



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE INGENIERÍA

75.06 Organización de Datos

Trabajo Práctico N°1

Análisis exploratorio de datos de Trocafone

Segundo Cuatrimestre de 2018

Grupo 26: El Big Datajuste

Repositorio: www.github.com/facundocenizo/datos_tp1

Alumno	Padrón	Mail
Cenizo, Facundo	96657	facundocenizo@hotmail.com
Cruz, Pablo	97553	cruzpa95@gmail.com
Villalba, Braian	97641	braian.e.a.villalba@gmail.com

Índice

1. INTRODUCCIÓN.....	1
1.1 Acerca de Trocafone	1
1.2 Objetivo.....	1
2. ANÁLISIS DEL SET DE DATOS	2
3. ANÁLISIS DE LOS USUARIOS.....	4
3.1 Geográfico	4
3.2 Temporal	5
• Actividad por día y por mes.....	6
• Actividad durante la semana.....	7
• Actividad por hora durante la semana.....	7
3.3 Dispositivos utilizados para ingresar al sitio	8
4. ANÁLISIS DE LAS FORMAS DE LLEGADA AL SITIO.....	10
4.1 Canales de ingreso	10
4.2 Ingresos a través de campañas publicitarias	11
5. ANÁLISIS DE LAS TENDENCIAS	12
6. ANÁLISIS DE VENTAS (CONVERSIONES)	13
6.1 Temporal	13
• Por día y por mes.....	13
6.2 Por características del equipo.....	14
• Por marca	14
• Por modelo	15
• Por color	16
• Por capacidad de almacenamiento	17
• Por estado	18
6.3 Ventas vs leads.....	19
7. CONCLUSIONES.....	20
8. FUENTES DE CONSULTA	21

1. INTRODUCCIÓN

1.1 Acerca de Trocafone

Es una empresa brasileña de tecnología creada en 2014 con operaciones en Brasil y Argentina. Por medio de la adquisición de bienes tecnológicos usados, y de un minucioso proceso de reacondicionamiento, les dan vida nueva a celulares y tablets que de otra manera se convertirían en basura electrónica. Comercializan más de 35 mil dispositivos por mes a través de su principal plataforma www.trocafone.com y en más de 2 mil tiendas físicas.



Más de 500 mil equipos
reacondicionados



Más de 2400 puntos de
recepción de equipos



Más de 200 empleados



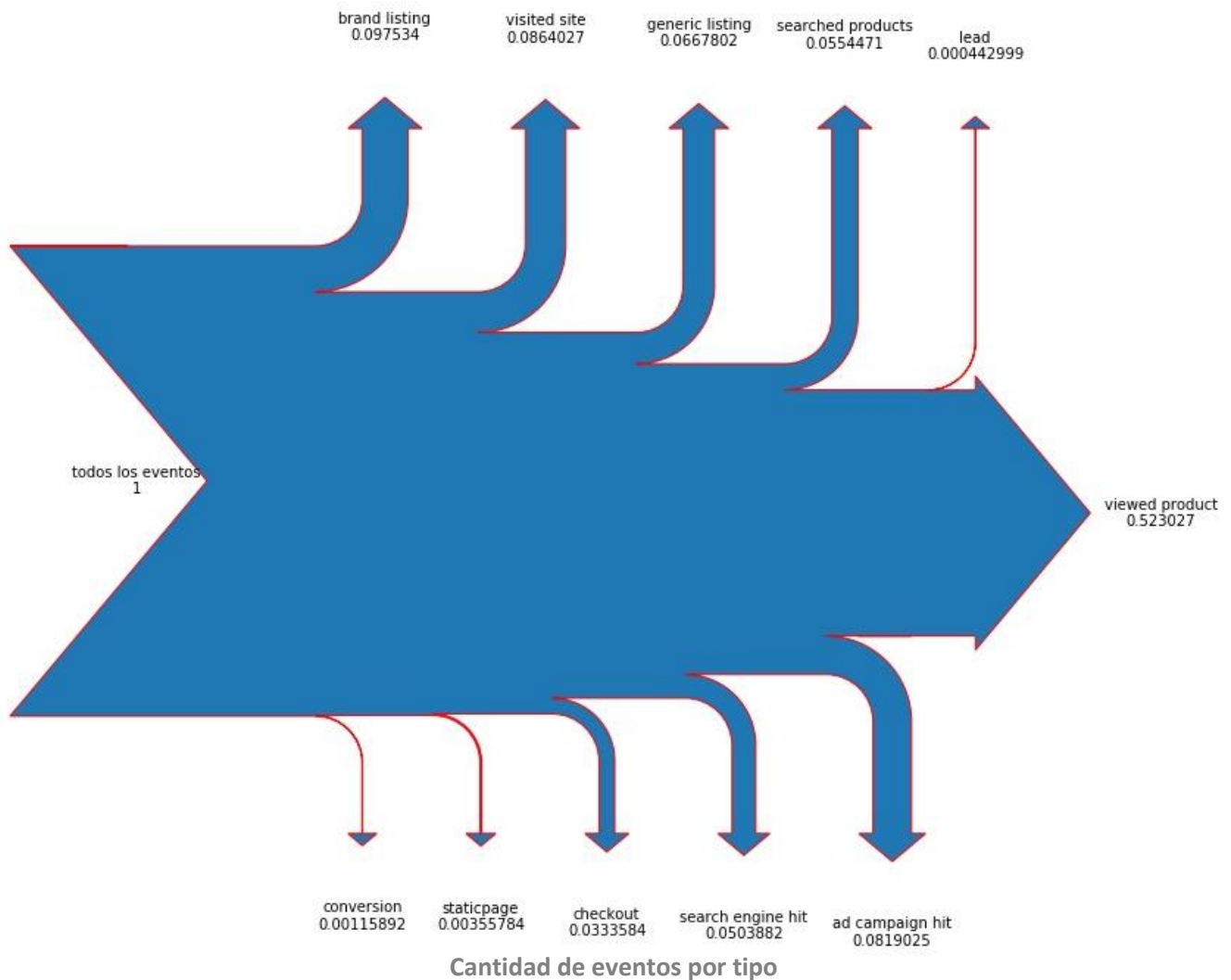
65 toneladas de
desperdicio electrónico
reducido

1.2 Objetivo

El trabajo práctico tiene como objetivo extraer la información que se considere más significativa para la empresa Trocafone a partir de un conjunto de datos recolectados a través de su plataforma de ecommerce de Brasil y realizar un análisis en profundidad que pueda resultarle útil a la misma.

2. ANÁLISIS DEL SET DE DATOS

A grandes rasgos se puede visualizar que los datos totales recibidos están organizados en primer lugar, en diferentes tipos de eventos.



A su vez cada evento está descripto por algunas de las siguientes columnas: Fecha y hora, tipo de evento, ID de usuario, URL (del sitio) visitada, producto relacionado, modelo, condición, cantidad de almacenamiento y color del dispositivo, motor de búsqueda del que proviene, tipo de dispositivo y sistema operativo utilizado por el usuario, entre otras.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1011288 entries, 0 to 1011287
Data columns (total 23 columns):
timestamp                1011288 non-null datetime64[ns]
event                    1011288 non-null object
person                   1011288 non-null object
url                      82756 non-null object
sku                      563838 non-null object
model                    564284 non-null object
condition                563836 non-null object
storage                  563836 non-null object
color                    563836 non-null object
skus                     221699 non-null object
search_term              48967 non-null object
staticpage               3598 non-null object
campaign_source          82796 non-null object
search_engine            50957 non-null object
channel                  87378 non-null object
new_vs_returning         87378 non-null object
city                     87378 non-null object
region                   87378 non-null object
country                  87378 non-null object
device_type              87378 non-null object
screen_resolution        87378 non-null object
operating_system_version 87378 non-null object
browser_version          87378 non-null object
dtypes: datetime64[ns](1), object(22)
memory usage: 177.5+ MB
```

Haciendo un pantallazo general de los datos con los que se cuenta, se observó que por ejemplo la columna “staticpage” contenía pocos datos no nulos en relación al total de los datos con los que se contaba, por tal motivo se decidió no realizar ningún tipo de análisis en ese sentido. Tampoco se creyó interesante detenerse en analizar las columnas “screen_resolution” y “browser_version”, pero sí la de “operating_system_version” para saber cuáles eran los sistemas operativos desde los que más se ingresaba.

3. ANÁLISIS DE LOS USUARIOS

Para llevar a cabo este análisis se realizó un recorte de los datos de tal manera que solamente se muestre un evento por usuario. A partir de allí se pudo observar que la cantidad total de usuarios que ingresaron a la web de Trocafone en el período en que fueron recolectados los datos fue de 27.624.

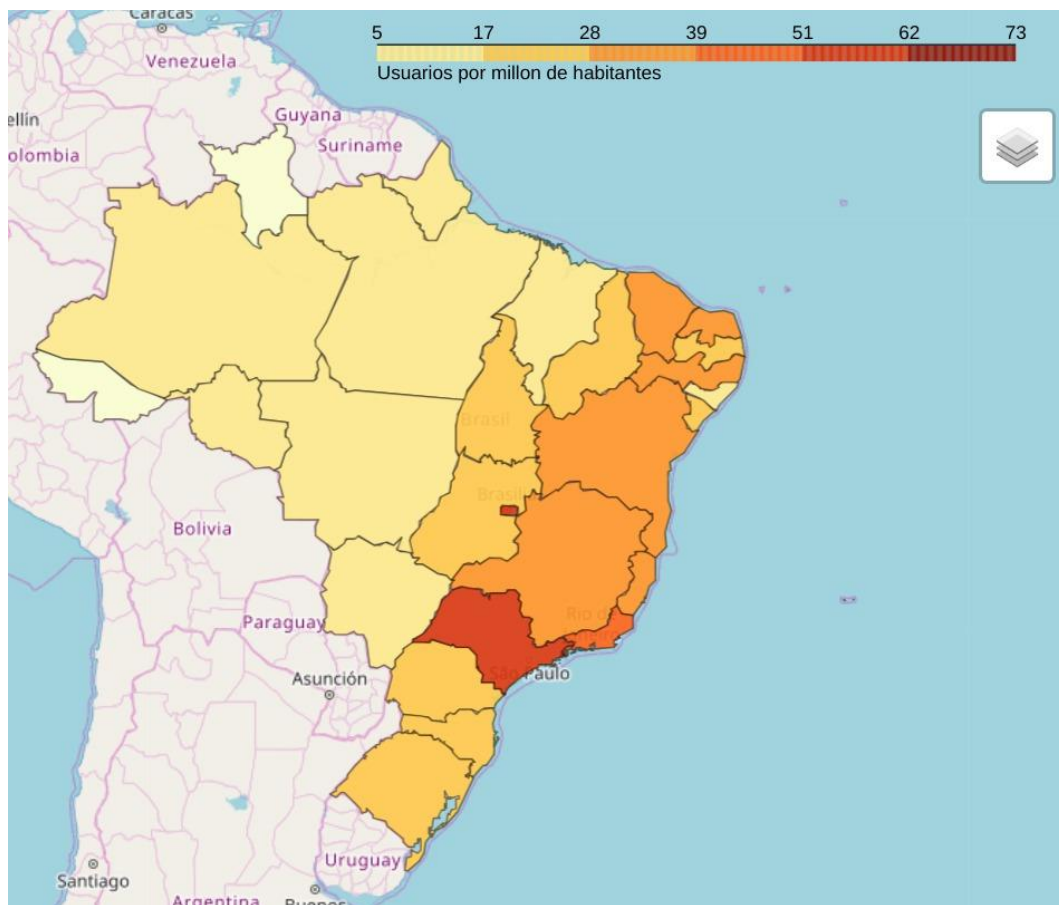
3.1 Geográfico

Ordenando los usuarios según su ubicación se observó que la cantidad proveniente de Brasil es significativamente mayor a la de los otros países.



Por esta razón se decidió analizar cómo se distribuyen los mismos dentro del territorio brasileño. Para poder realizar una comparación más representativa de los datos, se calculó la cantidad de usuarios cada un millón de habitantes por estado, ya que tomando únicamente la cantidad de usuarios por estado quedaban, obviamente en primer lugar, aquellos con mayor población. A continuación, se muestra el mapa coroplético resultante:

Distribución geográfica de los usuarios de Brasil



De aquí se puede concluir que Trocafone tiende a tener más usuarios (por millón de habitantes) en los lugares más poblados de Brasil, por ejemplo, en los estados de Sao Paulo, Rio de Janeiro, Bahía y Minas Gerais. Al mismo tiempo ocurre también que el segundo estado con más usuarios por millón de habitantes es el Distrito Federal a pesar de tener una población mucho menor, lo cual puede deberse a que es una ciudad desarrollada tecnológicamente ya que es la capital brasileña.

3.2 Temporal

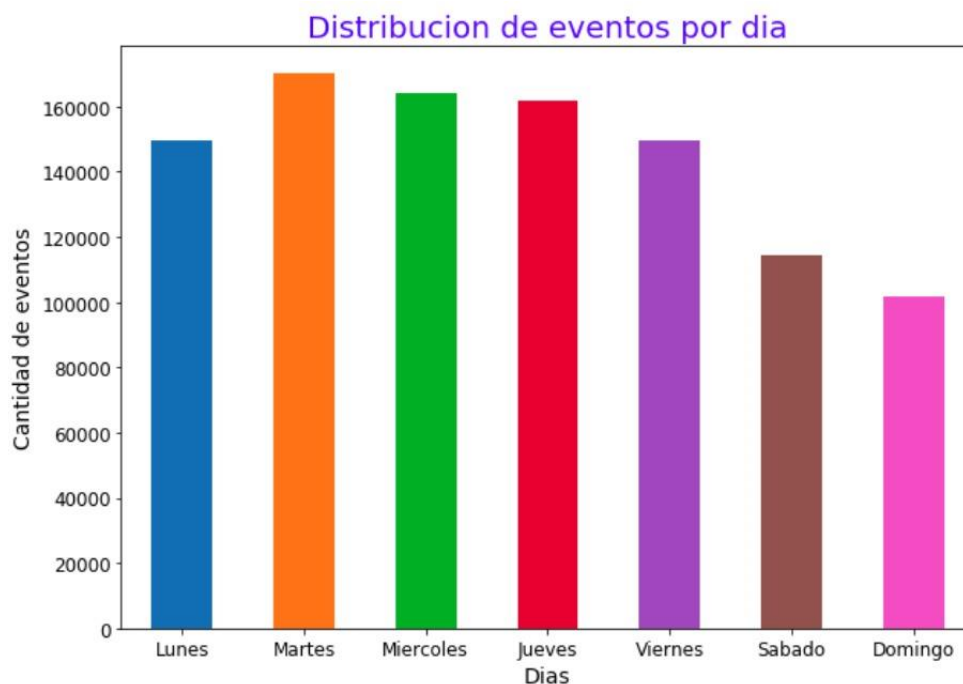
El set de datos contiene eventos que fueron realizados durante los primeros 6 meses del año 2018 (desde enero hasta los primeros 15 días de junio). Dentro de este período se observó la actividad en el sitio obteniendo los siguientes resultados:

- Actividad por día y por mes



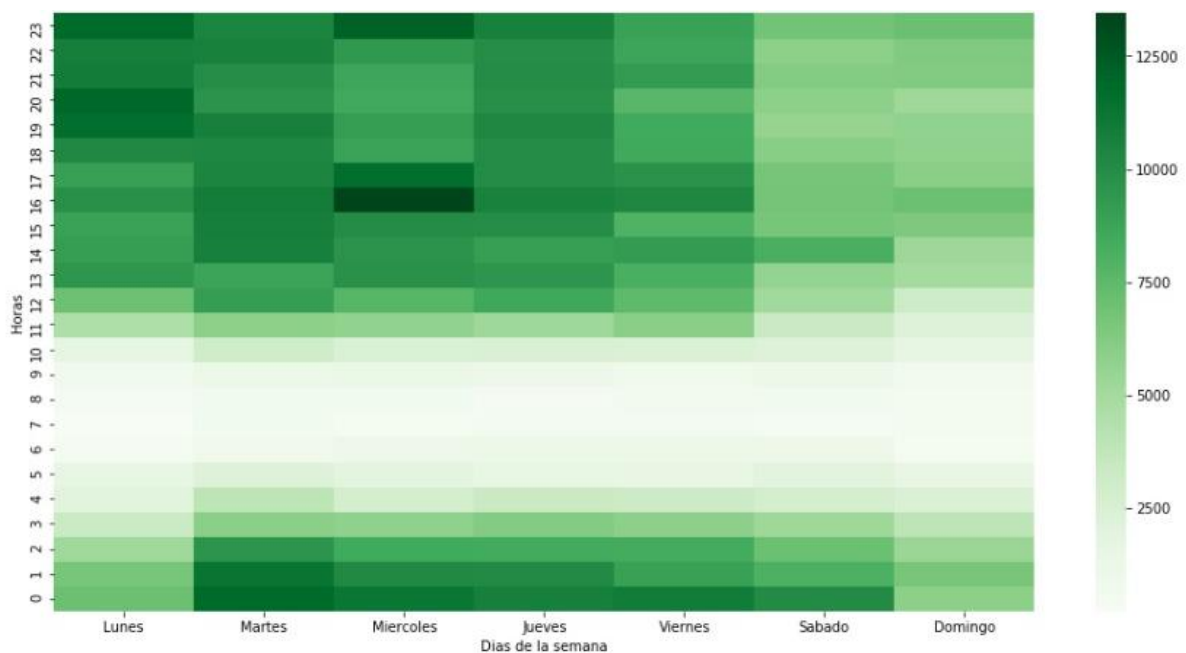
Se observa claramente en este diagrama que la actividad en la página web aumenta progresivamente luego de enero y febrero que son los meses de receso y que se da un salto importante cerca de mitad de mayo que se mantiene durante junio, sospechamos que el mismo puede atribuirse a una gran campaña publicitaria acompañada también de acontecimientos importantes que ocurren en este período los cuales se analizarán más adelante.

- Actividad durante la semana



Se ve en los diagramas que la mayor actividad se registra los días de semana, contrario a lo que puede imaginarse ya que los fines de semana se dispone de mayor tiempo libre para destinar a la navegación en este tipo de sitios.

- Actividad por hora durante la semana

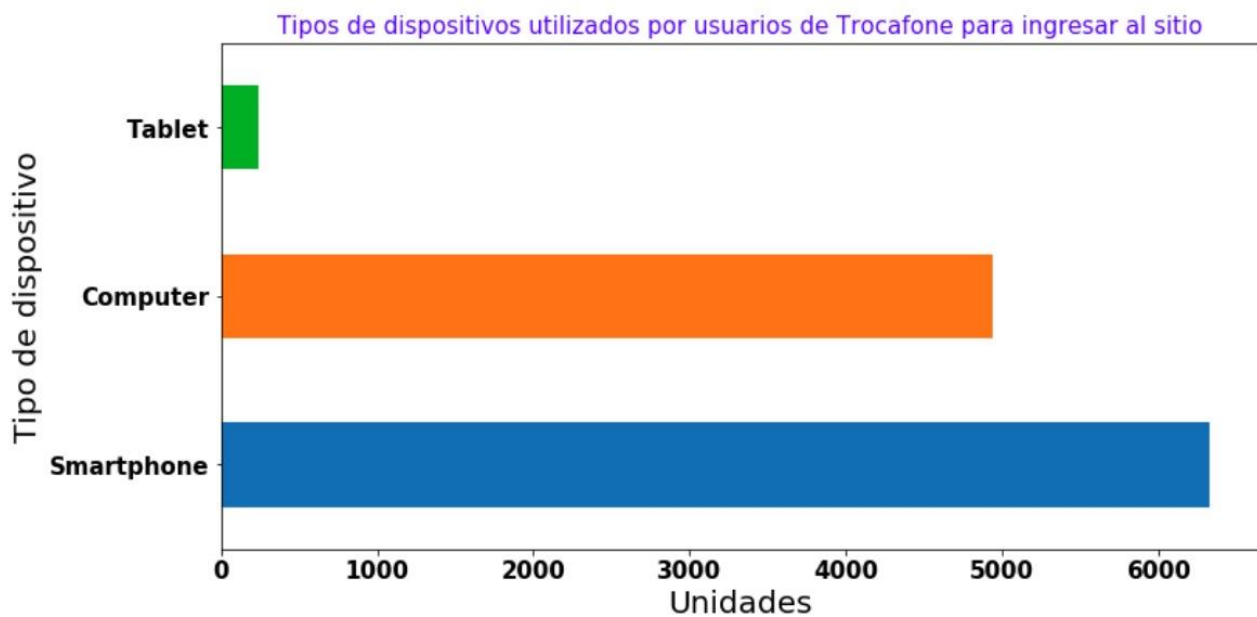


Se observa, como es de esperar, una disminución casi total de la actividad en las horas de la madrugada y un aumento de la misma hacia el mediodía. Llama la atención la gran cantidad de tráfico durante horas cercanas a la medianoche.

De este gráfico se puede concluir, por ejemplo, que el momento más efectivo para el lanzamiento de una campaña publicitaria o una oferta es el día miércoles a las 16 horas, que es el momento de mayor concentración de usuarios activos, por esta razón se lograría un mayor alcance, optimizando de esta manera la inversión.

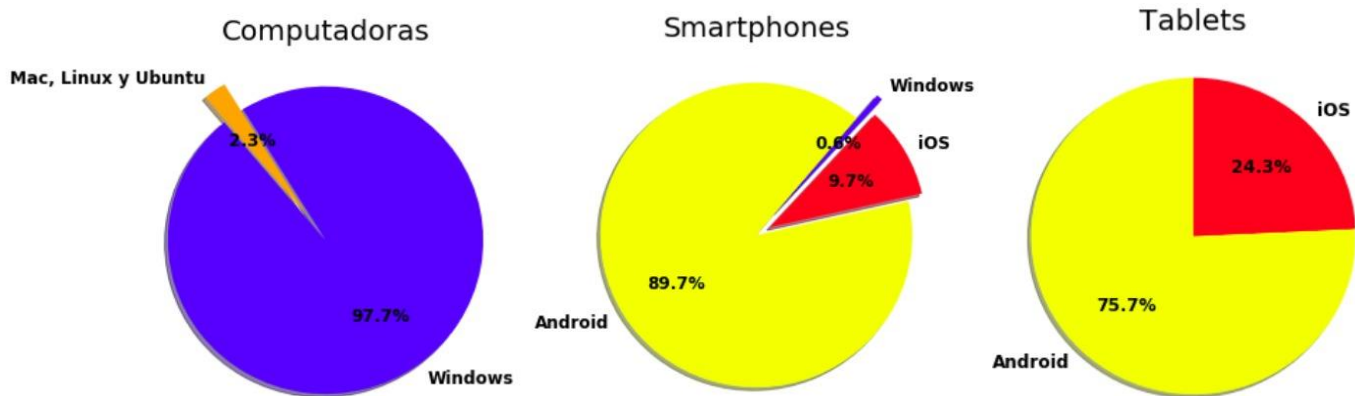
3.3 Dispositivos utilizados para ingresar al sitio

En cuanto a los tipos de dispositivos elegidos por los usuarios para ingresar al sitio se distinguen smartphones, computadoras y en mucho menor medida tablets.



Se puede observar que la cantidad de usuarios que ingresan a la web a través de smartphones es muy similar a la que ingresa por computadora. De aquí podemos resaltar la importancia de tener un sitio que esté debidamente adaptado a dispositivos móviles.

Para cada tipo de dispositivo diferenciamos cantidad de usuarios por sistema operativo. A continuación, se muestran los gráficos resultantes:



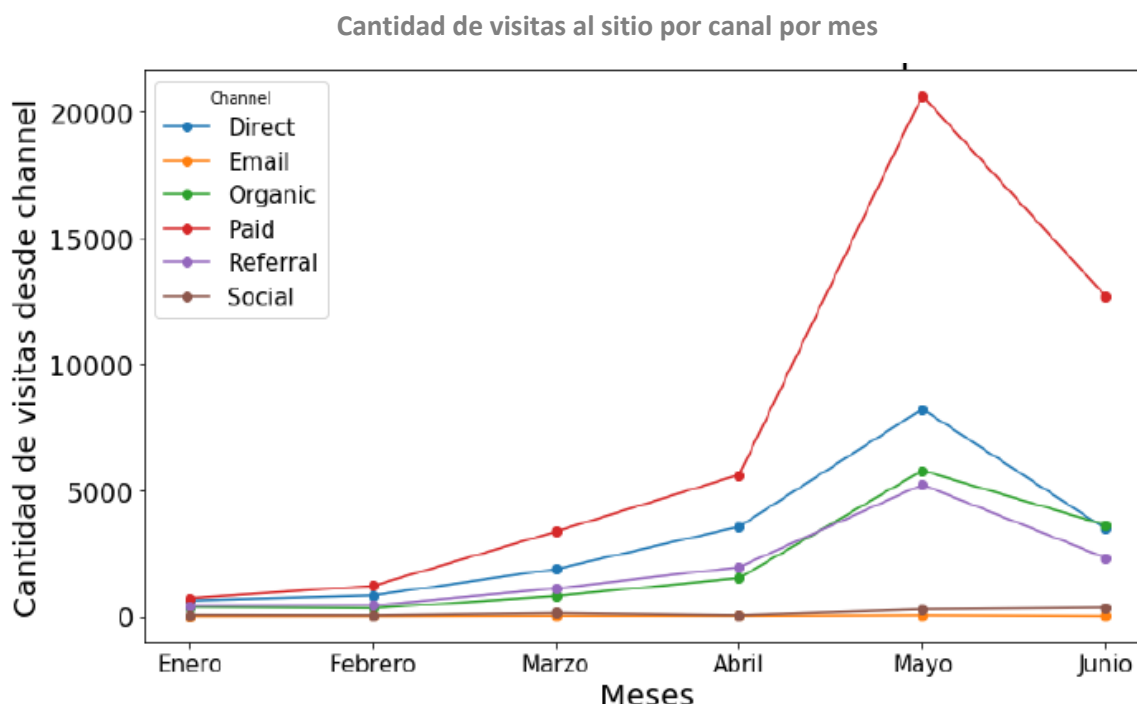
Analizando los pie-charts para los dos dispositivos móviles que utilizan los usuarios para ingresar al sitio, vemos que la mayoría lo hace desde el sistema operativo Android. Esto lleva a pensar que si en algún momento se quisiera lanzar una aplicación móvil de Trocafone lo más conveniente sería comenzar por una que fuera compatible con dicho sistema operativo.

En cuanto a los usuarios que ingresan desde una computadora se puede ver que la gran mayoría utiliza Windows, como era de esperar.

4. ANÁLISIS DE LAS FORMAS DE LLEGADA AL SITIO

A continuación, se hará un análisis sobre las distintas formas que existen para ingresar al sitio web de Trocafone.

4.1 Canales de ingreso



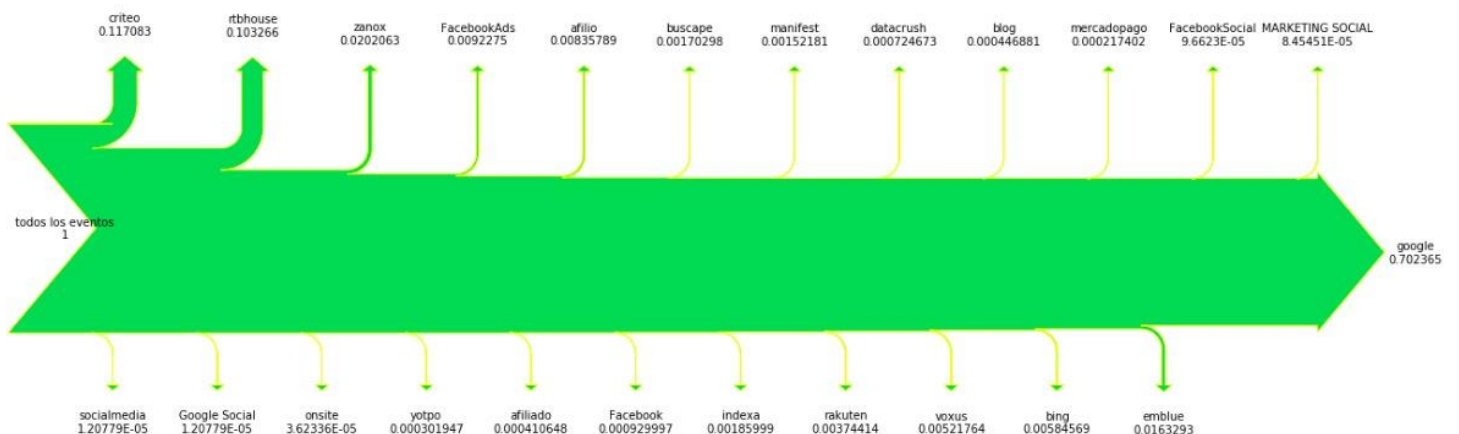
Aquí se analizaron los distintos canales desde donde los usuarios conocen al sitio. Lo que se observa es una predominancia absoluta de visitas generadas a través de un canal pago en todos los meses analizados, seguida por las visitas directas. La cantidad de ingresos a la web por mes es acorde al análisis sobre la actividad que se realizó anteriormente, pero lo que se puede resaltar de este esquema es que el incremento de visitas y de ventas que ocurrió durante el mes de mayo estuvo acompañado por una fuerte inversión en publicidad que consiguió un pico de ingresos al sitio generados por un canal pago.

Lo que se puede observar también es que el Email y el medio Social son prácticamente nulos como canales de ingreso y no aportan un número significativo de visitas. Podría sugerirse al equipo de marketing reforzar las campañas en esos sectores para aprovechar el alcance real que pueden tener.

4.2 Ingresos a través de campañas publicitarias

En esta sección se analizó las llegadas al sitio a partir de las campañas de marketing.

Ingresos al sitio desde campañas publicitarias



El gráfico muestra que, de todos los accesos al site que se realizaron a través de una campaña de marketing, el 70% pertenece a campañas de Google, lo cual tiene sentido ya que éste es el monstruo de la internet y con su servicio Google Ads tiene llegada a una extensa cantidad de sitios que no solamente son el buscador propio. Está claro que el sector de marketing debería centrar sus operaciones sobre la plataforma de Google ya que es la más efectiva.

5. ANÁLISIS DE LAS TENDENCIAS

Para este análisis se realizó un filtro por la columna “search_term” teniendo en cuenta los términos relacionados con modelos de celulares, despreciando los errores que puedan cometer los usuarios al buscar. Con estos datos se puede hacer una visualización de los modelos que más interesan al público resultando en el siguiente wordcloud:



Los resultados obtenidos reflejan las tendencias que se observan a nivel global en cuanto a los deseos de los compradores. La marca líder claramente es Apple con sus distintos modelos de su teléfono iPhone, llevándose la mayoría de las búsquedas de la web de Trocafone. Lo que puede resultar llamativo es que la marca Motorola predomina por encima de Samsung que es la otra marca líder del mercado. Más adelante, en la sección de ventas, se analizará si estas tendencias finalmente se traducen en conversiones.

6. ANÁLISIS DE VENTAS (CONVERSIONES)

Se realizó aquí un análisis del número total de equipos vendidos a través de la página web, teniendo en cuenta diferentes variables.

6.1 Temporal

- Por día y por mes

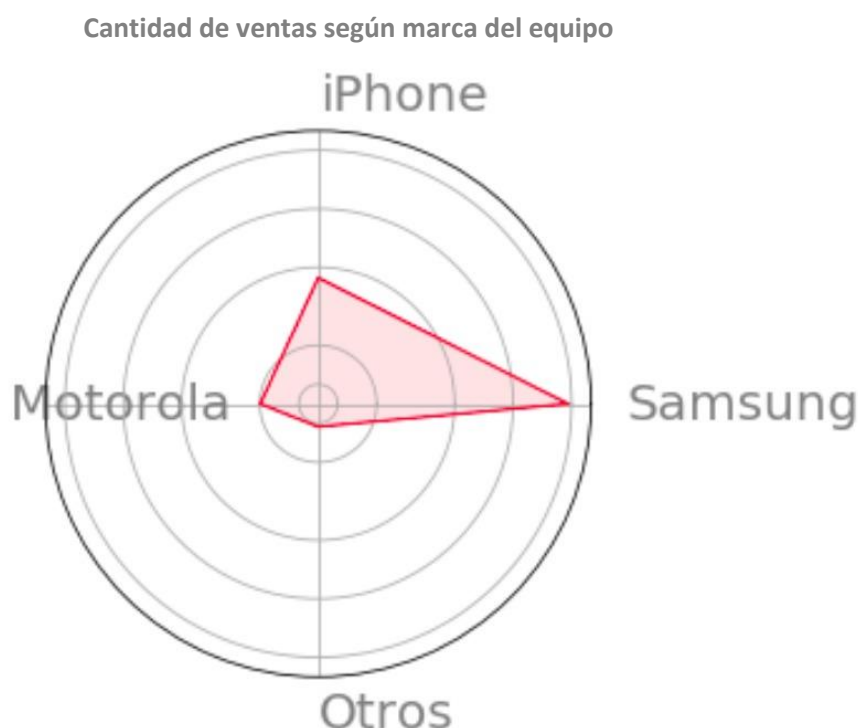


Se ve aquí que durante el mes de mayo hubo un gran aumento en las ventas que, como se anticipó anteriormente, puede deberse a la celebración del día de la madre que en Brasil es el día 12 de dicho mes. Se puede observar también que, en contraste con el gráfico presentado anteriormente de actividad vs tiempo, durante los primeros meses ambos eventos aumentan progresivamente alcanzando un tope antes de mediados de mayo. Lo que llama la atención es que mientras que la actividad durante el mes de junio se mantiene en niveles altos, las ventas disminuyen considerablemente.

6.2 Por características del equipo

- Por marca

Para este análisis se consideró interesante presentar un gráfico de radar. En este tipo de esquemas se define según escala, un valor a cada radio del radar y se divide la circunferencia en tantas partes como variables se quieren mostrar. Luego se le asigna el valor que alcanza cada variable y se genera el gráfico, en el cual se podrá ver a simple vista cuáles son las marcas preferidas por los compradores y su relación proporcional con las demás.

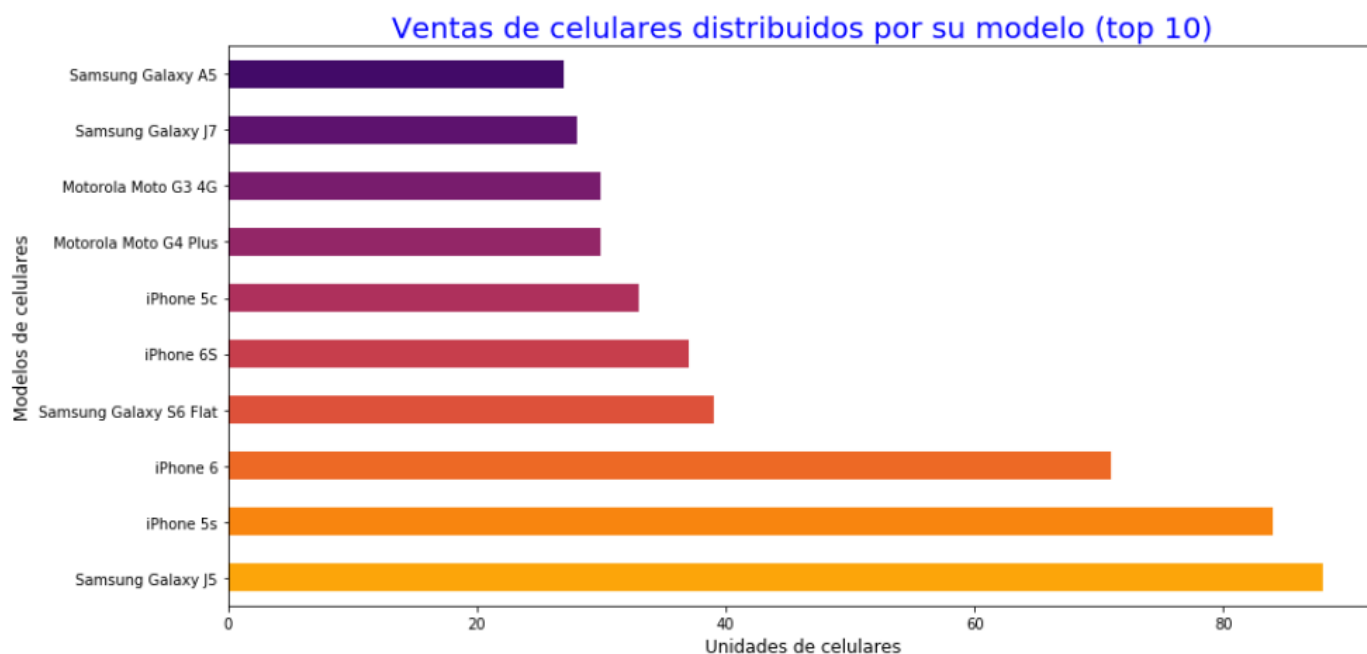


Acorde al mercado en general, en esta página los dos gigantes de la industria de la telefonía, Apple y Samsung, son las marcas que tienen la mayor cantidad de ventas. Motorola se consolida como la tercera opción superando en número de conversiones al resto de las compañías (LG, Sony, Lenovo, Asus) que se ubican muy por debajo.

Se observa que, a pesar de ser una página de celulares usados, cuando se trata de marcas, siguen predominando aquellas que son líderes a nivel global.

A su vez se puede analizar con respecto a lo expuesto anteriormente acerca de las tendencias del sitio, que ninguna de las marcas más buscadas encabeza el número de ventas. Este dato llamativo podría explicarse, por ejemplo, por una falta de stock de los modelos que más interesan al público o también porque a pesar de que la primera inclinación de los compradores sea hacia los equipos “de moda”, al momento de comprar entra en consideración otro factor determinante que es el precio de los mismos.

- Por modelo



Siguiendo con lo anterior, los modelos de celulares más vendidos son claramente dominados por Apple y Samsung a tal punto que dentro los teléfonos con más conversiones, corresponden a estas dos marcas 8 de los 10 equipos más elegidos.

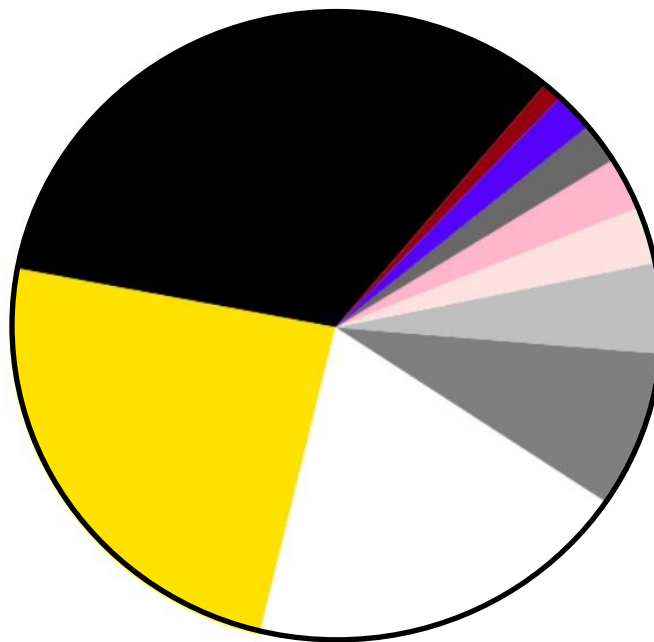
Se puede observar también que ninguno de los celulares más vendidos corresponde a equipos de alta gama sino más bien se corresponden con una categoría estándar en la actualidad.

Llama la atención también, en comparación con el gráfico de tendencias, que el Samsung Galaxy J5 siendo el modelo más vendido no figure dentro de los equipos más buscados. Esto puede deberse a que a pesar de no ser un celular “de moda” como lo son los iPhone, es un equipo de buen desempeño con un precio más accesible. Podría pensarse también que este gran número de ventas fue resultado de una buena oferta para este equipo en particular.

- Por color

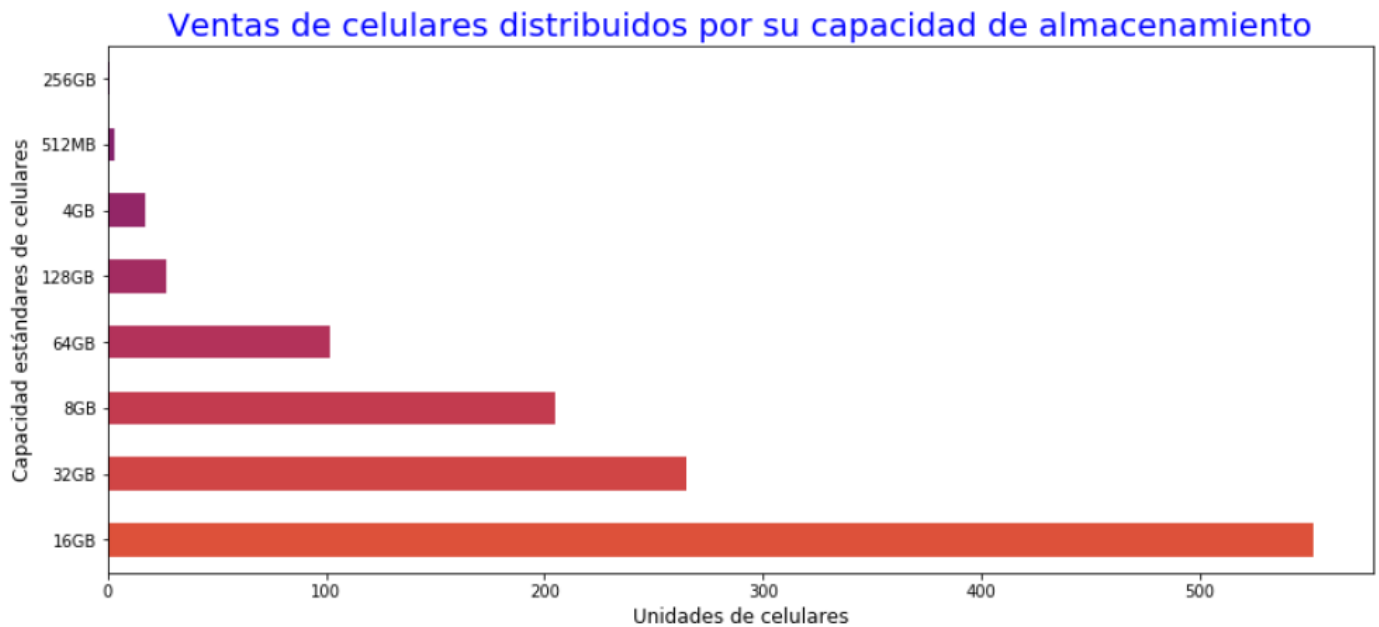
En este caso se generó un pie chart en el cual el color de cada porción representa el color real de los equipos vendidos.

Cantidad de ventas según color del equipo



Como era de esperarse, los equipos de color negro dominan las elecciones de los compradores, pero se ve que los usuarios escogen en segundo lugar el color dorado, incluso antes que el blanco, lo cual nos resulta curioso. Esto puede deberse principalmente a una moda o tendencia del momento y que también es un color muy elegido por las mujeres.

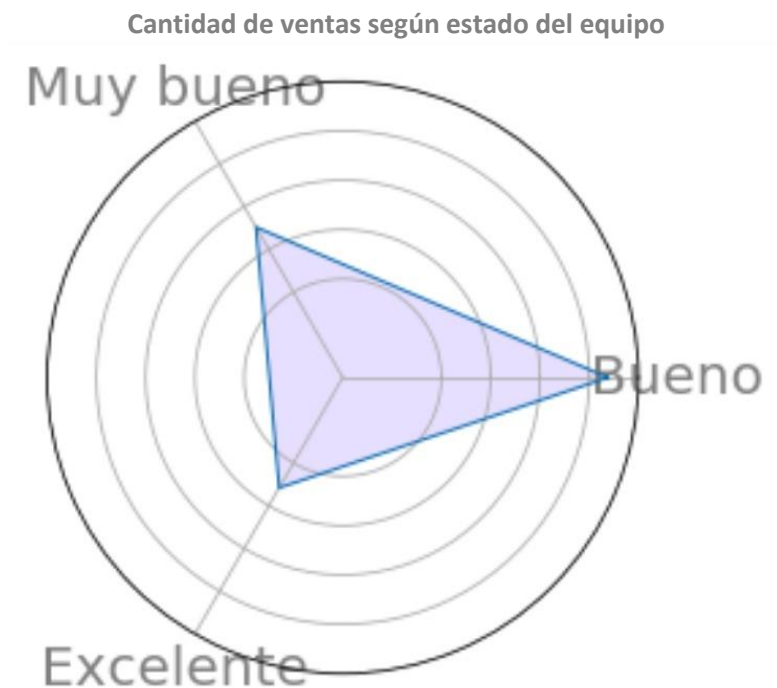
- Por capacidad de almacenamiento



Se puede decir que los clientes de Trocafone no consideran la capacidad de almacenamiento como una característica determinante al modelo de elegir un teléfono. Se observa que cerca del 50% de los equipos vendidos poseen 16 GB de memoria, cantidad que no es muy significativa en la actualidad.

- Por estado

Para hacer este gráfico más legible se decidió descartar los estados “Bom sem Touch ID” y “Novo” ya que las cantidades eran totalmente despreciables respecto a los demás estados.

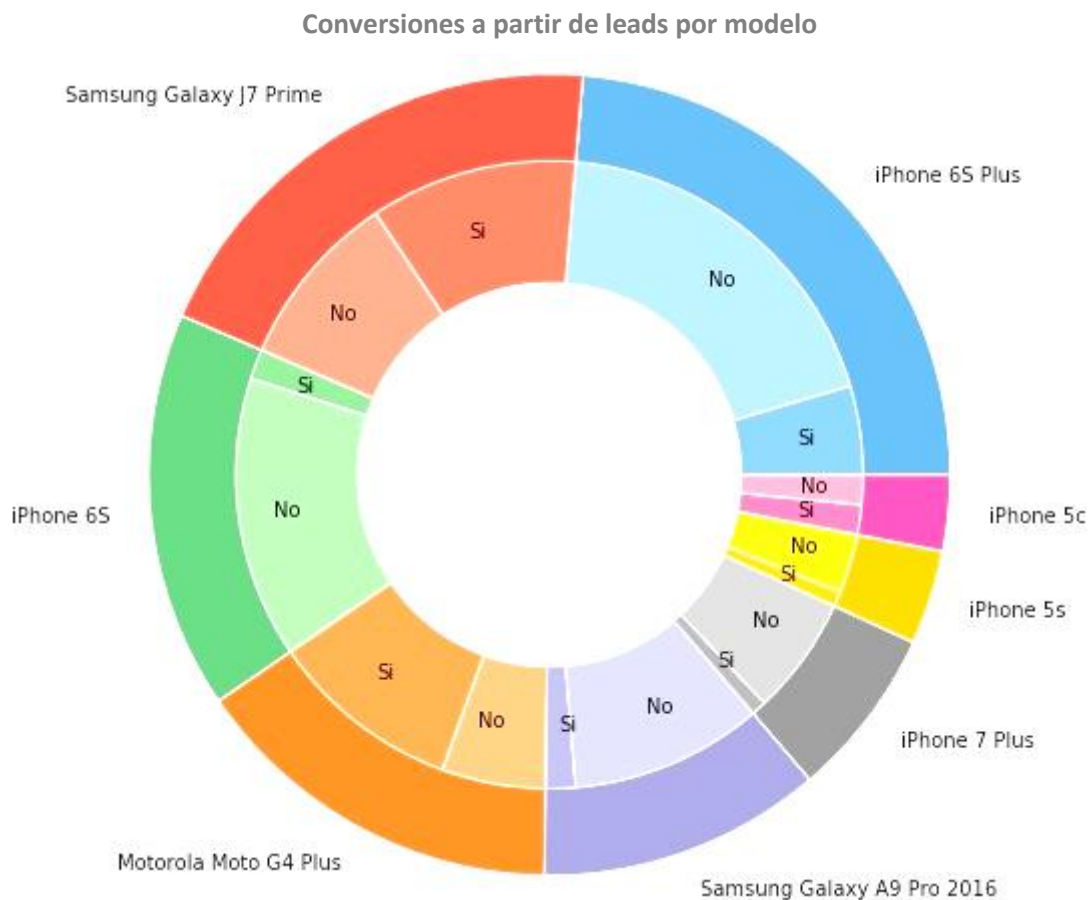


En este análisis se nota que la mayoría de los celulares elegidos por los compradores se encuentran en “buen” estado, no necesariamente los mismos requieren que estén en “muy buenas” o “excelentes” condiciones. Se puede afirmar entonces que los compradores de este site no son demasiado exigentes en cuanto al estado del equipo que van a comprar, sino que priorizan más otras variables como pueden ser el precio o la marca del teléfono.

6.3 Ventas vs leads

Se denomina “lead” al evento en el cual un usuario se registra para recibir una notificación cuando ingrese un determinado producto en stock.

Se tomaron los 8 modelos de celulares más solicitados y se comparó la cantidad de leads registrados para cada uno de ellos. Luego se realizó una distinción dentro de cada modelo según cuantas conversiones tuvo, siendo en el gráfico “Si” que se realizó una compra y “No” lo contrario.



Aquí se puede observar que una gran cantidad de los usuarios muestra interés en los celulares marca iPhone, como ya se vio anteriormente, pero después el porcentaje de ellos que los compra es mucho menor en comparación a aquellos que se interesan en las otras marcas más solicitadas. En cuanto al Samsung Galaxy A9 se produce el mismo fenómeno que con los iPhone, ambos casos pueden explicarse ya que son celulares de alta gama, potentes tecnológicamente y estéticamente atractivos, pero de un precio muy elevado. Lo contrario ocurre con los celulares Samsung J7 Prime y Motorola Moto G4 Plus que son dos modelos de los más baratos del mercado pero que no dejan de ser buenos aparatos y fieles en cuanto a su prestación.

7. CONCLUSIONES

A modo de cierre del trabajo práctico, luego de haber realizado los distintos análisis ya mostrados, se enunciaron a continuación algunas conclusiones generales:

Geográficamente, la mayor parte de los usuarios de Trocafone son habitantes de Brasil. Dentro de ese país, hay una estrecha relación entre la población de cada estado con la cantidad de usuarios por millón de habitantes de los mismos, a excepción de la ciudad capital que a pesar de no ser un estado de los más poblados sí tiene una gran concentración de usuarios.

Siguiendo con lo respectivo a los usuarios, se ve que en las épocas vacacionales (enero-febrero) se encuentran mayormente inactivos y que alcanza el pico de actividad en el mes de mayo y junio. En esto puede influir eventos como el día de la madre (12 de mayo) o quizá ofertas y campañas publicitarias que haya lanzado Trocafone en esos periodos.

Se observó que los días miércoles a las 16 horas es el momento de mayor concentración de actividad, quizá para Trocafone sería conveniente planificar ofertas para ser lanzadas cerca de ese rango horario para intentar maximizar la inversión.

Para ingresar a la plataforma web, los usuarios utilizan smartphones y computadoras prácticamente en igual medida, por lo que se resalta aquí la importancia de que el mismo se encuentre debidamente adaptado a ambos equipos electrónicos. También, analizando los dispositivos móviles, se pudo observar que la gran mayoría de los usuarios utilizan Android, de manera que, si la empresa quisiera lanzar una aplicación, sería aconsejable que comience por una que sea compatible con dicho sistema operativo.

Predomina como forma de llegada al sitio el buscador Google y los canales de acceso pagos. Está claro que el sector de marketing debería centrar sus operaciones sobre la plataforma de Google ya que es la más efectiva.

En cuanto a las preferencias de los usuarios a la hora de comprar se ve que las mismas no coinciden con las búsquedas que ellos realizan. Por ejemplo, se busca muchísimo los modelos de la marca iPhone, pero luego el más comprado es de la marca Samsung, modelo J5 que casi no aparece en las búsquedas.

Para finalizar las conclusiones, es interesante resaltar que los usuarios de Trocafone no priorizan al momento de realizar una compra los teléfonos de más alta gama sino más bien prefieren los de media, lo cual puede deberse a una cuestión de precio, pero esto no se puede afirmar porque no tenemos esos datos. En lo relacionado al estado de los teléfonos, la mayoría de los compradores nos buscan que se encuentren excelentes, es decir no son exigentes en ese sentido, por lo que a la empresa no le es necesario re acondicionar demasiado los equipos, simplemente alcanza con dejarlos en buen estado.

8. FUENTES DE CONSULTA

Trocafone – www.trocafone.com.br y www.trocafone.com.ar – Utilizada para obtener información acerca de la empresa.

StackOverflow – www.stackoverflow.com – Utilizada para despejar dudas.

Pandas Guide - <http://pandas.pydata.org> – Utilizada también para despejar dudas.

Gráficos - <https://python-graph-gallery.com/all-charts/> - Utilizada para consultar los diferentes tipos de gráficos que se pueden hacer tanto con Python, Matplotlib, Pandas, Seaborn, Folium, entre otras.