulta entonces más conveniente utilizar un umbral de **0.48 horas** ya que es con el que se obtiene un valor muy cercano al número de visitas previamente calculado.

Luego de haber identificado las distintas sesiones y de haber obtenido el tiempo correspondiente a cada una, se decidió confeccionar un *scatter plot* para poder visualizar la distribución de las sesiones segun el tiempo en horas de cada una. Los resultados obtenidos se hallan en la Figura 9.

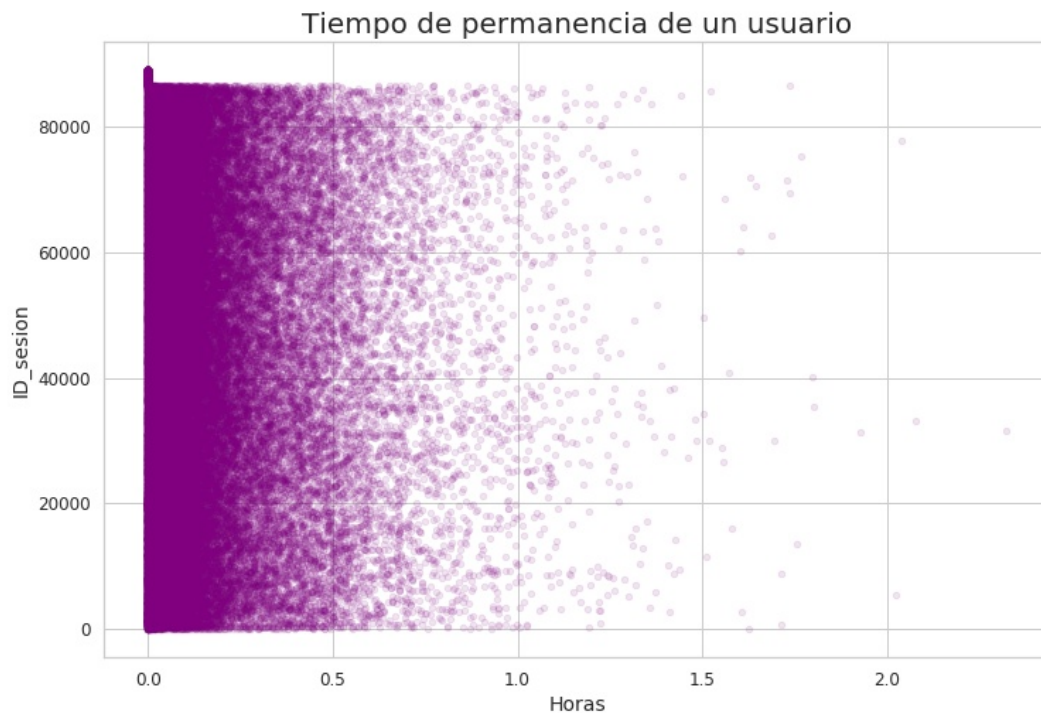


Figura 9: Distribución de las sesiones según la duración de cada una

Como podemos observar en el *scatter plot*, la mayoría de las sesiones se concentran en la franja de duración 0-0.25 horas (o bien 0-15 minutos). Hay muy pocas sesiones de duración mayor a 1 hora y se tornan despreciables las mayores a una hora y media. Notese arriba a la izquierda un conjunto de sesiones nulas para los números de ID más altos. Estas se corresponden con los IDs agregados manualmente para compensar aquellas sesiones de evento único que se eliminaron con el método utilizado.

Para conseguir un análisis más detallado se realizó un *describe* sobre los datos, de donde se obtuvo una duración promedio de 7.18 minutos.

Resulta de gran interés analizar el porcentaje de personas que visitaron la web y se marcharon en un lapso de tiempo corto. Es decir, medir la *tasa de rebote*.

2.3.1. Tasa de rebote

La tasa de rebote es una de las principales métricas para medir el *engagement*. En nuestro análisis consideraremos un tiempo de sesión satisfactorio de 30 segundos o más. Por lo tanto, se considerara como rebote aquellas sesiones menores a este lapso de tiempo.

Habiendo analizado previamente el tiempo de permanencia de los usuarios en la página resulta muy sencillo obtener la tasa mencionada. En la Figura 10 se presenta una comparación con gráficos de barra entre aquellas sesiones de duración mayor y menor a 30 segundos.

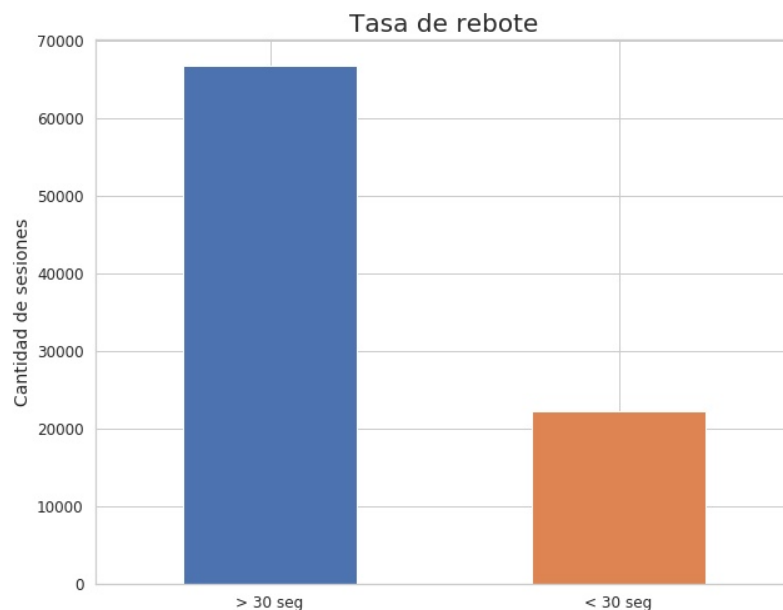


Figura 10: Comparación entre sesiones mayores y menores a 30 segundos

Tanto de los valores numéricos como del gráfico de barras podemos observar una tasa de rebote de aproximadamente el 25 %. Siendo más detallistas y utilizando los valores numéricos dados en el *notebook*, vemos que tenemos efectivamente un 33.3 % de rebote. Si bien este porcentaje es significativo, se encuentra dentro de los rangos ideales (<40 %), por lo que no debería ser un motivo por el cual preocuparse.

2.4. Usuarios por mes dependiendo el sistema operativo

En esta sección veremos como es el flujo de los sistemas operativos más comunes de los usuarios que ingresan en la página. Dado que estos sistemas operativos tienen muchas variaciones se decidió acotarlos por el nombre y sin sus versiones específicas para poder observar mejor qué tipo de usuarios entran en el sitio web.

Para computar los ingresos por sistema operativo elegimos los más comunes hasta la fecha (Android, Windows, iOS, OS X) usando el campo de *operating system* y separando los datos de los distintos sistemas para

después analizar sus comportamientos en la página (ver Sección 4). Luego decidimos analizar los ingresos por mes para ver si se mantenía una tendencia con algún sistema operativo, con el campo *timestamp* obtuvimos los meses de este registro y realizamos el gráfico de la Figura 11.

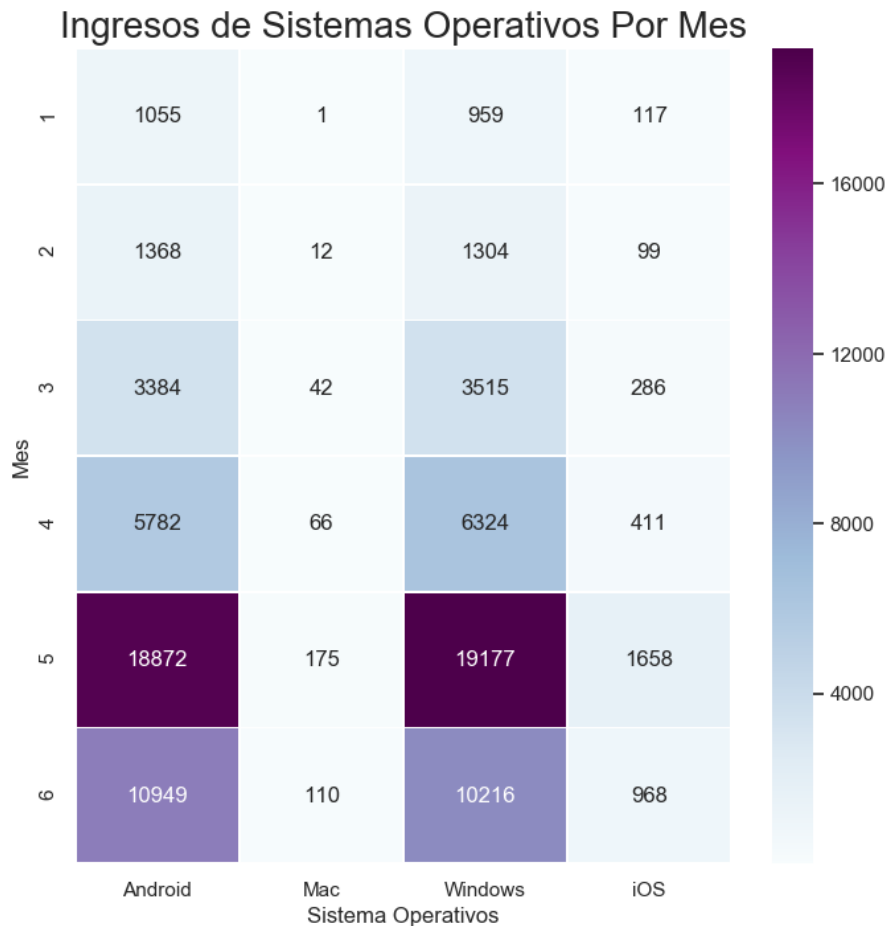


Figura 11: Mapa de calor de ingresos de los sistemas operativos

Lo que nos muestra es que el flujo de usuarios en la página es ampliamente mayor por parte de los usuarios de Android y de Windows en todo el período de los datos analizados por lo que nos lleva a hacer análisis más profundo a estos en otra sección.

2.5. Cantidad de ventas por día de semana.

Lo siguiente será analizar cómo es la distribución de ventas de equipos por día de la semana, a lo largo de los meses presentes en el set de datos. Recordar que, para indicar que un usuario finalmente compró un producto, en el dataframe se lo indica generando un evento *conversion*. Entonces, el procedimiento fue tomar las filas del set de datos cuyo evento fuera el mencionado, agruparlas por fecha en el que se dieron, y luego contar para cada uno de los 7 días de la semana cuantas conversiones de dieron. El resultado se puede ver en la Figura 12.

El mapa de calor parece ser una buena herramienta para poder visualizar un evento que llama la atención: los días con menor cantidad de ventas parecerían ser **Sábados y Domingos**. Mientras que, los días con mayor cantidad de ventas parecerían ser días de semana como **Martes, Miércoles o Jueves**.

Creemos que este análisis es útil principalmente para conocer más sobre la tendencia que tiene la gente a la hora de comprar un producto, ya que puede ser mas útil todavía a la hora de pensar estrategias de marketing, ventas, campañas, etcétera. Incentivar a los clientes a seguir comprando durante la semana, o realizar promociones o descuentos los días de fin de semana para aumentar las ventas son algunos ejemplos sobre posibles soluciones.



Figura 12: Ventas por día de la semana

2.6. Análisis de Campañas en el tiempo

Realizaremos un análisis general de las campañas, ya que observando por arriba los datos obtuvimos que el mayor porcentaje de los *ad campaign hit* son provenientes de la campaña de Google y gran cantidad de usuarios que usan otras campañas también usaron Google, por lo que decidimos hacer un enfoque general tomando a todos los usuarios que alguna vez entraron al sitio por medio de una campaña.

2.6.1. Visitas

Realizamos una separación en los datos con todos los registros que tenían alguna información sobre qué campaña produjo su ingreso al sitio y luego con el campo *timestamp* conseguimos la fecha en la que se ingresó por campaña. Para visualizar esto, se confeccionó el gráfico de la Figura 13.

Podemos observar un claro aumento en los meses con respecto a sus visitas, lo que en una primer idea nos haría a pensar que las ventas también tuvieron un aumento, aunque lo veremos con más profundidad en la siguiente sección. Por otro lado podemos observar que la campaña del mes de mayo logró atraer una gran cantidad de usuarios y beneficiar a la página.

2.6.2. Compras

Analizaremos a cualquier persona que tuvo contacto con algún tipo de campaña y que también realizó una compra en el sitio. Esto nos permite ver qué proporción de los que alguna vez entraron por campaña fueron un "usuarios de interés".

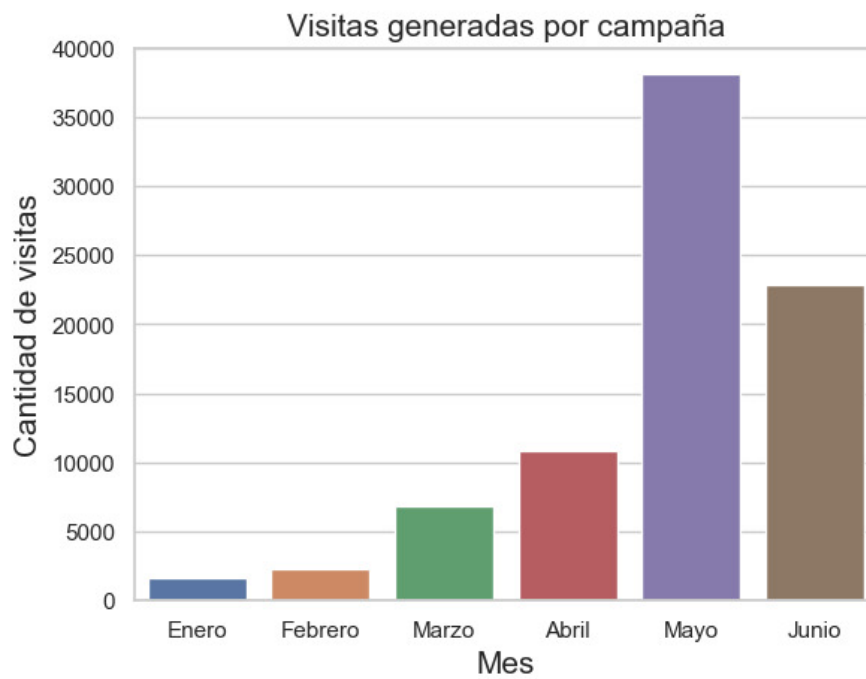


Figura 13: Cantidad de visitas generadas por campaña para los distintos meses

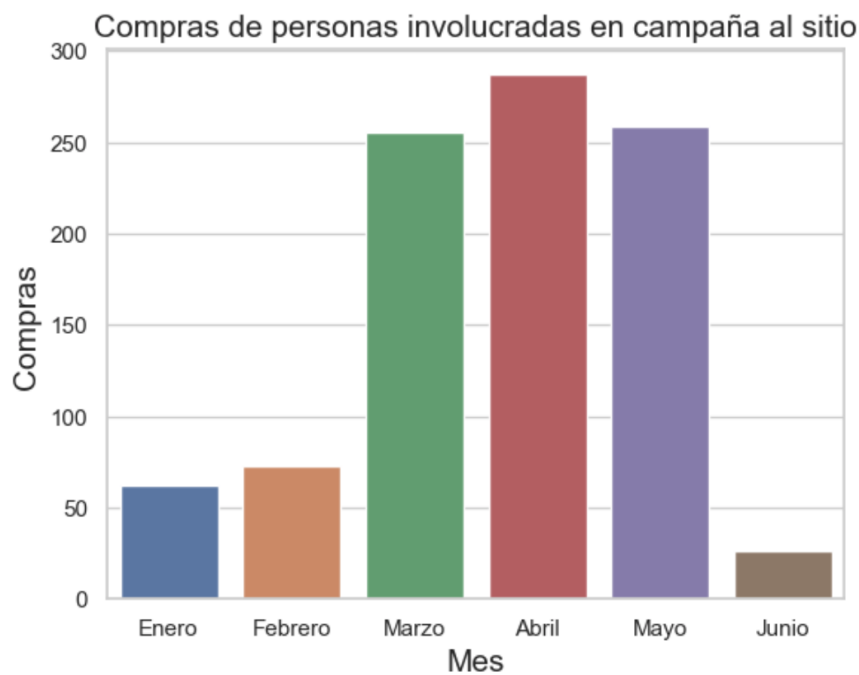


Figura 14: Cantidad de compras de usuarios que ingresaron por campaña para cada mes

Si nos detenemos en el gráfico de barras obtenido en la Figura 14 y lo comparamos con el de la Figura 13, vemos que la cantidad de compras no tiene una misma tendencia respecto a la cantidad de visitas generadas por campaña para la mayoría de estos meses. Es interesante notar que los meses de Marzo y Abril tuvieron campañas que significaron una buena cantidad de compras en relación a la cantidad de visitas generadas por este medio. Por otro lado, para el mes Mayo el gran incremento de visitas que trajo generó compras parecidas

a las de Marzo, la cual tuvo mucho menos visitas debidas a campañas. El mes Junio generó mucho menos compras que meses con menos visitas debidas a publicidad, por lo que nos hace pensar que en ese mes no hubo una campaña atractiva para los usuarios.

3. Análisis Demográfico

En esta sección del documento analizaremos el dataframe a trabajar desde su aspecto geográfico. El mismo presenta columnas de *región* y *country* que guardan información del lugar de donde las personas intentan acceder al sitio, generando un evento "*visited site*" (en la columna *event*). Creemos que este análisis es de útil conocimiento ya que, por ejemplo, es importante conocer las regiones donde Trocafone es mas popular, para poder encarar un desarrollo apropiado de la empresa en dichas zonas del mundo.

3.1. Cantidad de visitas al sitio en todo el mundo

Como se mencionó previamente, cada vez que un usuario en alguna parte del mundo intenta acceder al sitio, genera un evento '*visited site*', y se registra en las demás columnas del dataframe cierta información local del usuario. Para comenzar, tomamos la columna *country* para analizar desde qué parte del mundo acceden los usuarios y con qué frecuencia acceden desde dicho país. Nuestro criterio es, básicamente, ver cuantos eventos se generan desde los distintos países y luego ver que países generan más eventos o más visitas al sitio.

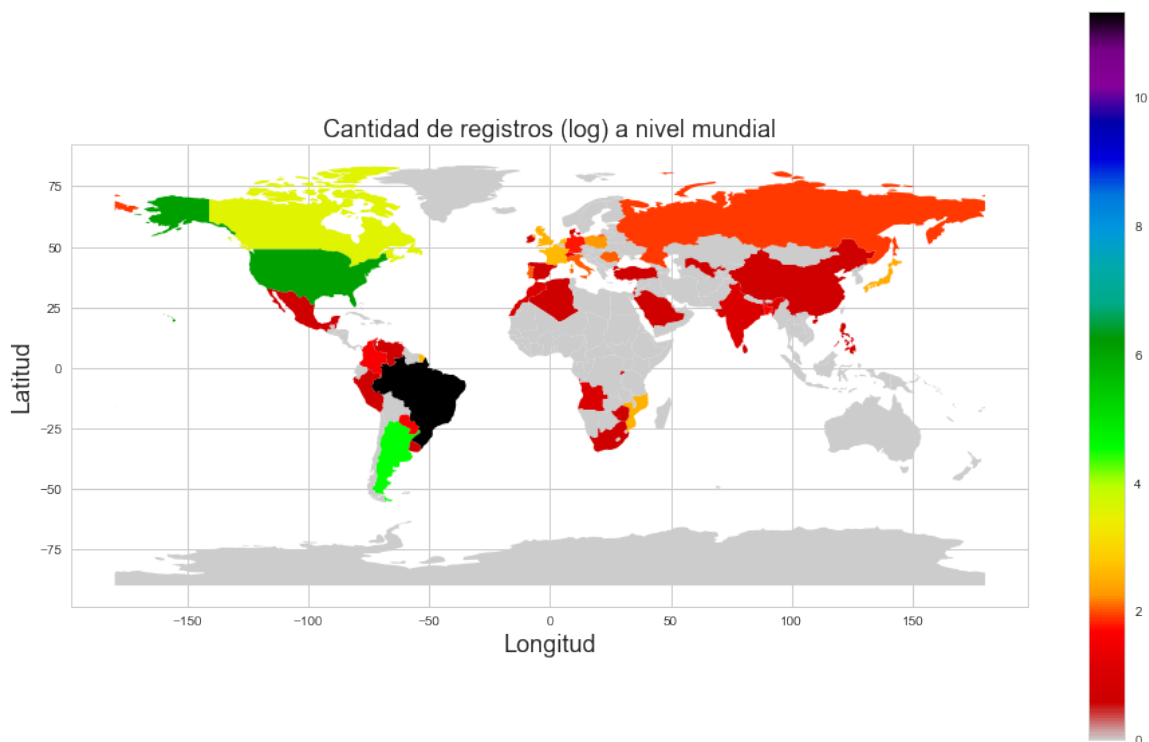


Figura 15: Cantidad de accesos por país

A partir de la Figura 15, se pueden dar algunas consideraciones importantes. Para empezar, el gráfico no muestra precisamente la cantidad de visitas por país sino que muestra el logaritmo de dicha cantidad. Esto se debe a que Brasil genera casi 200 veces más registros que el país que le sigue, lo que provocaría que la visualización no sea tan clara por la disparidad que ocasiona este hecho. Podría claramente deberse al origen de procedencia de la empresa: es lógico que la popularidad de Trocafone sea mayor en su país de origen. Pero, de la forma en que fue realizado el gráfico, se puede ver claramente que Brasil, Estados Unidos y Argentina ocupan el podio de los países con mayor cantidad de visitas al sitio. Por otra parte, se puede ver que casi todos los países en el continente americano tienen algunas visitas al sitio. También se registraron eventos procedentes de Europa (en España o Francia por ejemplo), Asia y África. No se registraron eventos en Oceanía.

Entonces, dado que Brasil y Estados Unidos son los países con mayor frecuencia de visitas al sitio, procederemos a analizarlos con consideración especial sobre ellos.

3.2. Análisis de visitas al sitio en Brasil

Como se analizó anteriormente, Brasil es el país donde la empresa parece ser mas popular entre sus habitantes. Entonces, haciendo un análisis similar, procederemos a obtener la información regional de nuestro país vecino, y volcarla en un mapa territorial.

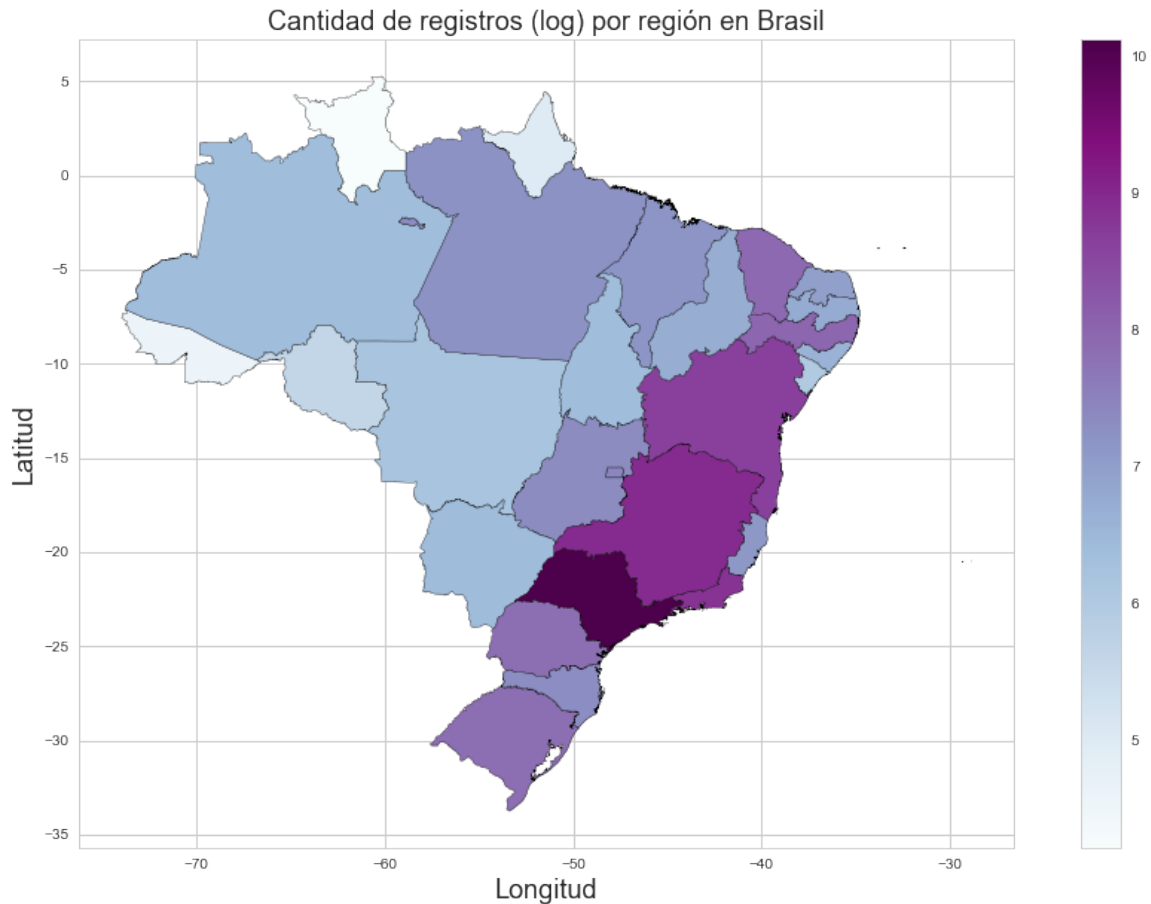


Figura 16: Accesos por estados en Brasil

Al igual que en la Figura 15, el gráfico muestra el logaritmo de la cantidad de registros por estados en el país. Esto se debe a que, al analizar los datos, nos encontramos con que la cantidad de registros en las regiones sur, sudeste y noroeste, las cuales contienen la mayor parte de la población, superan ampliamente la cantidad de registros en las regiones Norte y Centro-Oeste. Los estados con mayor cantidad de eventos generados son São Paulo, Minas Gerais, Bahia, Pernambuco, Ceará y Rio Grande do Sul.

Sería de gran importancia entonces promover el desarrollo de la empresa en las zonas mencionadas.

3.3. Análisis de visitas al sitio en Estados Unidos

El segundo país con mayor cantidad de registros en el Dataframe es Estados Unidos. De manera similar, obtenemos el siguiente gráfico.

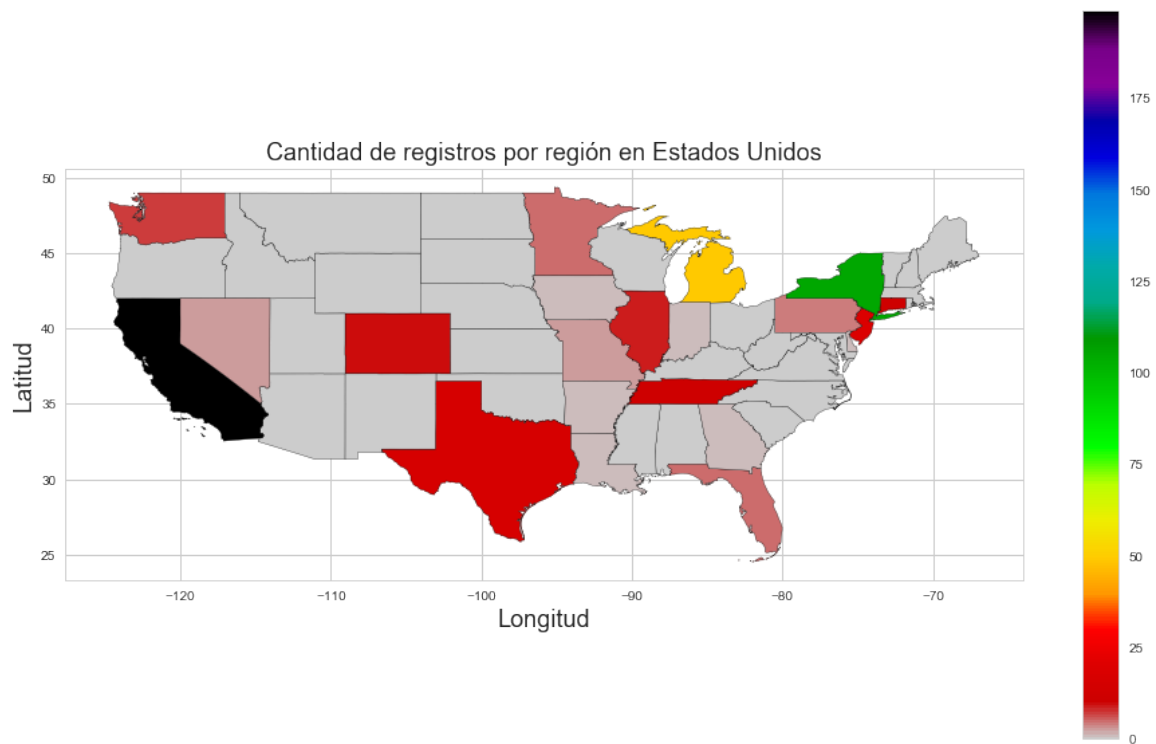


Figura 17: Accesos por estados en Estados Unidos

A partir de la Figura 17 podemos concluir que California, Nueva York y Michigan son los estados con mayor cantidad de registros del país americano. A diferencia con Brasil, en Estados Unidos no se tienen registros de todos los estados; de hecho, no se tiene registros de la mayor parte de ellos. Y también, debido a que esta vez no hay tanta diferencia de registros entre los distintos estados, este gráfico sí muestra la cantidad de registros por estado.

Como conclusión, además de sugerir lo mismo que para Brasil, se podría recomendar la incentivación, por medio de publicidad o campañas, a aquellos sectores de dicho país donde todavía la popularidad de Trocafone no es suficiente.

3.4. Tipo de acceso al sitio

En esta sección analizaremos desde qué dispositivos frecuentan acceder al sitio las personas en todo el mundo. Para ello, tomamos la columna *device type* del dataframe que divide los accesos en tres grupos principales: Smartphone, Computer y Tablet. En dicho análisis, separamos el dataframe por países y analizamos de qué manera acceden los usuarios de dichos países, dividiéndolos en dos grupos: Mobile (smartphones y tablet) o No Mobile (computer).

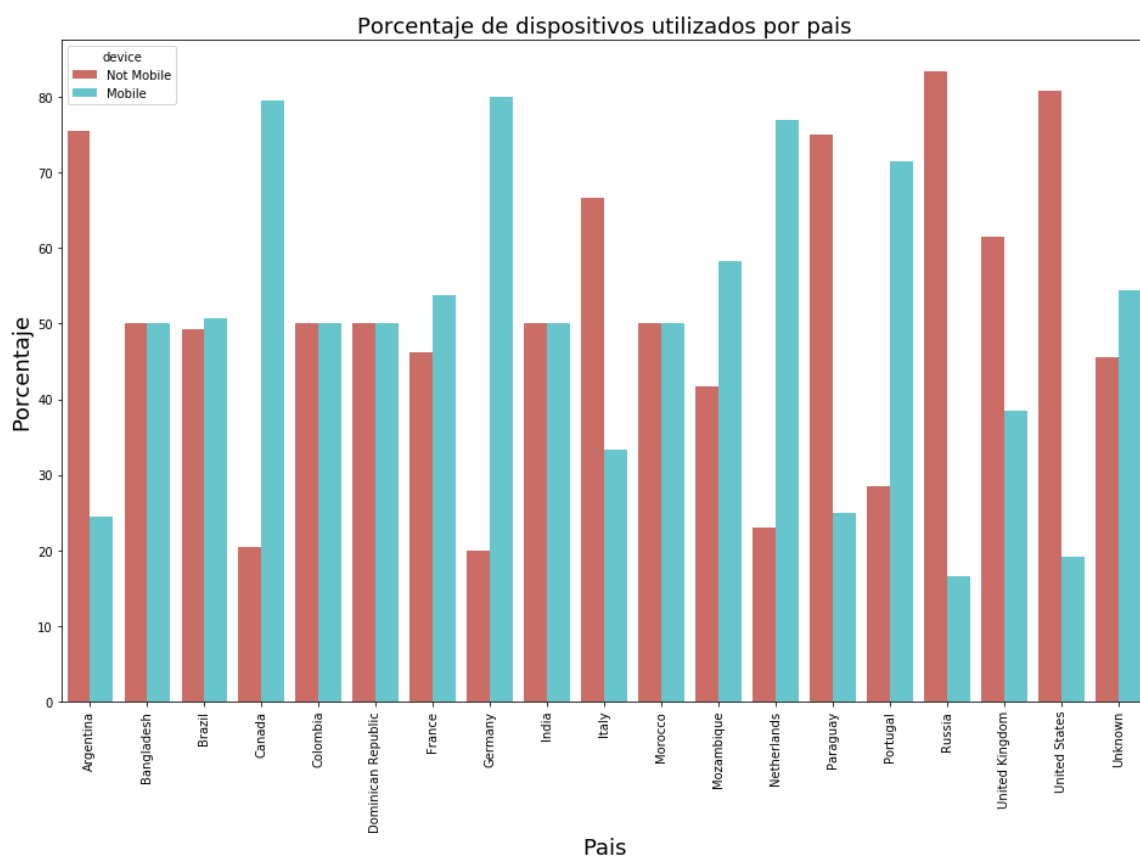


Figura 18: Tipos de accesos al sitio por país

La Figura 18 muestra el porcentaje de usuarios que ingresa por *Mobile* o *No Mobile*, por país. Si bien muestra diversos países, hay que tomar mayor consideración en los más relevantes. Para comenzar, Brasil, que tiene la mayor cantidad de registros, tiene apenas más ingresos por dispositivos móviles que por computadoras. Pero, para los casos de Estados Unidos y Argentina, se puede observar lo contrario: el ingreso por dispositivos móviles es ampliamente superado por el ingreso por computadoras.

Este análisis es muy útil a la hora de decidir qué recursos desarrollar en las distintas regiones. Por ejemplo, dado que en países de habla portuguesa es más frecuente el acceso por teléfonos y otros dispositivos móviles, sería más adecuado destinar recursos para esos tipos de dispositivos, como aplicaciones o sitio web optimizado. En cambio, en países como Estados Unidos o Argentina, sería conveniente desarrollar u optimizar los recursos para computadoras.

3.5. Teléfonos más buscados por país

Ahora analizaremos en distintos países cuáles son los productos más buscados en la página, por regiones. Utilizamos, básicamente, las columnas *region*, *country* y *model*. También buscamos cuales son los 5 productos más solicitados a nivel territorial.

3.5.1. Brasil

Para el territorio brasileiro, se buscaron cuales eran los 5 productos más visitados en la página. Luego, se dividió el análisis por Estados y se obtuvo el siguiente resultado:

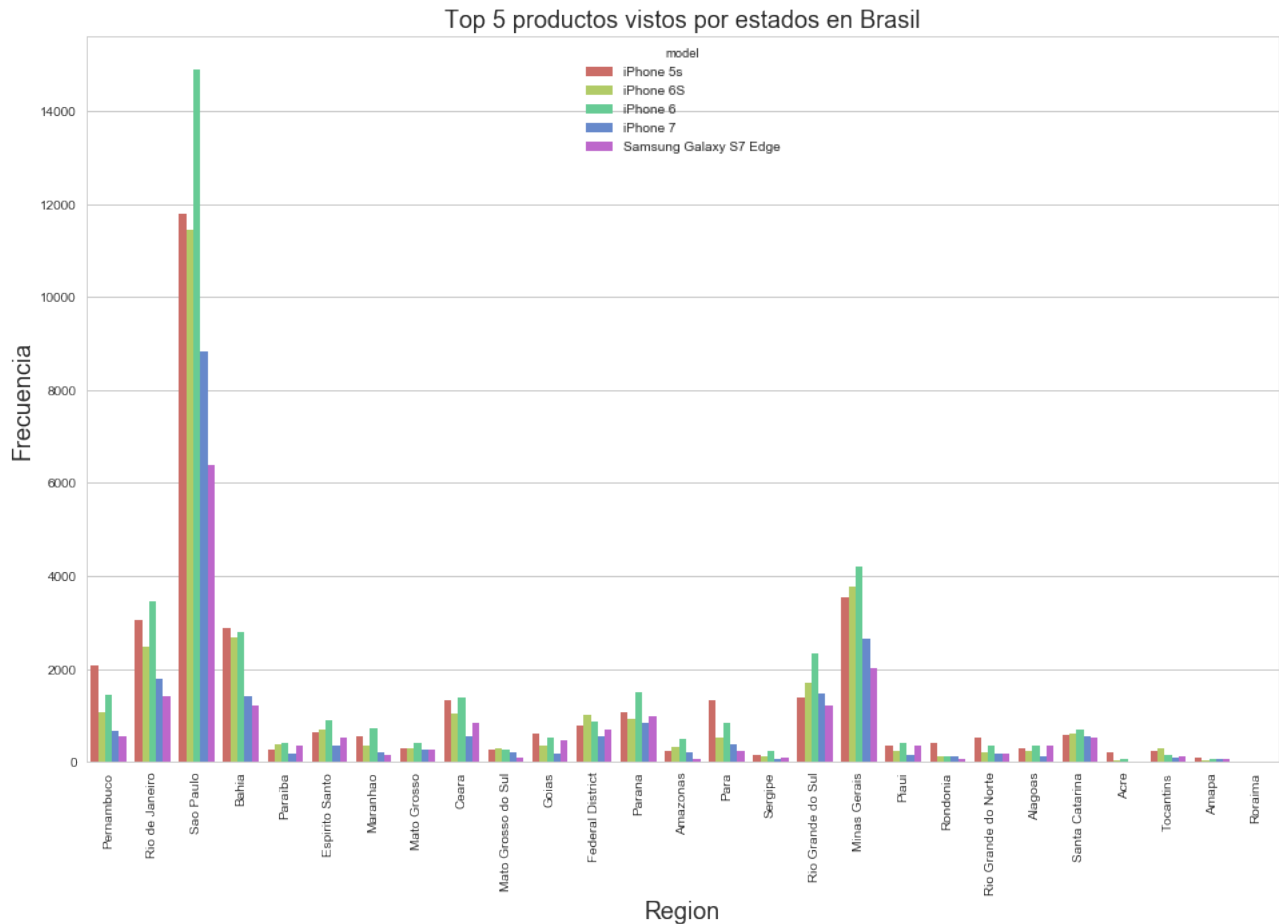


Figura 19: Top 5 de productos solicitados en Brasil

La Figura 19 muestra los 5 productos mas visitados del sitio (iPhone 5s, iPhone 6s, iPhone 6, iPhone 7 y Samsung Galaxy S7 Edge) en las distintas regiones del territorio brasileño. Sao Paulo, el Estado con mayor cantidad de usuarios registrados, tiene como cabeza de búsqueda el modelo iPhone 6, al igual que en Minas Gerais, Rio Grande do Sul y Rio de Janeiro.

3.5.2. Estados Unidos

De la misma forma, para Estados Unidos:

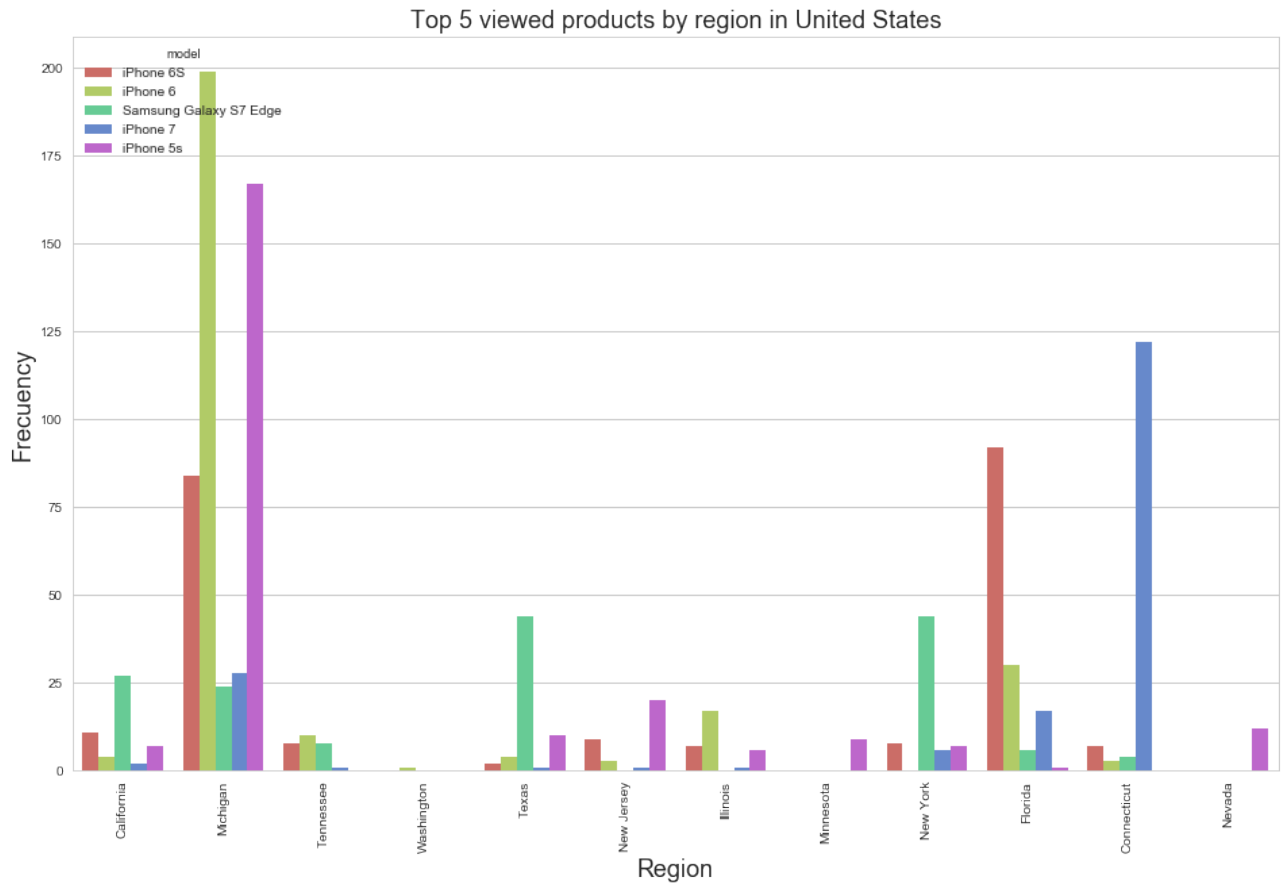


Figura 20: Top 5 de productos solicitados en Estados Unidos

Observando la Figura 20, podemos ver un dato curioso: Michigan tiene mayor cantidad de registros de búsqueda de los top 5 productos, y no así California, que se había observado que era el Estado con mayor cantidad de registros. Quiere decir que la gente habitante de dicho Estado no está tan interesada en estos 5 productos más buscados en Estados Unidos.

3.5.3. Argentina

Ahora, el análisis se centrará en el territorio argentino.

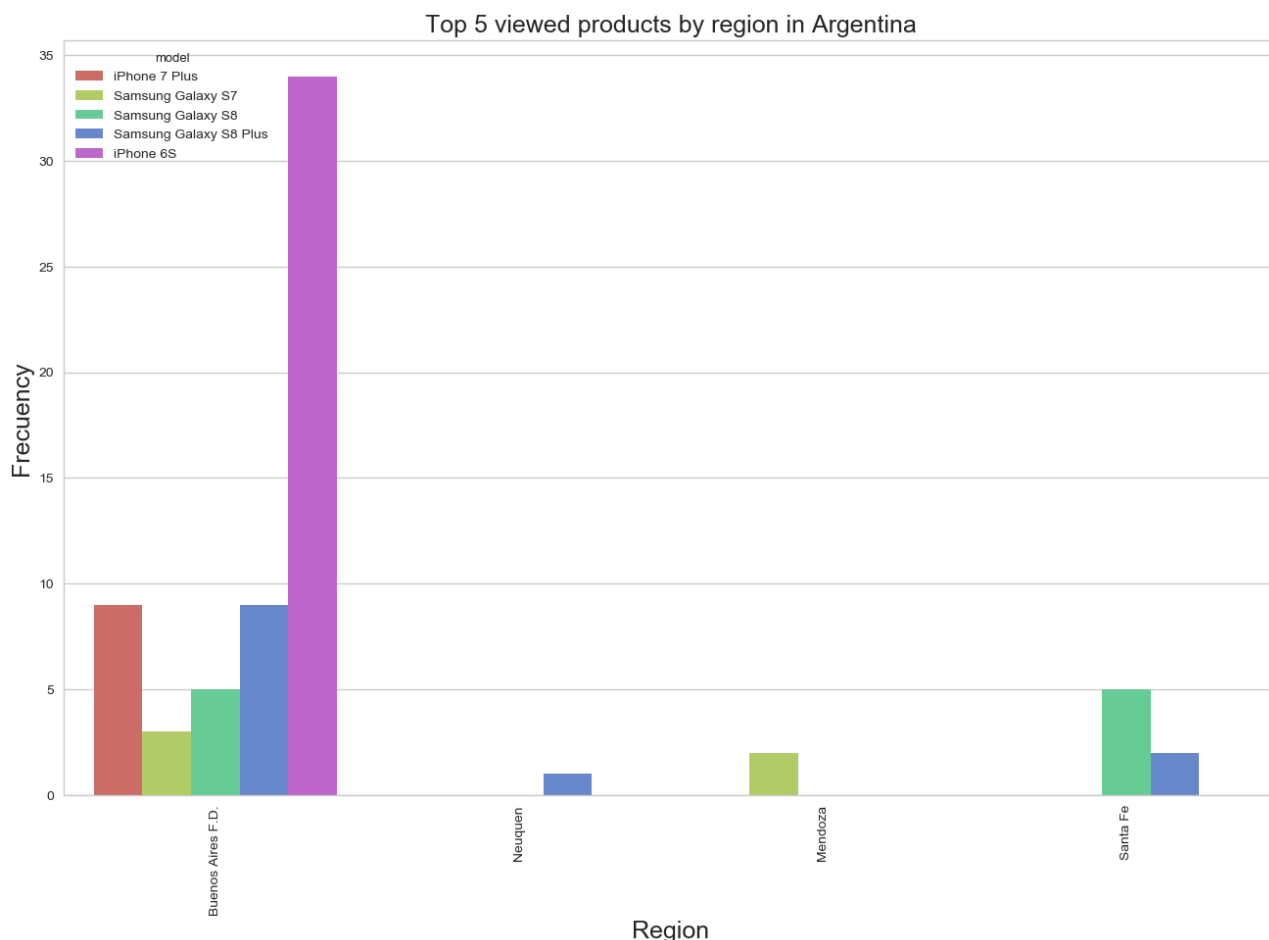


Figura 21: Top 5 de productos solicitados en Argentina

Este último análisis no es tan demostrativo como los dos anteriores, debido a la poca cantidad de registros disponibles para Argentina. Sin embargo, podemos destacar que Buenos Aires es la provincia con mayor cantidad de visitas a los dispositivos que aparecen en la Figura 21.

Otro aspecto a destacar es la frecuencia con la que aparecen los mismos dispositivos, sin importar que país estemos analizando. Por ejemplo, los teléfonos desarrollados por Apple son muy solicitados a nivel mundial. Los iPhone en sus versiones 6, 6S, 7 y 5S son los más buscados en el sitio. Más adelante se verá en qué estado los usuarios prefieren comprar su nuevo producto, pero por ahora podemos ver qué modelos son de su mayor interés.

4. Análisis de sistemas operativos

Del análisis hecho en la sección 2.4 sobre qué sistema operativo generaba más visitas al sitio web podemos conseguir cosas muy interesantes sobre los comportamientos según el sistema y así generar mejores recomendaciones en la página al usuario si se observara alguna tendencia.

Tendremos en cuenta la siguiente aproximación: si una persona entra por lo menos una vez con un sistema operativo tendremos en cuenta toda la información de todas sus acciones en el sitio sin importar que, por ejemplo, luego haya decidido meterse con otra plataforma. Decidimos tomar esta aproximación ya que el 95 % de los usuarios solo ingresaron por medio de un solo sistema operativo.

El análisis lo haremos en los tres sistemas mas usados en el sitio ya que los datos de los demás empiezan a ser pocos como para realizar un buen análisis, esto se ve claro en la Figura 22.

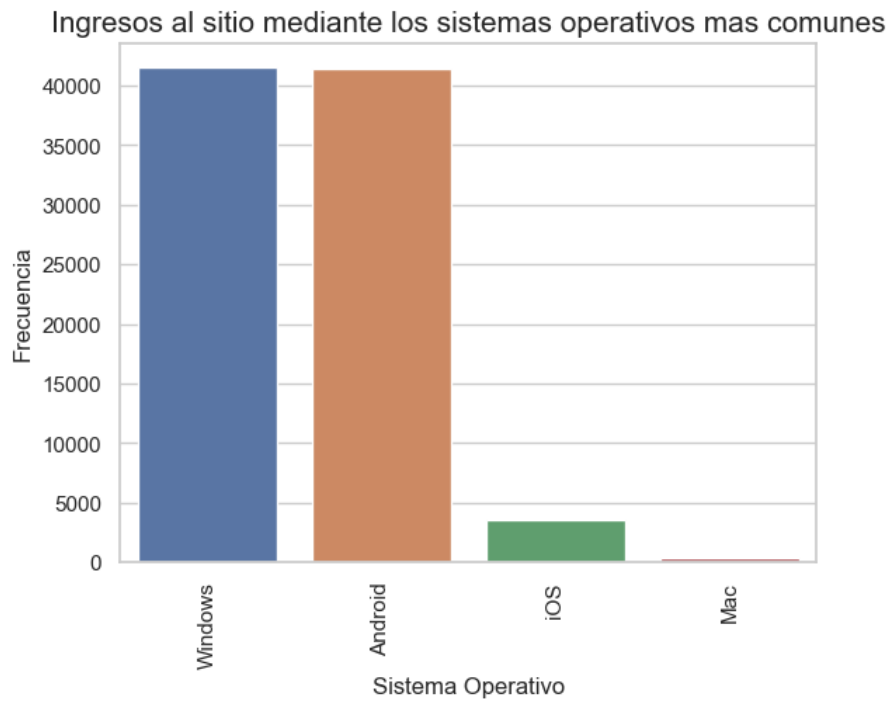


Figura 22: Ingresos al sitio mediante los sistemas operativos mas comunes

4.1. Usuarios de Android

Analizaremos cómo se comportaron los usuarios que entraron a la página por medio de un dispositivo con sistema operativo Android y buscaremos encontrar una relación entre los modelos buscados y su OS si la hubiese.

4.1.1. Relación de modelos

Iremos analizando del campo *event* cada caso que sea un *viewed product* y viendo cuáles fueron los modelos con más apariciones en los datos de esta categoría. De esta forma generamos la Figura 23. Puede observarse claramente que los usuarios de Android visitan más dispositivos que poseen un sistema operativo iOS que el mismo que tiene su dispositivo actual.

Haciendo los mismos pasos pero esta vez separando el evento de *conversion*, obtenemos la Figura 24. Comparando con lo anterior vemos que los dispositivos más comprados y más buscados por estos mismos usuarios no comparten la misma tendencia, lo que nos muestra que los más vistos no son necesariamente los más comprados y que los usuarios suelen conformarse con teléfonos de gama más baja que la de los que visitan.

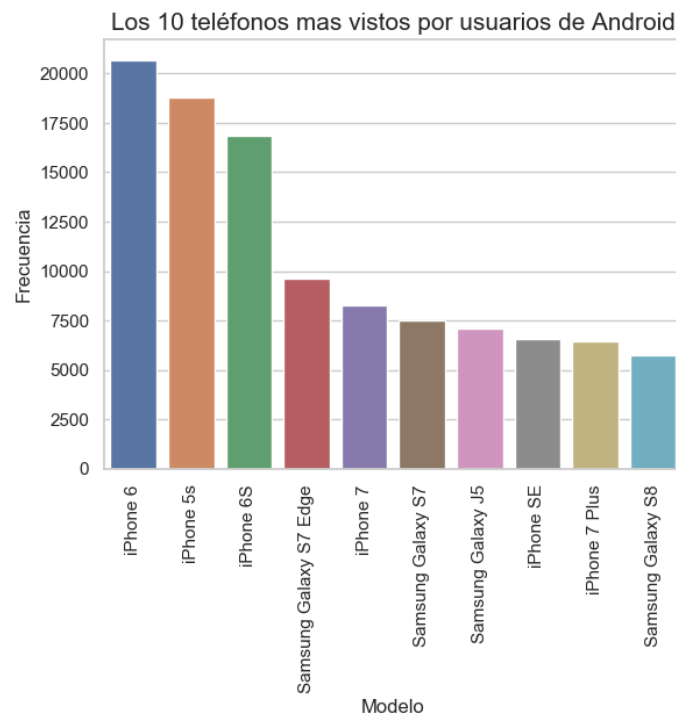


Figura 23: Los 10 teléfonos mas visitados por usuarios de Android

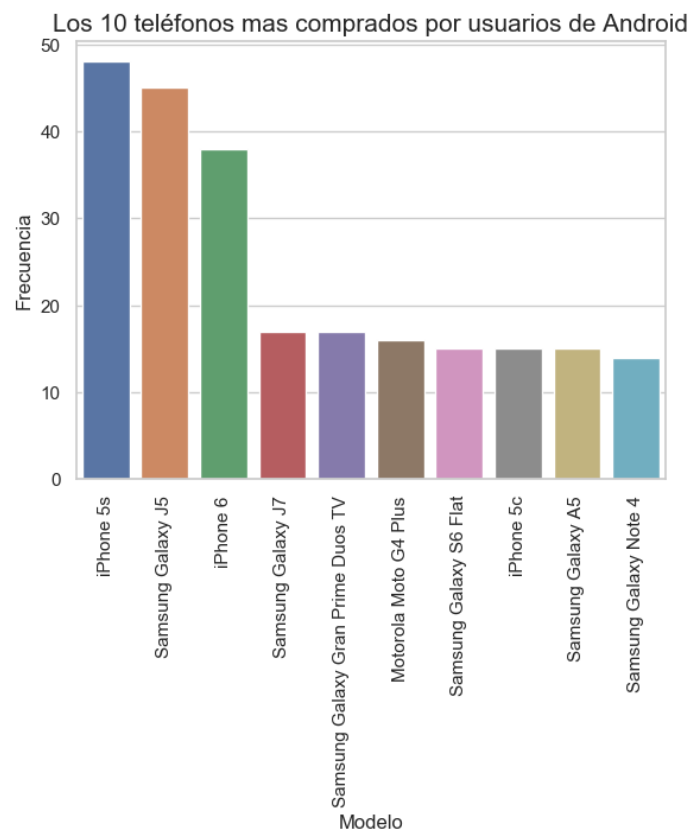


Figura 24: Los 10 teléfonos mas comprados por usuarios de Android

4.2. Usuarios de Windows

Ahora veremos como se comportan los usuarios de Windows realizando el mismo análisis que hicimos con los de Android en 4.1, observando si tienen una relación con su comportamiento.

4.2.1. Relación de modelos

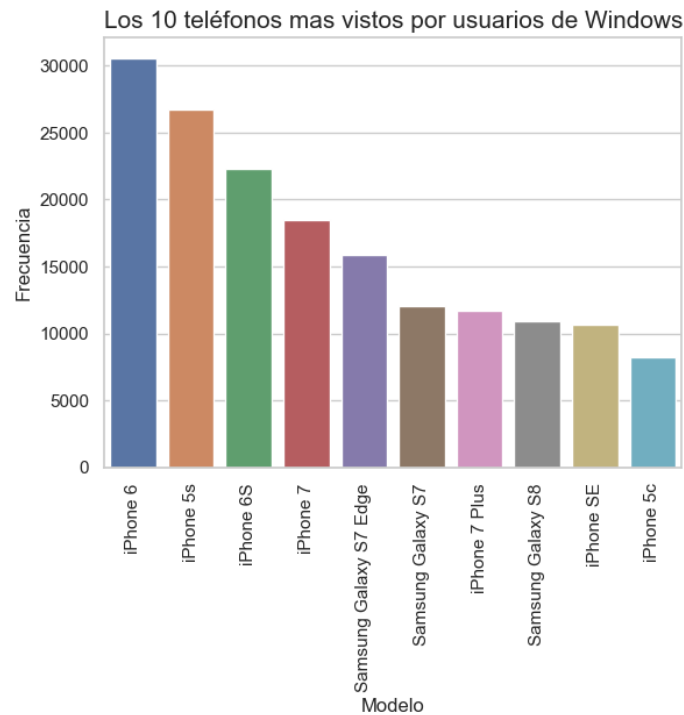


Figura 25: Los 10 teléfonos mas visitados por usuarios de Windows

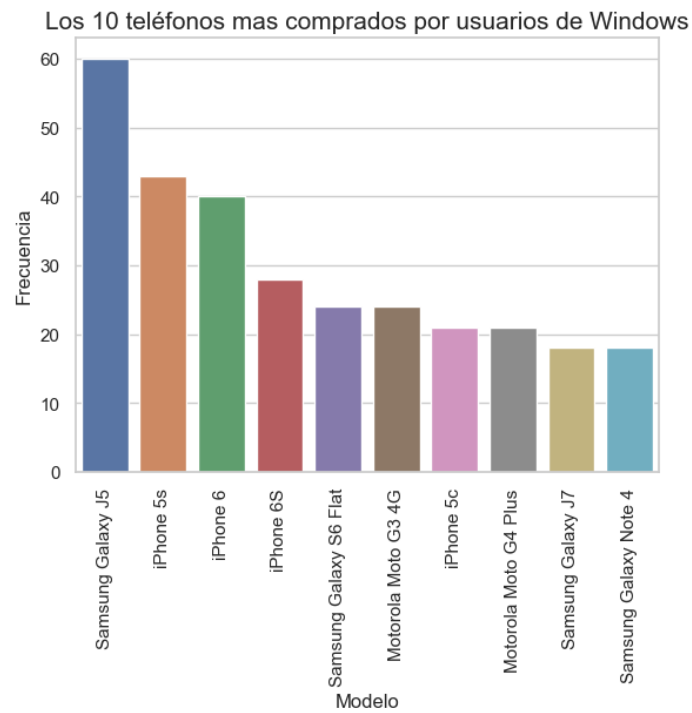


Figura 26: Los 10 teléfonos mas comprados por usuarios de Windows

Para este caso se nota una tendencia similar entre los productos visitados y los comprados. Se observa que el *iPhone 5s* deja de ser el más vendido y pasa a serlo el *Samsung Galaxy J5* que esta en la posición 16 de los más buscados (no figura en el gráfico de la Figura 25). En adición, más teléfonos con Android aparecieron en los más comprados por lo que podemos ver que los usuarios de Windows tienen un comportamiento parecido de observar teléfonos pero conformarse con productos de gama mas baja.

4.3. Usuarios de iOS

Decidimos agregar el análisis de iOS aunque la cantidad de usuarios y páginas vistas sean muy bajas, ya que podríamos ver si estos también tienen algún tipo de tendencia a la hora de comprar dispositivos.

4.3.1. Relación de modelos

Repetiremos el análisis hecho en 4.1 y 4.2 (ver Figuras 27 y 28). Vemos que claramente tienen una preferencia en buscar otros de su mismo sistema operativo, y mucho mas notorio que en los análisis anteriores.

Sin embargo, al momento de comprar vemos que aparecen algunos teléfonos Android, aunque no en la misma proporción que para los usuarios que accedieron desde Android o Windows. Al tener pocos datos de compras es mas difícil analizar en este caso, pero vemos que prefieren quedarse con el mismo sistema operativo, excepto para algunos casos.

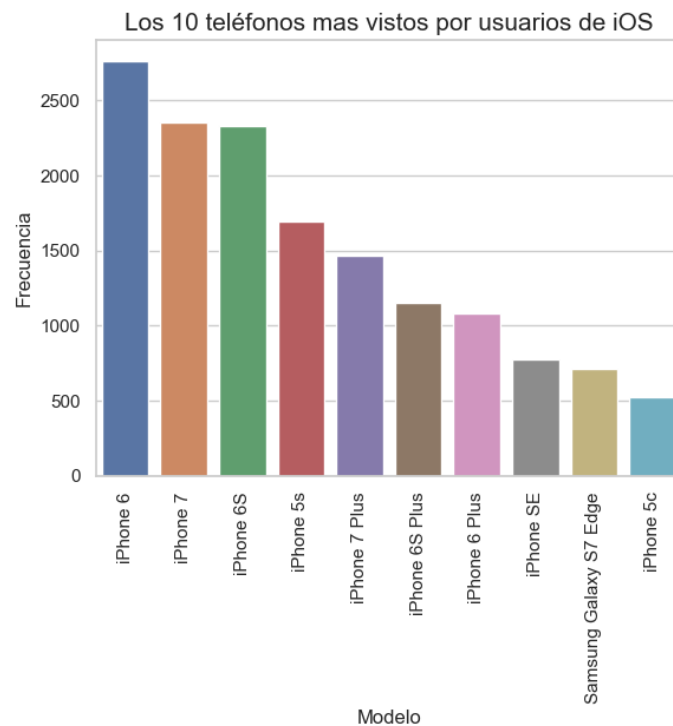


Figura 27: Los 10 teléfonos mas visitados por usuarios de Android

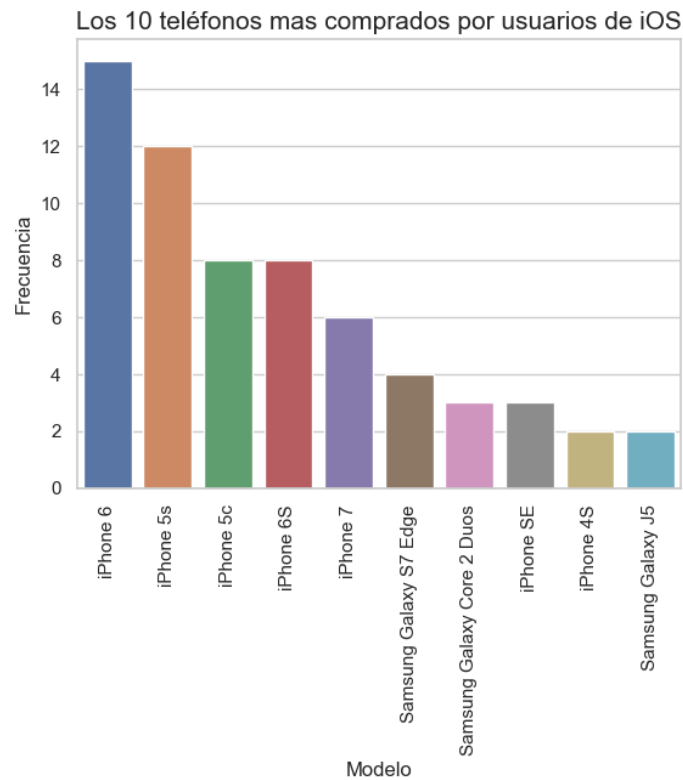


Figura 28: Los 10 teléfonos mas comprados por usuarios de Android

5. Análisis sobre los productos

En esta sección del informe nos centraremos en evaluar ciertas características sobre los productos, como el estado en el que se encontraban a la hora de ser vendidos, cuáles fueron los mas vendidos, etcétera.

5.1. Productos más vendidos

Para empezar con este análisis, veremos cuáles fueron los 20 productos más vendidos por Trocafone. Para hacerlo, tomamos solo las filas del campo *event* que registraron una *conversion* e hicimos un conteo de los productos que se vendieron. El resultado se puede ver en la Figura 29.

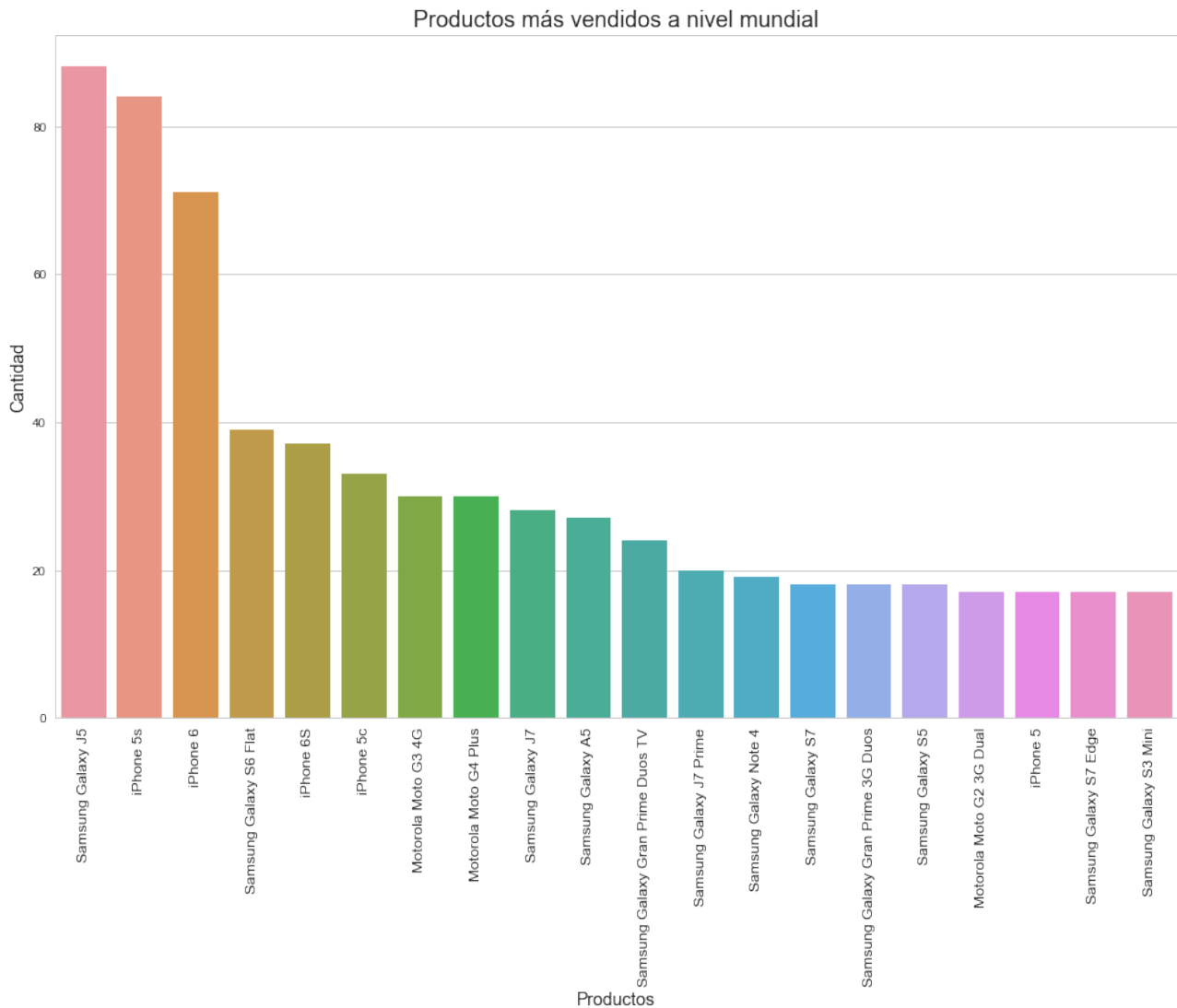


Figura 29: Los 20 productos más vendidos a nivel global

Entonces, podemos ver que el mercado esta repartido entre dispositivos Samsung, Apple y Motorola. Conocer qué productos son los más vendidos es útil para saber en dónde centrar las estrategias de marketing, promociones, campañas y demás. También para ver que tipo de gama de teléfonos es la que más se vende.

5.2. Condición de los teléfonos mas vendidos

Ahora, profundizando un poco más, sabiendo que Trocafone es una empresa especializada en un modelo de negocio como es el *ReCommerce*, queremos saber en qué estado están los productos que son vendidos a los distintos clientes. Para ello, hacemos una distinción sobre la columna *condition* del set de datos, una vez que se realizó un evento *conversion*. Juntando toda esa información, el resultado es visible en la Figura 30.

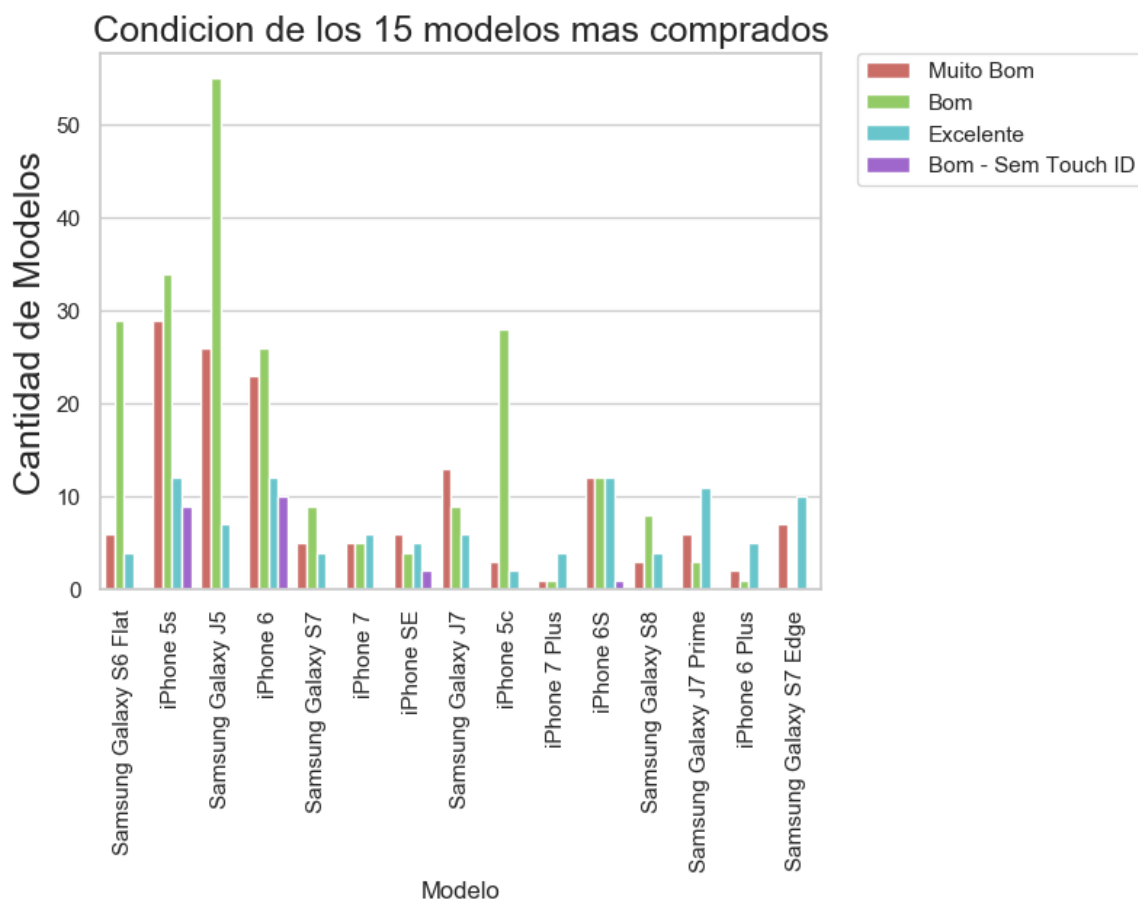


Figura 30: Condición de los 15 productos más vendidos a nivel global

En líneas generales, salvando ciertos casos aislados, el gráfico nos permite ver que parecería ser que la gente opta por comprar su nuevo equipo no en su mejor condición.

Aún más todavía: se puede ver que, entre los 15 teléfonos más vendidos ninguno fue adquirido en condición *Novo*. Y, en condición *Excelente*, solo algunos. La mayor cantidad de clientes los prefiere en estado Bom o Muito Bom.

Esta información es muy importante para la empresa. Es bueno saber que su política de ventas tiene cierto nivel de confianza por parte de sus clientes. También le es útil para seguir desarrollándose en ese ámbito.

5.3. Análisis del feature lead

El evento *lead* es cuando el usuario se registra para recibir una notificación de disponibilidad de stock, para un producto que no se encontraba disponible en ese momento. Nos pareció interesante analizar si este evento que muestra el interés del usuario por comprar logra obtener buenas ventas.

Lo que hicimos para analizar esto fue localizar qué personas realizaron este evento y luego buscamos cuales de ellas tuvieron algún evento de *conversion*. Los resultados se hallan en la Figura 31.

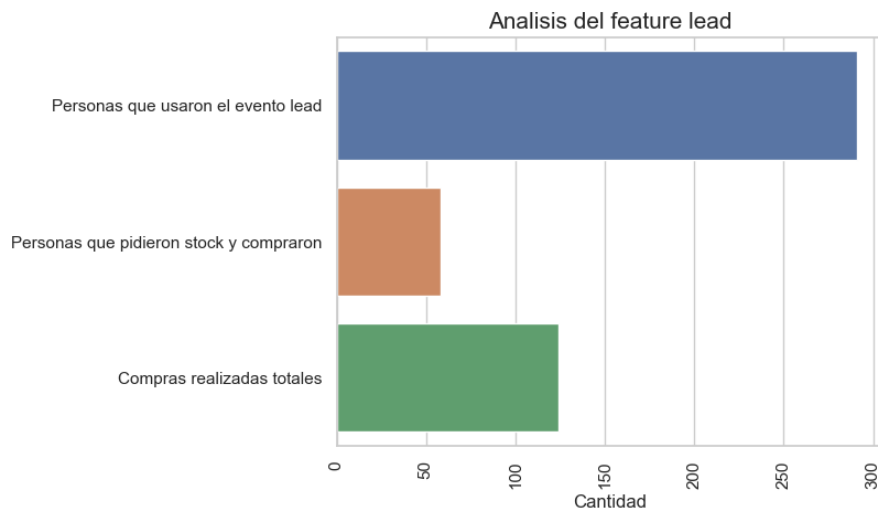


Figura 31: Resultados del análisis en lead

Podemos ver que estos usuarios generaron un poco mas que un 10% de las compras en el sitio web, lo cual nos hace pensar que este "feature" es muy bueno y al ser personas que muestran interés por productos podríamos ofrecerles dependiendo del modelo que eligieron una lista de dispositivos similares, esta puede estar formada de dispositivos que compraron los usuarios que mostraron interés por este teléfono o por teléfonos que compraron los usuarios que mostraron un interés similar al usuario que pidió el stock.

Ahora observaremos cuales son los modelos mas solicitados en este evento del sitio buscando a ver si este genero alguna relación con los que finalmente fueron comprados por estos mismos usuarios.

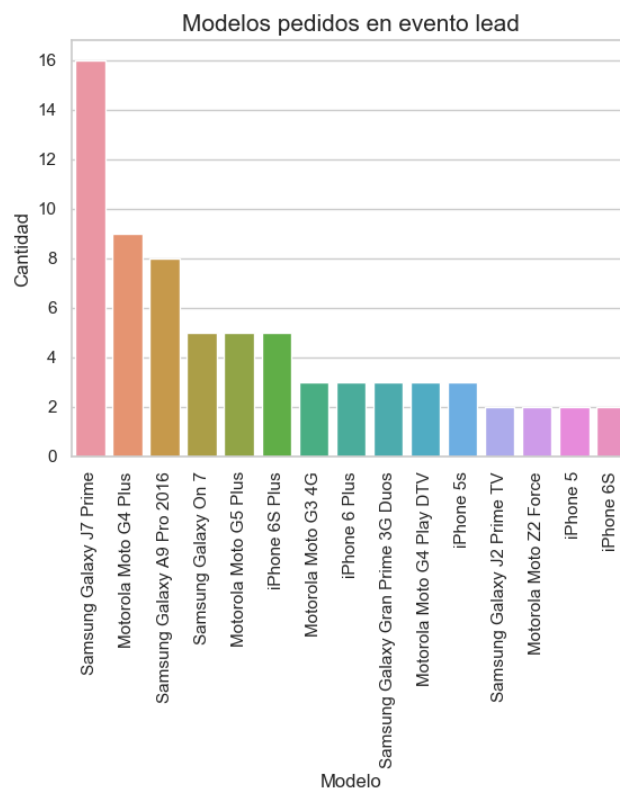


Figura 32: Modelos más solicitados con evento *lead*

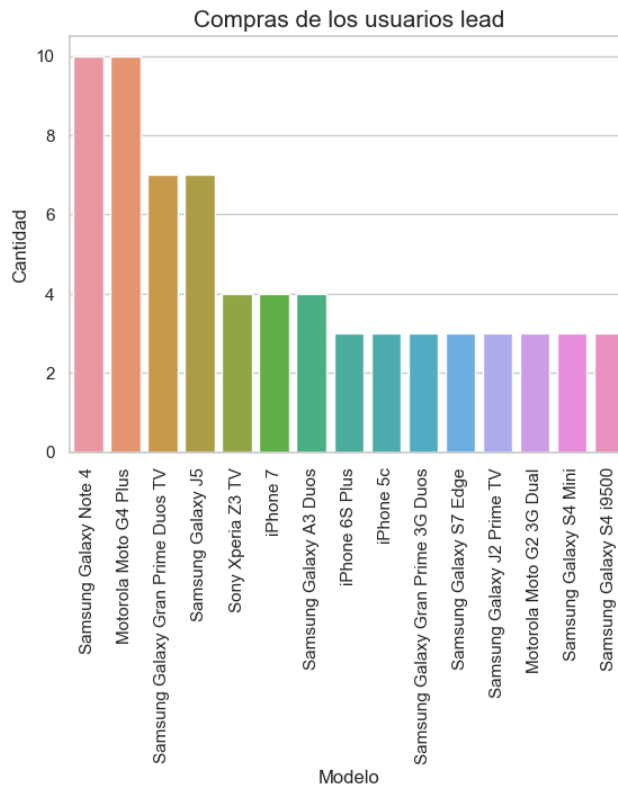


Figura 33: Modelos comprados con evento *lead*

Podemos observar que estos gráficos tienen un parecido simétrico muy interesante, aunque la mayoría no termina comprando estos teléfonos que aparecen en los mas buscados cuando no tienen stock. De todas formas esto muestra que hubo interés en comprar y terminaron eligiendo otro celular que los pudo satisfacer.

5.4. Análisis de ventas

En este apartado, consideramos únicamente las consultas que finalizaron en una venta. Entendemos la venta como el principal indicador de confianza, o al menos, el más objetivo. A diferencia del análisis de búsqueda donde hablamos de productos atractivos; aquí, hablamos de productos confiables. Observamos los modelos más vendidos y segmentamos intentando responder si un modelo genera mayor o menor confianza de acuerdo a su marca o de acuerdo a su año de lanzamiento.

5.4.1. Ventas por modelo

Agrupamos las ventas por modelo y seleccionamos un *top 8* de los más vendidos, no sólo porque el sitio ofrece demasiados modelos sino porque hay mucha diferencia de ventas entre los primeros de la lista y los últimos. Por lo tanto, con un *top* ya tenemos un grupo representativo del conjunto, sin excluir a ningún modelo con cantidad significativa de ventas.

Como primer acercamiento a los datos, es interesante ver de manera porcentual qué porción de la torta de ventas pertenece a cada marca disponible.

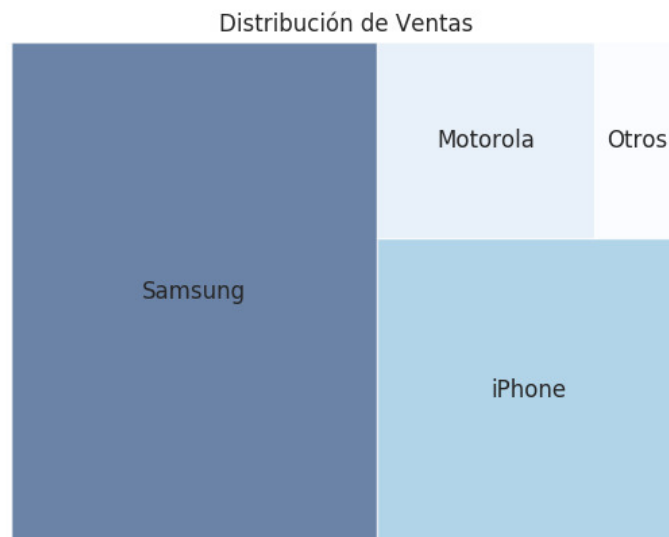


Figura 34: Proporción de las ventas según marca

A primera vista, parecería que la marca tiene una gran incidencia en la decisión de compra de los clientes, siendo Samsung la imperante, seguida por iPhone y Motorola.

Ahora bien, para saber si la confianza radica en el mero hecho de llevar la marca o si inciden más factores inherentes al modelo, exploramos el registro de ventas de cada uno de sus modelos.

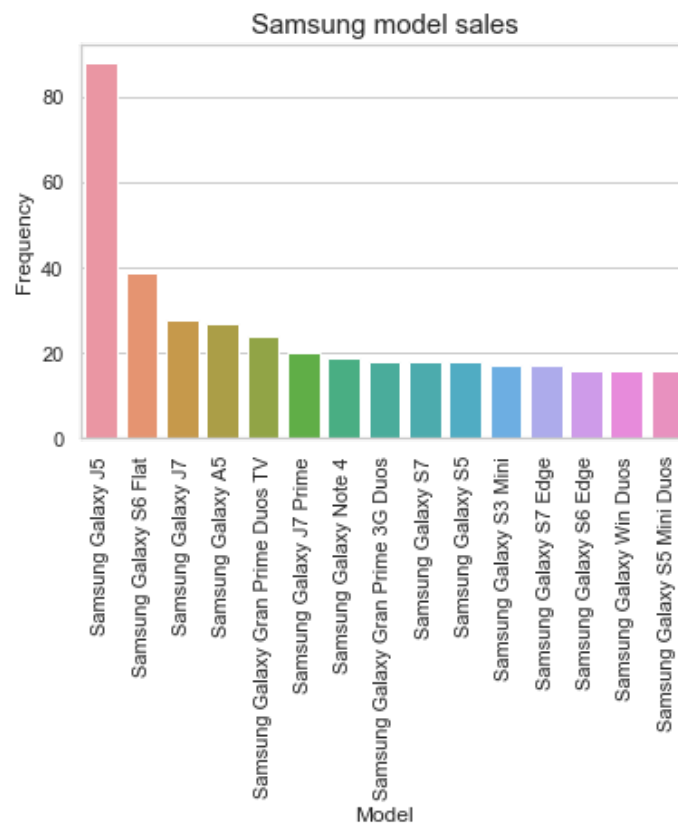


Figura 35: Ventas de modelos Samsung

En la Figura 35, vemos que Samsung tiene un piso de ventas bastante fijo y asegurado, independientemente del modelo. Esto podría indicar que la marca juega un rol clave.

De todas formas, es preciso analizar el caso particular del Samsung Galaxy J5 dado que no sólo supera ampliamente el piso de ventas para modelos Samsung, sino que encabeza la lista de modelos más vendidos.

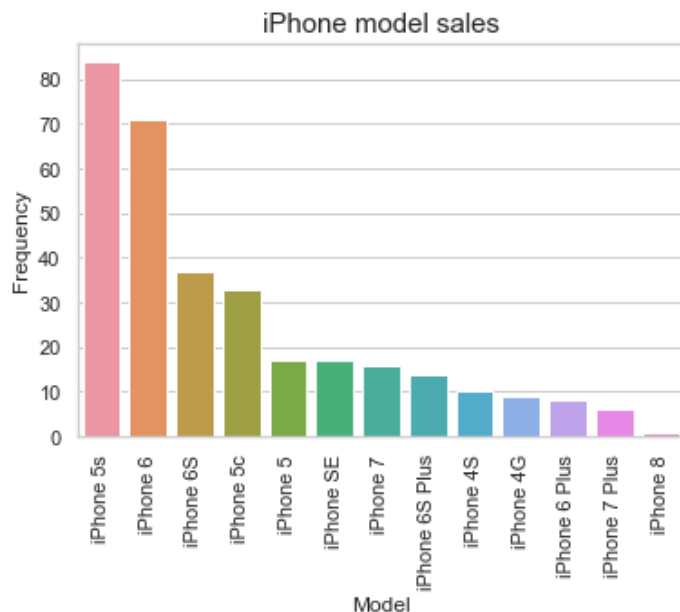


Figura 36: Ventas de modelos iPhone

El gráfico de iPhone (Figura 36) nos presenta un escenario distinto. El piso de ventas por modelo no está tan fijo ni asegurado, depende más del modelo. Acá, arriesgaremos a decir que no es por una cuestión de desconfianza en la marca sino más bien por el precio. Probablemente los dos más vendidos hayan sido los únicos cuya oferta se cruzó con la demanda.

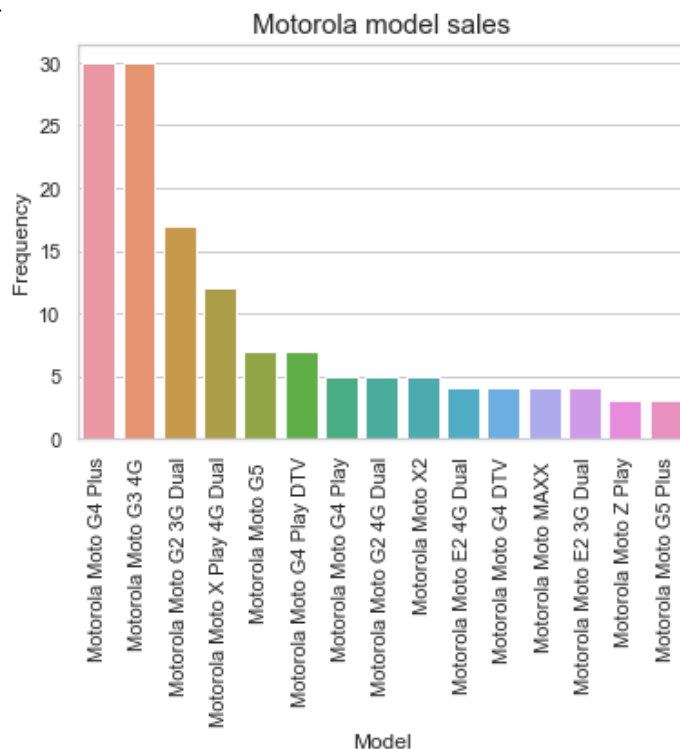


Figura 37: Ventas de modelos Motorola

Para Motorola (Figura 37), el escenario es muy similar al de Samsung pero con un piso de ventas un poco más bajo.

5.4.2. Año de lanzamiento

Dado que dentro de las tres principales marcas observamos que un modelo o dos se llevaban la gran parte de las ventas, decidimos explorar una característica de los mismos: el año de lanzamiento; no necesariamente de fabricación sino de circulación; desde cuando se escucha ese modelo en el mercado.

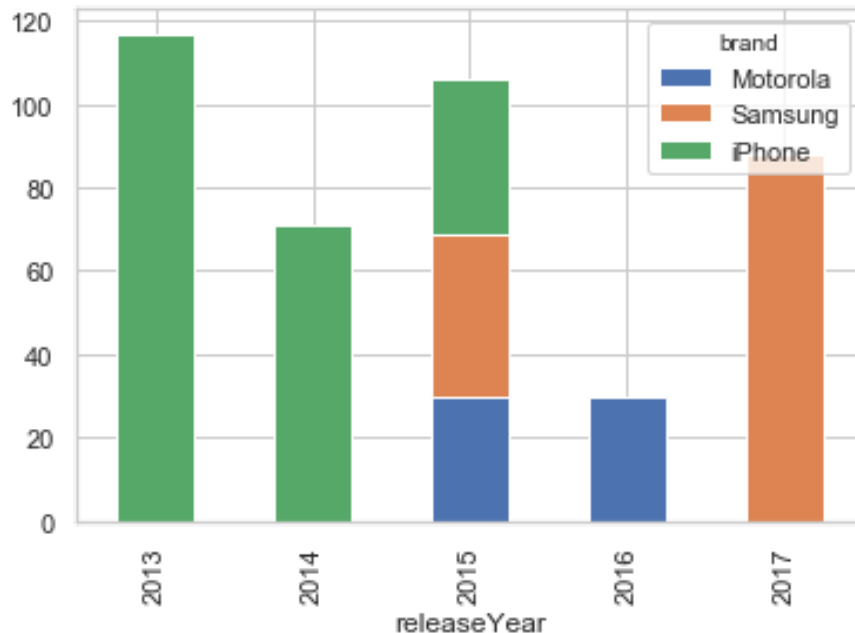


Figura 38: Año de lanzamiento de los modelos

No sorprende que, dentro de los modelos más vendidos, no haya ninguno de muchos años de antigüedad. Tampoco sorprende que la marca que genera más confianza, a la hora de comprar productos de 5 años de antigüedad en el mercado, es iPhone. Mientras que para Samsung y iPhone el grosor de las ventas está en modelos más recientes.

5.4.3. Pauta de ventas

La marca pesa. Cada una representa un contrato de confianza distinto en el imaginario colectivo. Podríamos decir que el contrato con marcas como Samsung y Motorola establece que se confía ciegamente en la calidad de sus productos de los últimos 3 años. Para productos anteriores, tal vez algo de confianza hay pero no valen la pena arriesgar. En cambio para iPhone, el contrato reconoce modelos de hasta 5 años de antigüedad con total aceptación.

6. Conclusiones finales

Para terminar con nuestro análisis, a continuación se presentarán algunas conclusiones generales sobre las secciones anteriores.

- A nivel global, la popularidad de Trocafone se centra especialmente en América, especialmente en países como Brasil, Estados Unidos, Argentina y Canadá. En otros continentes solo se registraron la minoría de los registros.
- A nivel regional, Brasil es el país con mayor interacción con el sitio. Dentro de este territorio, la zona sur, sudeste y noroeste es la que más eventos genera en el sitio. Por otra parte, en Estados Unidos, los estados con mayor interacción son California, Nueva York y Michigan.
- Hablando un poco sobre la característica de negocio de la empresa, parece ser que la estrategia de *Re-Commerce* parece tener éxito, ya que pudimos ver que la mayor cantidad de ventas de los teléfonos más vendidos son de dispositivos usados y que no están en las mejores condiciones.
- En líneas generales, los usuarios de la página buscan muy frecuentemente dispositivos de las marcas Apple, Samsung o Motorola.
- A lo largo de los meses, la cantidad de usuarios que visitaron el sitio se fue incrementando, tomando valores pico entre los meses Mayo y Junio. En cuanto a las conversiones, el efecto es casi similar, pero de todas formas son de muy baja proporción comparado con la cantidad de visitas.
- La tasa de rebote está dentro del rango aceptable, por lo que no debería ser un motivo de preocupación para la empresa.
- Es importante que la empresa contemple que el tiempo de sesión promedio de un usuario no supera los 8 minutos y lo tenga en cuenta sabiendo que es el tiempo que en general dispone para captar la atención del usuario.
- Los usuarios ingresan, en mayor proporción, desde dispositivos con sistemas operativos Android y Windows.
- Durante los meses, los días donde se efectuaron mayor cantidad de ventas fueron días de semana, y hay muy pocas ventas los días Sábado y Domingo.
- Las campañas realizadas parecen ser efectivas en cuanto a la cantidad de visitantes que generan. Se pudo ver que a medida que pasa el tiempo, la cantidad de visitas por mes generadas por el evento *ad campaign hit* aumentan. Sin embargo, si bien las ventas también aumentan, siguen siendo bajas en proporción con las visitas. Se pudo ver que las ventas entre Marzo y Mayo, si bien son mas altas, aproximadamente se mantuvieron mes a mes, mientras que las visitas aumentaron abruptamente.
- También se presentó un análisis de los tipos de accesos al sitio por país. Se puede ver que en algunos casos, como Brasil o Canadá, los ingresos por dispositivos móviles son mayores que por computadora, y en otros casos, como Argentina o Estados Unidos, la situación es al revés. Este análisis es útil a la hora del desarrollo del sitio especializado por país.
- Los usuarios de Android, Windows y iOS visitan, con gran frecuencia, teléfonos de la marca Apple. Entre los mas visitados, están los teléfonos iPhone en sus versiones 5s, 6, 6s y 7. Pero luego se ve una tendencia a comprar teléfonos de menor gama, aunque hayan visitados productos de mayor gama.