

Contents

1	Procesamiento y análisis de datos	2
1.1	Datos utilizados	2
1.2	Sobre la empresa	2
1.3	Lenguajes y librerías utilizados	2
1.4	Repositorio de Github	3
2	Análisis temporal	4
2.1	Distribución de eventos por hora	4
2.2	Tipos de eventos por hora	5
2.3	Evolución de la cantidad de eventos a lo largo del mes	7
2.4	Distribución de compras por hora del día y día del mes	9
2.5	Distribución de eventos por día de la semana	10
2.6	Distribución de compras por día de la semana	11
2.7	Evolución de la cantidad de eventos a lo largo del año	12
2.8	Evolución de la cantidad de compras a lo largo del año	13
2.9	Usuarios nuevos	15
3	Análisis de eventos	16
3.1	Cantidad de eventos por usuario	16
3.2	Eventos generados por usuarios	19
3.3	Modelos más vendidos	20
3.4	Capacidad de los equipos	24
3.5	Compras por condición	26
3.6	Cantidad de eventos por tipo de dispositivo	28
3.7	Navegadores preferidos	28
3.8	Proporción de usuarios que ingresan por un buscador	30
3.9	Buscadores más utilizados	30
4	Análisis Geográfico	31
4.1	Ciudades con más usuarios	31
4.2	Ciudades con más eventos	32
4.3	Países con más usuarios	34
4.4	Países con más eventos	35
5	Análisis del usuario	38
5.1	Cantidad de productos vistos por persona	40
6	Conclusiones generales	41

1 Procesamiento y análisis de datos

En esta sección se introduce brevemente el producto a analizar y las herramientas que se utilizaron para realizar dicho análisis.

1.1 Datos utilizados

Se estudiaron datos provistos por la empresa Trocafone, analizando un conjunto de eventos de web analytics de usuarios que visitaron www.trocafone.com, su plataforma de e-commerce de Brasil.

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 (e-commerce, marketplace y tiendas físicas).

Para conocer más de su modelo de negocio, pueden visitar el siguiente artículo:

<https://medium.com/trocafone/el-maravilloso-mundo-de-trocafone-5bdc5761856b>

1.3 Lenguajes y librerías utilizados

- Para el procesamiento de datos se utilizó Python 3, con la librería Pandas.
- Para las visualizaciones, se utilizaron las librerías Matplotlib y Seaborn, y en particular, para el análisis geográfico, la librería Folium.
- Para codificar las ciudades como coordenadas se utilizó la librería geopy
- Como editor se utilizó Jupyter Lab (o Jupyter Notebook)

1.4 Repositorio de Github

Para el trabajo en conjunto del equipo, se utilizó un repositorio en github, donde se encuentran todos los archivos necesarios del análisis y este informe propiamente dicho.

Link: <https://github.com/emabrea/7506-DATOS-TP1.git>

2 Análisis temporal

Se procede a analizar los datos en relación con el tiempo, y encontrar patrones en los eventos.

2.1 Distribución de eventos por hora

Los posibles tipos de eventos son:

- **“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.

2.2 Tipos de eventos por hora

Se presenta un gráfico de barras apiladas, diferenciando los tipos de eventos:

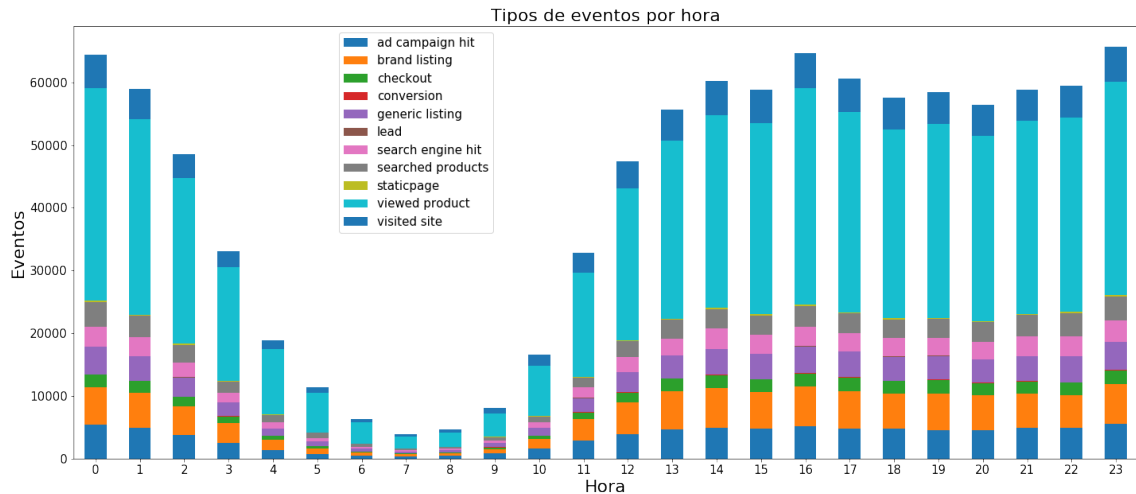


Figura 1.

Como se observa, a partir de las 01:00 hay una reducción importante en la cantidad de eventos, los cuales entre las 14:00 y las 00:00 se mantienen casi constantes y en valores altos.

La hora del día con más eventos es entre las 22:00 y las 23:00

Adicionalmente, este gráfico de barras apiladas evidencia el predominio de **"viewed product"** frente al resto de los eventos, siendo de aproximadamente un orden de magnitud mayor al evento que lo sigue en cantidad, podríamos filtrar a dicho evento para poder apreciar mejor la relación de los demás.

event	ad campaign hit	brand listing	checkout	conversion	generic listing	lead	search engine hit	searched products	staticpage	viewed product	visited site
Hour											
0	5415	5954	2035	54	4358	19	3166	3964	240	33943	5230
1	4918	5529	1924	65	3897	5	2986	3484	171	31106	4893
2	3826	4506	1549	47	3049	4	2361	2806	162	26410	3816
3	2449	3239	1048	44	2140	1	1617	1730	90	18233	2527
4	1340	1667	625	12	1210	1	900	1220	70	10375	1418
5	772	913	362	11	723	2	531	789	54	6371	807
6	439	497	228	5	423	3	298	493	16	3394	461
7	327	374	153	3	279	7	209	224	14	1948	372
8	443	391	167	6	311	10	250	256	19	2319	451
9	898	646	264	9	659	15	428	515	42	3679	922
10	1636	1536	523	10	1162	25	903	906	72	8050	1816
11	2907	3427	1035	32	2263	28	1641	1604	124	16581	3242
12	3888	5095	1544	42	3220	37	2382	2496	205	24177	4340
13	4697	6094	1944	74	3594	28	2744	2955	216	28350	4988
14	4949	6298	2096	57	4089	26	3241	3121	221	30713	5400
15	4787	5794	2018	74	3985	24	3036	3125	221	30483	5305
16	5159	6342	2072	72	4269	23	3145	3264	235	34559	5497
17	4766	6038	2140	84	4085	31	2923	3088	213	31925	5267
18	4794	5593	1973	74	3835	25	2929	2952	218	30053	5168
19	4576	5864	2083	97	3774	29	2805	3064	198	30898	5010
20	4520	5659	1908	71	3665	30	2814	3081	198	29537	4895
21	4928	5497	1890	77	3958	24	3075	3455	201	30773	4989
22	4868	5253	2009	73	4144	24	3185	3655	197	30966	5038
23	5525	6429	2145	79	4442	27	3388	3826	201	34088	5526

Tabla 1: Tabla de frecuencias de eventos por hora

En el siguiente gráfico se pueden apreciar mejor las similitudes entre algunos eventos. Mostramos un gráfico ilustrando algunos de los eventos de mayor ocurrencia durante el día (excluyendo a **viewed product**).

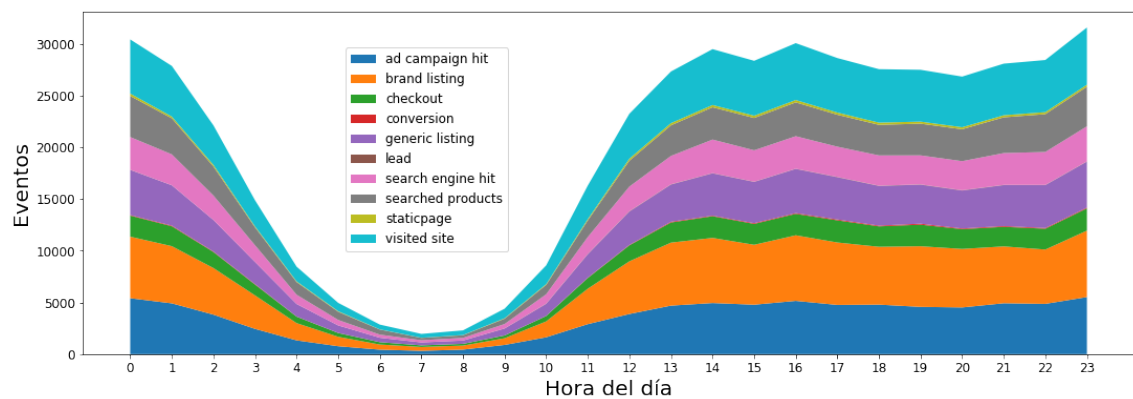


Figura 2.

En el gráfico se aprecian los diferentes eventos de búsqueda, pudiendo ver una similitud en su tendencia temporal.

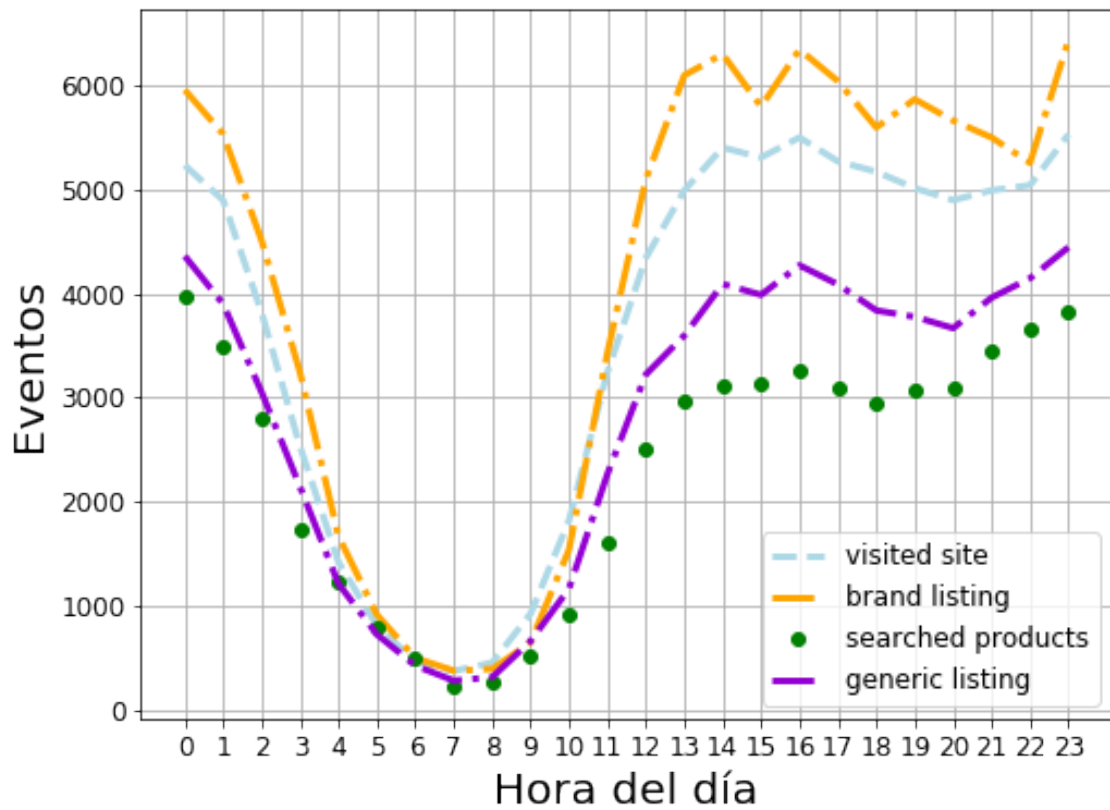


Figura 3.

2.3 Evolución de la cantidad de eventos a lo largo del mes

Se presenta un gráfico con la cantidad de eventos por día del mes:

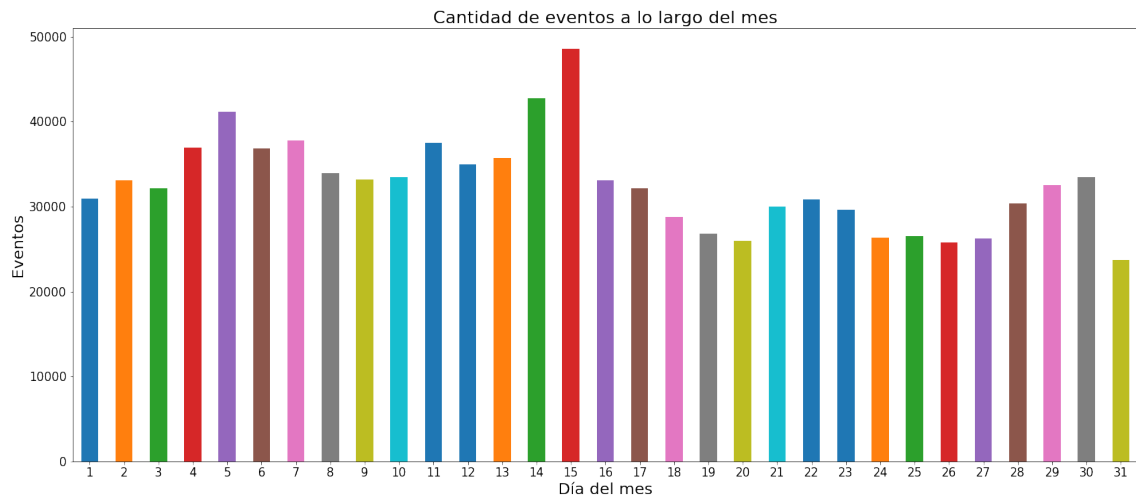


Figura 4.

Lo interesante del grafico es que el día de más eventos es el día 15 (mitad de mes), y a partir de ahí hay un gran descenso en la cantidad de eventos. Se puede pensar que la gente al cobrar a principio de mes, tiende a gastar más en la primer quincena.

2.4 Distribución de compras por hora del día y día del mes

Continuando con el análisis temporal, se realizó un Heatmap para evidenciar la distribución de eventos durante el mes.

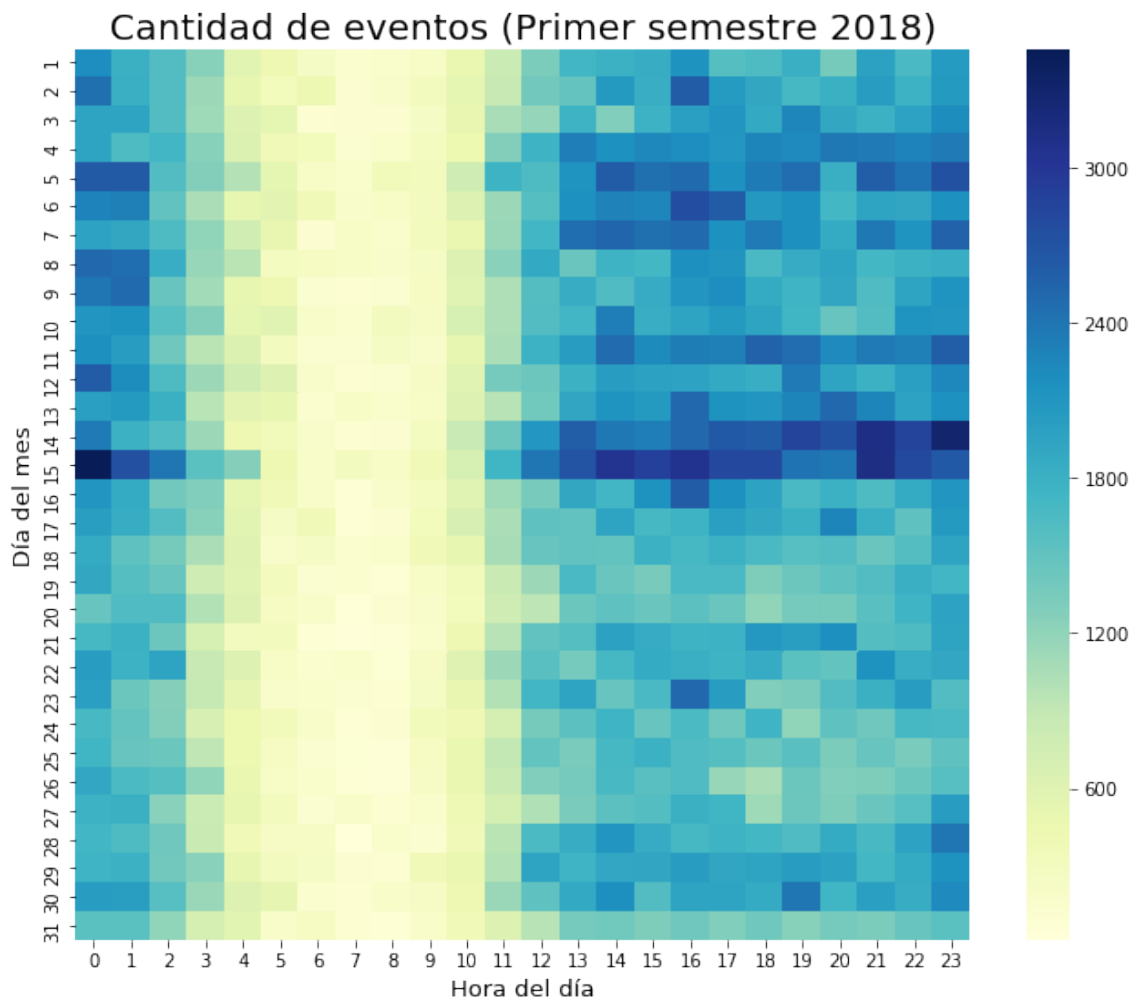


Figura 5.

Se puede observar un descenso en la cantidad de eventos en las horas de la mañana y un incremento particularmente interesante en los días 14 y 15 del mes, lo cual podría ser de utilidad a la hora de elegir la fecha para las ofertas o promociones mensuales.

2.5 Distribución de eventos por día de la semana

A continuación, se analizarán qué días de la semana tienen más eventos.

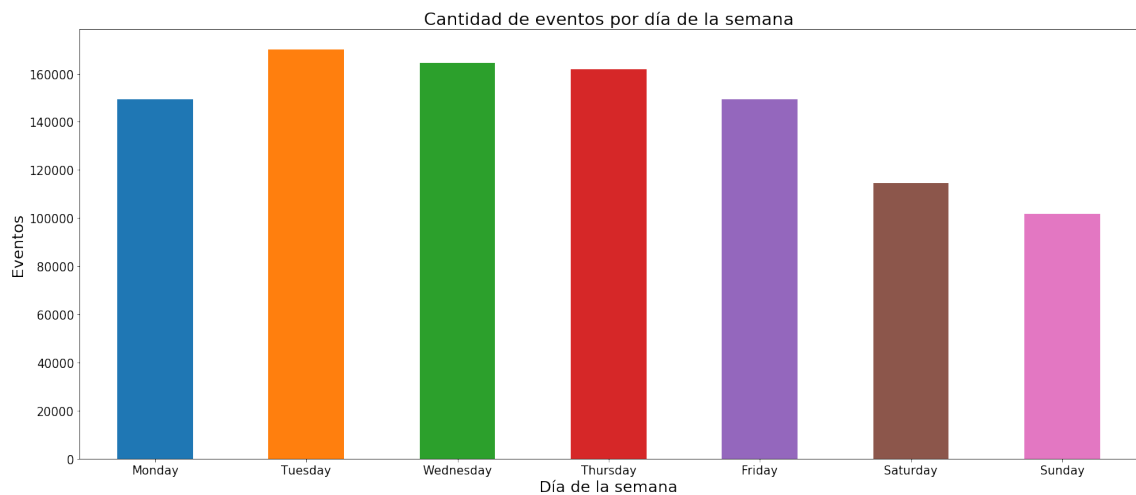


Figura 6.

Se observa que el día martes es el más ocupado, y a partir de ahí se reduce gradualmente la cantidad de eventos, para luego aumentar nuevamente el lunes.

Es de destacar que los fines de semana hay muy poca actividad en relación con los demás días.

El promedio de visitas de lunes a viernes es 158993

El promedio en los fines de semana es 101709.

2.6 Distribución de compras por día de la semana

Continuando con el análisis de la sección anterior, se filtran los eventos que resultaron en compras (conversión).

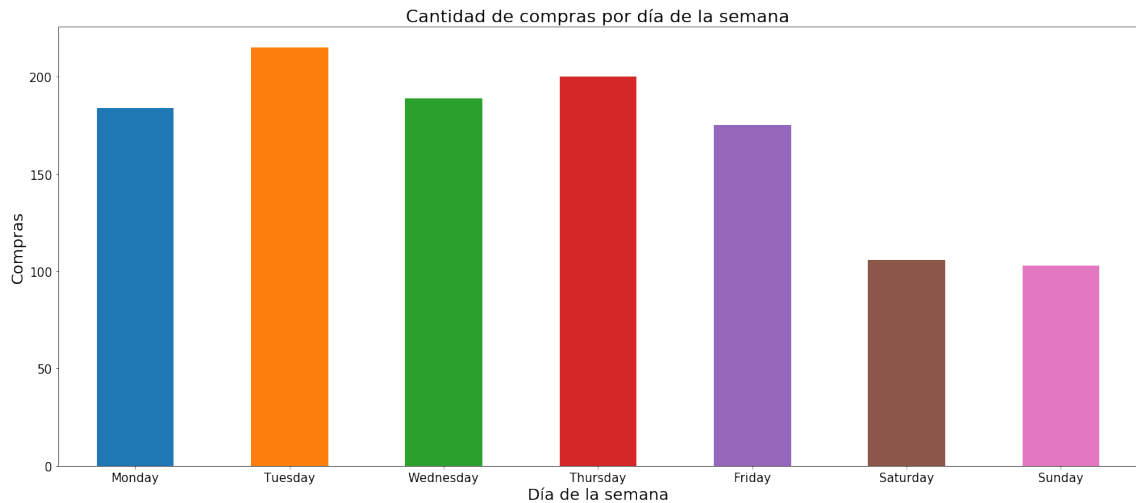


Figura 7.

Se observa que la cantidad de compras mantiene la relación con la cantidad de eventos.

Es por esto que los martes son los días donde hubo más compras, mientras que en los fines de semana se registraron pocas compras.

2.7 Evolución de la cantidad de eventos a lo largo del año

Se procede a realizar un histograma mostrando la evolución en la cantidad de eventos a lo largo del año.

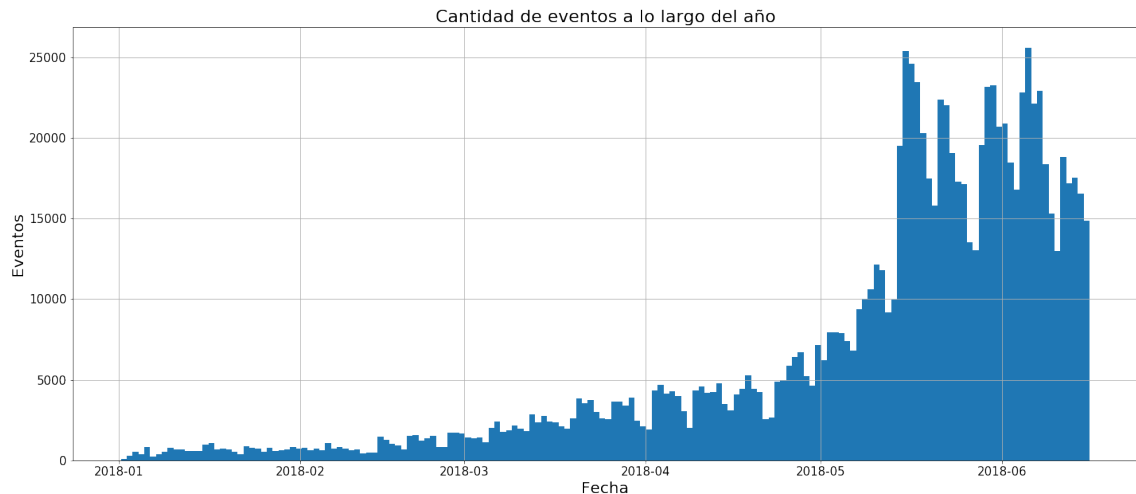


Figura 8.

A partir del gráfico se deduce que, o bien los primeros 4 meses no hubo una cantidad significativa de eventos, o estos no fueron registrados.

Igualmente, en los meses de mayo y junio se observa un gran incremento en la cantidad de eventos, con un gran salto a mitad del mes de mayo.

Se necesitarán más datos para predecir el progreso en los futuros meses, pero parece indicar que se mantendrán constantes la cantidad de eventos.

2.8 Evolución de la cantidad de compras a lo largo del año

Se muestra un histograma de las compras a lo largo del año:

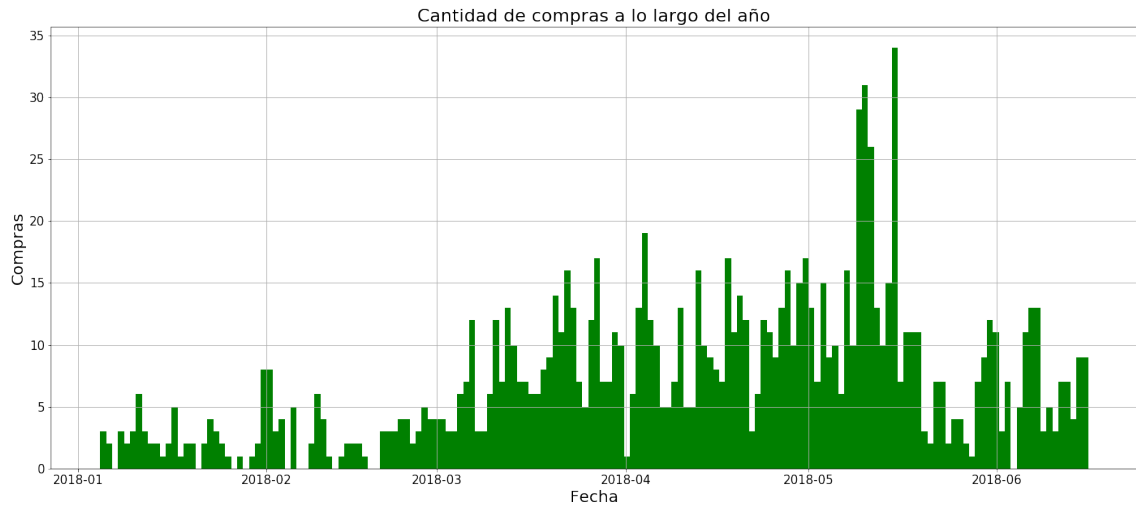


Figura 9.

Filtrando por compras, se observa que a partir del mes aumentaron lentamente, manteniendo un umbral de 20 compras diarias.

Sin embargo, a mitad del mes de mayo hubo picos de compras, durante 3 días seguidos se vendieron más de 25 celulares por día (tal vez debido a una promoción).

En los últimos días, se volvió a los niveles anteriores (menores a 20 compras diarias), tal vez indicando la necesidad de una nueva promoción.

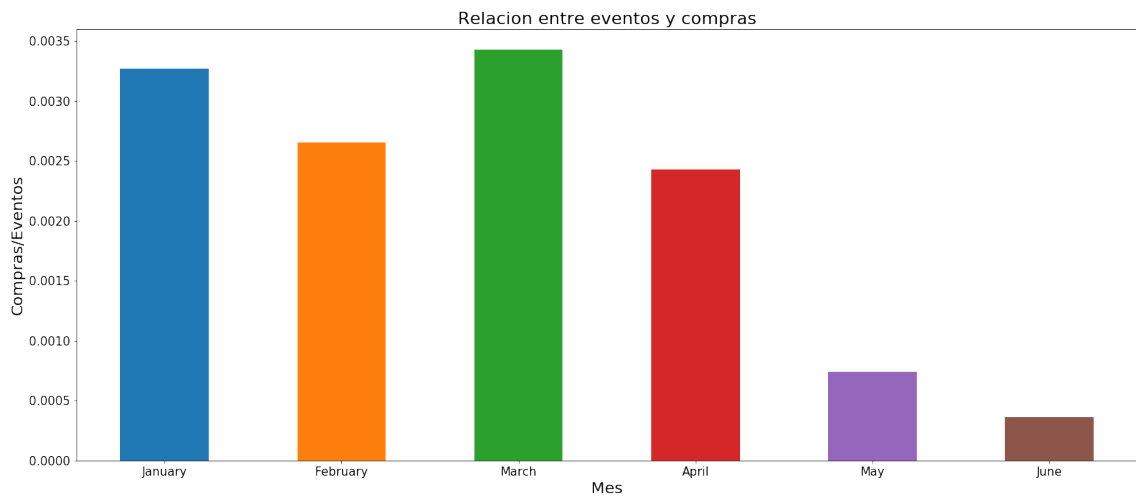


Figura 10.

En este gráfico observamos que si bien las visitas aumentaron en los últimos meses, las compras no lo hicieron con la misma intensidad, y por lo tanto, la relación entre eventos y compras fue disminuyendo.

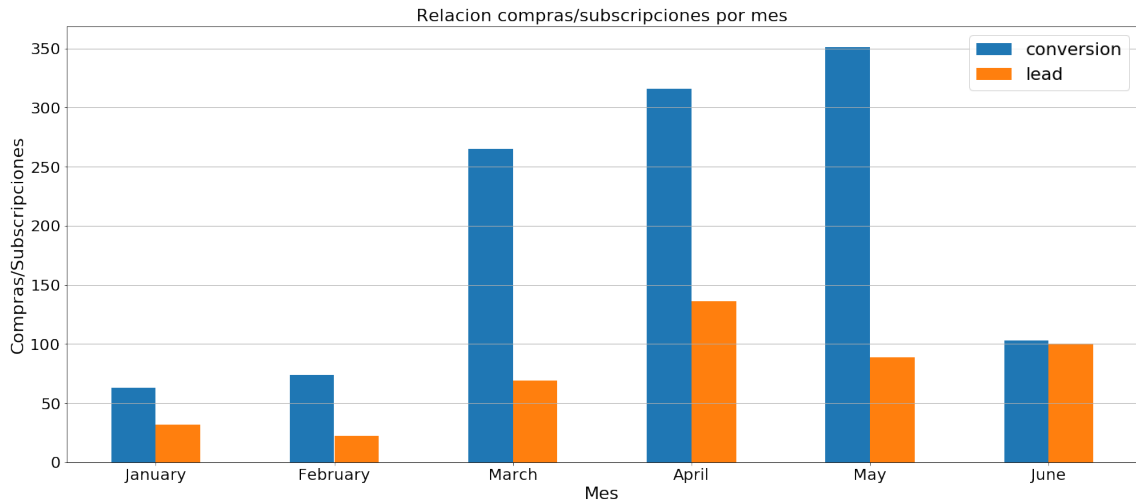


Figura 11.

Por otro lado, aunque las compras aumentan lentamente mes a mes, las subscripciones no mantienen una tendencia fija, variando mes a mes.

Un dato interesante es que en lo que va del mes de junio, prácticamente todas **las compras provocaron subscripciones nuevas**, lo cual indica que hubo una mejora en ese aspecto.

2.9 Usuarios nuevos

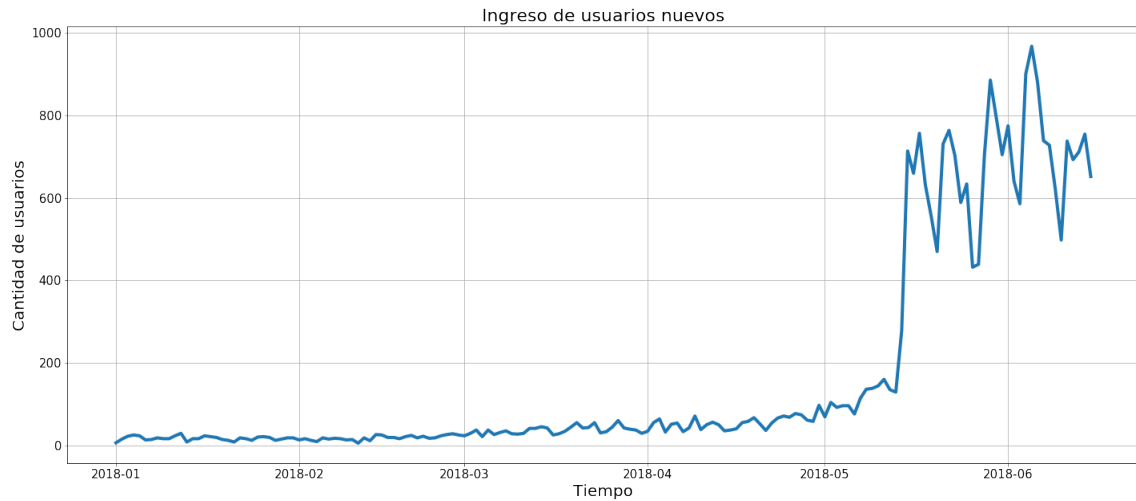


Figura 12.

Durante los primeros 4 meses, la cantidad de usuarios nuevos se mantuvo constante, por debajo de las 100 personas, pero a mediados de mayo, hubo un gran aumento repentino, que se mantuvo constante durante las ultimas semanas, y por lo que se observa, al haber picos cada vez mayores, la tendencia es alcista.

3 Análisis de eventos

A continuación, vamos a analizar cuáles son los eventos más importantes y a partir de ello, ver cómo se relacionan con los demás datos y entre ellos, ya sea la cantidad de eventos que una persona genera, el tipo de los mismos, la cantidad de eventos promedio antes de una “conversion”.

Debido a que la cantidad de clases de evento es de once, un número con el que no habrá problemas en el procesamiento de los mismos y las visualizaciones se decidió trabajar con la totalidad de estos.

Esto nos brinda un panorama general frente a los análisis y nos ahorra el descarte de información que podría ser de utilidad.

3.1 Cantidad de eventos por usuario

En un primer análisis, buscamos tener un panorama general sobre el comportamiento de los usuarios en el sitio. Para eso, agrupando por usuario, y teniendo en cuenta todos los eventos, tenemos la cantidad de eventos que cada uno generó.

Cabe aclarar que el análisis consecuente se hará con una muestra de 27624 usuarios, que generaron un total de 1011288 eventos.

Una mirada rápida de este set de datos nos permite obtener las siguientes características:

	promedio	media	std	max	min
eventos por persona	36.609036	14.0	80.595628	2771	1

Tabla 2: Características de eventos por persona.

Veamos que la media de la cantidad de eventos generados por usuario es de 14, teniendo un desvío estándar de aproximadamente 80, esto indica la variabilidad considerable que posee nuestra muestra, y nos da un primer indicio de que podría ser conveniente no analizar la totalidad de los usuarios, sino un subconjunto de estos.

Veamos que, pese a que la media es de 14 eventos por usuario, nuestro valor máximo es de 2771, lo que nos advierte que nuestra muestra podría

contener outliers, es decir, usuarios (en general en poca cantidad) que generan una cantidad de eventos que no obedecen a los estimadores de nuestra muestra, particularmente en nuestro caso, se trata de usuarios que generan una gran cantidad de eventos.

Por ultimo pero no menos importante, una moda de 4 eventos, con un valor de 3248 usuarios nos sugiere una distribución muestral sesgada hacia la izquierda. Pese a que nuestro valor máximo es de 2771, si graficamos la cantidad de usuarios por eventos generados obtenemos:



Figura 13.

Por motivos de visualización, y al observar que la muestra es decreciente, decidimos no graficar la totalidad del soporte, ya que la zona con mas cantidad de usuarios sería demasiado estrecha para sacar conclusiones.

Elegimos en particular el primer decimo del mismo, que ya de por si nos esclarece las suposiciones previas que habíamos tenido en cuenta.

En particular, se observa una clara disminución en el número de usuarios con más de 75 eventos. Esto nos indica que la cantidad de usuarios con más eventos, va a ser despreciable frente al resto (en particular del orden de las centenas para valores menores a 25 eventos)

Este grafico de línea nos sugiere desestimar dichos valores. Para poder tener una mejor observación de la muestra. Una vez realizado el filter, se presentan dos violin plots que facilitan nos presentan de forma visual una idea más precisa de nuestra muestra.

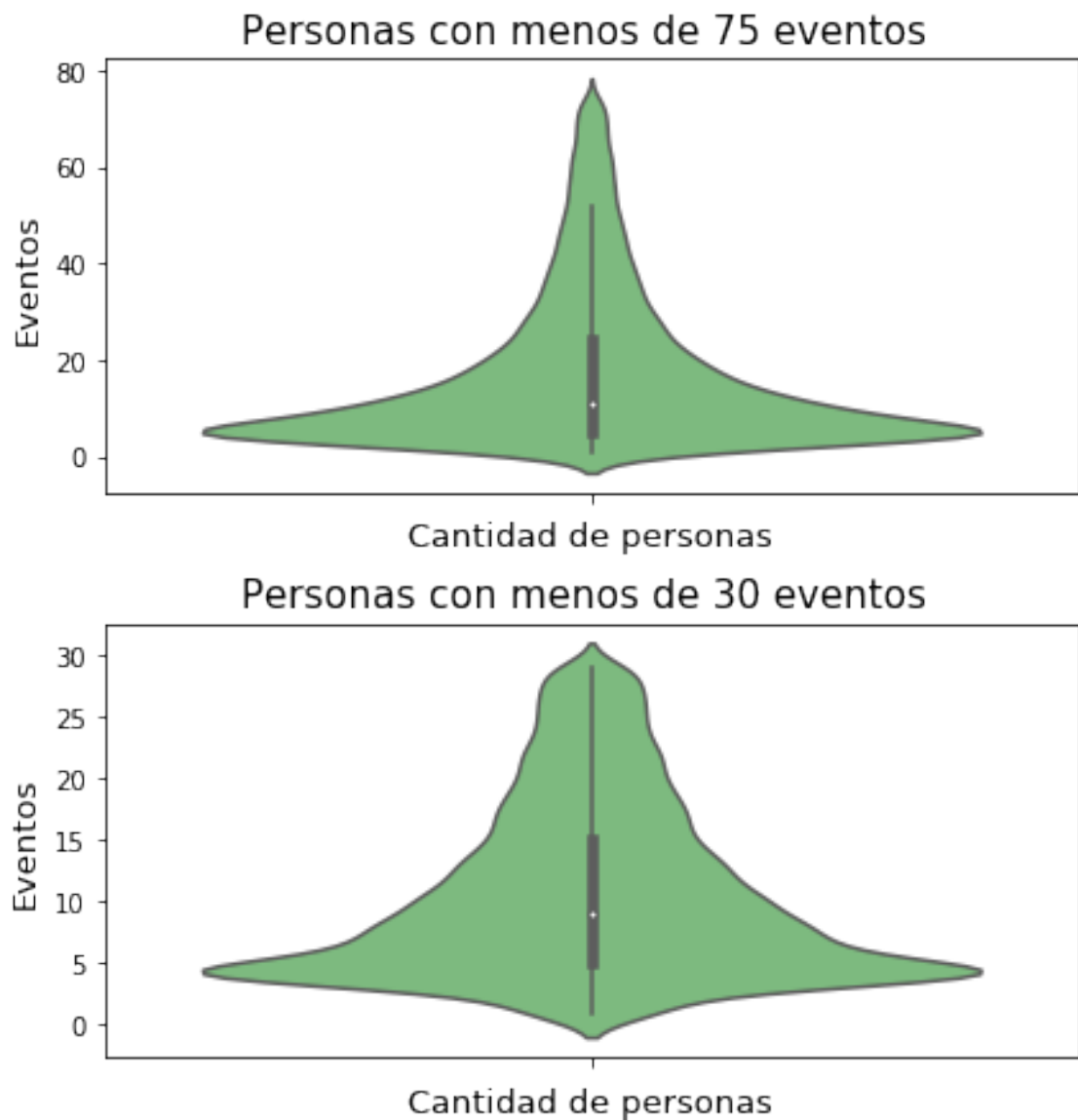


Figura 14.

Se filtraron correspondientemente las personas con mas de 75 y 30 eventos. Se eligieron estos valores meramente con fines visuales.

En el primer caso, por la poca cantidad de valores superiores a dicho rango (obteniendo un intervalo aproximadamente hasta: $\text{media} + \text{std}$)

En el segundo caso se busco principalmente presentar un intervalo de valores simétricos con respecto a la media.

En el interior de cada plot se ve la distribucion de los valores y su media.

3.2 Eventos generados por usuarios

De ahora en adelante tendremos en cuenta los tipos de eventos generados por dichos usuarios y su correlación, con el fin de detectar cualquier tipo de similitud o patrón de los mismos.

A continuación, presentamos un gráfico de barras con los distintos eventos y la frecuencia absoluta de los mismos:

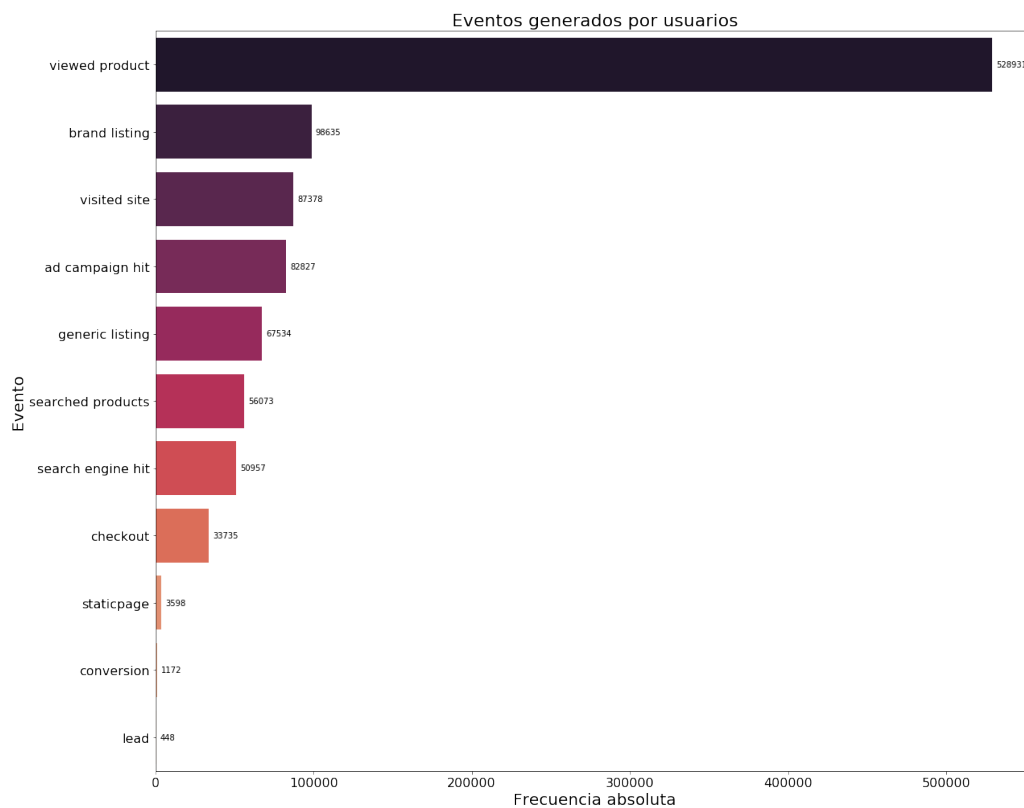


Figura 15.

Se puede observar una dominación absoluta del evento “viewed product”, debido a que muchos de los demás eventos se generan junto con el mismo en una secuencia de eventos.

Algo que parece importante destacar es que solamente el 3% de las veces que se ingresa al “checkout” se termina realizando una “conversion”, es decir, efectuando una compra.

Esto podría deberse, entre otros, a estos dos factores:

- El proceso de “checkout” no transmite la seguridad necesaria.
- El proceso desde “checkout” hasta “conversión” es excesivamente complejo.

3.3 Modelos más vendidos

A continuación se presenta una lista con los modelos más vendidos, la información que se extrae de la misma podría ser de utilidad para reforzar las ofertas y publicidades sobre estos modelos, y disminuir a aquellos en los cuales no se están produciendo una cantidad considerable de ventas.

Samsung Galaxy J5	88
iPhone 5s	84
iPhone 6	71
Samsung Galaxy S6 Flat	39
iPhone 6S	37
iPhone 5c	33
Motorola Moto G4 Plus	30
Motorola Moto G3 4G	30
Samsung Galaxy J7	28
Samsung Galaxy A5	27
Samsung Galaxy Gran Prime Duos TV	24
Samsung Galaxy J7 Prime	20
Samsung Galaxy Note 4	19
Samsung Galaxy S7	18
Samsung Galaxy S5	18
Samsung Galaxy Gran Prime 3G Duos	18
Motorola Moto G2 3G Dual	17
iPhone SE	17
Samsung Galaxy S3 Mini	17
iPhone 5	17

Tabla 3: Modelos más vendidos.

En un primer análisis general, se observa que celulares de gama media son los más frecuentes, destacando la tendencia a los modelos de Samsung, Apple y Motorola.

Solo acercamos una conclusión rápida por el momento, ya que el análisis con respecto a las Marcas se realizara en otra sección con más profundidad.

Otra observación es que en el Top-3 de modelos vendidos, los valores presentan una gran similitud. En contraste a la relación con sus relegados seguidores, cuyas ventas son menores que la mitad de las de estos.

Como complemento visual a la presentación de tabla, se incluye un gráfico de barras de dicho set:

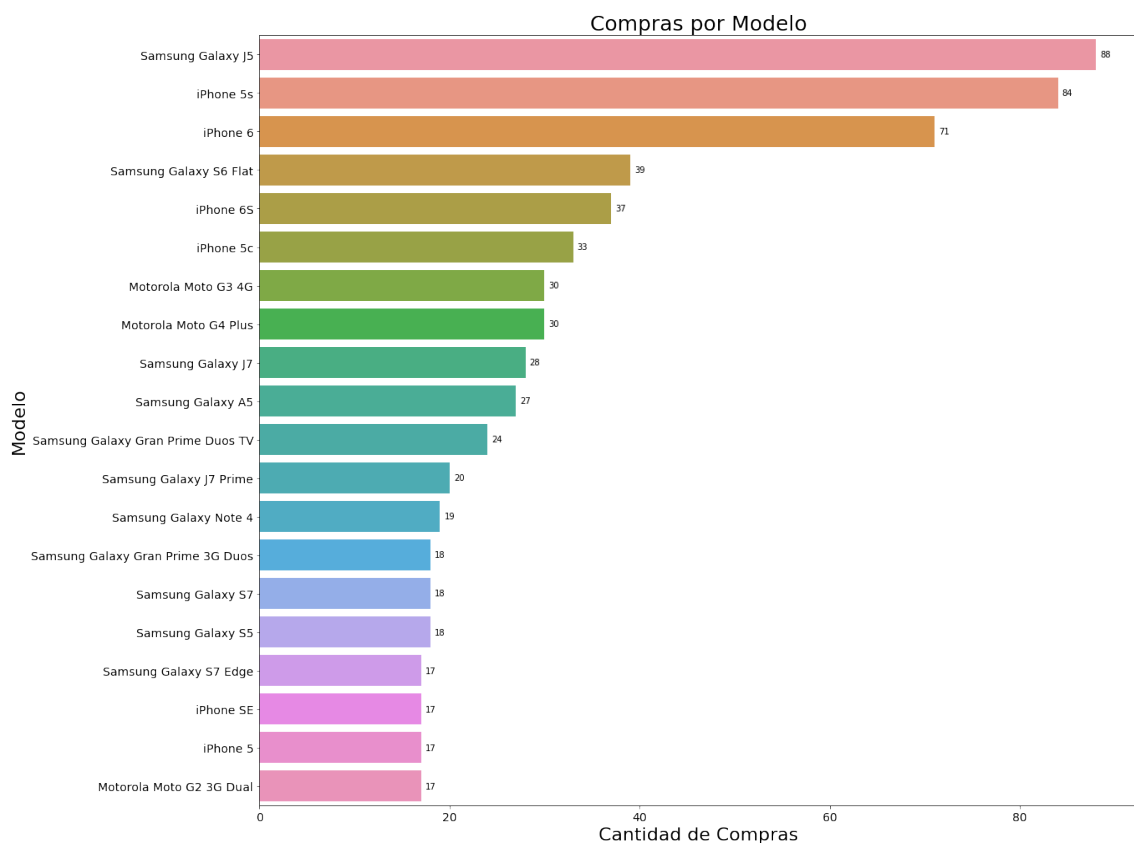


Figura 16.

Adentrándonos en la observación rápida sobre las ventas por marcas, se presenta un gráfico con la cantidad de ventas por marca:

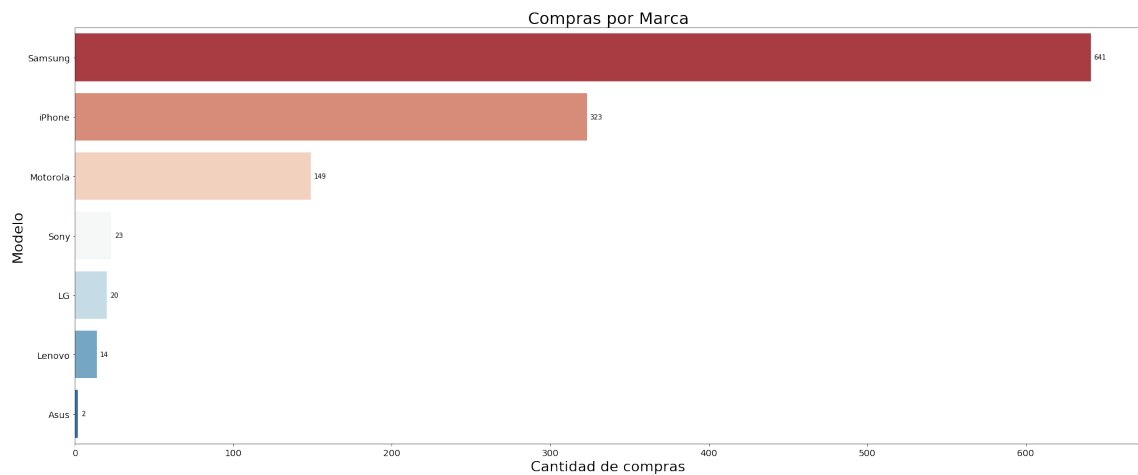


Figura 17.

El gráfico reafirma las conclusiones previas situando a Samsung, Apple (iPhone) y Motorola como los líderes en este rubro.

Adicionalmente el gráfico nos permite ver la supremacía de dichas marcas con respecto a sus competidoras, cuyos ordenes de ventas son como a lo sumo un 10% de las líderes.

Otro aspecto importante para analizar es la cantidad de visitas por marca, lo cual, en conjunto con el gráfico previo, puede evidenciar la tendencia que tienen los usuarios al considerar una marca en particular, pero luego no realizar la compra.

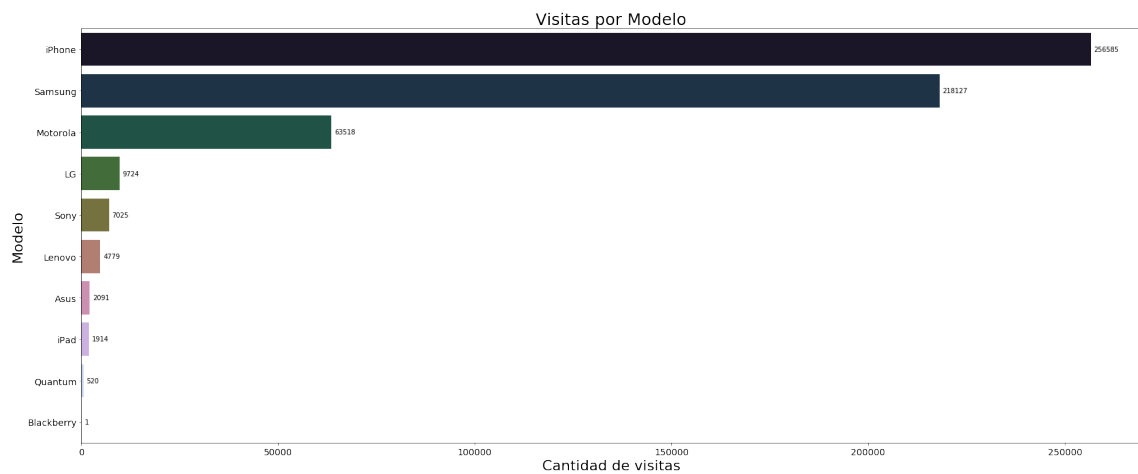


Figura 18.

Se puede observar que los celulares iPhone son altamente visitados, pero luego el rango de compra disminuye de forma mas abruptamente que las marcas Samsung y Motorola, a continuación se presentan la razón entre la cantidad de ventas y visitas por marca:

- **iPhone:** 794.380805
- **Samsung:** 340.603744
- **Motorola:** 426.29530

Teniendo en cuenta dicha razón, se puede concluir que los usuarios que visitan modelos Samsung tienen mayor tendencia a comprar que las demás marcas.

Cabe aclarar que no se consideraron las demás marcas debido a la baja cantidad de ventas de estas con respecto a las tres mencionadas, evitando caer en un mal uso de la ecuación de de-Moivre, y en conclusiones erróneas.

Para continuar el análisis de estas tres marcas se propone observar la evolución de sus ventas a lo largo de los meses, con el fin de observar tendencias en las mismas:

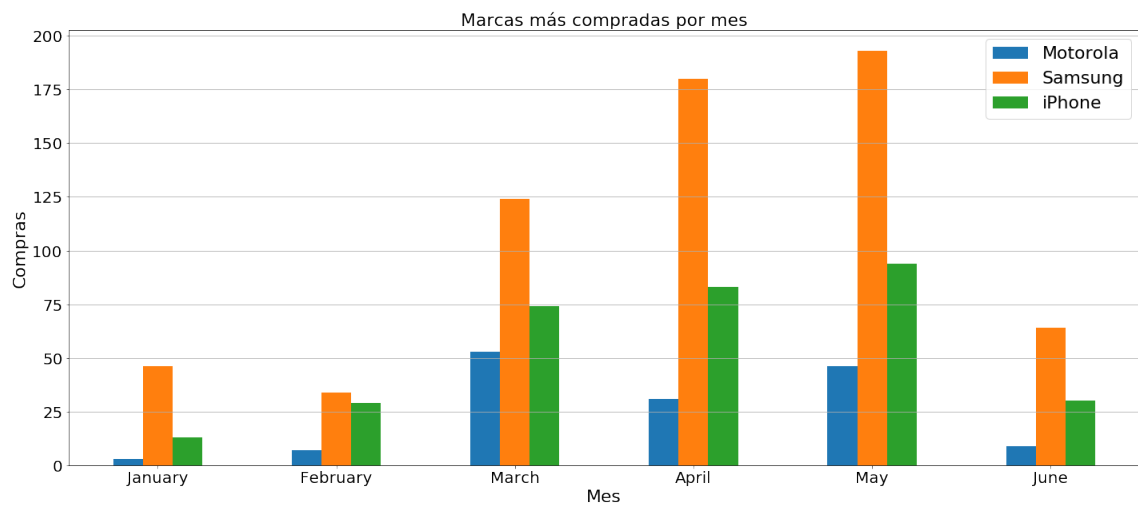


Figura 19.

Se puede observar un crecimiento paulatino en las marcas Samsung y Apple(iPhone) en los primeros cinco meses del año. Una explicación para la disminución de las ventas en enero, febrero y junio podría ser la falta de promociones y difusión del sitio, o el lanzamiento de nuevos modelos en los meses con mas ventas.

Se ve un comportamiento errático de la marca Motorola, en torno a las 25 ventas por mes.

3.4 Capacidad de los equipos

Otro aspecto relevante es la Capacidad de los equipos, es conocida la tendencia al aumento de la misma en los nuevos modelos, debido a aumentos en los tamaños de las actualizaciones y Sistema Operativo y mayor flujo de multimedia a almacenar.

Usualmente, las capacidades de 16GB y 32GB son las más solicitadas, debido a su comodidad y precio frente a otras mayores.

Dicha tendencia se evidencia en el siguiente gráfico:

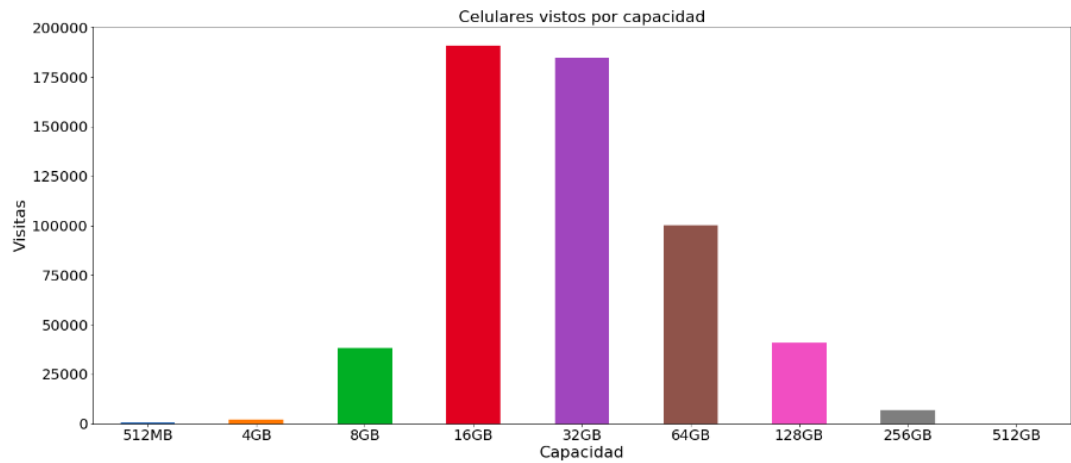


Figura 20.

Si adicionamos al mismo el gráfico de Ventas por capacidad, se observa una moda en los celulares de 16GB, que posiblemente provenga a la armonía capacidad/precio que en la actualidad ofrecen.

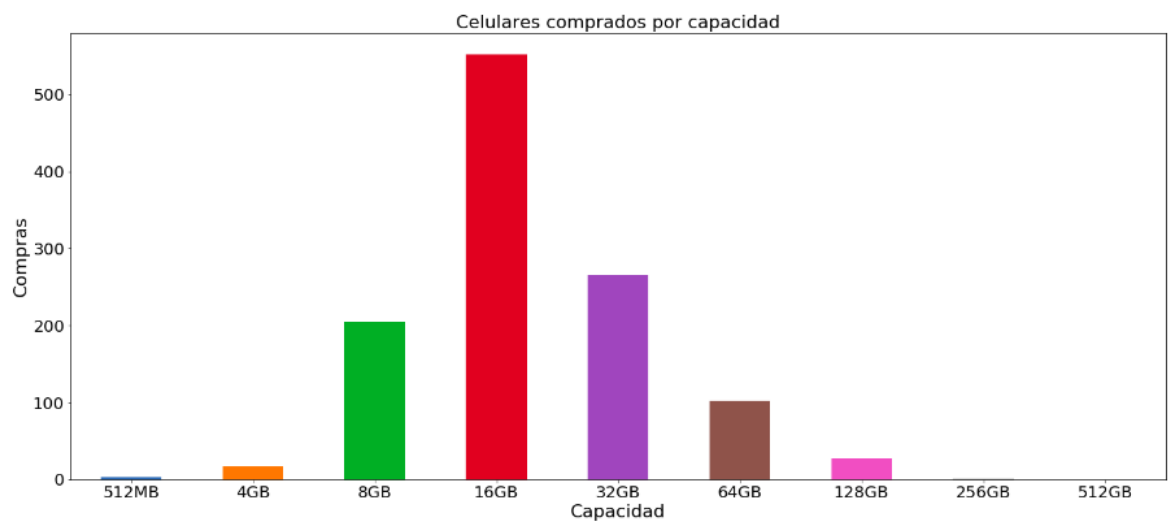


Figura 21.

3.5 Compras por condición

Otra faceta interesante a presentar es las vistas de los celulares en función de su condición:

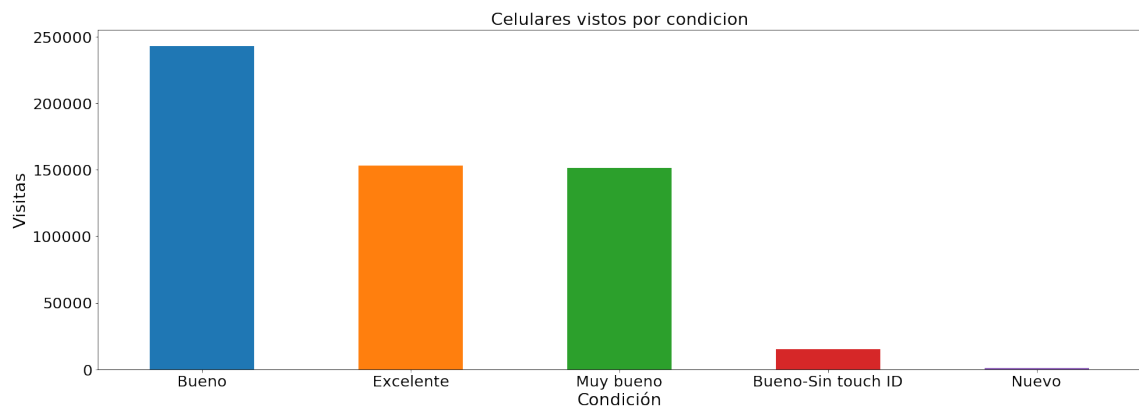


Figura 22.

La desestimable cantidad de vistas en celulares nuevos remarca el espíritu del sitio, mientras que los celulares buenos son los más observados, posiblemente por una interesante disminución de su precio en un producto todavía funcional y duradero.

Si además, como en los anteriores análisis, agregamos un gráfico de compras, podremos ver la clara tendencia a celulares buenos, disminuyendo mientras mejor condición presentan los mismos:

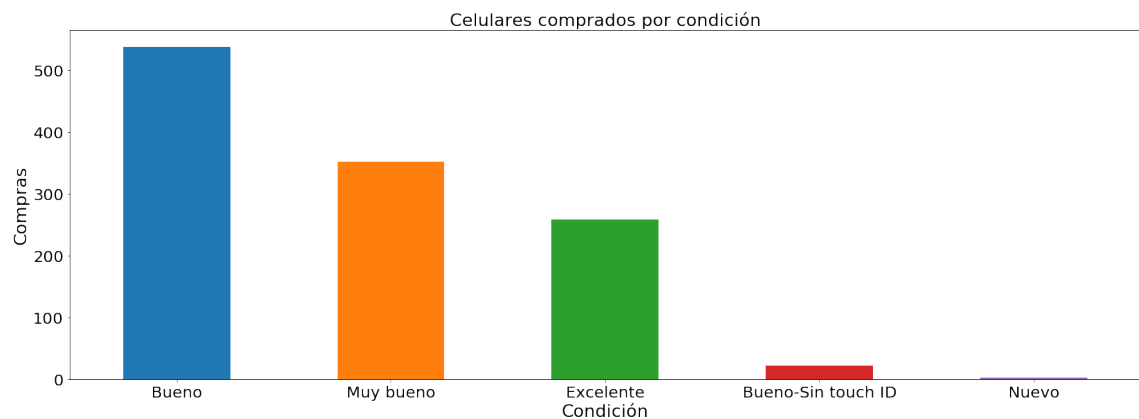


Figura 23.

Adicionalmente, es de interés conocer el origen de los usuarios, en este gráfico se pretende exhibir la cantidad de usuarios provenientes de una campaña a lo largo de los meses del set:

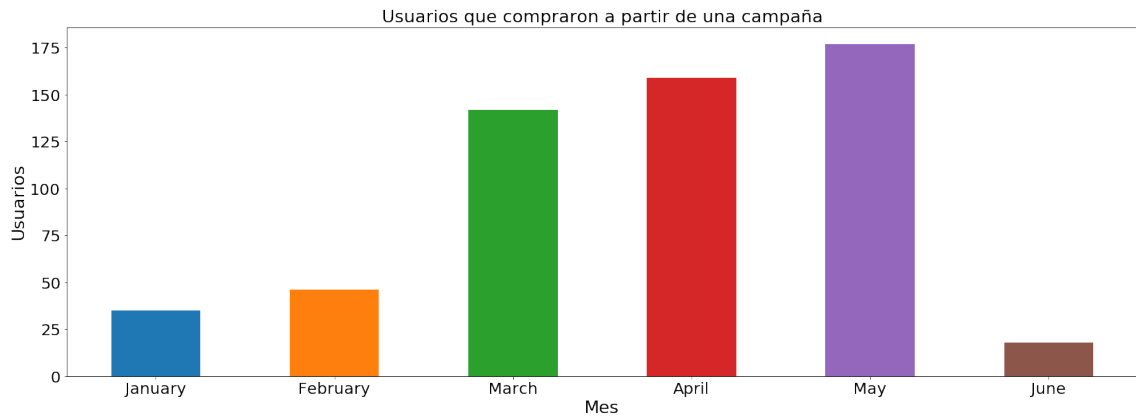


Figura 24.

Esto podría relacionarse con la venta de equipos de las marcas líderes (Samsung, iPhone y Motorola), cuyos picos de venta ocurrieron en estos meses. El conjunto de ambos gráficos podría servir como una iniciativa para intentar incrementar los usuarios que provienen de campañas.

3.6 Cantidad de eventos por tipo de dispositivo

En este párrafo nos centraremos a diferenciar los diferentes tipos de dispositivos y la cantidad de eventos en los que estuvieron involucrados:

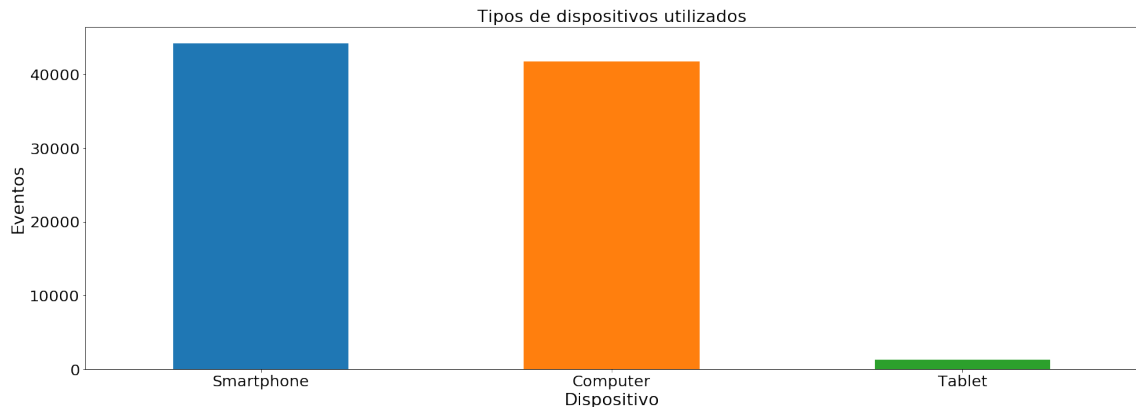


Figura 24.

Se observa una tendencia marcada a la elección de Smartphones y Computadoras, siendo Tablet un tipo de dispositivo muy poco utilizado. Esto puede deberse, entre otras cosas, a la poca reutilización del producto, la carencia de información de los usuarios de estos en el sitio, o la menor popularidad del mismo frente a los demás dispositivos.

3.7 Navegadores preferidos

En esta sección se pretende hacer un análisis de los Navegadores Web más utilizados. Para comenzar se muestra una lista de los mismos, y un gráfico de barras que permite evidenciar mejor la relación entre los navegadores mas usados:

	browser_version	percentages
Chrome	38081	43.581908
Chrome Mobile	36780	42.092975
Mobile Safari	3230	3.696583
Firefox	2566	2.936666
Samsung Internet	1699	1.944425
Facebook	1415	1.619401
Edge	893	1.021996
Opera	608	0.695827
IE	398	0.455492
Android	347	0.397125

Tabla 4: Navegadores preferidos.

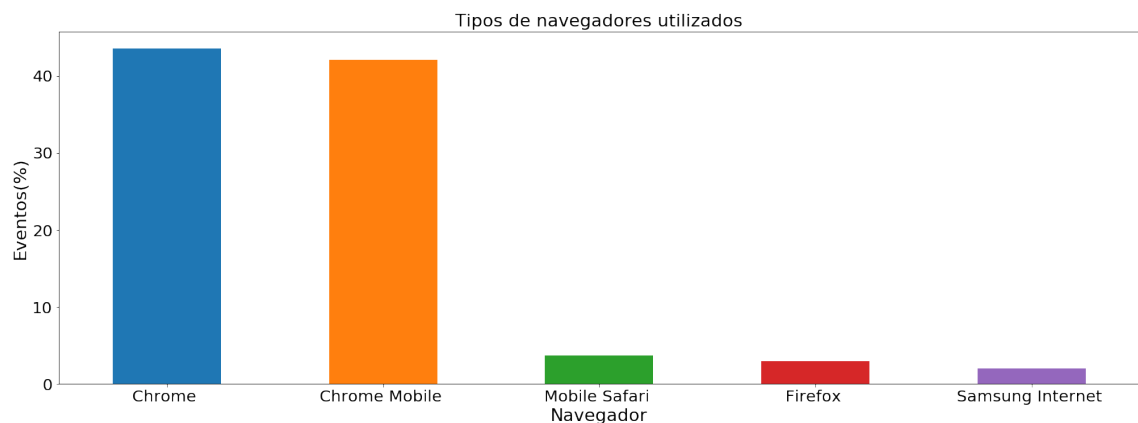


Figura 25.

Es clara la supremacía de Chrome frente a otros navegadores, esto podría deberse a la inmensa cantidad de dispositivos Android que cuentan con dicho navegador como predeterminado, y su vinculación directa con el buscador de Google.

3.8 Proporción de usuarios que ingresan por un buscador

Otro aspecto relacionado con los navegadores es la proporción de usuarios que ingresan mediante un buscador.

En el caso del set de datos se calculó que el 5% de los usuarios ingresa mediante un buscador, lo que puede significar en parte a usuarios redireccionados mediante publicidades u otros sitios web.

3.9 Buscadores más utilizados

Teniendo en cuenta los buscadores utilizados, se observa:

	frecuencia	porcentaje
Google	50240	98.592931
Bing	424	0.832074
Yahoo	287	0.563220
Ask	6	0.011775

Tabla 5: Buscadores más utilizados.

Esto nos permite concluir que la elección del Navegador está íntimamente relacionada con el buscador, teniendo a Google como el buscador predeterminado tanto de Chrome como de Chrome Mobile.

4 Análisis Geográfico

A continuación se realizará un análisis de la distribución geográfica de los usuarios y los eventos generados por ellos, tanto a nivel de ciudad como de país.

4.1 Ciudades con más usuarios

	cantidad	porcentaje
São Paulo	4297	14.783596
Rio de Janeiro	1379	4.744375
Belo Horizonte	938	3.227138
Salvador	932	3.206496
Brasília	535	1.840639
Curitiba	441	1.517237
Fortaleza	429	1.475951
Osasco	403	1.386500
Recife	386	1.328012
Porto Alegre	375	1.290167

Tabla 6: Ciudades con más usuarios.

4.2 Ciudades con más eventos

	frecuencia	porcentaje
São Paulo	11711	16.365517
Rio de Janeiro	3538	4.944172
Belo Horizonte	2568	3.588647
Salvador	2314	3.233695
Brasília	1530	2.138096
Osasco	1148	1.604271
Curitiba	1132	1.581911
Fortaleza	1084	1.514834
Recife	913	1.275870
Goiânia	853	1.192023

Tabla 7: Ciudades con más eventos.

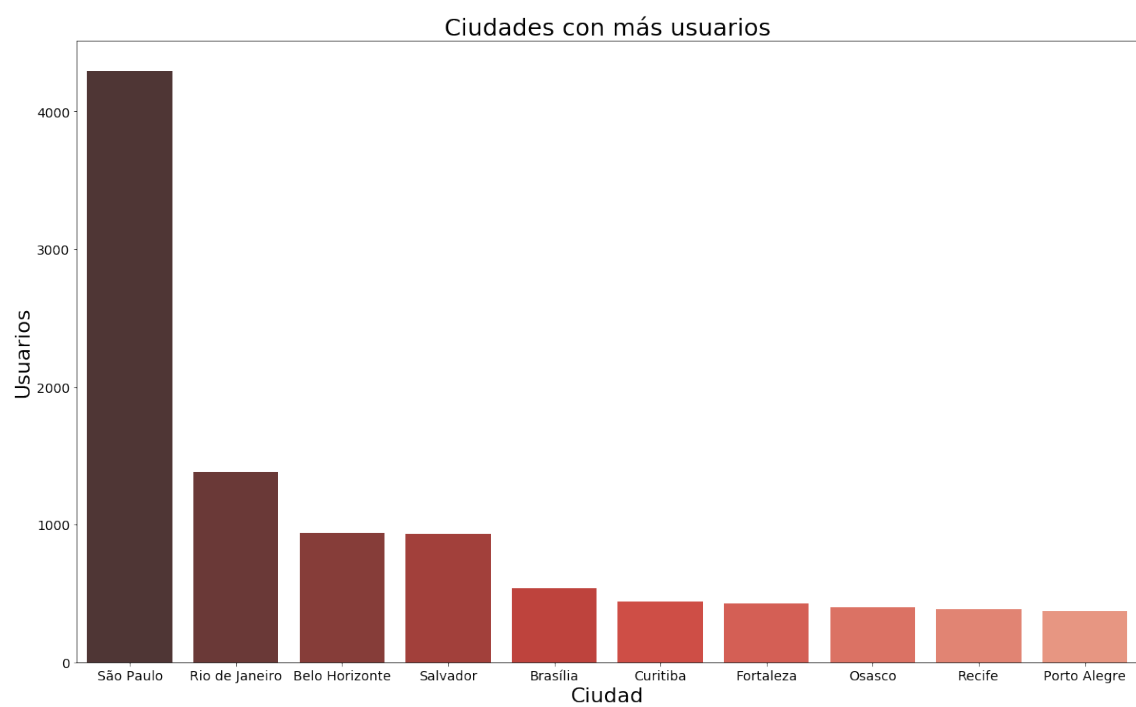


Figura 26

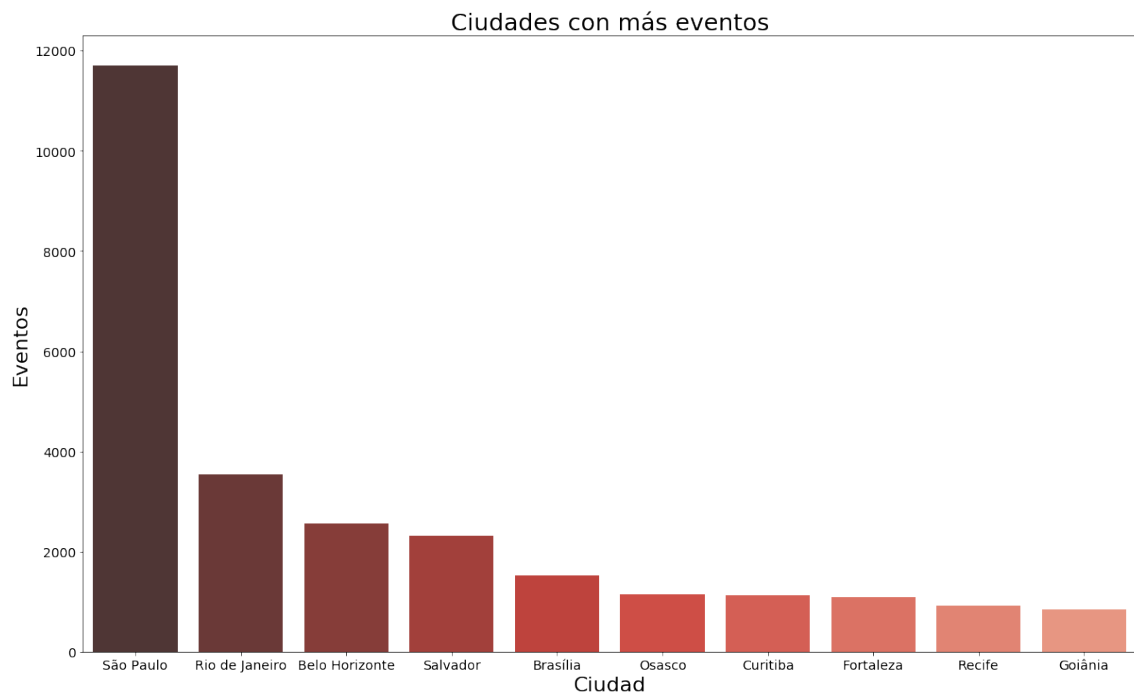


Figura 27

Tanto si tenemos en cuenta la cantidad de usuarios como la cantidad de eventos, se destacan las ciudades más importantes de Brasil: San Pablo, Rio de Janeiro, Belo Horizonte, Salvador, Brasilia, etc. Es esperable la correlación entre cantidad de usuarios y cantidad de eventos por ciudad, no se detectan anomalías, por lo que se puede decir que no hay clientes que destaquen por sus volúmenes de compra.

4.3 Países con más usuarios

	cantidad	porcentaje
Brazil	25872	98.215777
United States	322	1.222383
Argentina	25	0.094905
Canada	20	0.075924
France	10	0.037962
Netherlands	10	0.037962
Mozambique	9	0.034166
United Kingdom	8	0.030370
Portugal	7	0.026574
Italy	5	0.018981

Tabla 5: Países con más usuarios.

La mayoría de los usuarios son de Brasil, el porcentaje del resto de los países es depreciable frente a este.

4.4 Países con más eventos

Países con más eventos

	cantidad	porcentaje
Brazil	84308	99.110092
United States	465	0.546641
Argentina	94	0.110504
Canada	35	0.041145
Netherlands	13	0.015282
France	13	0.015282
United Kingdom	12	0.014107
Mozambique	12	0.014107
Japan	11	0.012931
Poland	9	0.010580

Tabla 5: Países con más eventos.

Presentamos un mapa con las visitas a la plataforma:

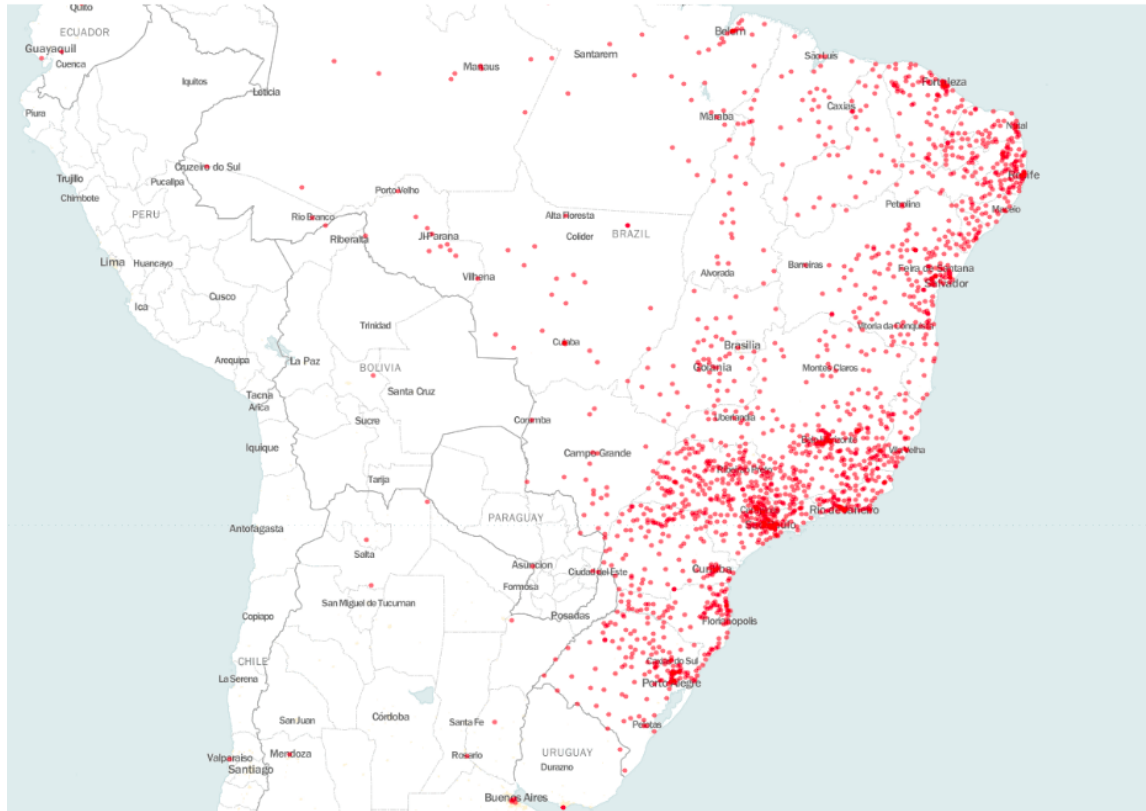


Mapa 1: Mundo.

Se observan visitas aisladas de todo el mundo, pero claramente se concentran en el este de Brasil. También es destacable EEUU, distribuido en varias ciudades, Buenos Aires en Argentina, y zonas de la península Iberica.

Se nota muchisima actividad en las zonas de habla portuguesa e hispana, que son los dos idiomas en los que esta disponible la pagina. Seguramente **sería beneficioso contar con un sitio traducido al ingles para seguirse expandiendo por Estados Unidos.**

Se presenta un mapa de Brasil con las visitas al sitio:



Mapa 2: Brasil.

Si nos enfocamos en Brasil propiamente dicho, se puede ver que los usuarios están bien distribuidos por todas las zonas pobladas del país, concentrándose especialmente en los principales centros urbanos.

5 Análisis del usuario

Se muestran algunas cifras referidas a la actividad de los usuarios:

- Cantidad de personas que compraron: 716
- Promedio de visitas de personas que compran: 1.6368715083798884
- Máximo de compras de una persona: 15

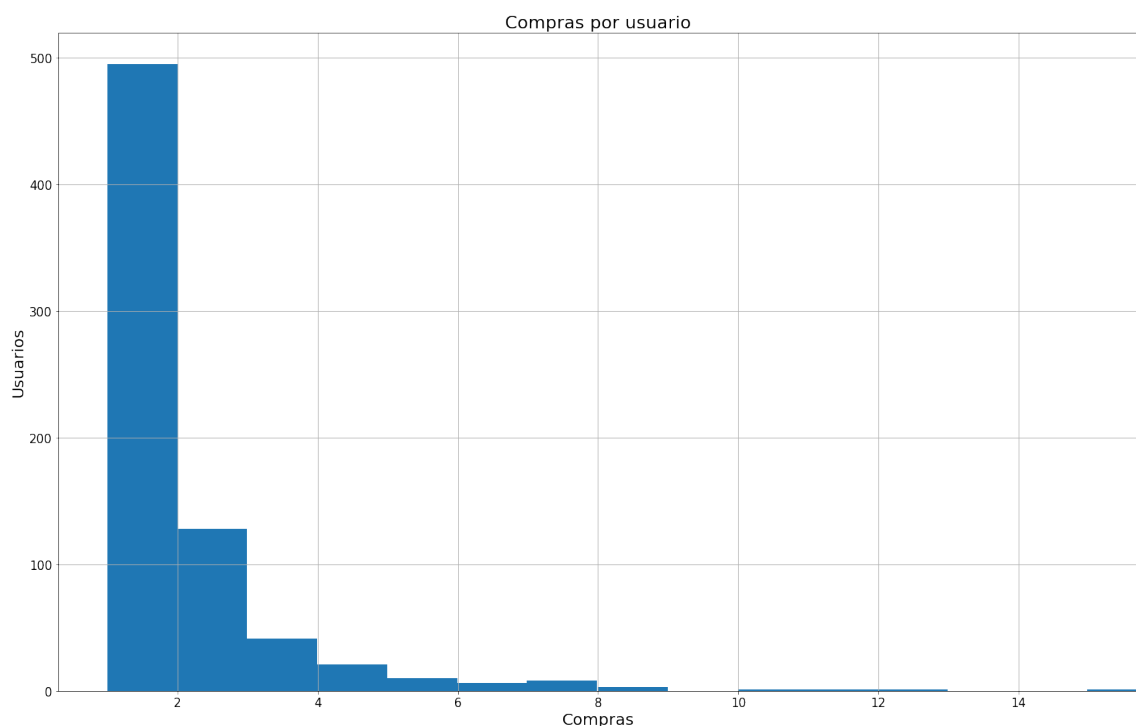


Figura 28.

En este histograma se observa que la mayoría de las personas compran una sola vez, poco más de 100 compraron 2 veces, y a partir de ahí son valores muy bajos, aunque curiosamente un cliente realizó 15 compras.

Porcentaje de usuarios que realizaron solo una compra: 69.1340782122905

Cantidad de personas que realizaron ckeckouts: 27624

Podemos ver algo interesante, porque si se calcula la cantidad de usuarios distintos, resulta ser 27624. Esto nos indica que el sistema reconoce a

usuarios como distintos recién después de hacer un checkout. Esto es muy importante para entender mejor el sistema de web analytics.

Promedio de checkouts por persona: 1.22122067767159

Relación entre personas que hicieron checkouts y personas que compraron 2.591949029829134

Esto demuestra que mucha gente realizó un checkout, pero solo el 2,5 % efectivamente compraron un celular.

Relación entre total de checkouts y total de compras 3.4741366533274047

5.1 Cantidad de productos vistos por persona

	promedio	media	std	max	min
productos vistos por persona	11.242362	4.0	21.306081	552	1

Tabla 10: Cantidad de productos vistos por persona

Usuarios que ingresan por una campaña publicitaria: 8.190248475211808

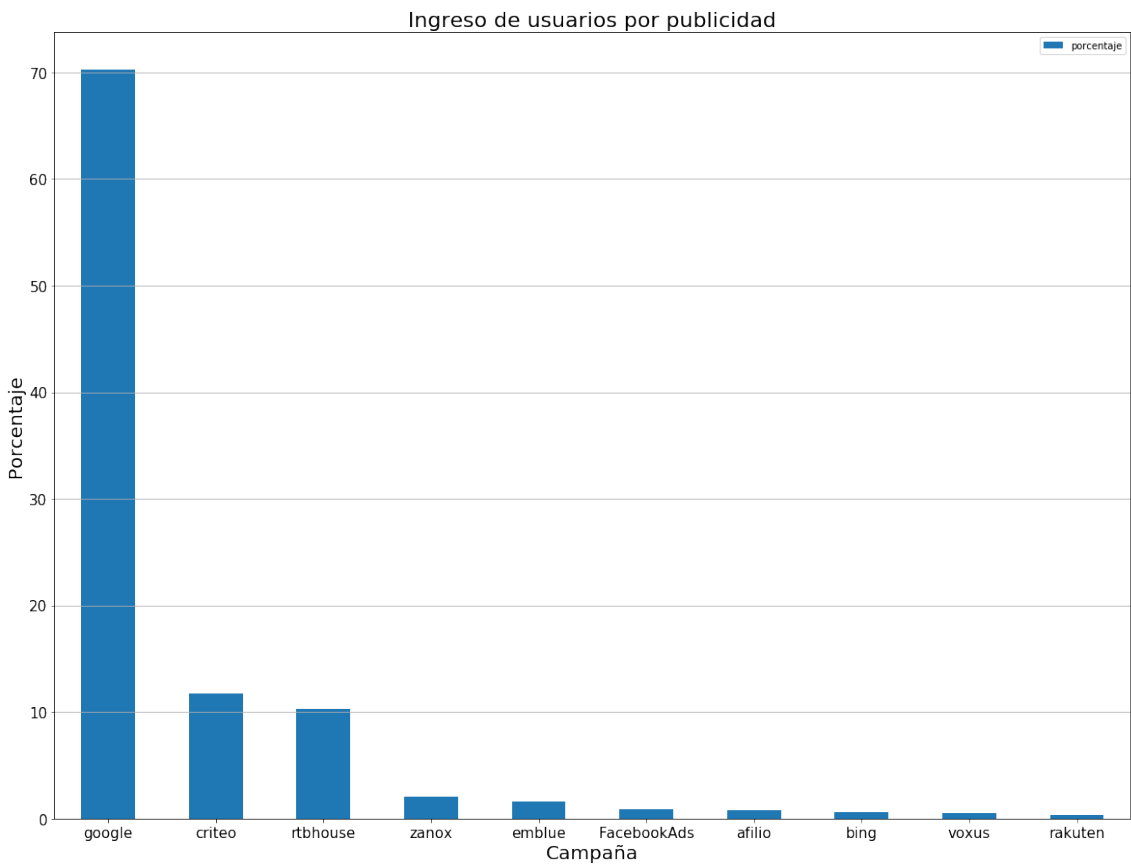


Figura 28.

Como vimos anteriormente, Google es el principal buscador usado, y por lo tanto tiene sentido que la mayoría de las campañas publicitarias provengan de ahí. Es de notar que las campañas en Facebook no están resultando.

6 Conclusiones generales

Al finalizar el presente trabajo, nos propusimos destacar ciertos aspectos del mismo, de tal modo que permitan a la empresa tener una visión más clara de su progreso.

Los principales puntos a tratar son:

- Las ventas de celulares vienen aumentando con los meses, pero lentamente, aproximadamente 50 ventas más por mes.
- Las campañas y avisos publicitarios deben publicarse durante la semana, y no en el fin de semana. Además, dichas campañas deben mostrarse a partir de las 16:00, recordando que la mayor actividad ocurre a la noche.
- El día martes es un buen candidato para publicitar, tal vez porque los lunes son muy cargados de trabajo, y recién el martes uno tiene más libertad.
- Continuar con la campaña de subscripciones, luego de realizar una compra, pues en el último mes viene dando resultados.
- Si bien la marca más comprada es Samsung, deberían profundizar en la venta de iPhone, que son los más visitados, pues están de moda por ser considerados de alta gama. Esto no implica no seguir con los modelos de Samsung, simplemente abrir el panorama.
- Si bien los celulares de 16gb son los más vendidos, prestar atención con los de 32gb, que podrían ser los próximos protagonistas. A aquellos con 8gb, no brindar tanto soporte como antes.
- La condición “Bueno” sigue siendo la más visitada y comprada, tener esto en cuenta. Cada mes más usuarios compraron a partir de una campaña, pero como se vió en el análisis, la única con resultados significativos es la de Google. Por lo tanto, creemos que debería continuarse con esta modalidad, pero habría que evaluar la relación costo beneficio de las otras campañas publicitarias.
- Dar soporte especializado al navegador Google Chrome, tanto en su versión de escritorio como en su versión mobile, los otros no son tan significativos a nivel cantidad de usuarios.

- Por su parte, las tablets están en desuso, por lo tanto, enfocarse en los smartphones principalmente, que van a superar las computadoras de escritorio pronto, en términos de uso.
- Si bien la mayoría de los usuarios son de Brasil, se noto una creciente actividad en los Estados Unidos. Por lo tanto, una propuesta es traducir la página web al inglés, para promover la visita de personas de dicho país. También notamos un posible negocio futuro en Argentina, donde se debe hacer hincapié en la diferencia de precio con respecto a otros vendedores.
- Por último, se debe volver a hacer un análisis de datos en unos meses, y verificar el progreso de la empresa en base a las recomendaciones propuestas, para determinar el correcto camino.

References

- [1] Trocafone website, www.trocafone.com.
- [2] NumPy - NumPy, <http://www.numpy.org/>.
- [3] Python Data Analysis Library, <https://pandas.pydata.org/>.
- [4] Matplotlib: Python plotting — Matplotlib 3.0.0 documentation, matplotlib.org.
- [5] seaborn: statistical data visualization — seaborn 0.9.0 documentation, seaborn.pydata.org
- [6] Folium information,
<https://github.com/python-visualization/folium>
- [7] GeoPy's documentation,
<https://geopy.readthedocs.io/en/stable/>