

U.B.A. FACULTAD DE INGENIERÍA
Departamento de Computación
Organización de Datos 75-06
Ing. Informática

Trabajo Práctico N°1
Análisis exploratorio
“Trocafone”

Alvarez, Natalia Nayla	90928
Flores Sosa, Zoraida	87039
Fonzalida, Miguel Angel	86125
Porras Carhuamaca, Sherly	91076

Link al GitHub: <https://github.com/miguelAfonzalida/75.06-Datos-TP->

Fecha de Entrega : 24/09/2018

Fecha de Aprobación :

Grupo:20

Firma de Aprobación :

Introducción

Se pide evaluar un set de datos sobre el análisis de datos sobre un conjunto de eventos de web analytics de usuarios que visitaron www.trocafone.com, su plataforma de ecommerce de Brasil. Trocafone es un side to side Marketplace para la compra y venta de dispositivos electrónicos que se encuentra actualmente operando en Brasil y Argentina.

El objetivo de este trabajo es realizar un análisis exploratorio de los datos del set, para obtener resultados interesantes que le sean relevantes a la empresa.

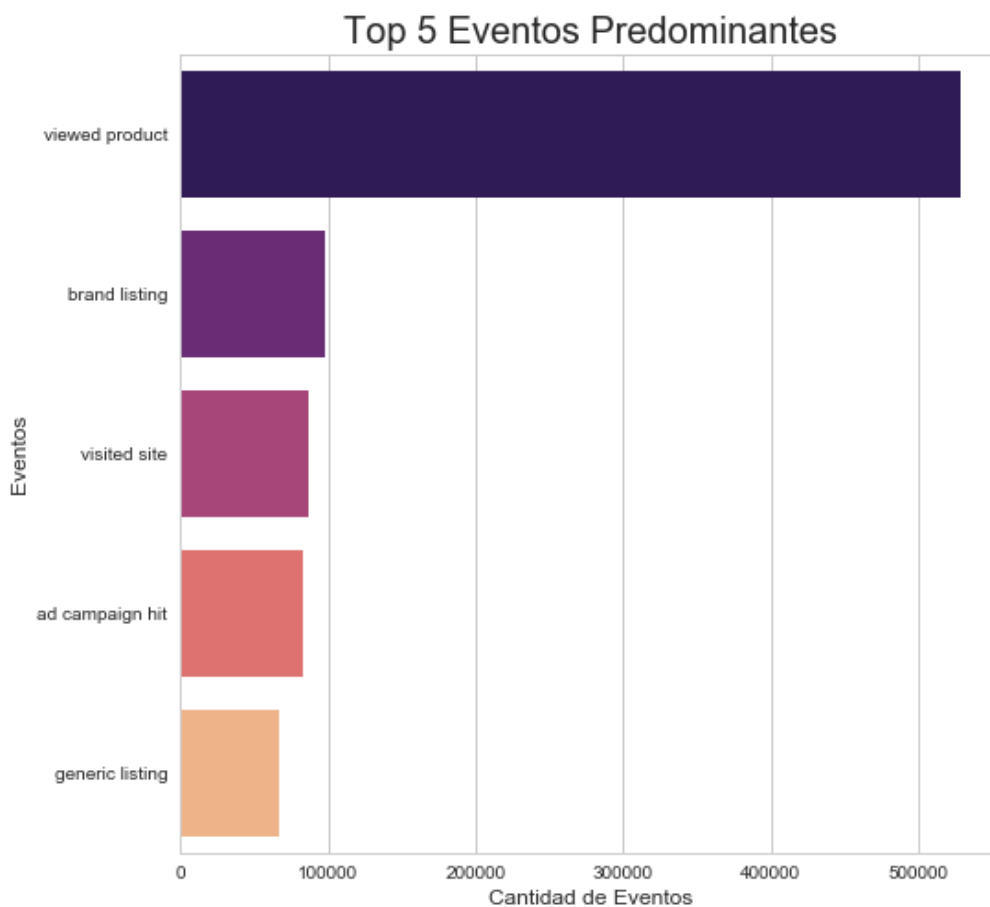
The screenshot shows the Trocafone website interface. At the top, there is a search bar with the text "Buscá Celulares Usados con Garantía" and a magnifying glass icon. Below the search bar, there are navigation links for various brands: Samsung, Nokia, LG, Motorola, Sony, and Huawei. The main banner features a large image of a Samsung Galaxy J7 smartphone with the text "Show de OFERTAS" and "SAMSUNG GALAXY J7 BLANCO DESDE: \$5.179,00". Below the banner, there is a row of four icons representing different benefits: "Compra Segura", "12 cuotas sin interés", "90 días de Garantía", and "Envíos gratis". The section "Ofertas del día" is followed by "Smartphones exclusivos", which displays a grid of six smartphone offers. Each offer includes a product image, the brand and model, storage capacity, color, starting price, and financing options.

Smartphone	Brand	Model	Storage	Color	Starting Price (ARS)	Financing Options
Samsung Galaxy S6 Edge+	Samsung	S6 Edge+	64GB	Dorado	13.569	12x sin interés de \$1.131
Moto Z Play	Motorola	Z Play	32GB	Negro	9.549	12x sin interés de \$796
Sony Xperia XA1	Sony	Xperia XA1	32GB	Blanco	9.489	12x sin interés de \$791
Sony Xperia XA1	Sony	Xperia XA1	32GB	Negro	9.259	12x sin interés de \$772
Samsung Galaxy A5 (2017)	Samsung	Galaxy A5 (2017)	32GB	Dorado	9.199	12x sin interés de \$767
Samsung Galaxy A5 (2017)	Samsung	Galaxy A5 (2017)	32GB	Negro	9.199	12x sin interés de \$767

Desarrollo

Para comenzar a realizar este trabajo, nos enfocamos a explorar las columnas del dataset haciéndonos preguntas para obtener alguna conclusión interesante que sea productivo para nuestro análisis.

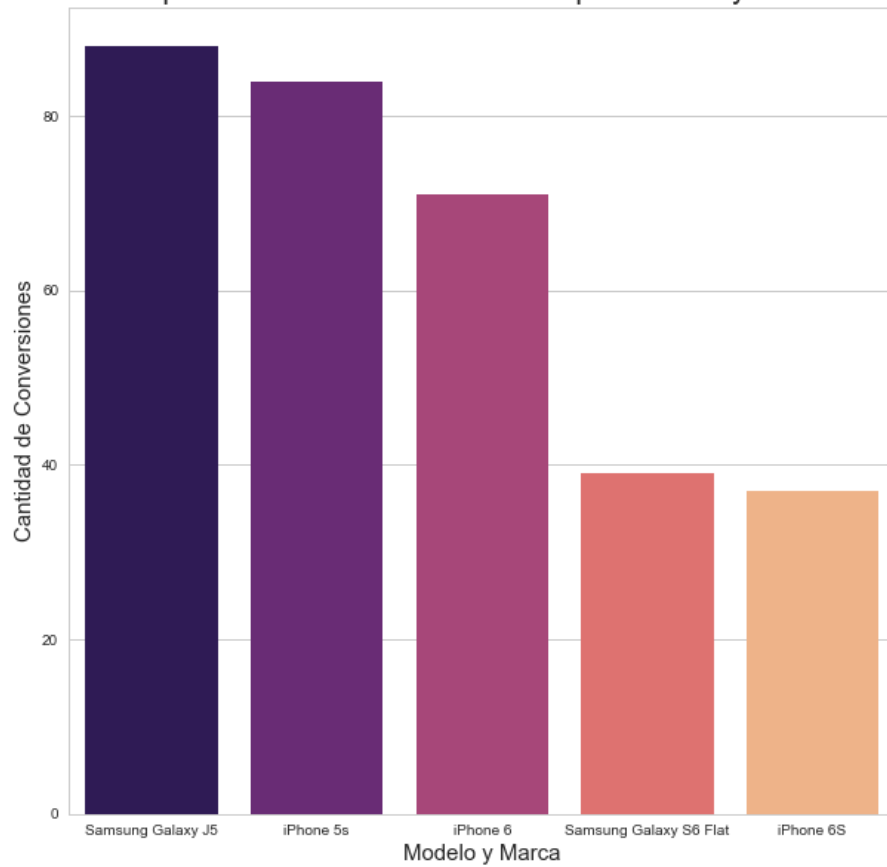
- **¿Cuáles son los eventos que más predominan en el set?**



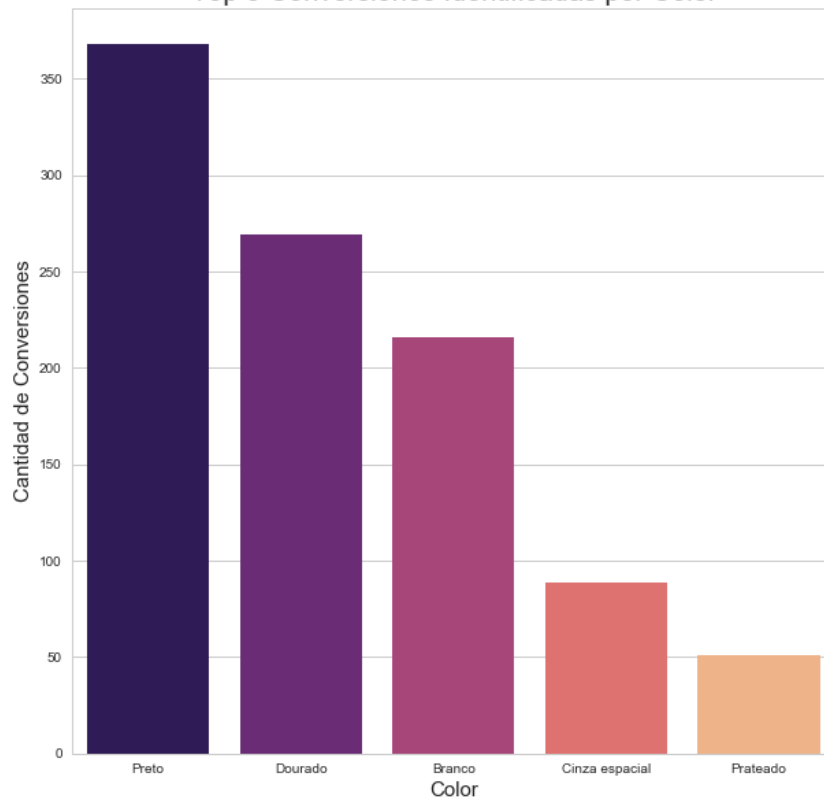
Como se puede observar en el gráfico de arriba, el evento que más predomina es el “viewed product”, que significa que los usuarios visitan la página de un determinado producto.

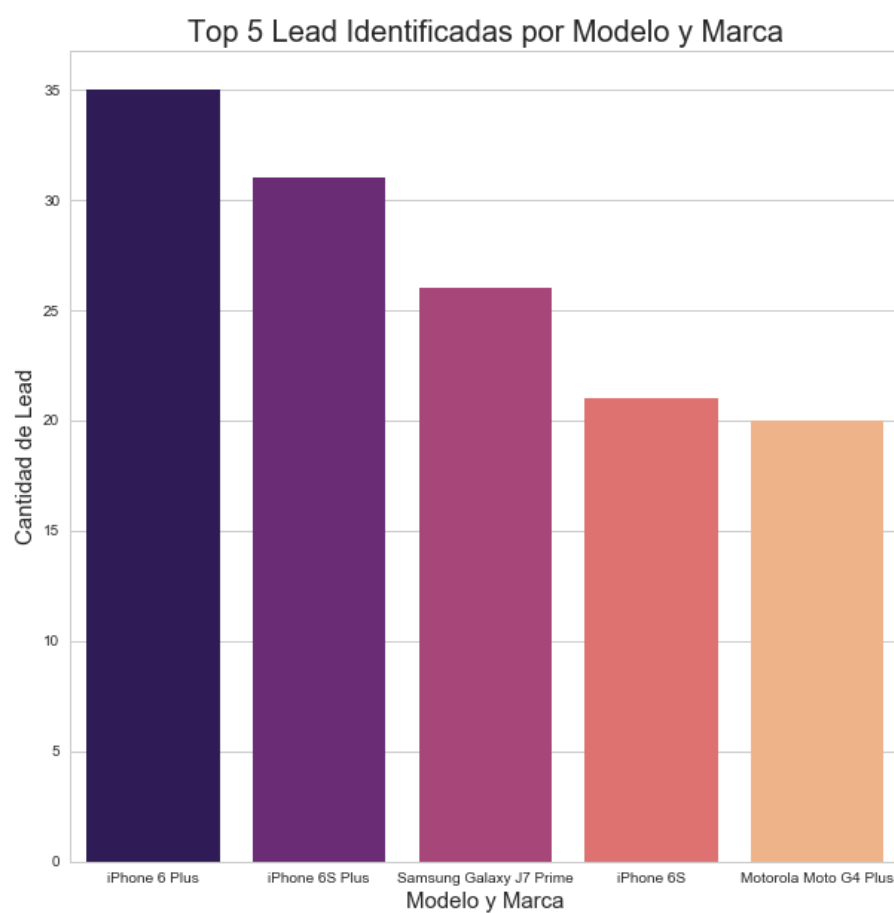
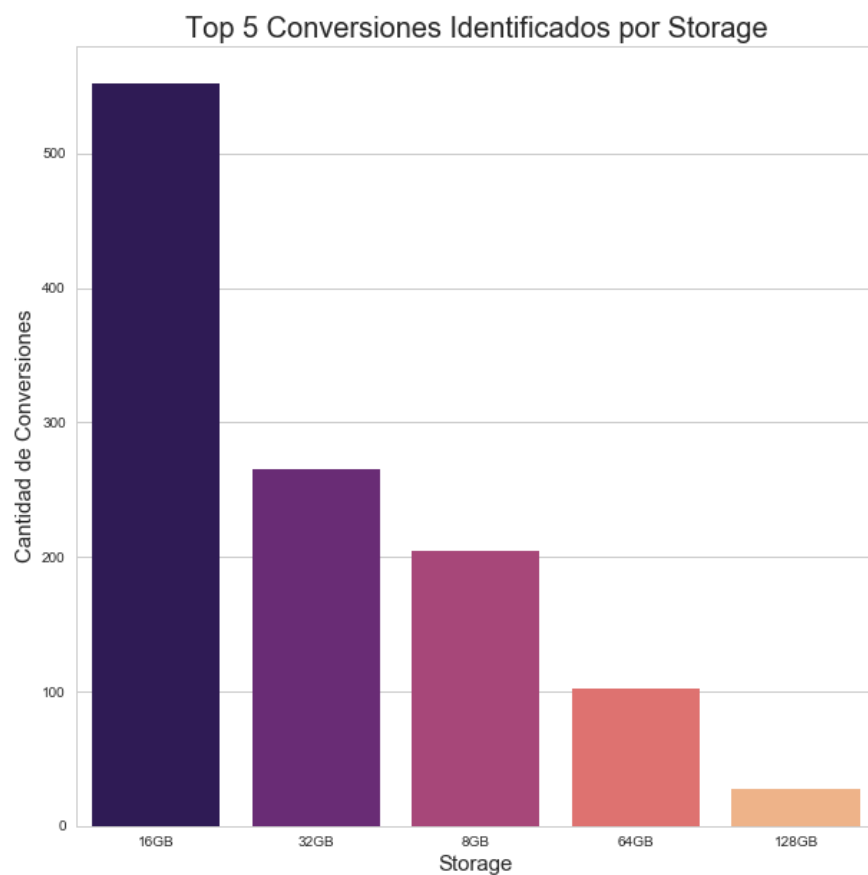
- **¿Cuál es el producto o los productos según su modelo, marca, capacidad y color que tuvieron más ventas(conversiones)?**

Top 5 Conversiones Identificadas por Modelo y Marca



Top 5 Conversiones Identificadas por Color

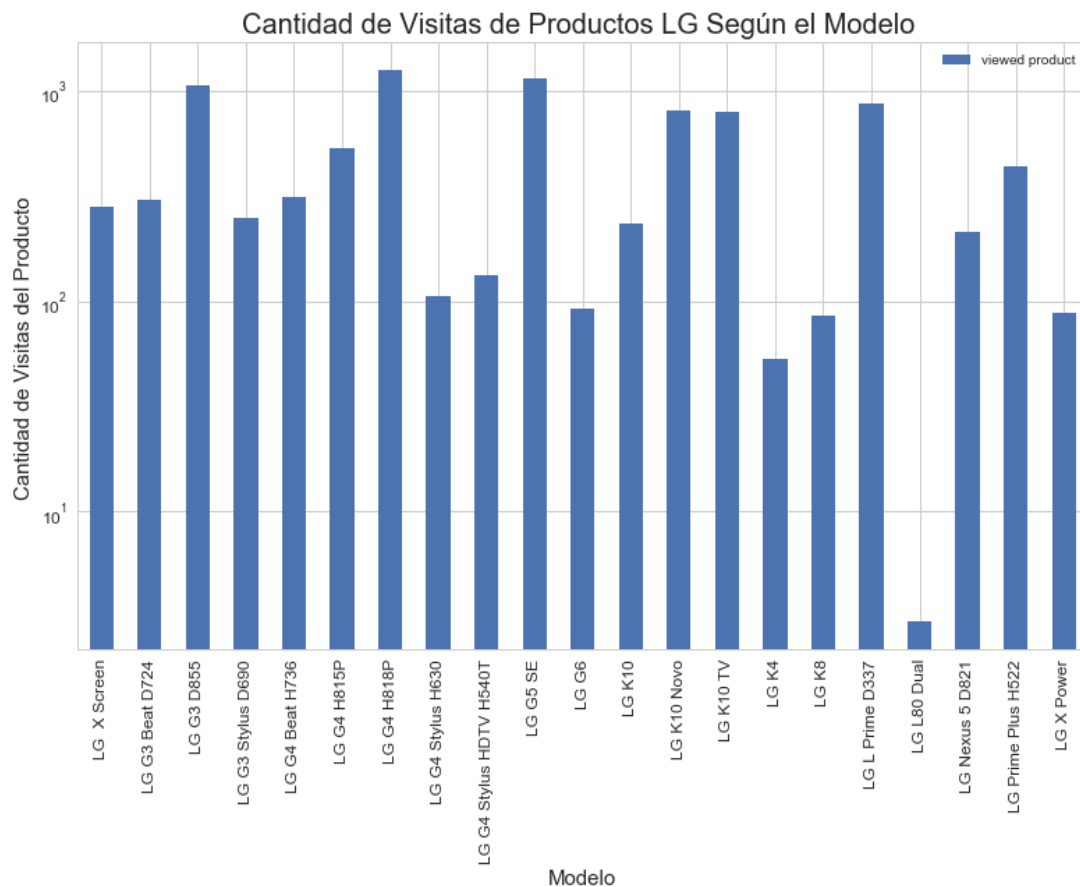




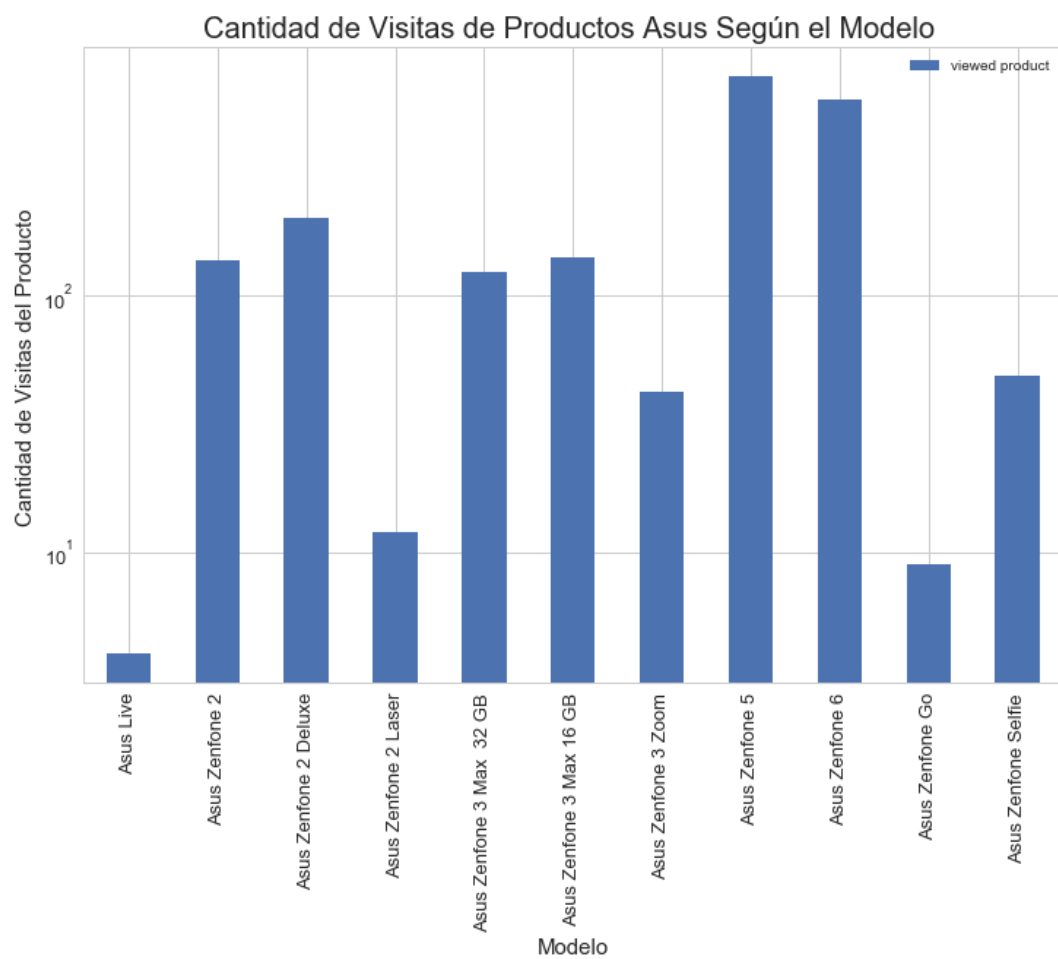
Efectivamente el producto más vendido es el Samsung galaxy J5 aunque le sigue el iphone5s, preferentemente de color negro y con un almacenamiento de 16gb. Esta información servirá para saber cual es el producto que tiene más demanda dentro del mercado de celulares para tenerlo en cuenta a la hora de reacondicionar ese tipo de producto y así aumentar las ventas del mismo.

A su vez, podemos ver que el modelo iphone 6 plus es uno de los productos que no se encuentra en stock por lo que los usuarios se registran para recibir una notificación.

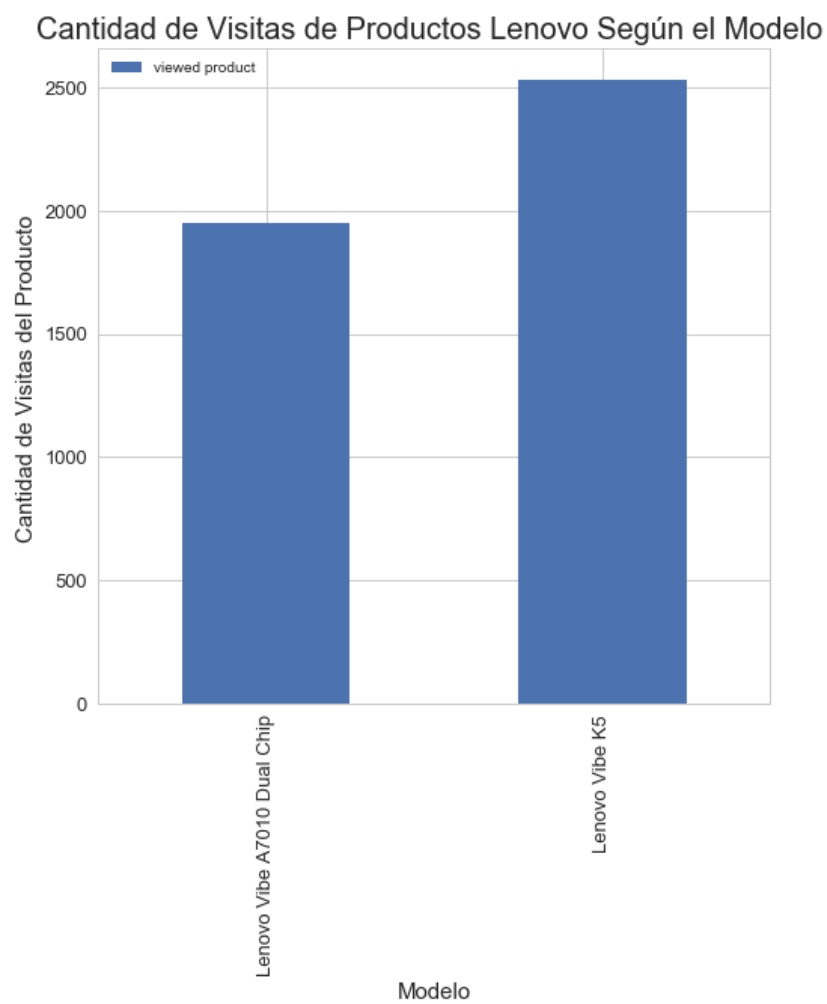
- **¿Cuáles son los productos que más se visitan según su modelo?**



En el caso de los productos Lg, los modelos más visitados son el G3, G4, G5.

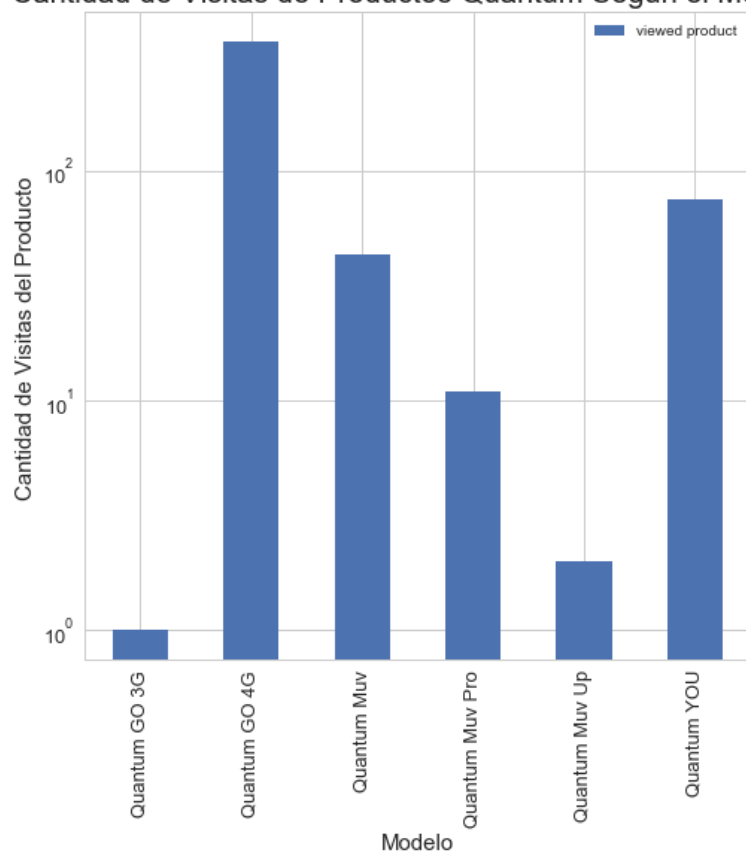


En el caso de asus, son el Zenfone 5 y el Zenfone 6.

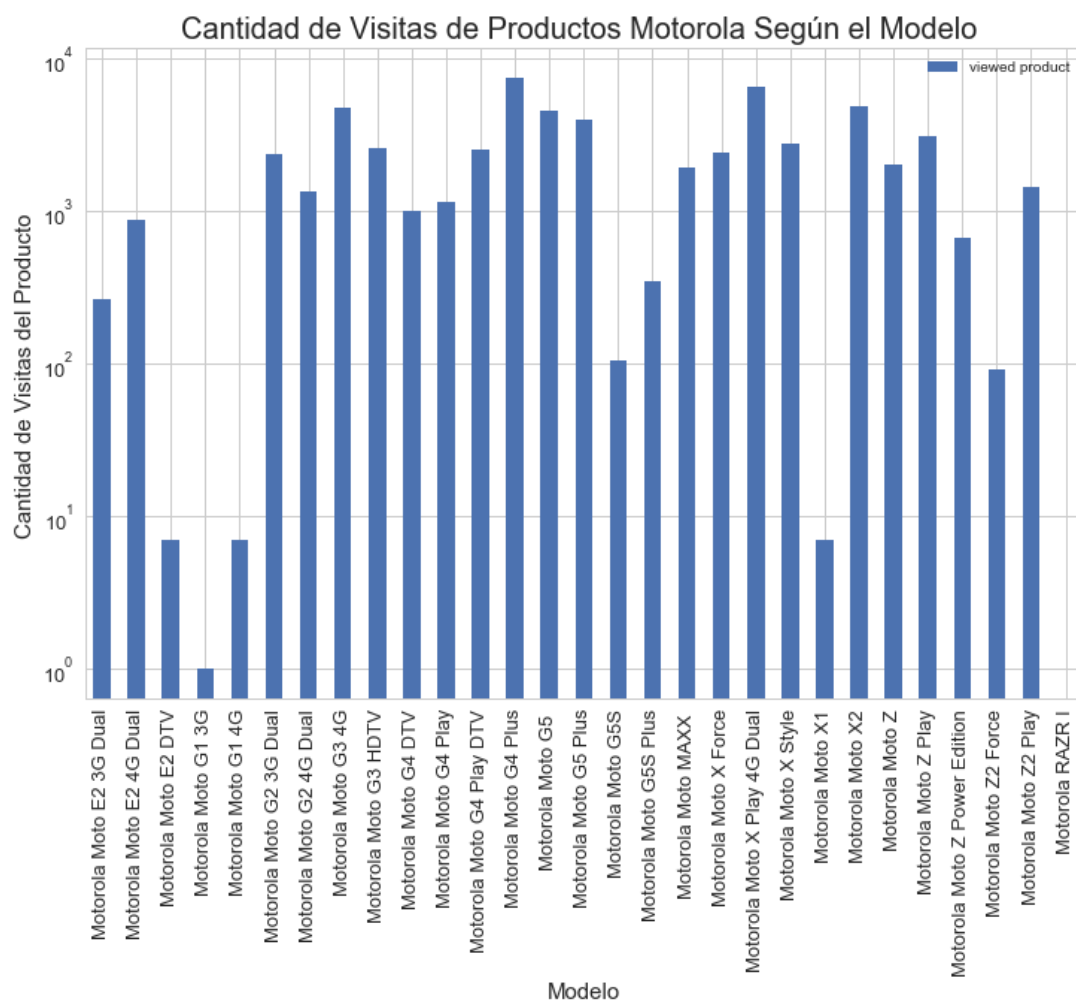


Los productos de Lenovo son el Vibe A7 y el K5.

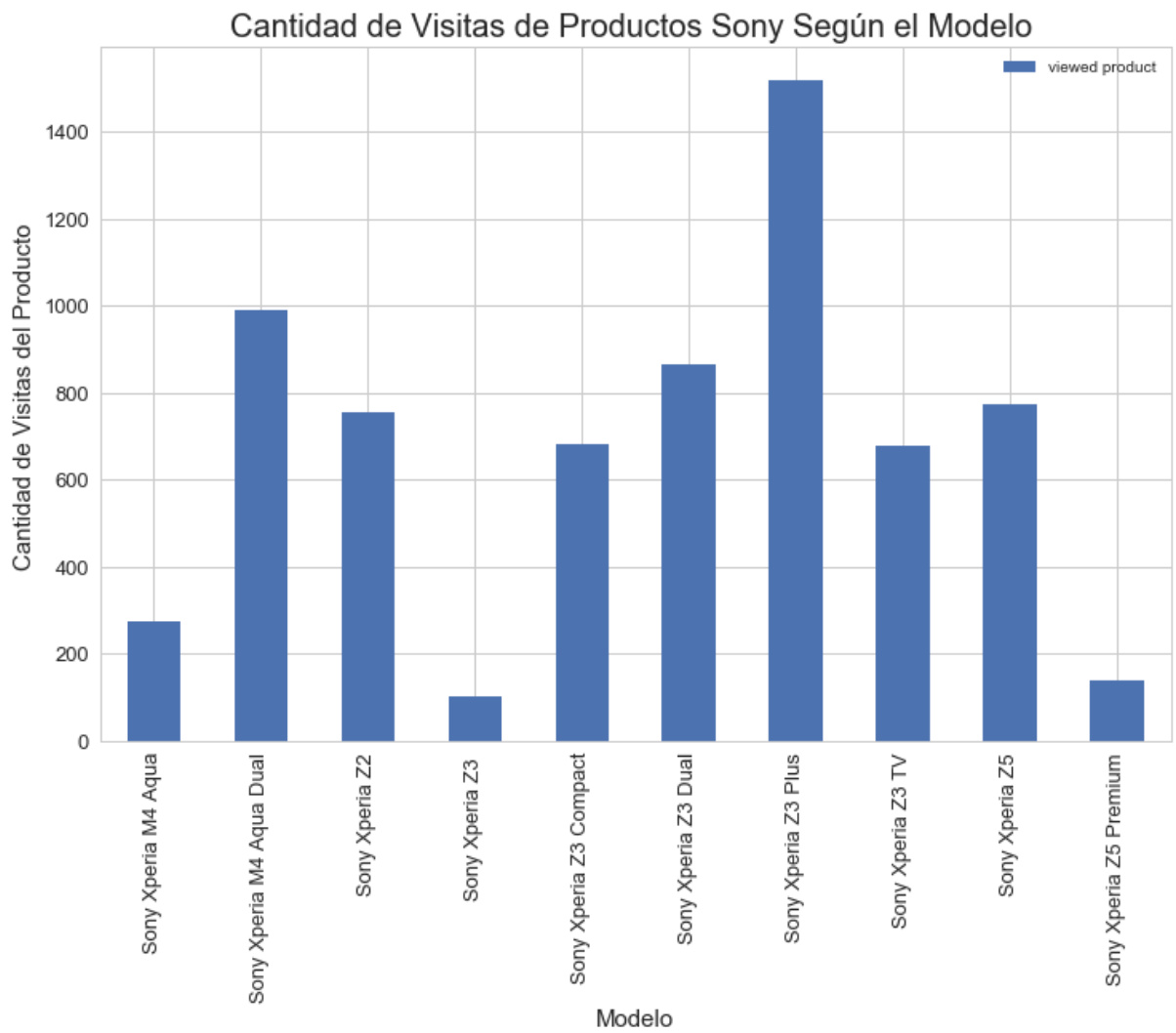
Cantidad de Visitas de Productos Quantum Según el Modelo



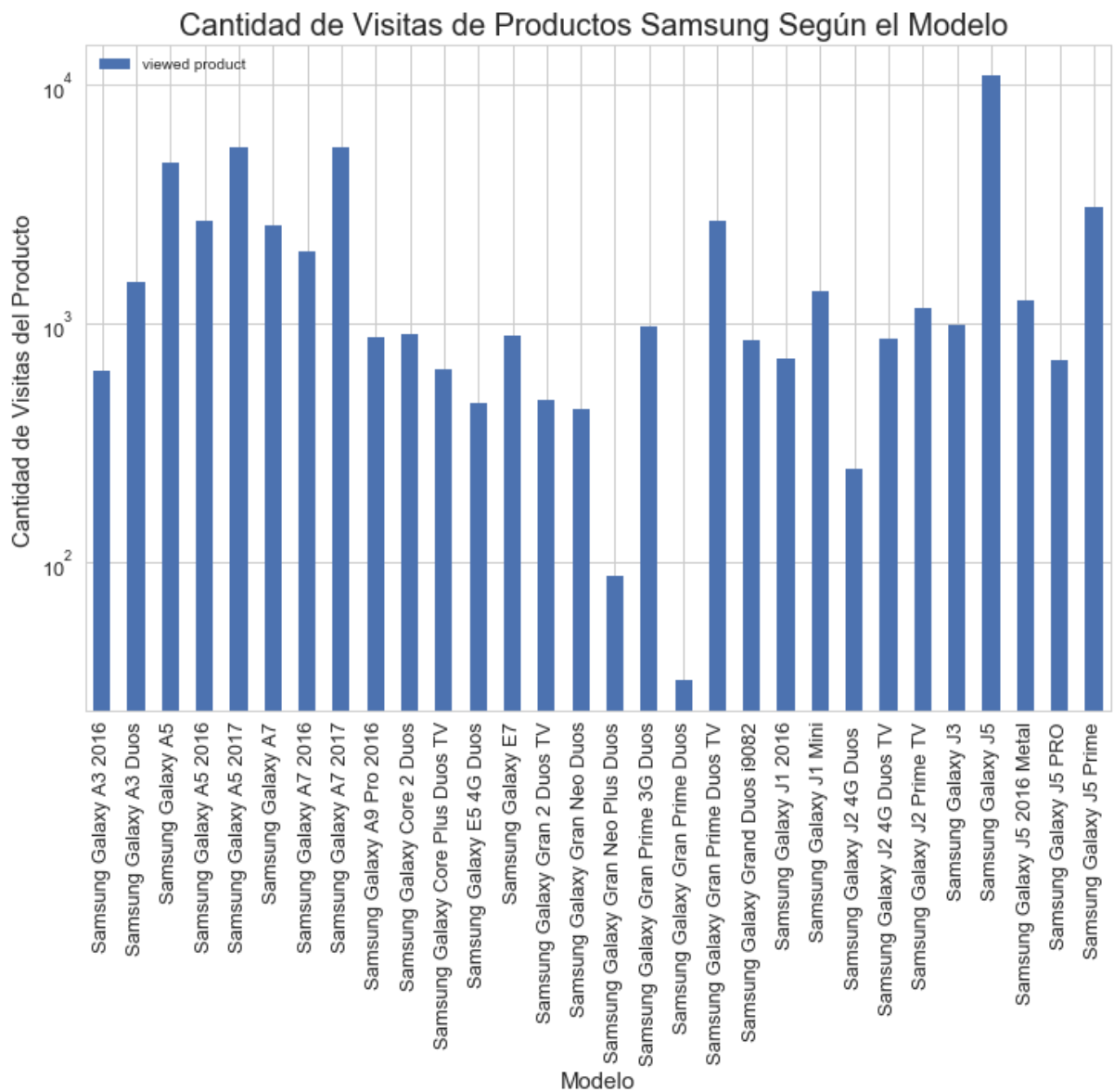
Para el caso de Quantum, como se puede observar en el gráfico el más visto es el Quantum GO 4G.



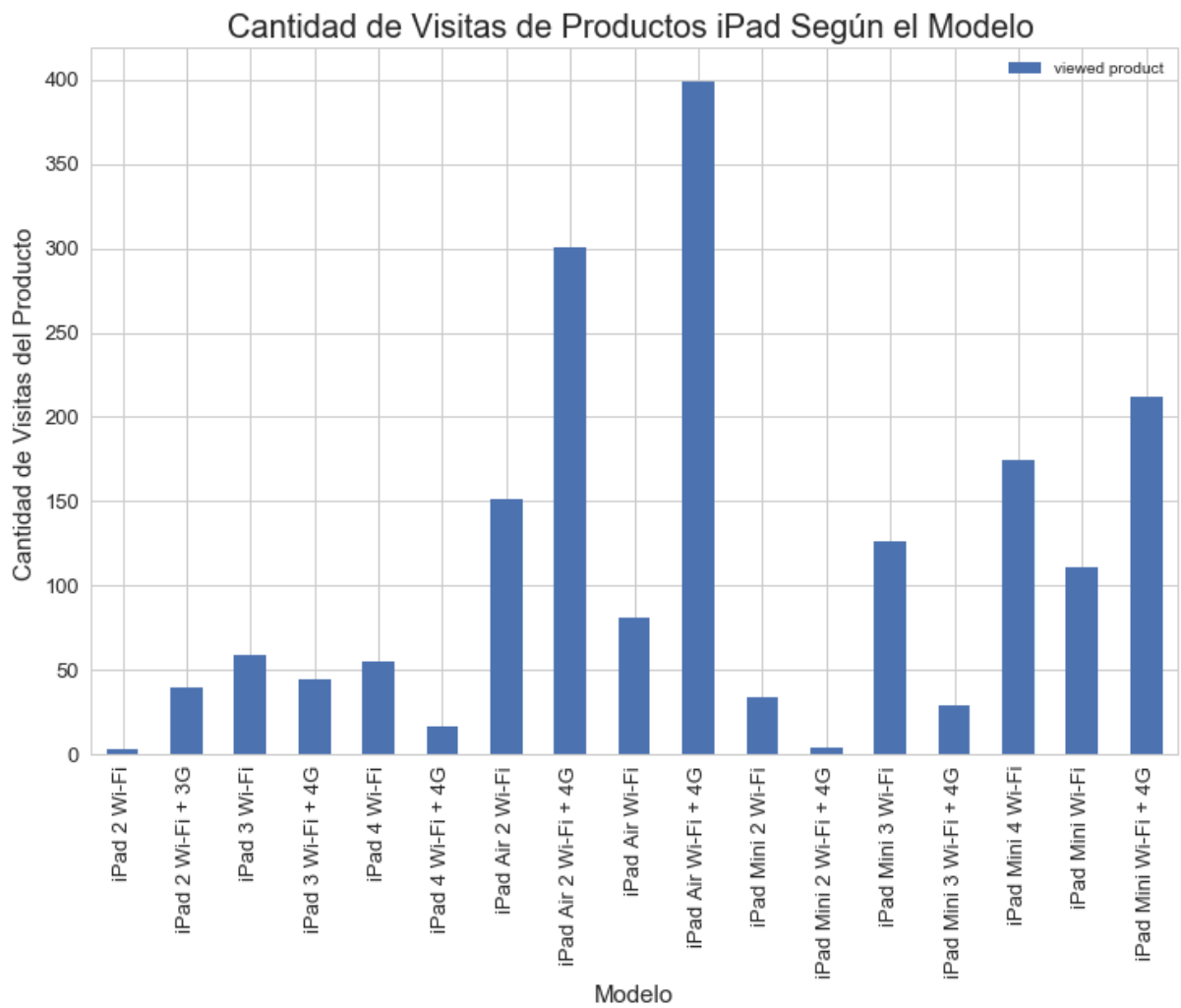
Para motorola, los modelos mas vistos son: el moto G4 play y el Moto X Play.



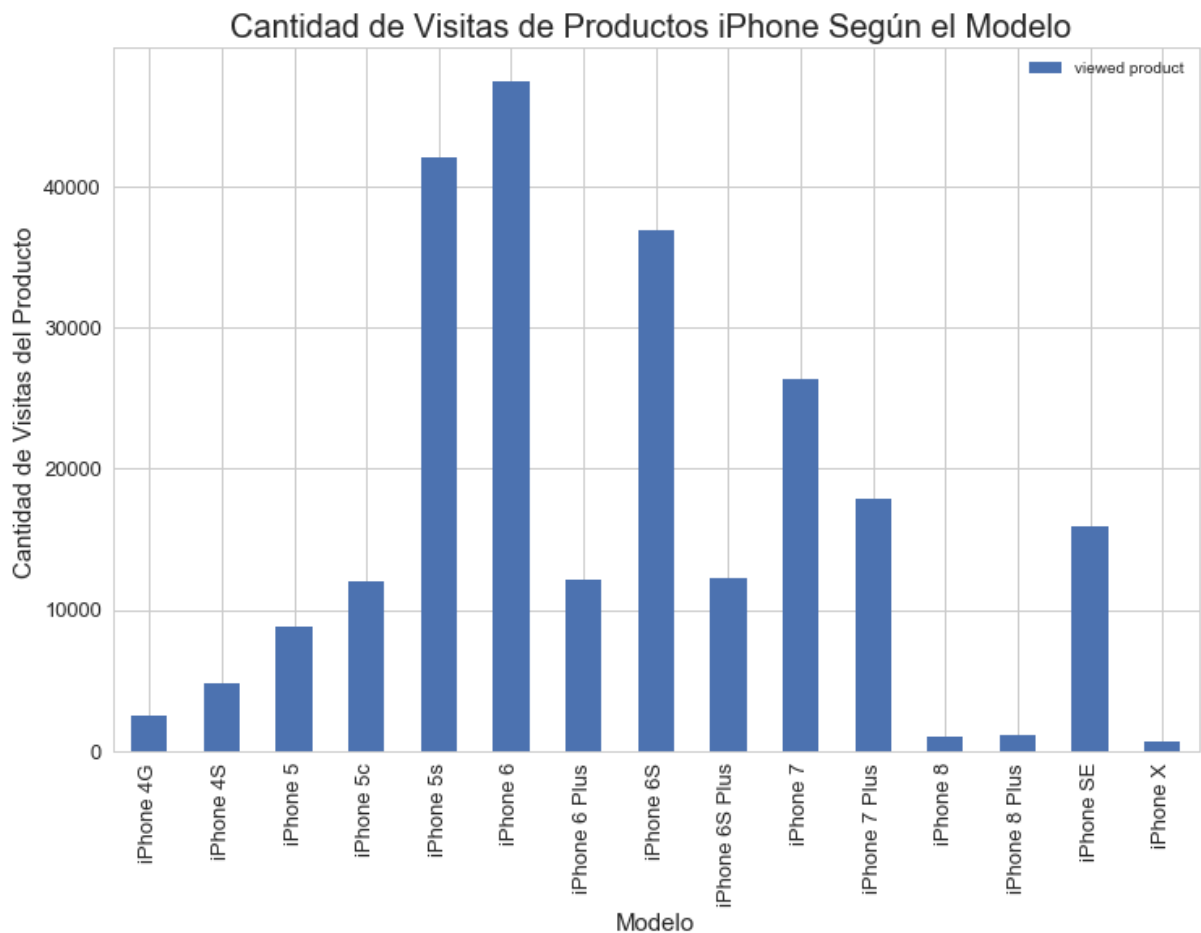
En el caso de Sony, el producto con más vistas es el Xperia Z3 Plus.



y para Samsung, el producto que más se ve es el Samsung Galaxy J5.



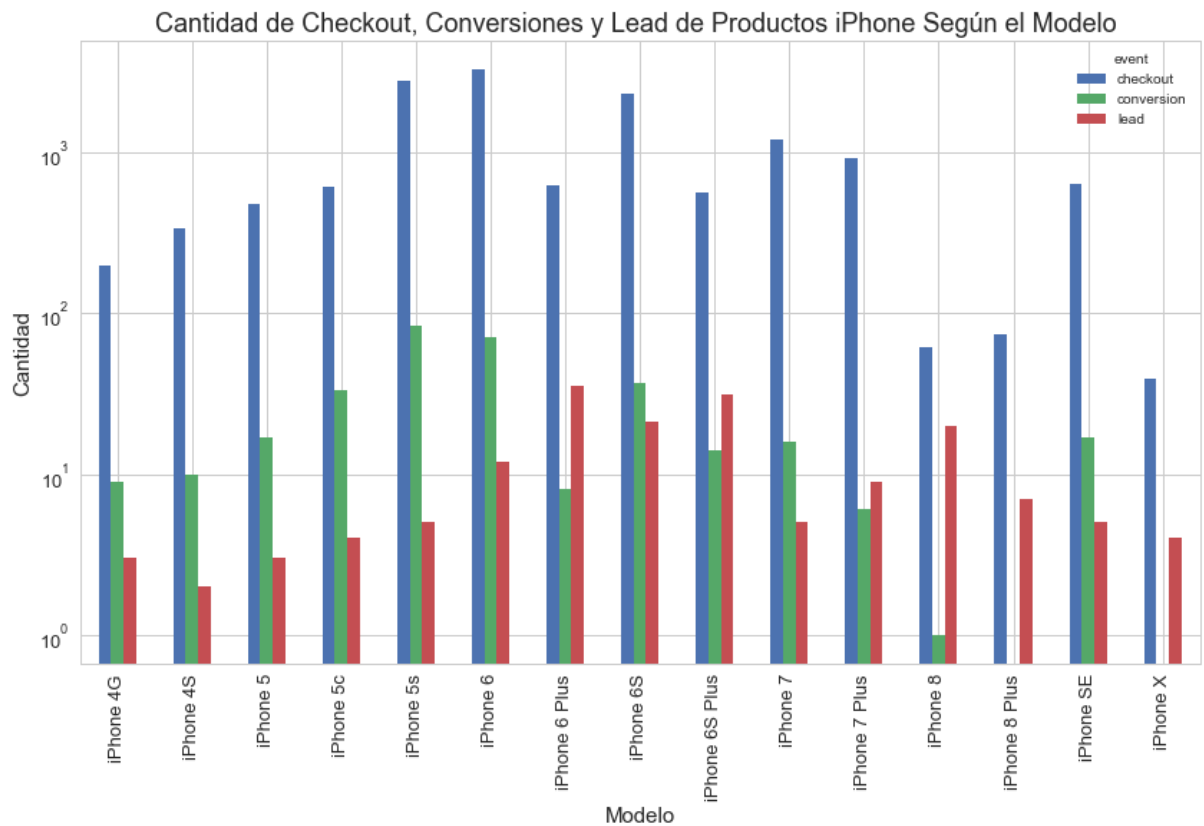
Para Ipad, el que más predomina en las visitas que el usuario busca es el Ipad Air Wifi, mientras que para Iphone es el Iphone6, como se muestra en el siguiente gráfico.



- **¿Cual es la cantidad de checkout, conversiones y lead que se realiza para una determinada marca?**

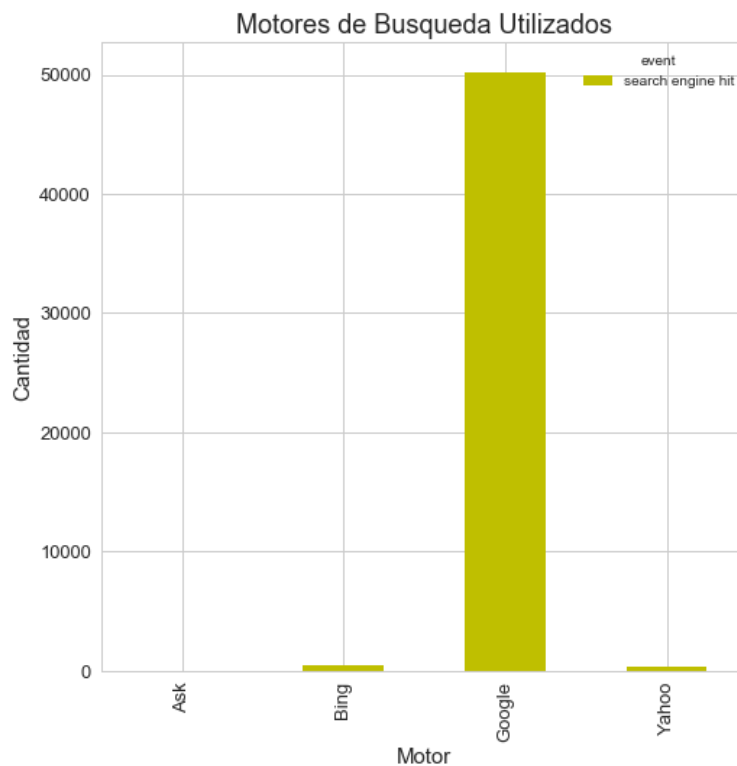
Hicimos el análisis para cada marca y modelo, pero solo mostraremos a modo de ejemplo el caso del iphone donde el iphone6 tuvo un mayor número de checkouts, pero el Iphone 5 tuvo, por poca diferencia, un mayor número de conversiones, El Iphone 6 plus fue el que, junto con el iphone 6s plus tuvieron mayor lead.

Con esto, no pudimos observar un patrón sobre los checkout, conversiones y lead por lo que solo incluimos un solo caso para mostrar que no se correlacionan estos eventos.

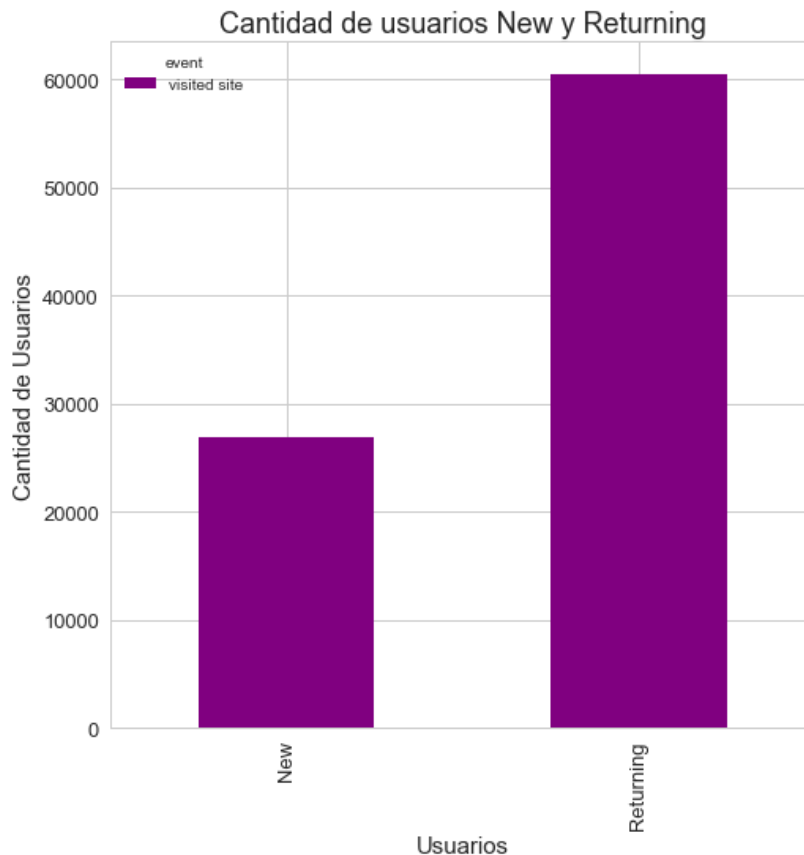


- **¿Cual es el motor de Búsqueda más utilizado en la plataforma?**

Como se puede apreciar en el siguiente gráfico, el más utilizado es Google y esto se debe a la velocidad y comodidad al utilizar este buscador.



- **¿Quienes visitan un sitio: un nuevo usuario o uno que previamente lo visitó?**

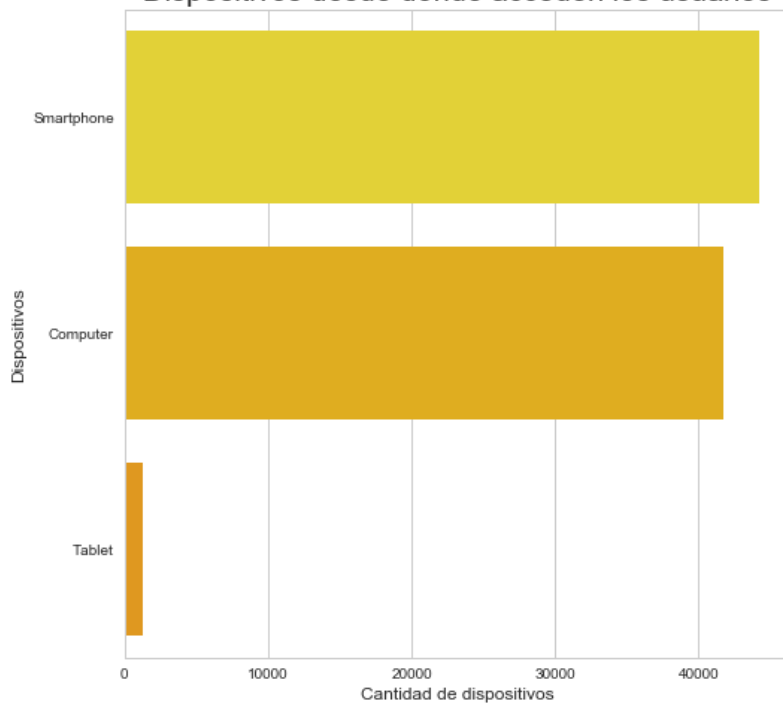


Casi el triple de los usuarios, son los que previamente visitaron el sitio. De igual manera, esto es muy interesante ya que son muchos los usuarios nuevos que visitan el sitio. Esto podría servirle al departamento de publicidad para atrapar a nuevos usuarios, ya sea ofreciendo un voucher o un descuento si invitas a un amigo, por ejemplo.

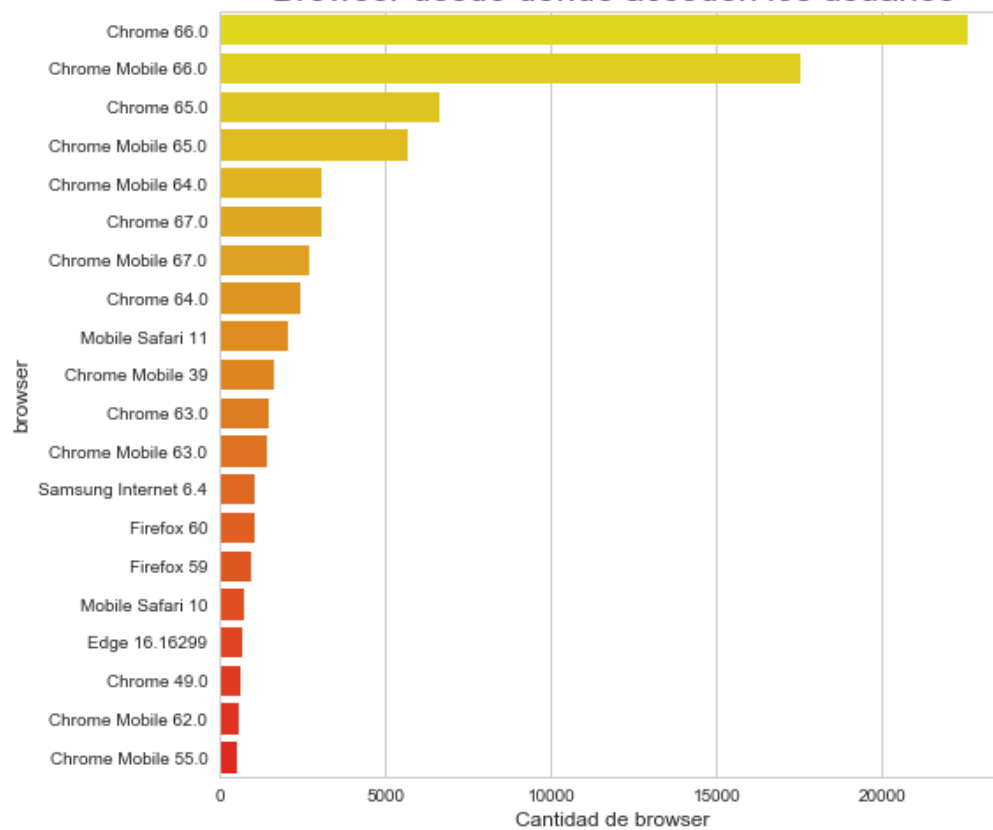
- **¿Desde dónde acceden los usuarios (dispositivos, sistema operativo, browser)?**

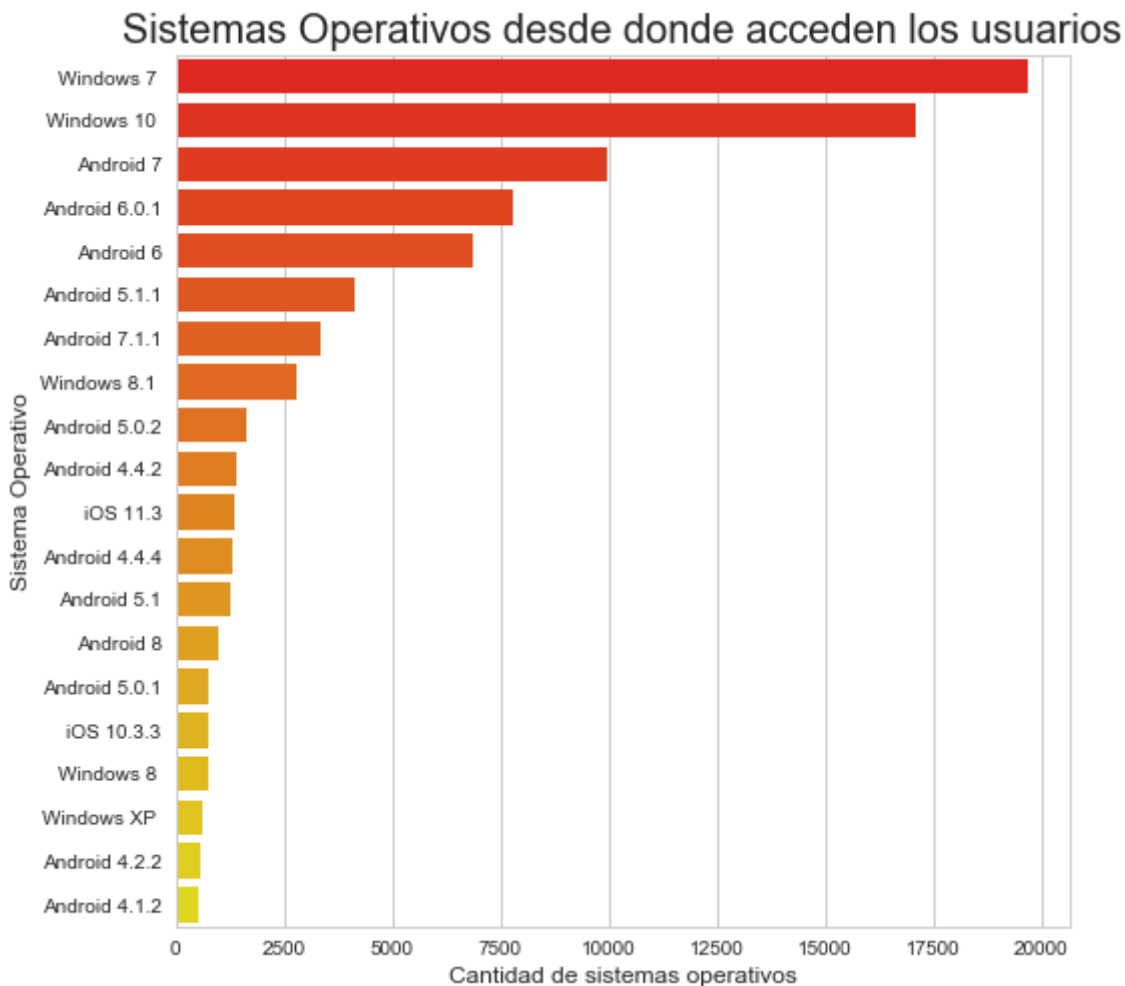
Los usuarios acceden, en su mayoría desde los celulares con android 7 usando chrome mobile y desde las computadoras con windows 10 o 7 usando Chrome como browser.

Dispositivos desde donde acceden los usuarios



Browser desde donde acceden los usuarios

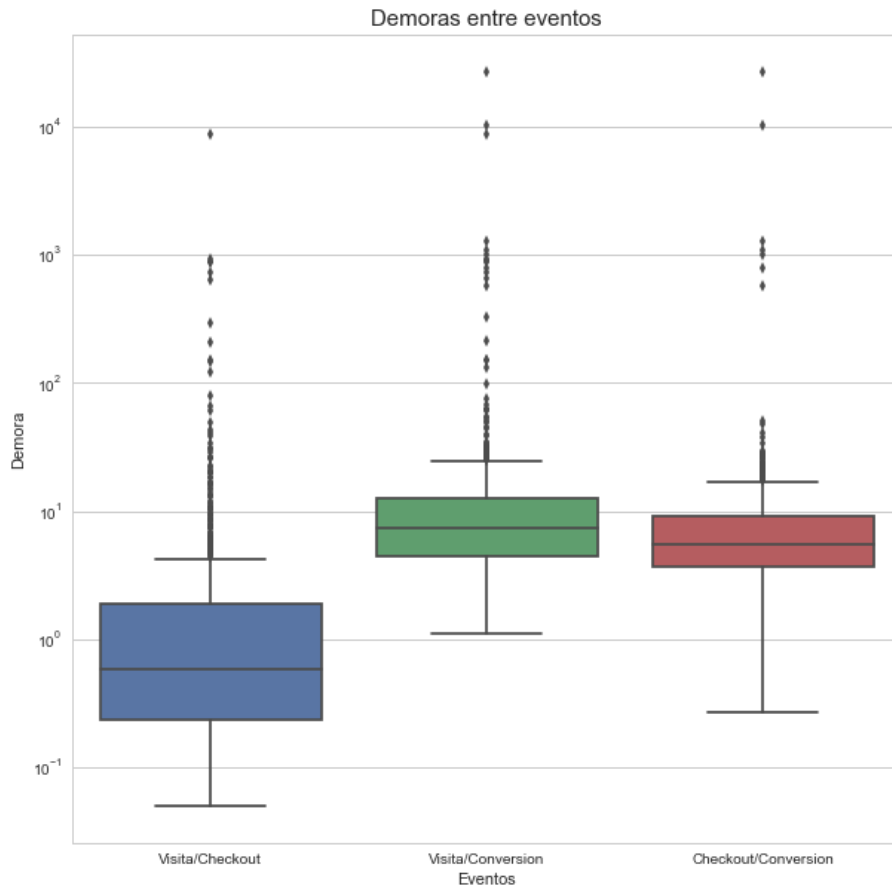




● **Análisis De Correlación de Demoras:**

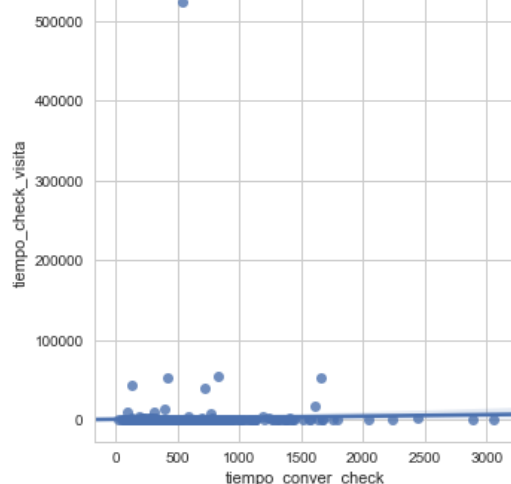
¿Para los usuarios que realizaron una compra de un producto, existe alguna correlación entre lo que tarda un usuario desde que ve un producto, realiza el checkout y finalmente la compra?

Para esto primeramente lo que realizamos fue tomar todos aquellos usuarios compradores que para una cierta compra tengan completos los eventos de Visita, Checkout y Conversión. Luego para una misma Sku (identificador del producto) los colocamos en un mismo Data Frame. Nos quedamos con los últimos eventos de Visita y Checkout para cada usuario para cada compra de un determinado producto. Obtuvimos el siguiente gráfico de demoras entre eventos(en segundos).



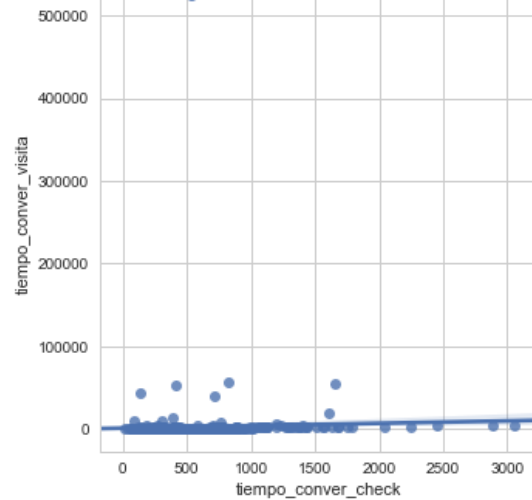
Como podemos observar hay una persona que entre el checkout y la conversión tarda unos 100000 segundos = 70 días. Es lógico que una persona se tome su tiempo desde la última visita hasta efectuar un checkout o una compra pero entre un checkout y una compra no. por lo que decidimos establecer una demora máxima entre checkout y compra de 6 horas (nos pareció un rango más que considerable para esta demora). Luego pasamos a analizar las correlaciones entre las demoras.

Correlación lineal entre: Tiempo de Visita-Checkout y Tiempo de Checkout-Conversión



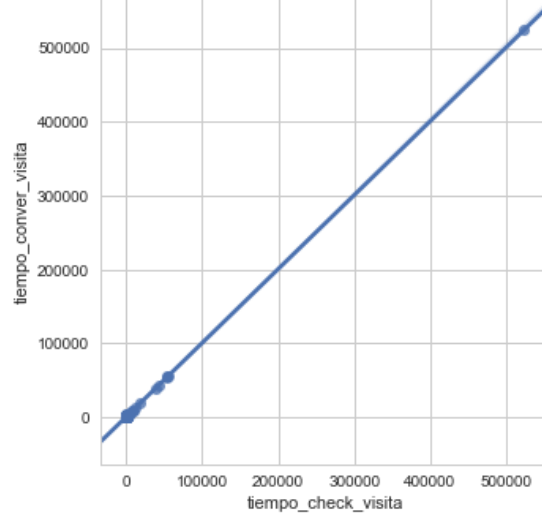
Grado de correlación (método Pearson) $R = 0.03319997627278722$

Correlación lineal entre: Tiempo de Visita-Conversión y Tiempo de Checkout-Conversión



Grado de correlación (método Pearson) $R = 0.05043429858159153$

Correlación lineal entre: Tiempo de Visita-Checkout y Tiempo de Visita-Conversión

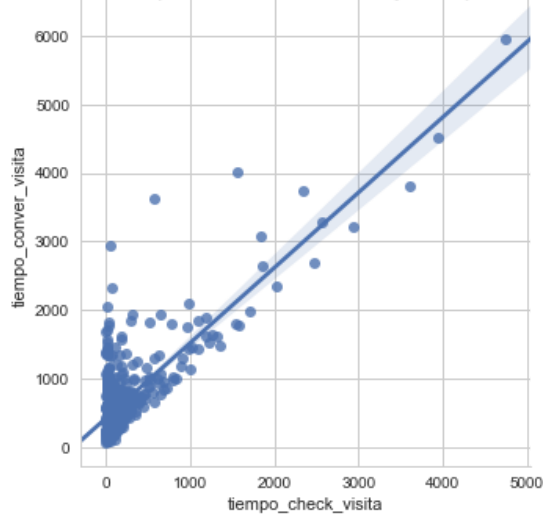


Grado de correlación (método Pearson) $R = 0.9998512288959148$

Teniendo en cuenta que el grado de correlación es un indicador de cuánto depende una variable de otra y que va de -1 a 1, para este último caso vemos una correlación bastante directa.

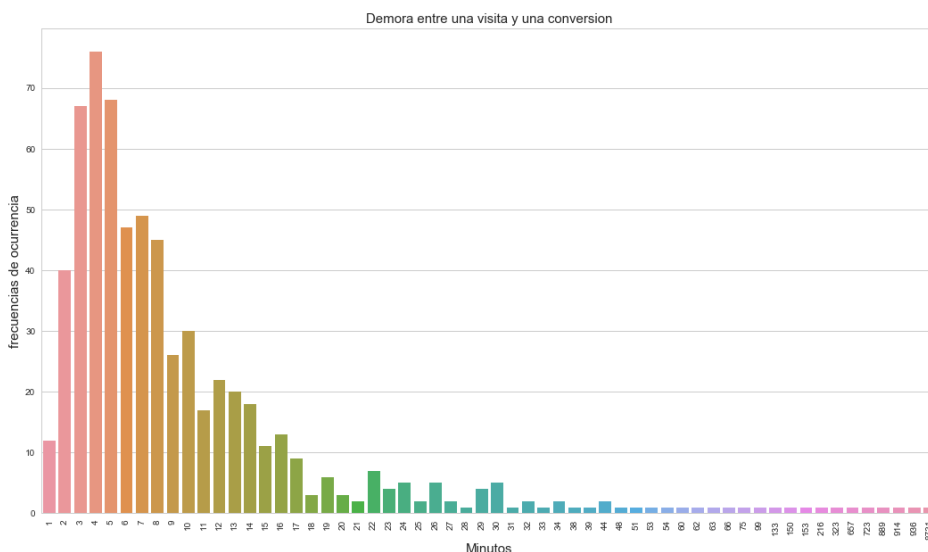
Si hacemos un acercamiento en los tiempos más bajos donde hay más acumulación observamos:

Correlación lineal entre: Tiempo de Visita-Checkout y Tiempo de Visita-Convesion



Conclusión:

Se puede apreciar la para los compradores de un cierta producto que el tiempo que tardan entre comprar el producto desde que lo vieron por última vez es muy similar al tiempo que demoran entre que lo vieron por última vez y efectuaron el checkout. Y que todo esto corresponde mayormente para tiempos bajos en una compra como podemos apreciar en el siguiente gráfico:



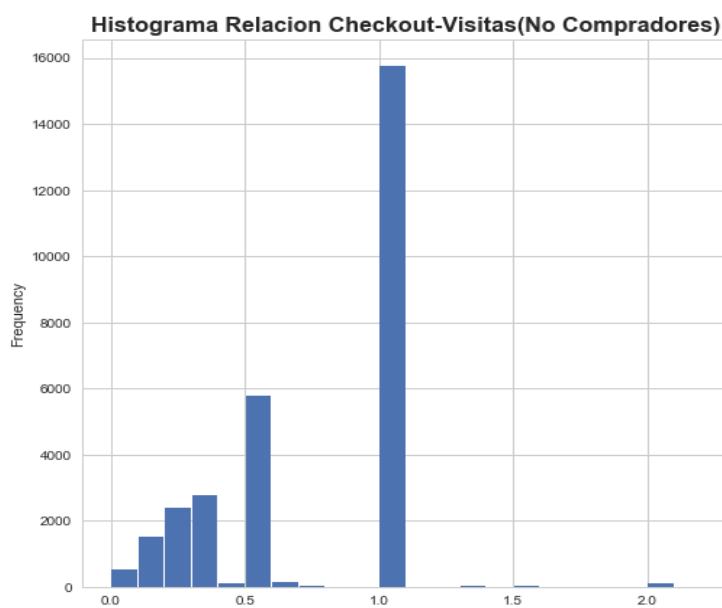
En este gráfico se puede apreciar que los mayores tiempos entre una última visita a un producto y su posterior compra se dan entre los 3 y 5 minutos.

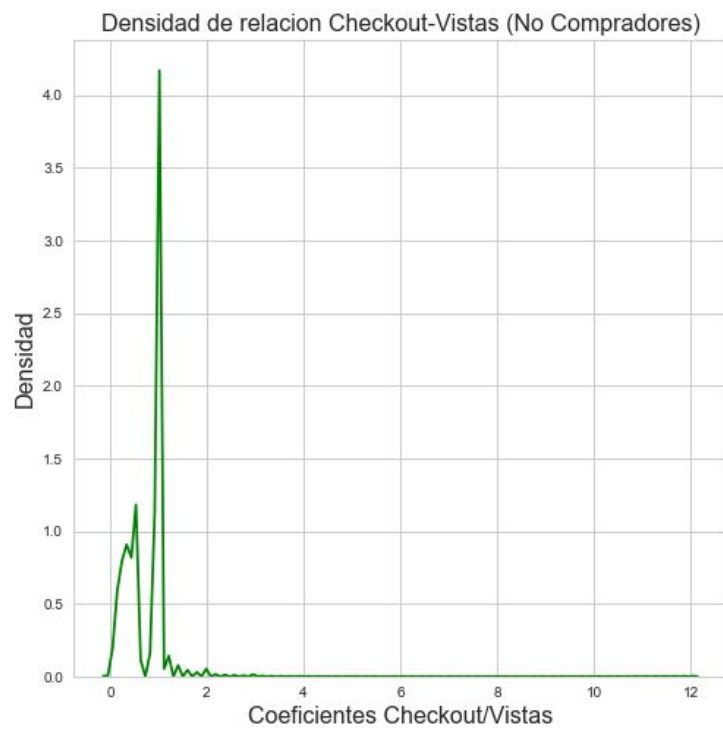
¿Para los usuarios que realizaron una compra de un producto, y los que no realizaron la compra existe alguna característica que podamos encontrar en relación con las demoras de los eventos que realizan?

Juntamos por un lado los usuarios que no realizaron una compra y calculamos, para cada uno de ellos, un coeficiente entre la cantidad de Visitas a un determinado producto (Sku) y la cantidad de checkout que realizaron sobre dicho producto.

Coeficiente = número de Checkout / número de Visitas

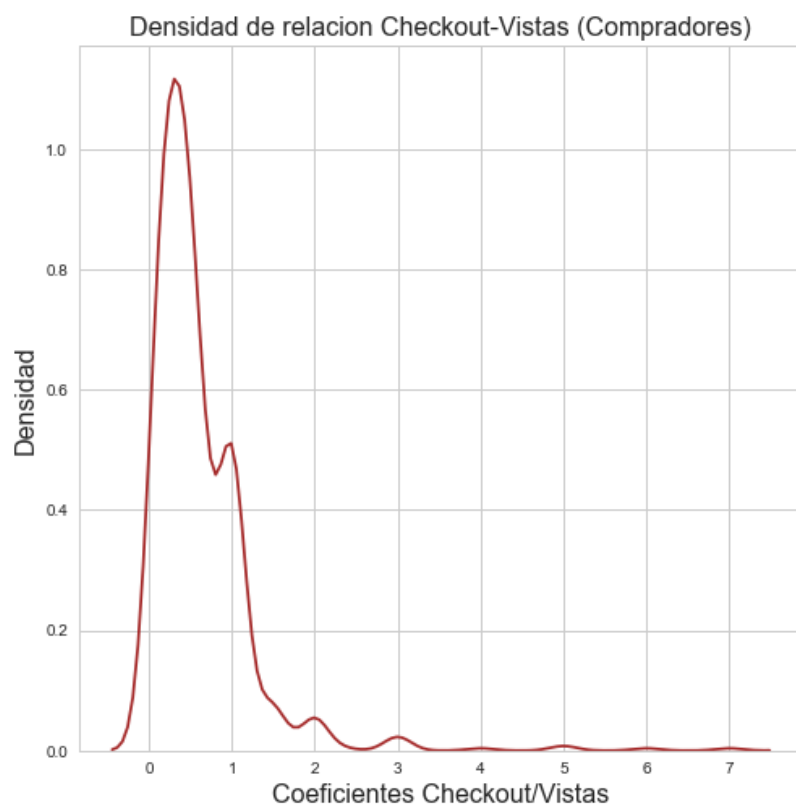
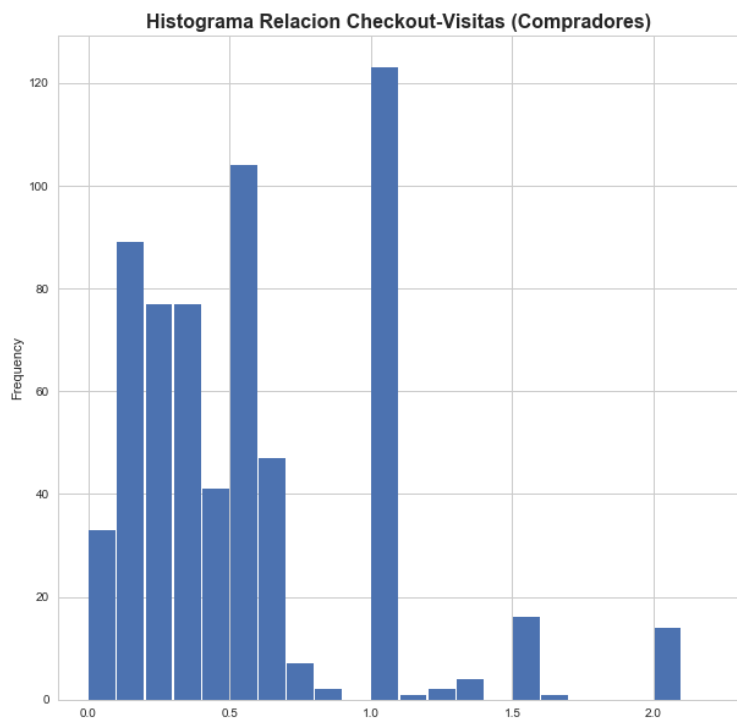
En base a esto se obtuvo:





Media de Coeficiente ($\text{num_Checkout}/\text{num_Visitas}$) = 0.7239155623707983

Se calculó el mismo coeficiente para los compradores y se obtuvieron los siguientes resultados:



Media de Coeficiente(num_Checkout/num_Visitas) = 0.59330408335437

Conclusión:

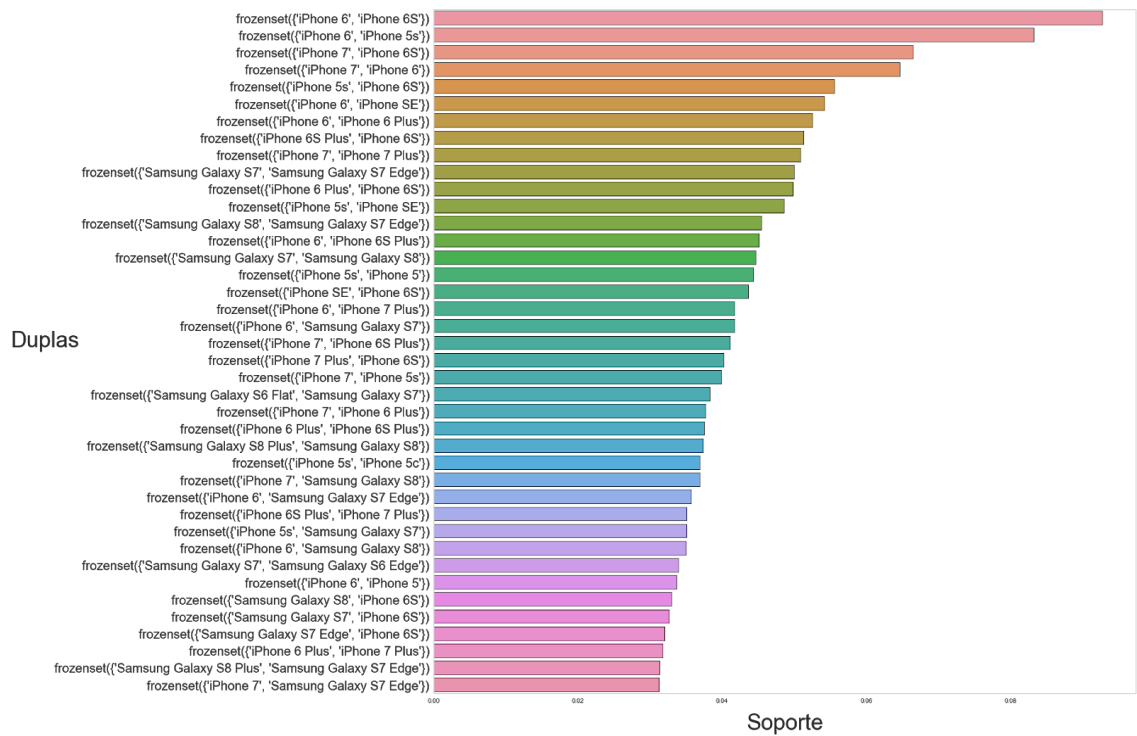
Se puede apreciar la para los compradores el número de visitas suele ser mayor que el número de checkout dando por resultado una menor media, más cercana al 0.5 (osea un promedio de un checkout por cada dos visitas). Mientras que para el caso el los no compradores se obtuvo una media más cercana al 1 (osea un promedio de un checkout por cada visitas)

- **¿Cuales son las duplas de productos con mayor ocurrencia (ya sea, visita, checkout o compra) entre los usuarios?**

Para encontrar las duplas con mayor soporte utilizamos el algoritmo Apriori que es un algoritmo popular para extraer conjuntos de elementos frecuentes con aplicaciones en aprendizaje de reglas de asociación. El algoritmo apriori ha sido diseñado para operar en bases de datos que contienen transacciones, como las compras realizadas por los clientes de una tienda. Un elemento se considera como "frecuente" si cumple un umbral de soporte especificado por el usuario. Por ejemplo, si el umbral de soporte se establece en 0.5 (50%), un conjunto de elementos frecuente se define como un conjunto de elementos que ocurren juntos en al menos el 50% de todas las transacciones en la base de datos.

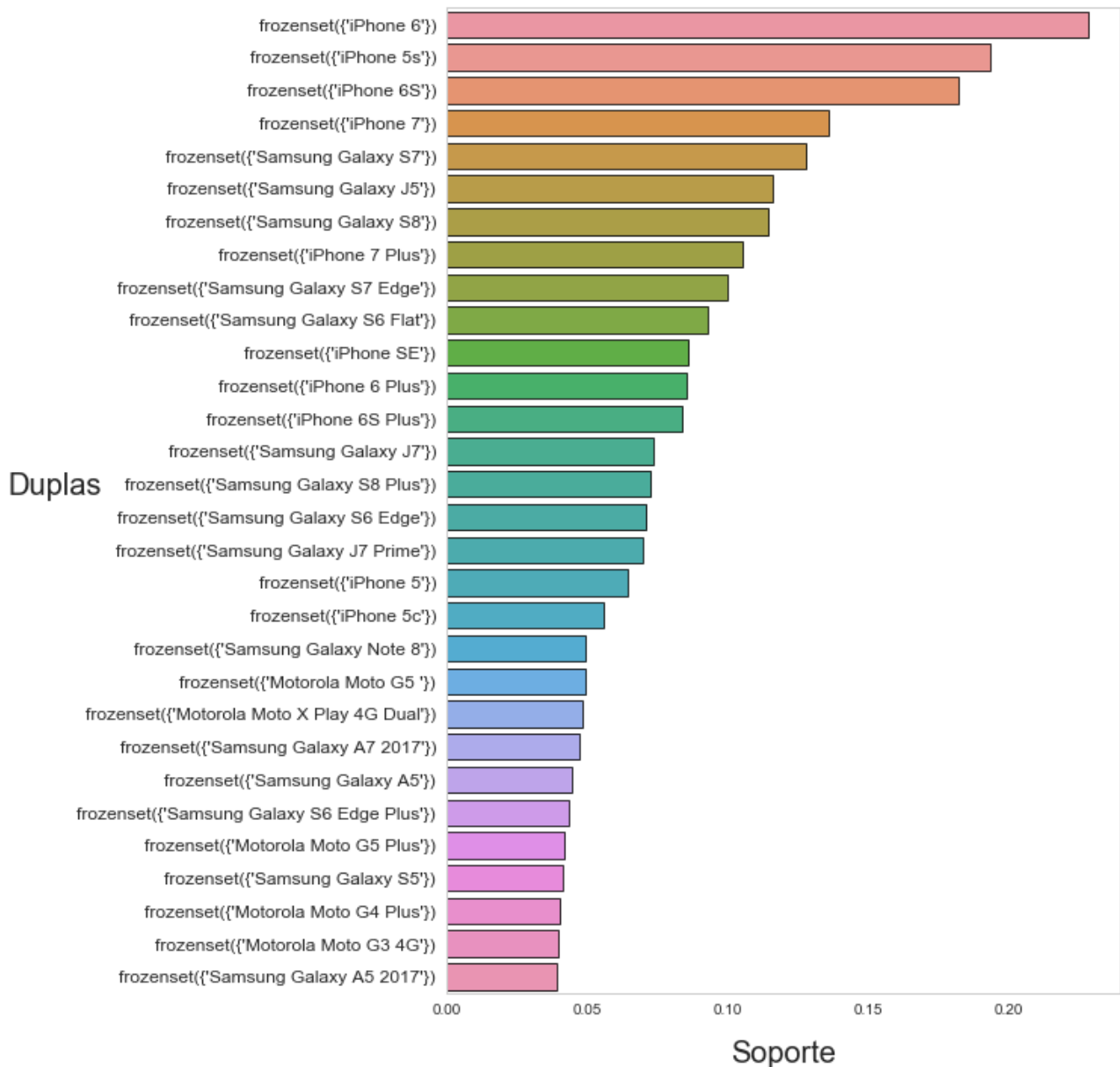
Se obtuvo para las mejores 40 duplas los siguientes resultados:

Duplas de Modelos con mayor soporte (Con más ocurrencia por los Usuarios)



También aplicando el algoritmo Apriori para obtener los 30 productos con más soporte por parte de los usuario vemos que:

Duplas de Modelos con mayor soporte (Con más ocurrencia por los Usuarios)

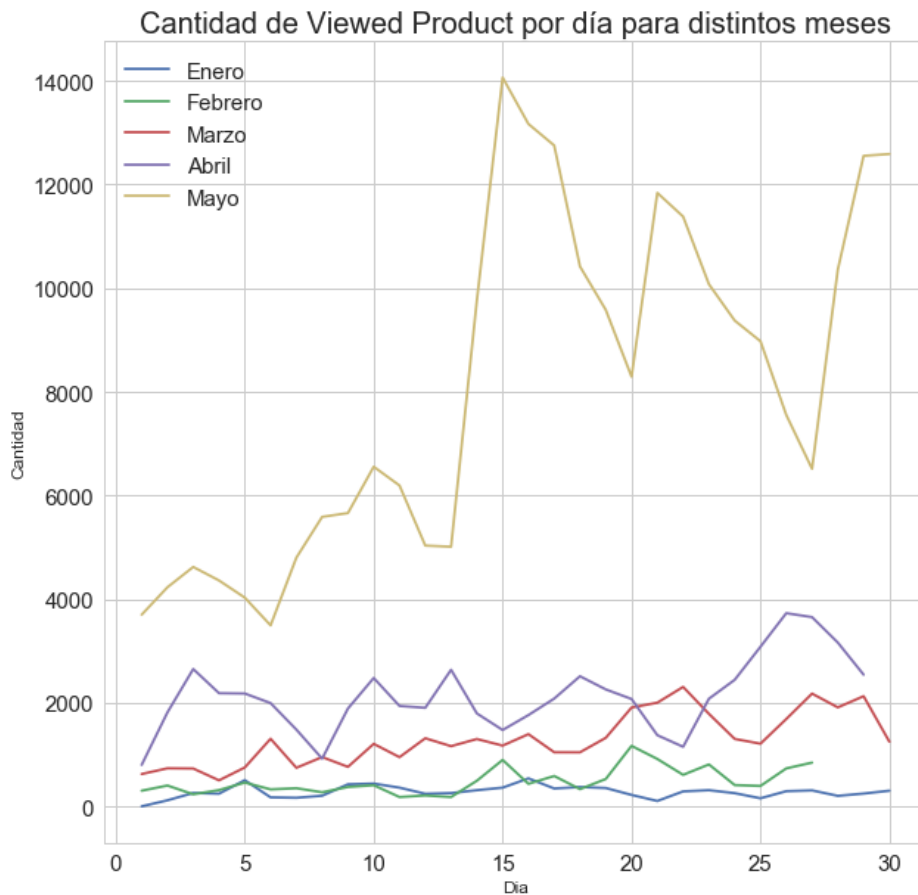


Conclusión:

Podemos observar que en el mayor de los casos ocurre que el que las duplas corresponden a iPhone con sus distintos modelos y que para las duplas con mayor soporte suele ser de la misma marca cosa que vemos también en Samsung.

- ¿Cómo evolucionan las visitas durante el periodo (Enero hasta mayo) con el que se dispone de los datos completos de los meses?

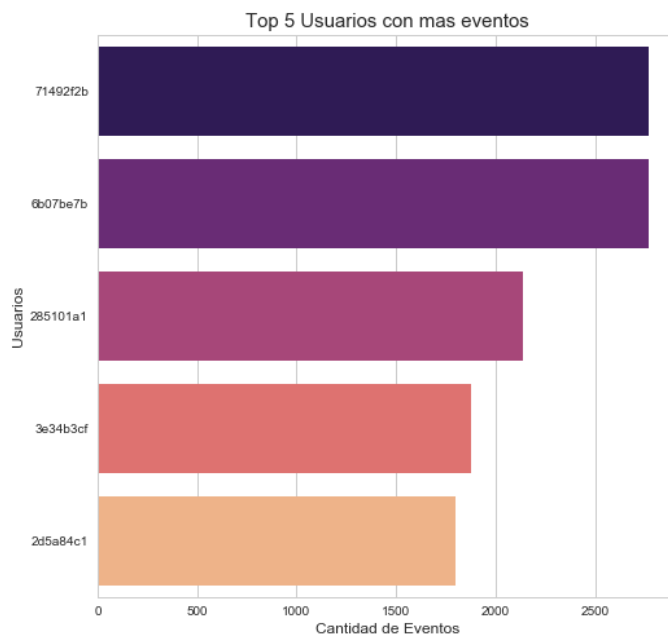
Aclaración: se descarta Junio por no contar con los datos de todo el mes.



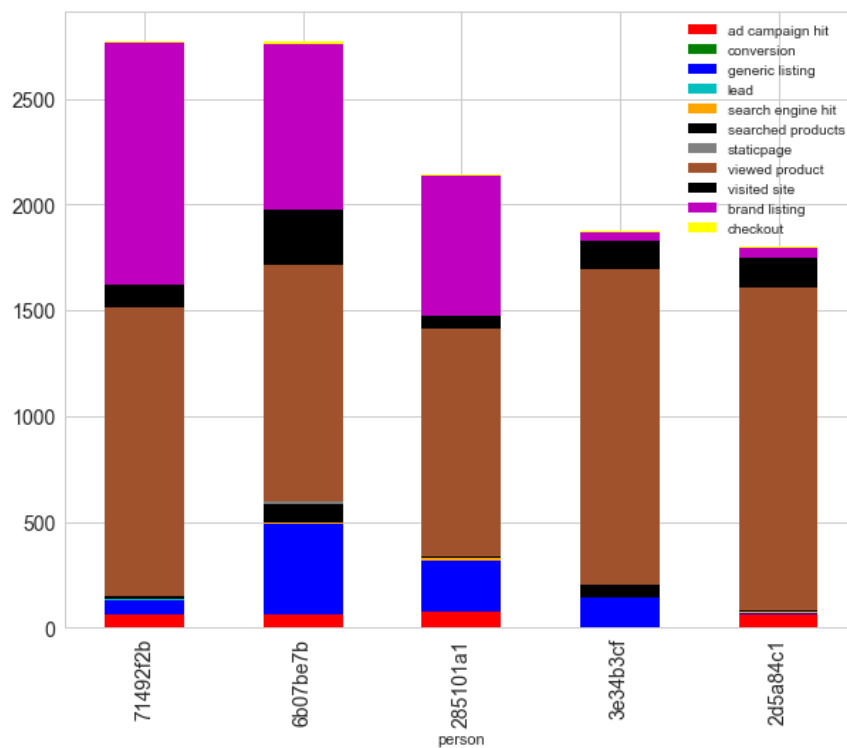
Conclusión:

Se observa un marcado aumento de las visitas por parte de los usuarios.

- ¿Cuales son los 5 usuarios que mayor cantidad de eventos realizan?



¿Cuáles son esos eventos?



Vemos que hay usuarios que realizaron más de 2500 eventos entre enero y mitad de junio

Como podemos ver el país con mayor ingreso de usuarios al sitio son de Brasil, que tiene un total de 84308 ingresos, una cifra realmente muy alta comparada a los Estados Unidos y Argentina



Como era de esperarse las ciudades que tienen la mayor cantidad de usuarios son pertenecientes a Brasil, y la que resalta es la de Sao Paulo que contiene un total de 11711 usuarios.

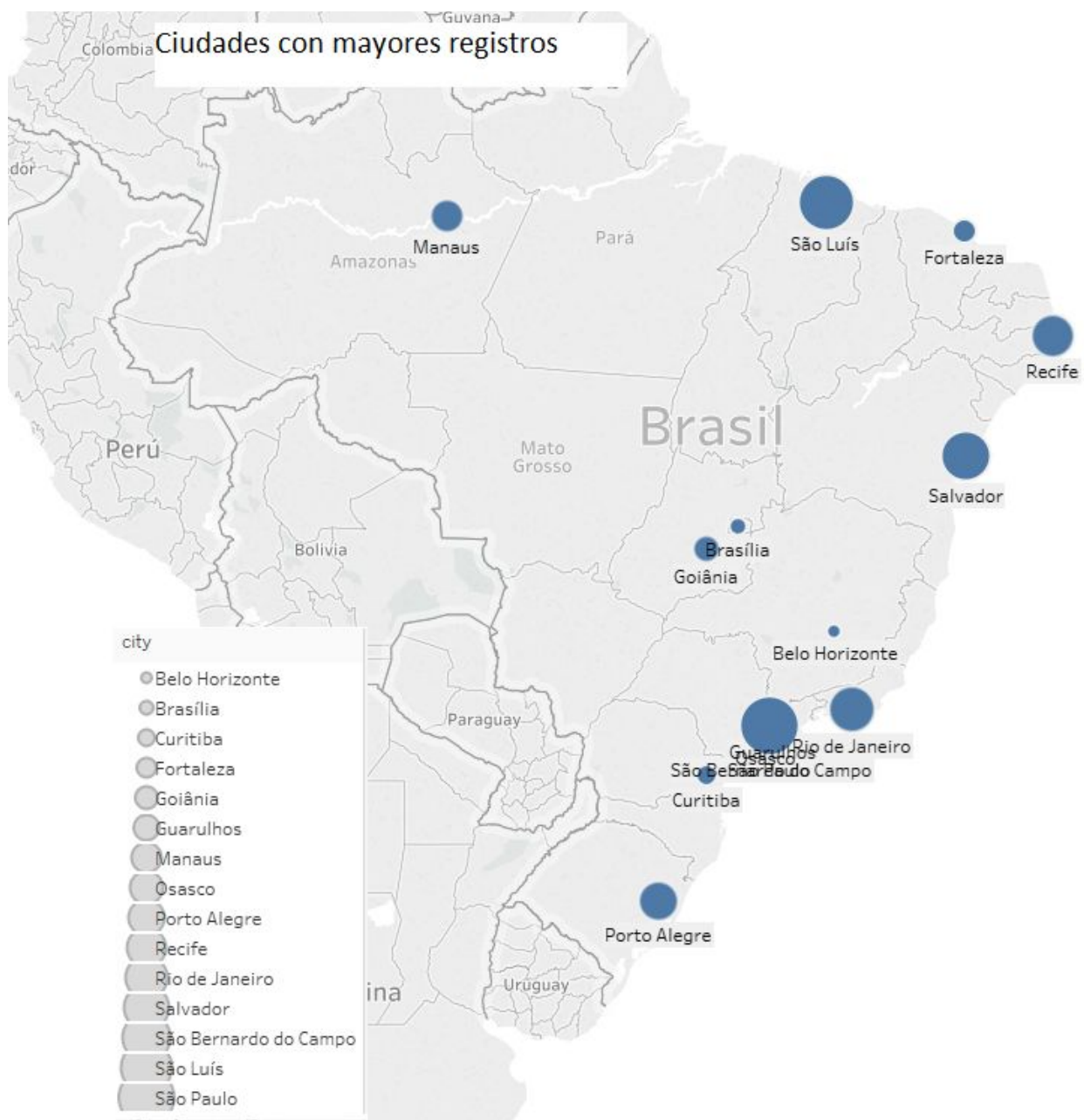
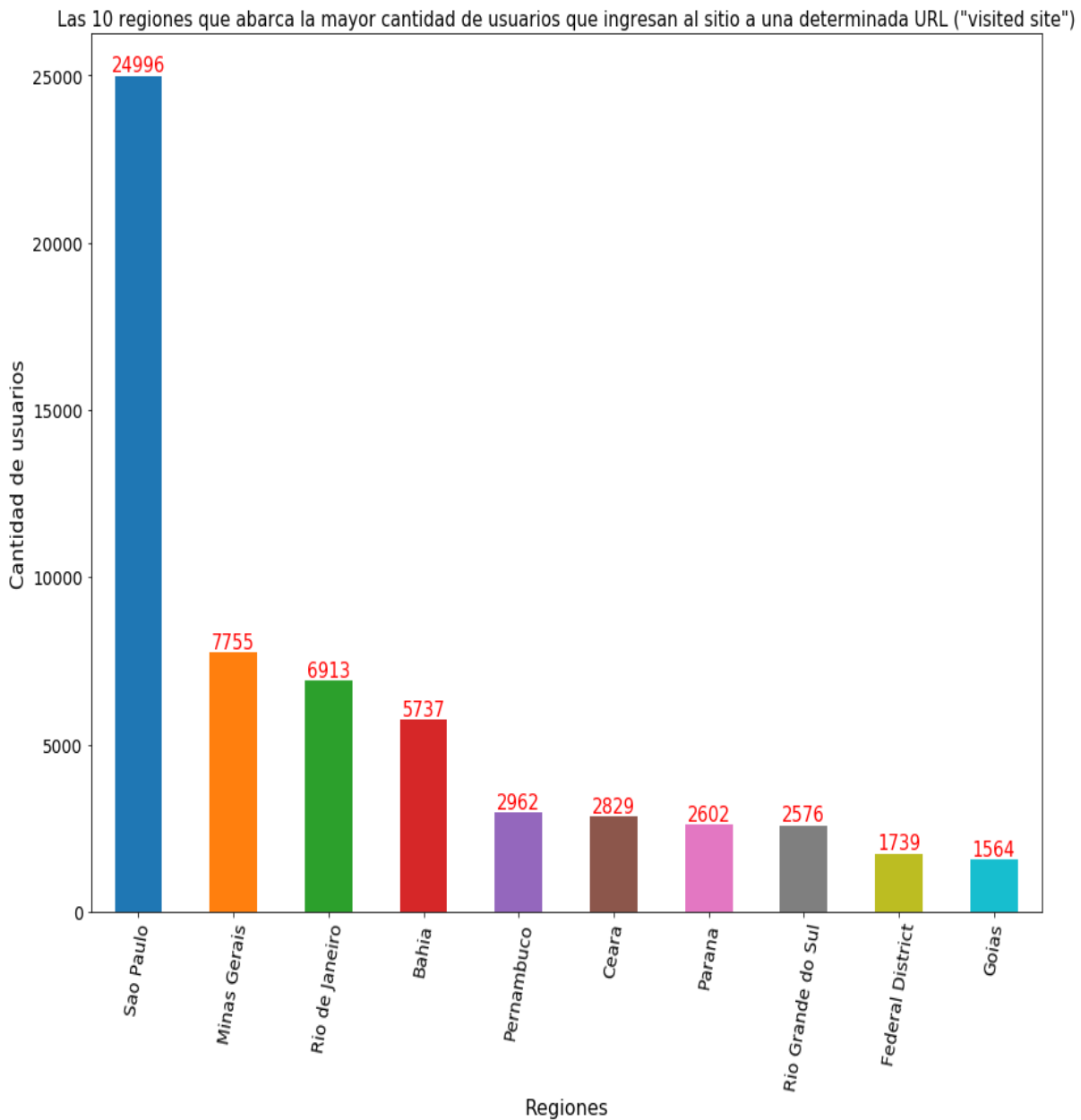


Gráfico realizado con Tableau.

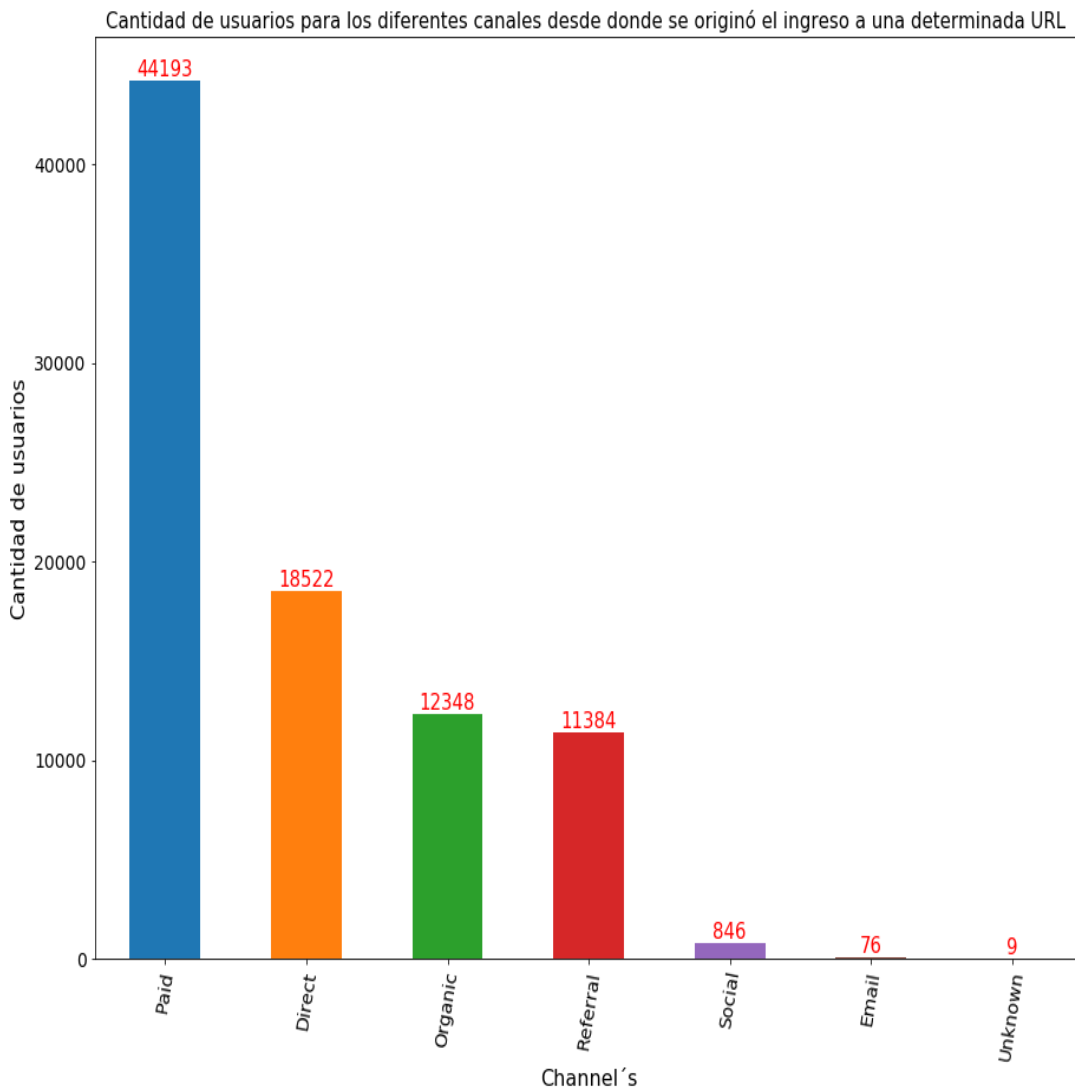
- **Análisis: Cantidad de usuarios que ingresan al sitio para las distintas Regiones**

Como era de esperarse las regiones que tienen la mayor cantidad de usuarios son pertenecientes a Brasil, y la que resalta es la de Sao Paulo con 24996 usuarios.

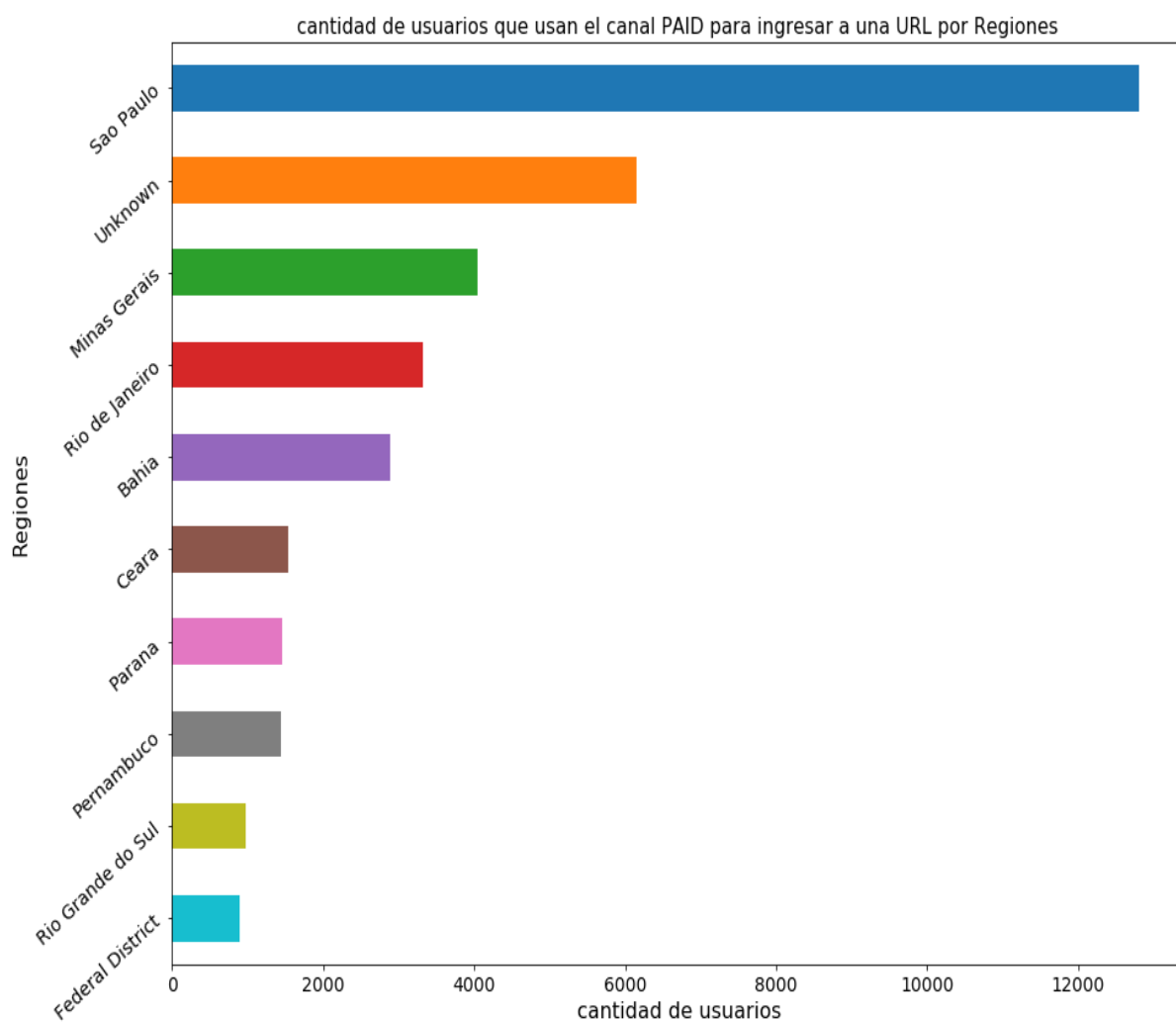


- **Análisis: Veremos cuales son los canales más usados para el ingreso de los usuarios al sitio**

Como podemos observar el canal más usado por los usuarios es el PAID que quiere decir que la empresa invierte o apuesta a la publicidad externa para dar a conocer sus productos e incrementar las ventas de los mismos.



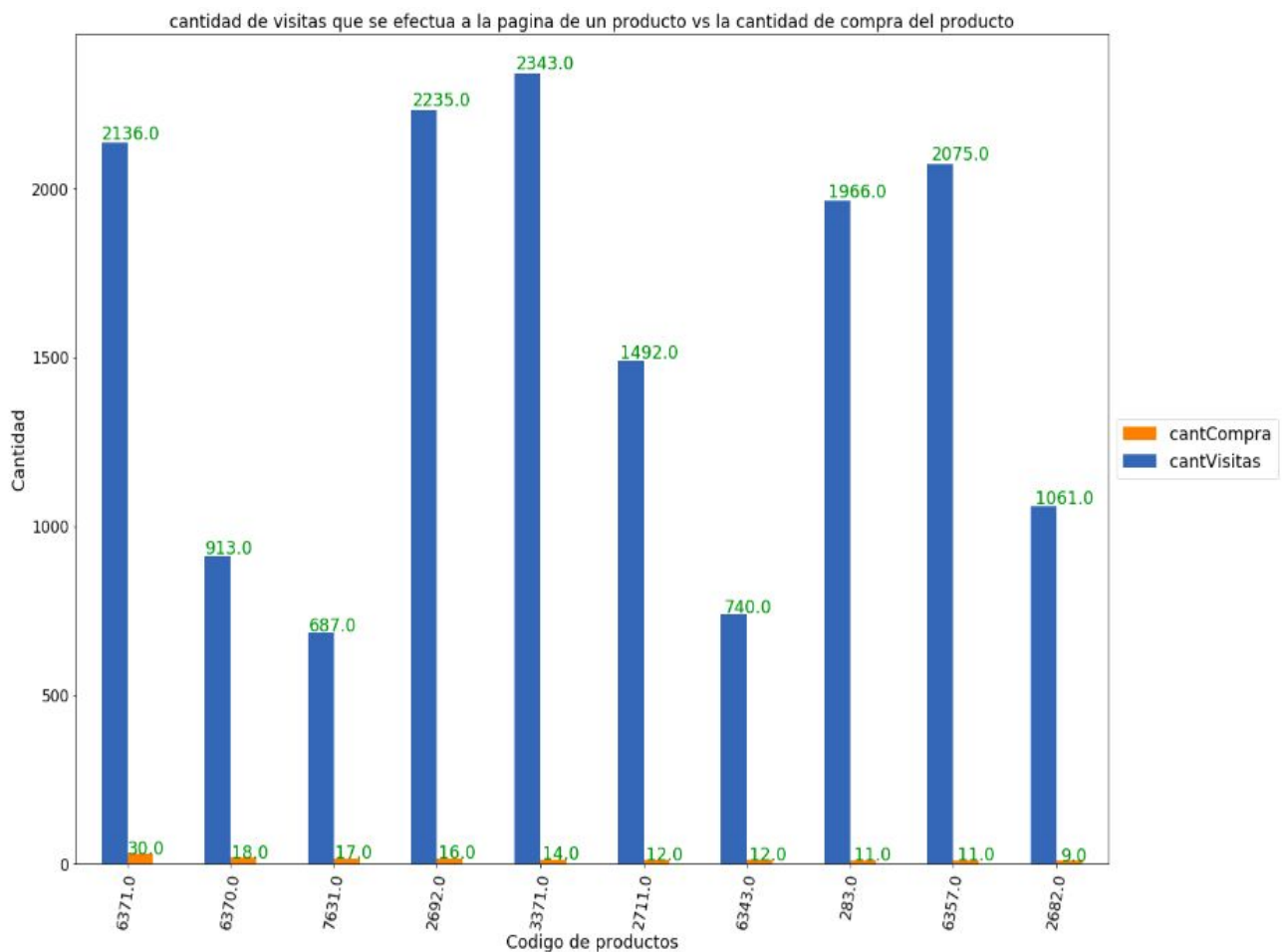
- **Cantidad de usuarios que usan el PAID por Regiones.**



- **Análisis: La mayor cantidad de visitas de un producto corresponde a la mayor venta del mismo?**

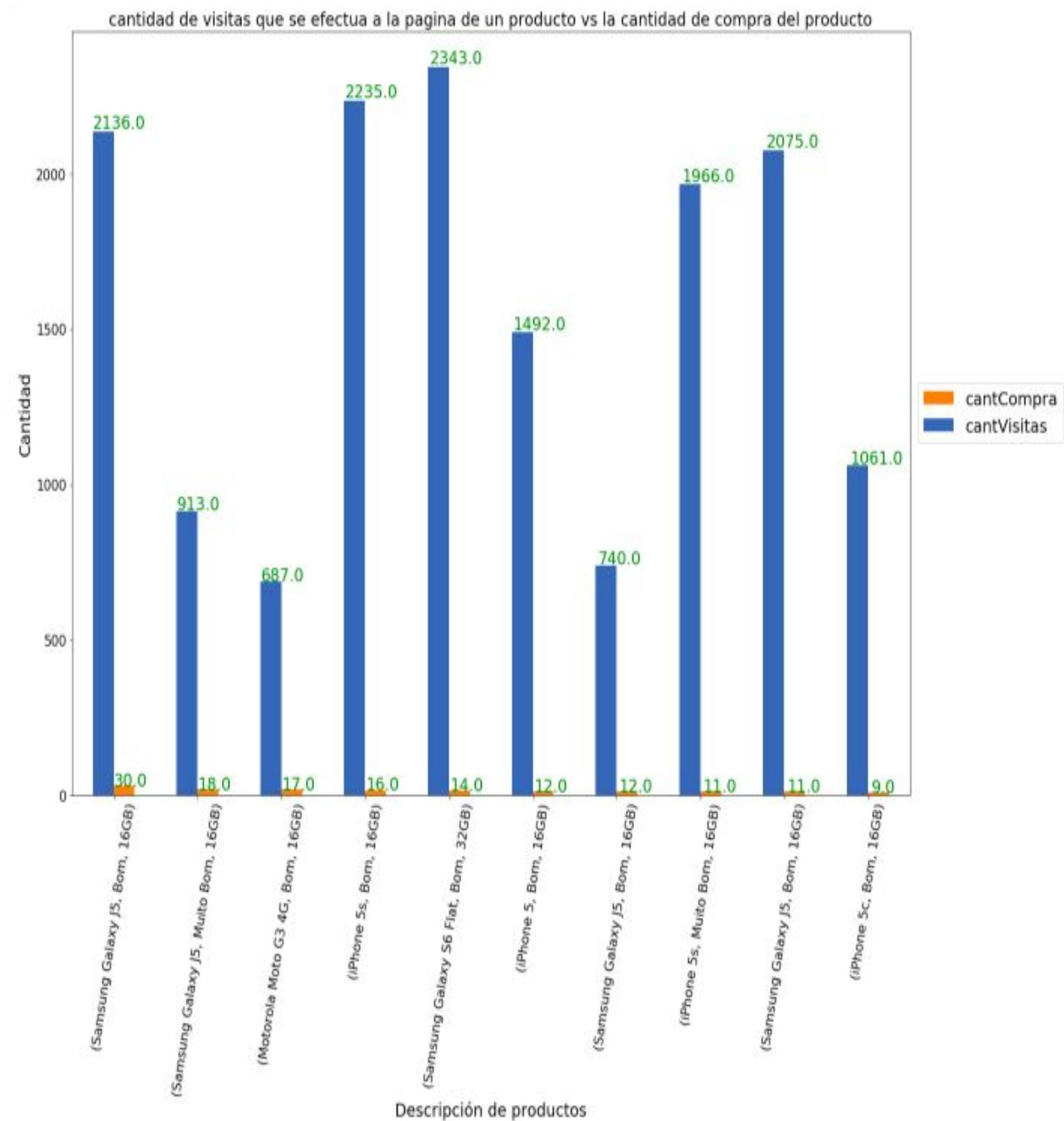
Queríamos mostrar la relación que pueda existir entre la vista de los productos y la compra que se pueda producir, sacamos los códigos de todos los productos que fueron visitados frente a los que se concretaron la compra y fuimos analizando por los códigos de los productos.

En el gráfico que mostramos es la cantidad de compra y vistas para los principales productos comprados



Conclusión: Podemos notar que los productos más visitados son en su mayoría los que menos se vendieron.

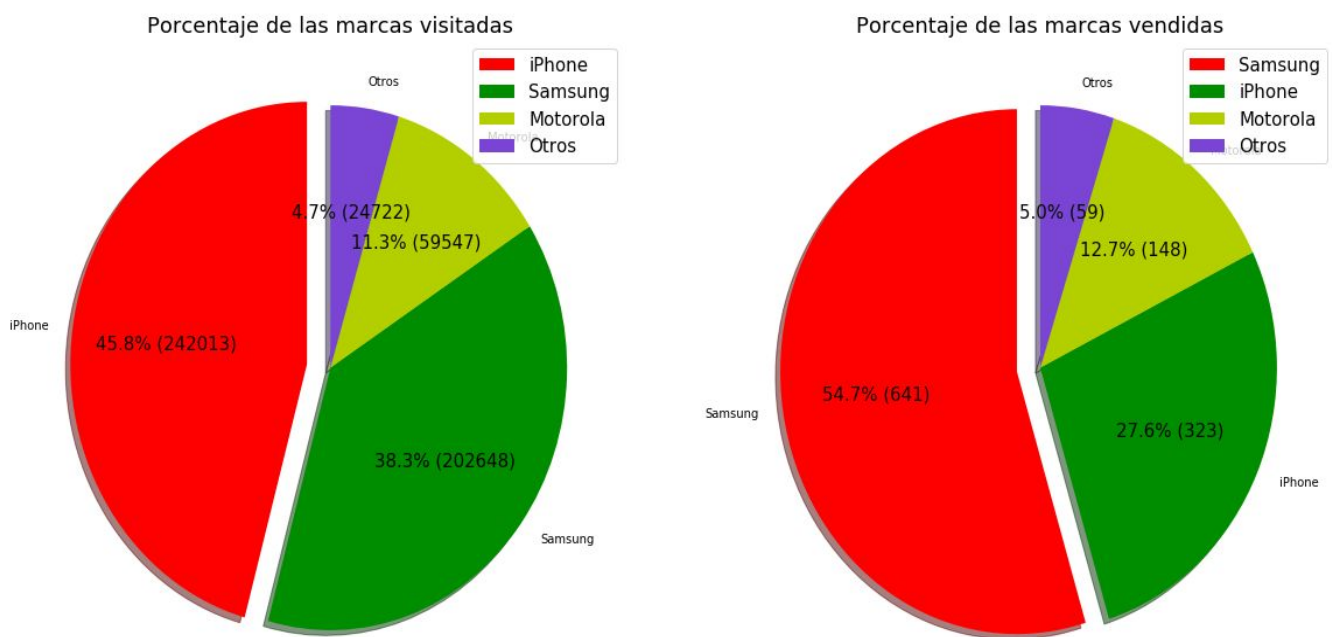
- Como tenemos los códigos de productos pero en una visualización los códigos no nos muestra nada por lo que en la siguiente visualización mostramos la descripción de los productos.



Conclusión: Los productos más vistos es el SAMSUNG GALAXY S6, cuya condiciones son BUENAS (BOM) y posee 32 gb de almacenamiento, pero este mismo equipo es uno de los pocos que se vendio.

- **Análisis: Nos interesa saber qué marca se visita más con respecto a las que se venden.**

En este análisis lo que hicimos fue separar la información que existía en el campo **model**, creando las columnas marca y versión (**brand_model**; **version_model**) tanto para los eventos de visita y ventas con eso filtramos por marcas. Los gráficos a continuación lo muestran.



Conclusión: Lo interesante que podemos observar es que una de las marcas más vistas es la de samsung, donde en las antiguas visualizaciones, denotaban que el producto más visto pertenecía a dicha marca, pero también vemos que el “**IPHONE**” en realidad fue una de las marcas más vistas y la segunda más vendida.

- **Mostramos las marcas más vistas**

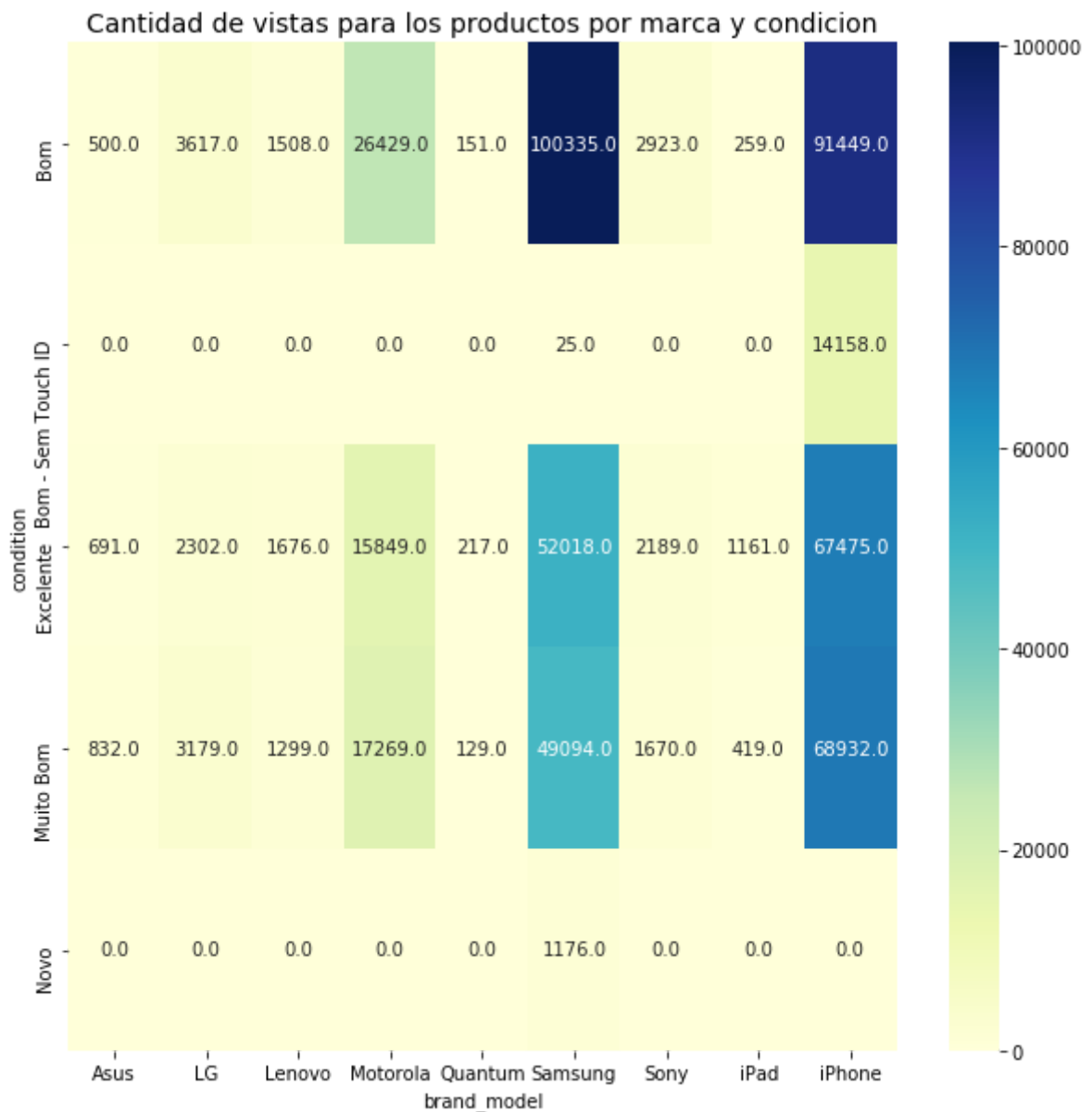
Como era de esperarse las que predominan son SAMSUNG, IPHONE, MOTOROLA, nos pareció interesante mostrarlo mediante este tipo de visualización “WORDLES”.

Las marcas más vistas



- **Análisis: Cantidad de productos vistos por marca y condición del mismo**

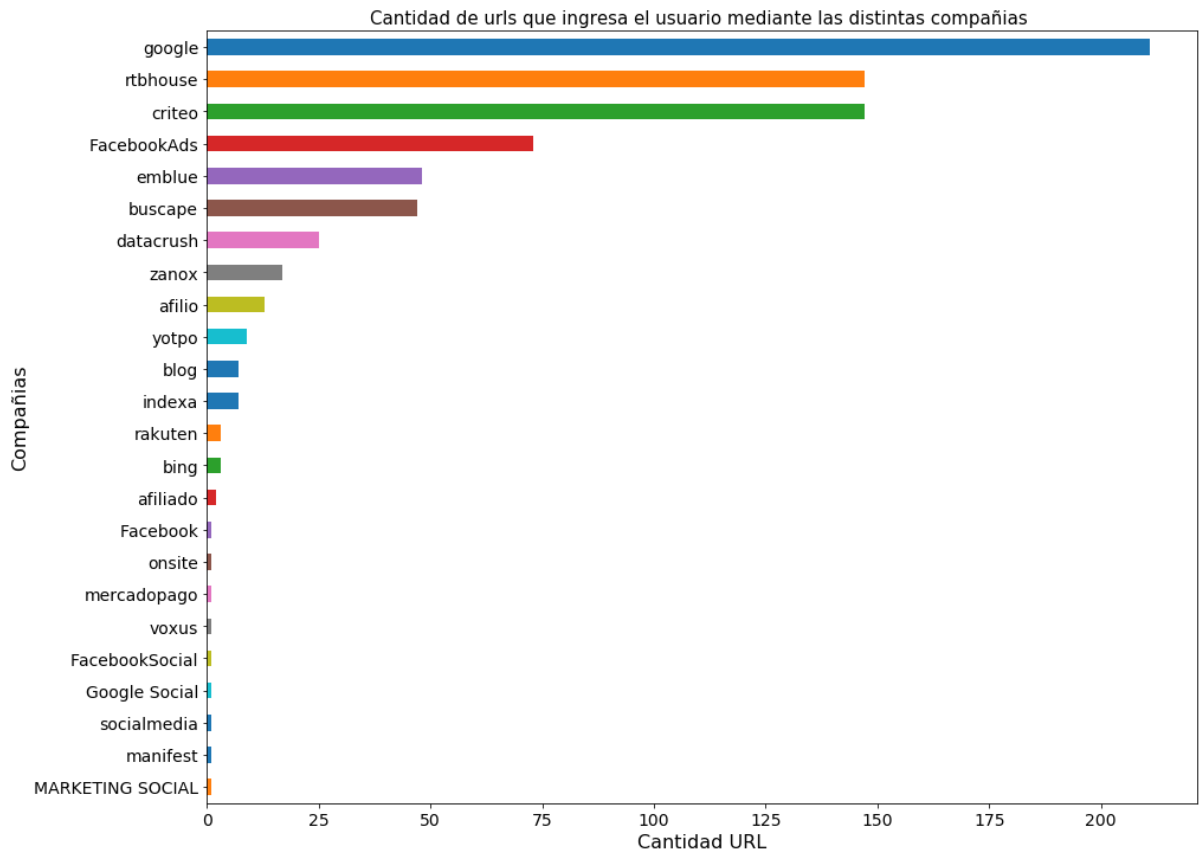
Si bien habíamos mostrado que la marca más vista es el iphone lo que queríamos ver es también en qué condiciones se encuentra el producto visto, entonces mostramos la cantidad de vistas de los productos bajo las condiciones que estaban por las marcas más importantes.



Conclusión: Es interesante ver que la cantidad de vistas de equipos nuevos solo se da en las marcas SAMSUNG.

- **Análisis:** Cantidad de URL'S que el usuario visitó mediante una campaña de marketing online realizado por las distintas compañías.

Nos interesó ver la cantidad de url's de campañas que fueron originados por las distintas compañías.



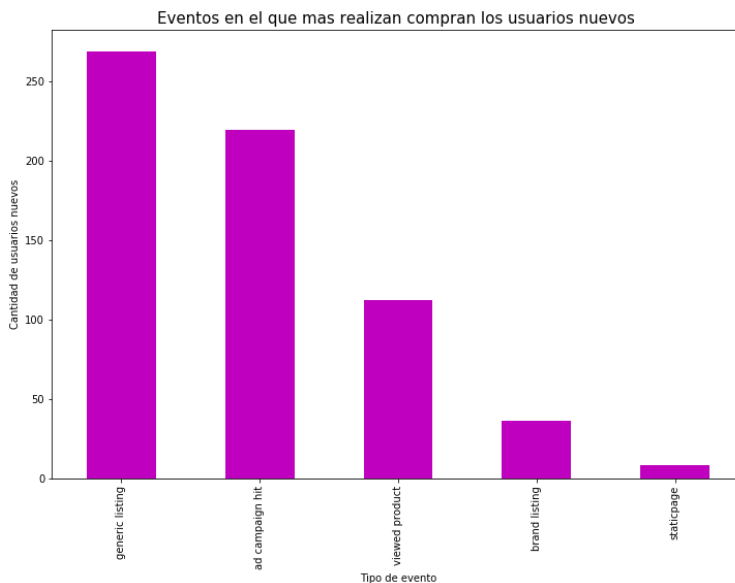
Conclusión: Como podemos ver google origina mayor tráfico de url's visitados por el usuario.

- **Análisis: ¿Cuál es el tipo de evento que más compran realizan los usuarios nuevos?**

Para este análisis nos hemos dado cuenta que existe una gran cantidad de usuarios que no están registrados como nuevos o como usuarios que hayan retornado sus visitas, por ese motivo para las

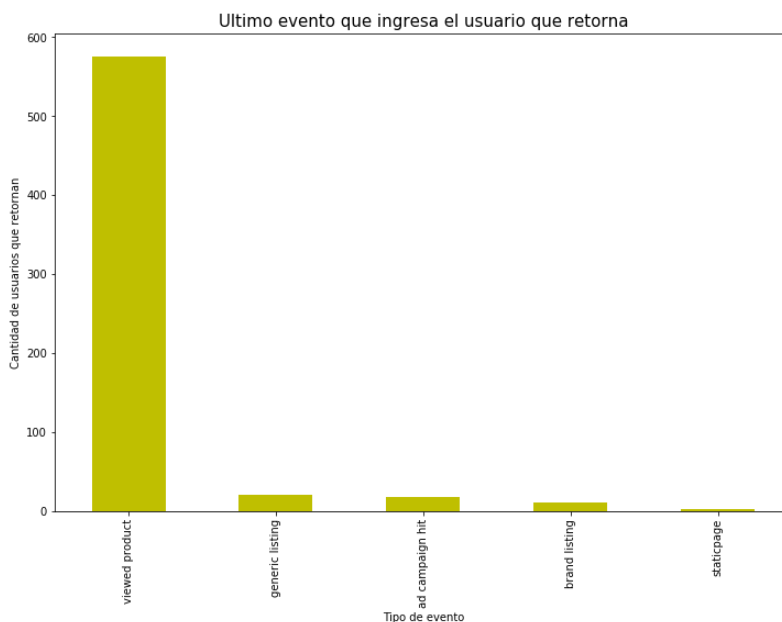
preguntas que nos hacemos, consideramos el campo person como datos más certeros para nuestro análisis, ya que ello significa identificar un cliente sumado a que no existe nan's en ese campo

Conclusión: Se puede observar que la mayoría de los usuarios nuevos realizan su compra cuando visitan la homepage.. Por otro lado se ve que hay una fuerte campaña de marketing que los coloca como segunda opción



- **Análisis: Los usuarios que retornan para comprar, ¿Por que tipo de evento es el que más realizan?**

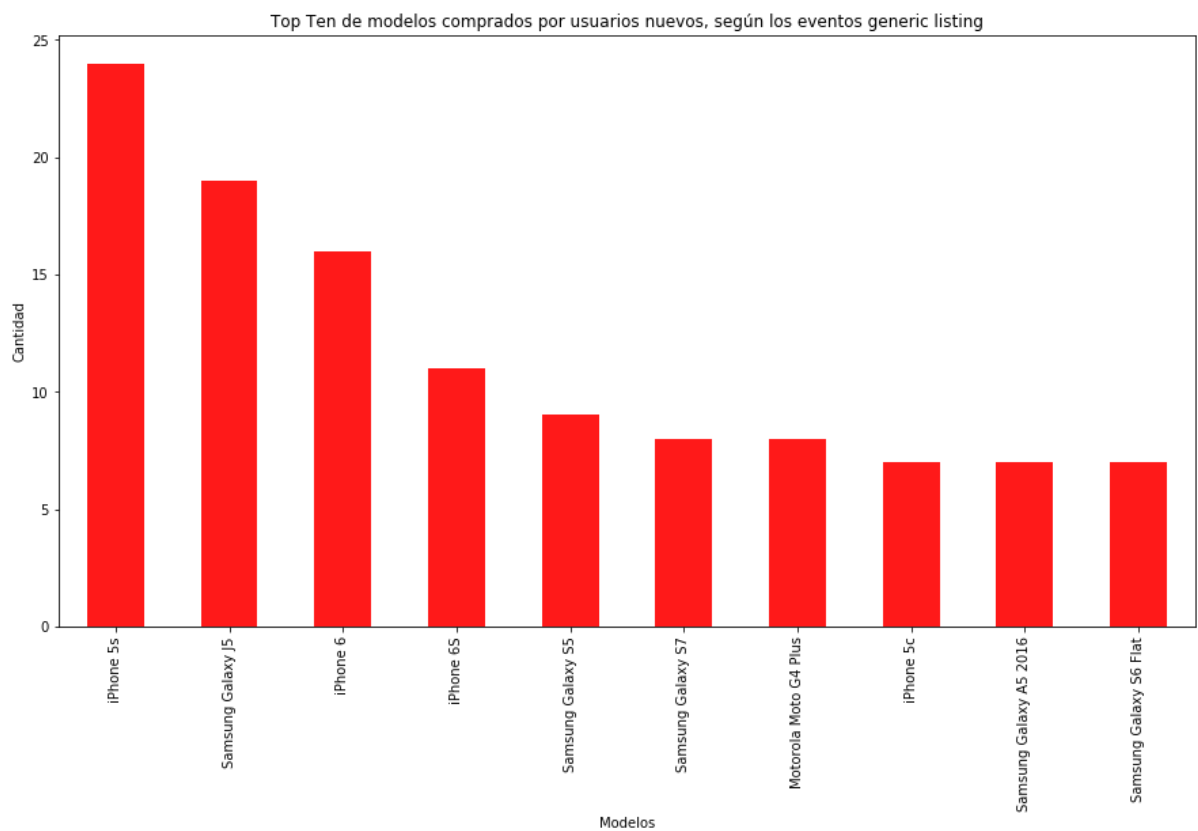
(Se considera solo la última compra que realiza el usuario que retorna)

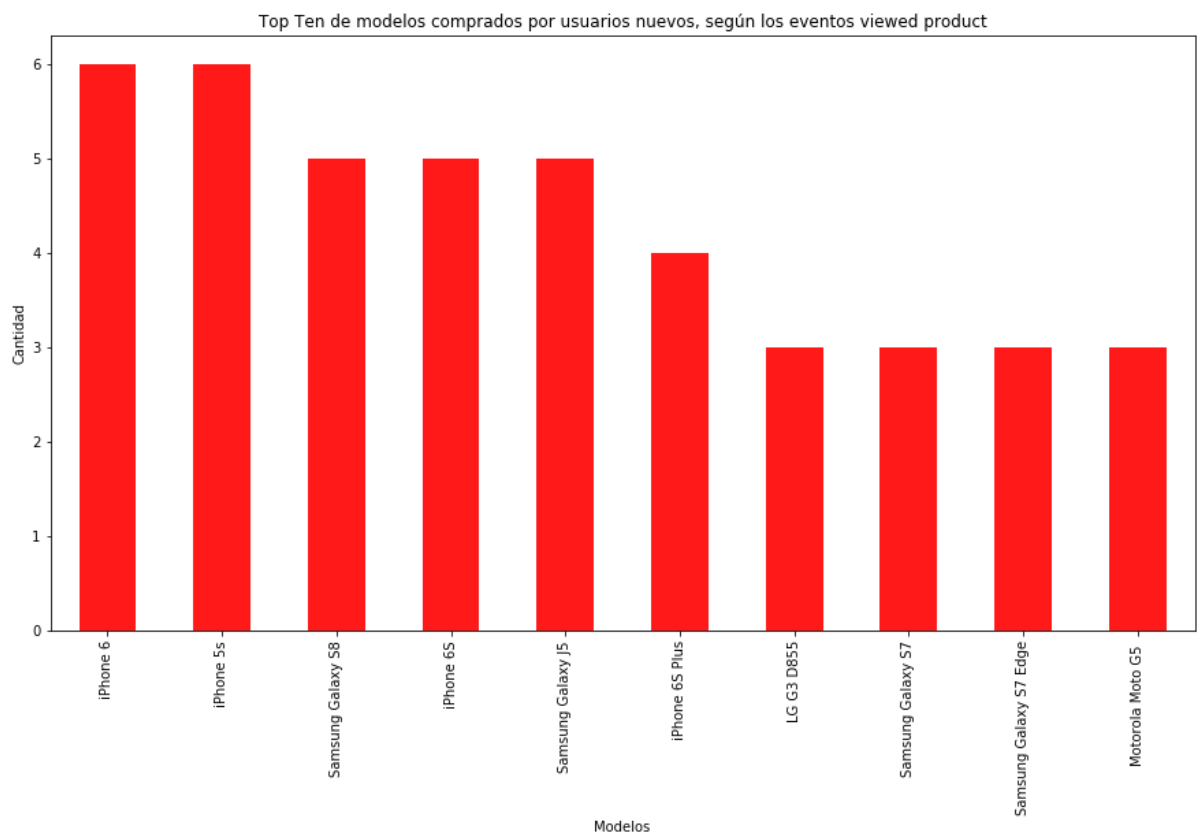
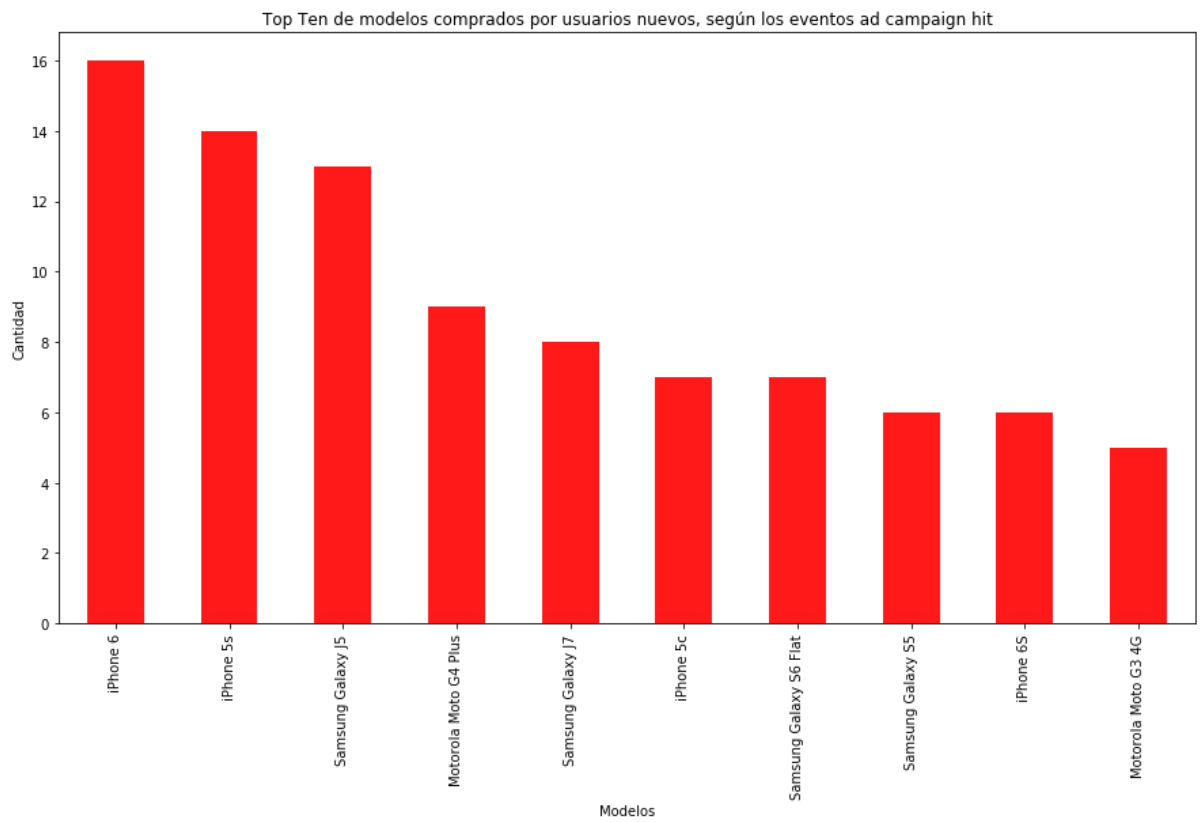


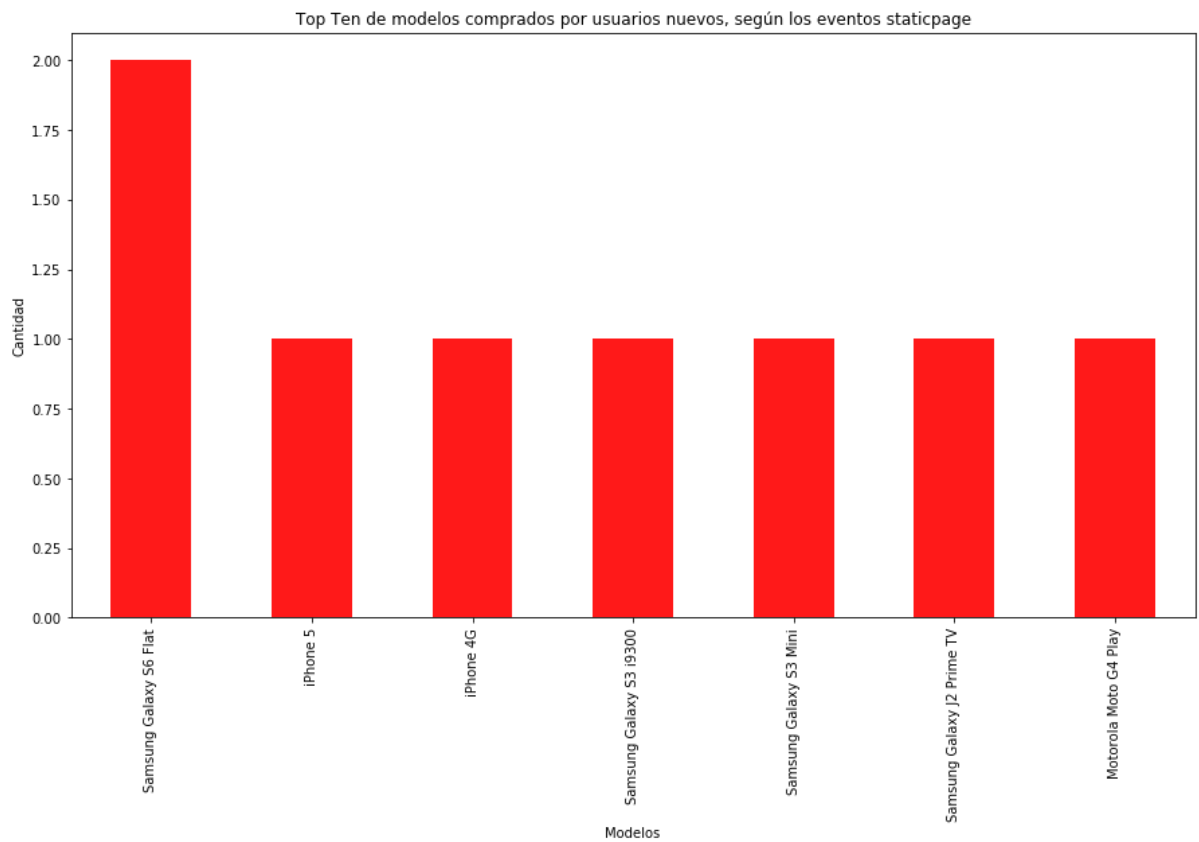
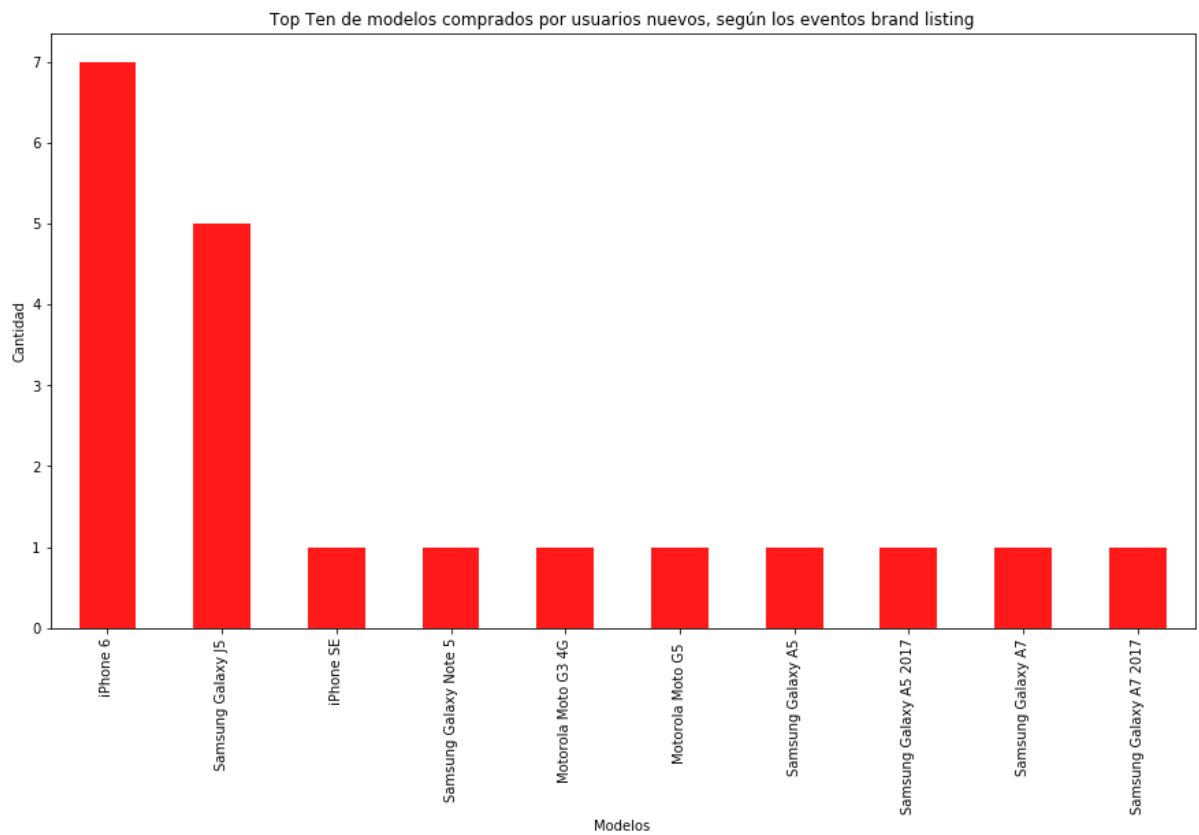
Conclusión: los usuarios que han retornado en su mayoría ha vuelto a comprar

- **Análisis:** ¿Los usuarios nuevos que han realizado una compra, que tipo de producto se llevan?

Conclusión: en la mayoría de los casos, los productos más comprados son los iphone 6 y iphone 5s



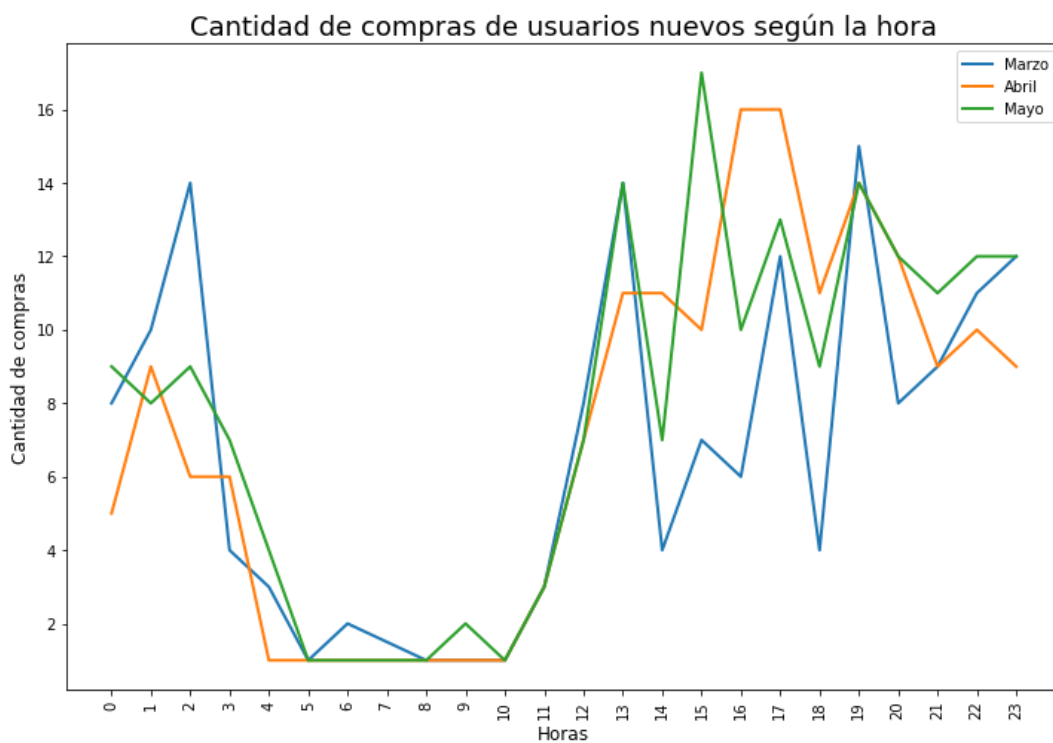


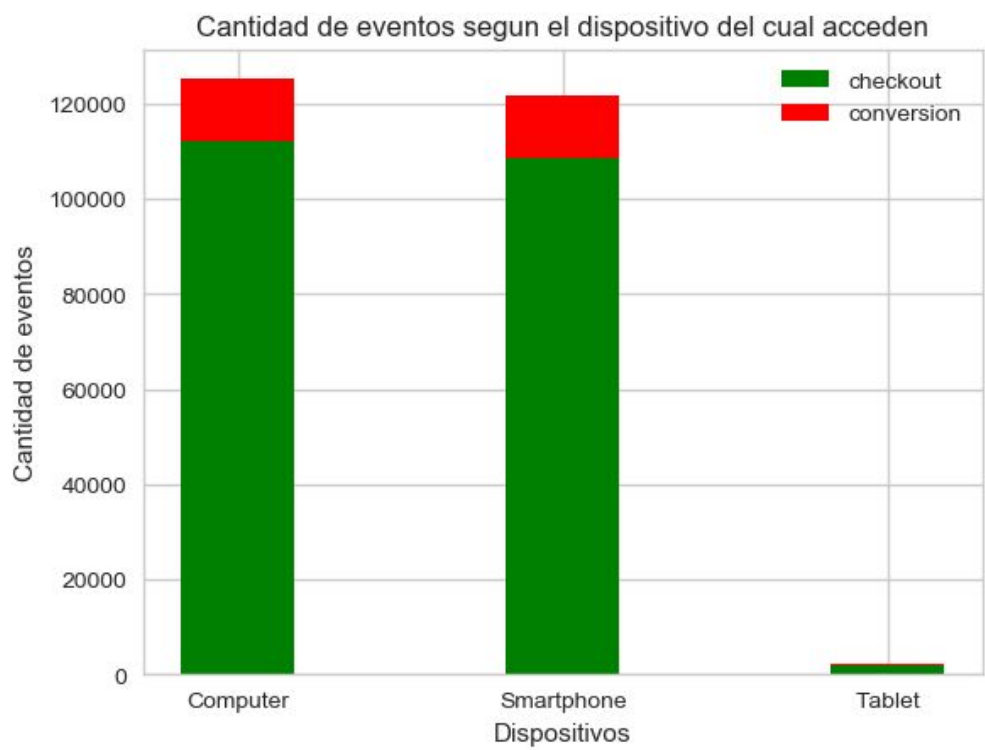


- **Análisis: ¿Los usuarios nuevos, en que horario realizaron sus compras?**

(Para el análisis de la compra de usuarios nuevos respecto a la hora, no tomaremos en cuenta los meses 1,6 y 2, ya que representa un número menor del 28%)

Conclusión: Podemos verificar que no influye la hora para concretar una compra en los usuarios nuevos





Conclusiones Generales

Aunque en cada pregunta que nos realizamos en los distintos análisis se realizaron apreciaciones y conclusiones, queríamos remarcar algunos puntos que podrían tener una utilidad real, porque al fin y al cabo, ese sería el objetivo de este análisis.

Como pudimos ver en el análisis de la correlación temporal entre las demoras que una persona realiza entre los eventos Visita, Checkout y Conversión(Compra), existe un altísimo grado de correlación entre el tiempo que pasa entre que una persona visita un producto hasta que termine comprándolo y el tiempo en que dicha persona demora entre visitar el producto y hacer el checkout previo a la compra, siendo este tiempo predominante entre tres a 5 minutos. Esto nos podría servir de utilidad para analizar a posibles compradores teniendo en cuenta si una persona hace un checkout rápido debería hacer una compra rápida, ya que si se demora en un tiempo superior(al que tardó entre visitar y hacer el checkout) disminuirían las chances de efectuar una compra.

También observamos que una persona que finalmente compra un producto suele visitar el mismo más de una vez antes de hacer un checkout mientras que las que visitan una sola vez el producto difícilmente lo compran.

Pudimos calcular que las duplas de soporte para los modelos son predominantes para la marca Iphone y que es más frecuente que una persona que visita una marca visite otro modelo de la misma marca.

Vimos que el soporte para un solo producto pertenece a la marca Iphone seguida por Samsung y Motorola.

En cuanto a las campañas contratadas por trocafone se ve que los resultados se dan mayormente por Google, RTBHouse, FacebookAds, Criterio y Emblue y Buscape y que no vale la pena gastar dinero en las demás compañías.