




Organización de Datos 75.06/95.58
Trabajo Práctico N° 1

GRUPO: 15

Apellido y Nombre	Padrón	Correo electrónico
Marinero Santiago	97969	santiagomarinero1@gmail.com
Opizzi Juan Cruz	99807	juanopizzi@gmail.com
Lasanta Tito	99162	titonoolvida@gmail.com
Lizarraga Martin Augusto Juaquin	99636	augusjoaliza@hotmail.com

Github: 

<https://github.com/Titolasanta/bufanda>

Índice

1. Introducción	2
2. Limpieza	2
3. Análisis de datos	2
3.1. Frecuencia de Eventos	2
3.2. Frecuencia de Eventos a lo largo del tiempo	3
3.3. Frecuencia de los modelos	5
3.4. Frecuencia de los modelos en el tiempo	5
3.5. Frecuencia de los colores para los modelos	9
3.6. Colores según mes	14
3.7. Condición del Celular	19
3.8. Hits de search engines y ads	21
3.9. Relación de Regiones por compras	25
3.10. Canales de ingreso	28
3.11. Términos de búsqueda	29
3.12. Usuarios Nuevos vs Usuarios que regresan	33
4. Relación entre eventos	34
4.1. Navegadores y Modelos	35
4.2. Navegadores y Estado	38
4.3. Checkout vs Compras	40

1. Introducción

El objetivo planteado en este trabajo practico es el de explorar los datos provistos para encontrar o corroborar hechos, el mejor modo de transmitir estos hallazgos es mediante visualizaciones lo cual implica que se tuvo que aprender a expresar los datos de este modo las herramientas utilizadas para cumplir estos objetivos fueron las siguientes:

- Pandas: Donde se cargo la información, se le aplicaron filtros, y finalmente se le dio forma lista para graficar
- Seaborn: Programa utilizado para graficar los datos formados en pandas
- Matplot lib: Programa utilizado para graficar los datos formados en pandas

2. Limpieza

Los datos provistos a analizar vinieron en muy buenas condiciones, por lo tanto el proceso de filtrar o traducir información fue corto, principalmente las modificaciones realizadas al set de datos fueron las de convertir colores descriptos como “azul super galáctico” en “azul” o bien, quitar la versión del tipo de browser utilizado para acceder a la página, adicionalmente para poder realizar observaciones del tipo que modelo se vio más en que bowser se mergearon sobre el id de los usuarios diferentes eventos.

3. Análisis de datos

3.1. Frecuencia de Eventos

En esta sección se quiso analizar la frecuencia con la que pasan los distintos eventos del .csv. Para poder ver el evento mas frecuente.

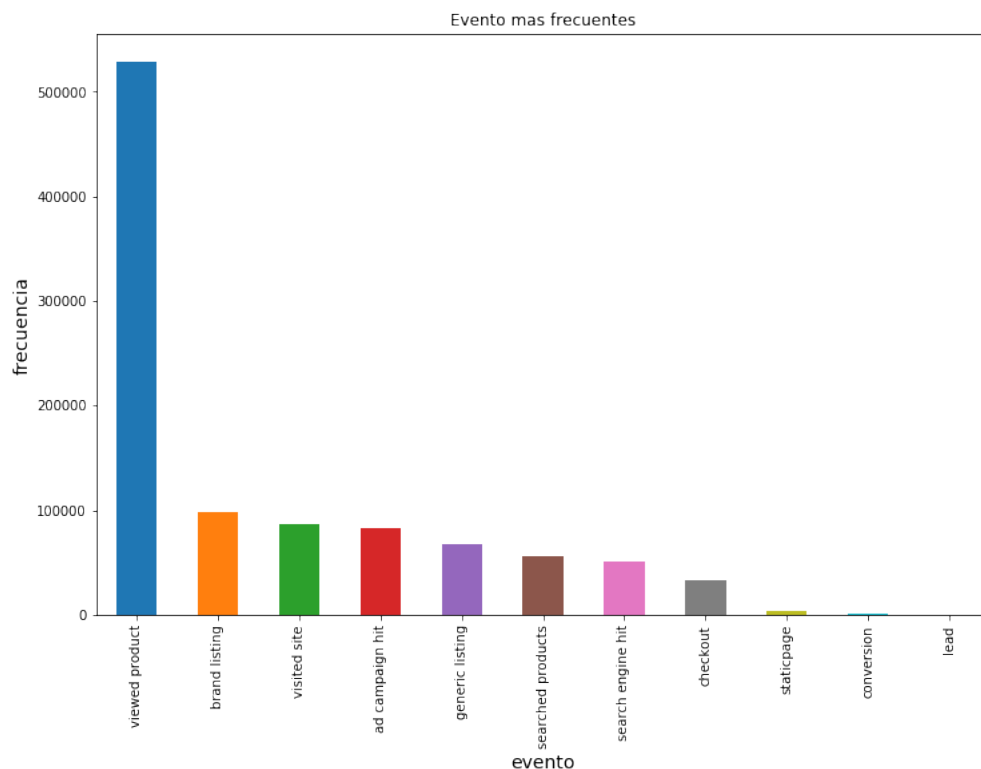


Figura 1: Gráfico de barra de la frecuencia de eventos

Como resultado de este análisis se puede concluir que el evento “viewed product” es el mas frecuente de todos.

3.2. Frecuencia de Eventos a lo largo del tiempo

Aquí se puede ver como el sitio web creció con el paso del tiempo, teniendo en Mayo una explosión en popularidad, con posibilidad de doblar el número de visitas para Junio, ya que los datos facilitados tienen fecha hasta la mitad de tal mes. Los datos están normalizados, habiendo sido divididos por las visita de cada mes.

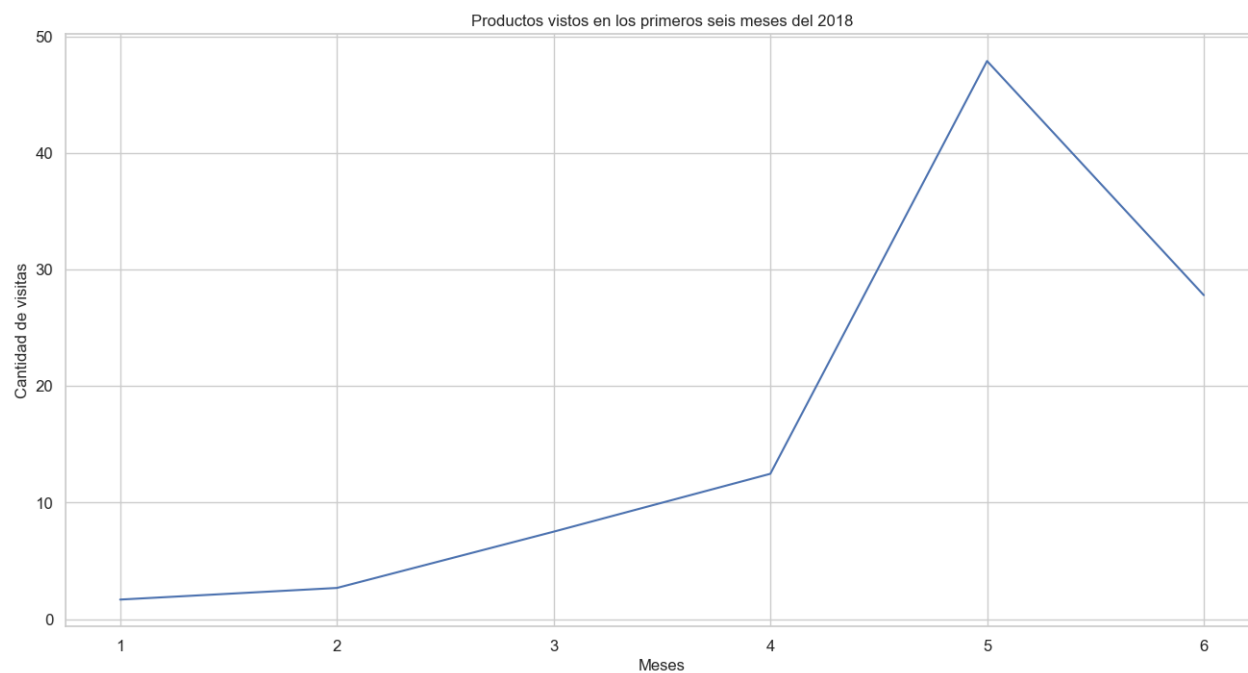


Figura 2: Productos vistos a lo largo del tiempo

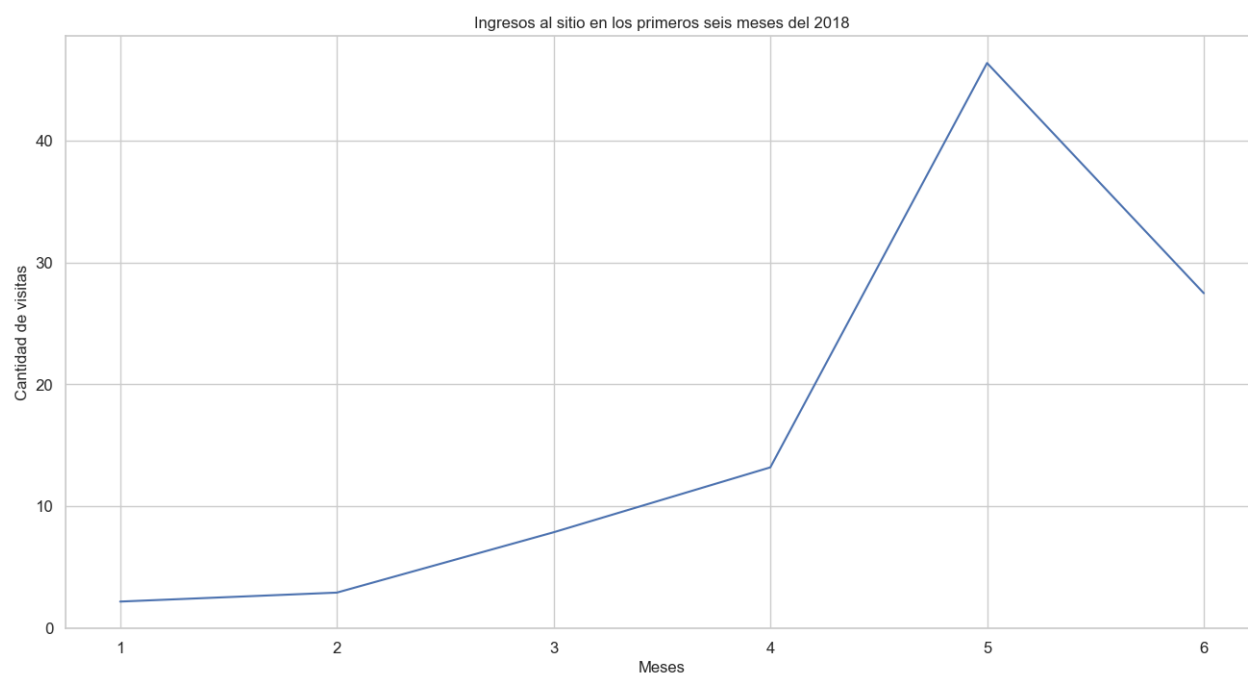


Figura 3: Ingresos al sitio a lo largo del tiempo

3.3. Frecuencia de los modelos

En esta sección se vio cuales eran los modelos mas vistos por las personas en el mundo

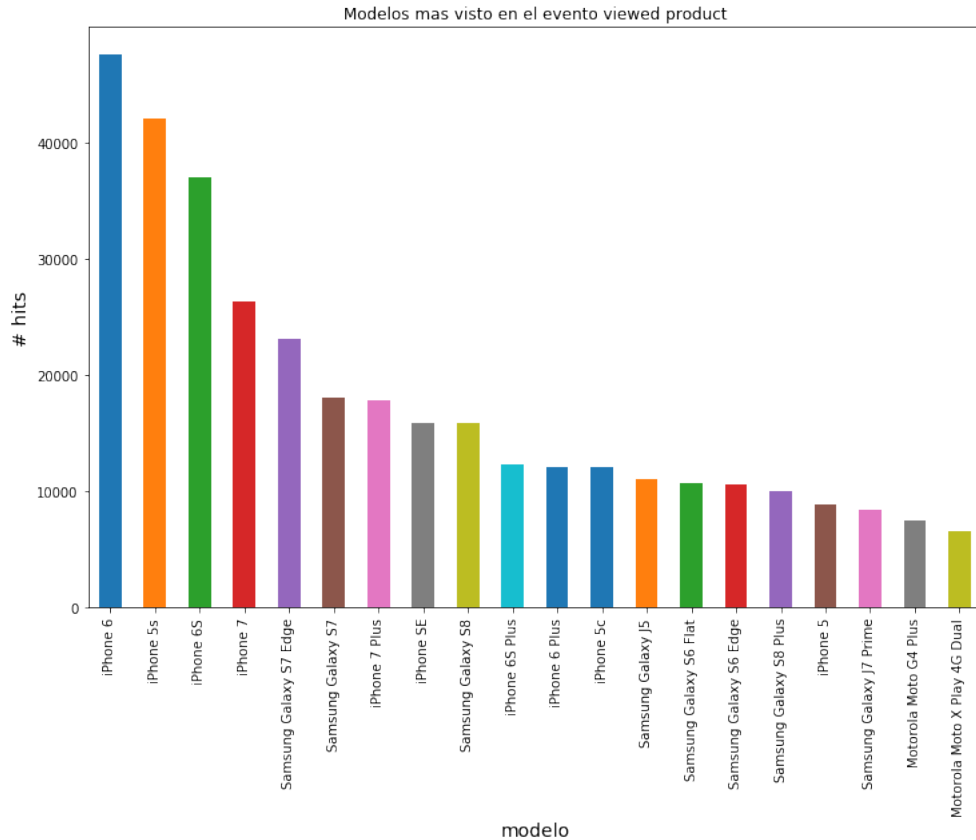


Figura 4: Gráfico de barra de la frecuencia de eventos

El resultado los modelos mas vendidos los iPhone, en particular el iPhone 6. Siguiéndolos se encuentra el Samsung Galaxy.

3.4. Frecuencia de los modelos en el tiempo

En esta sección como la frecuencia de los distintos modelos evoluciono en el tiempo

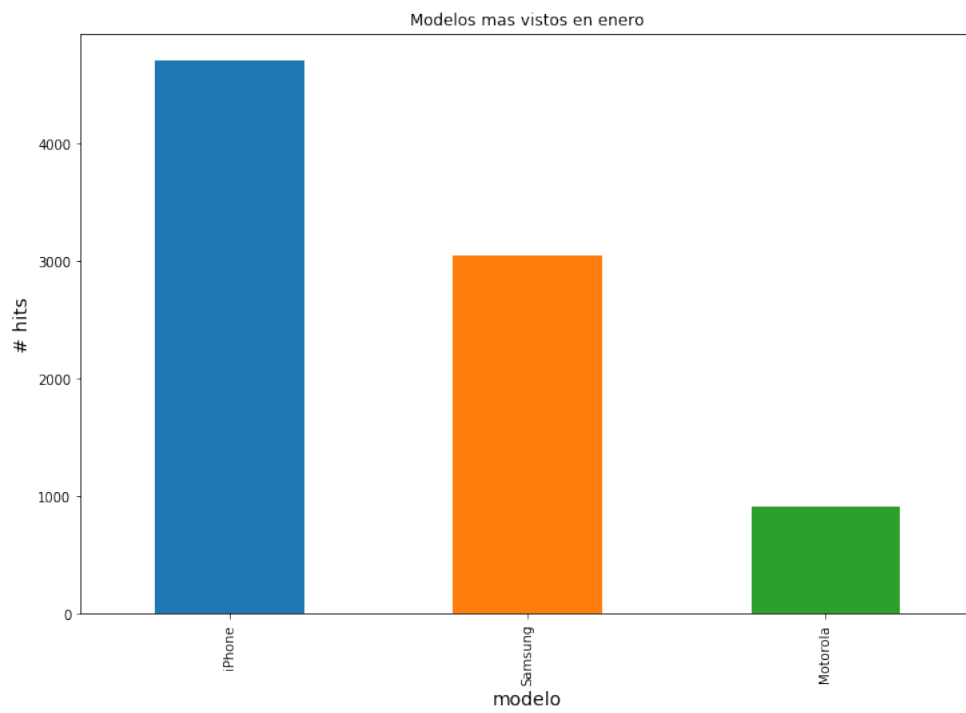


Figura 5: Modelos mas frecuentes en Marzo

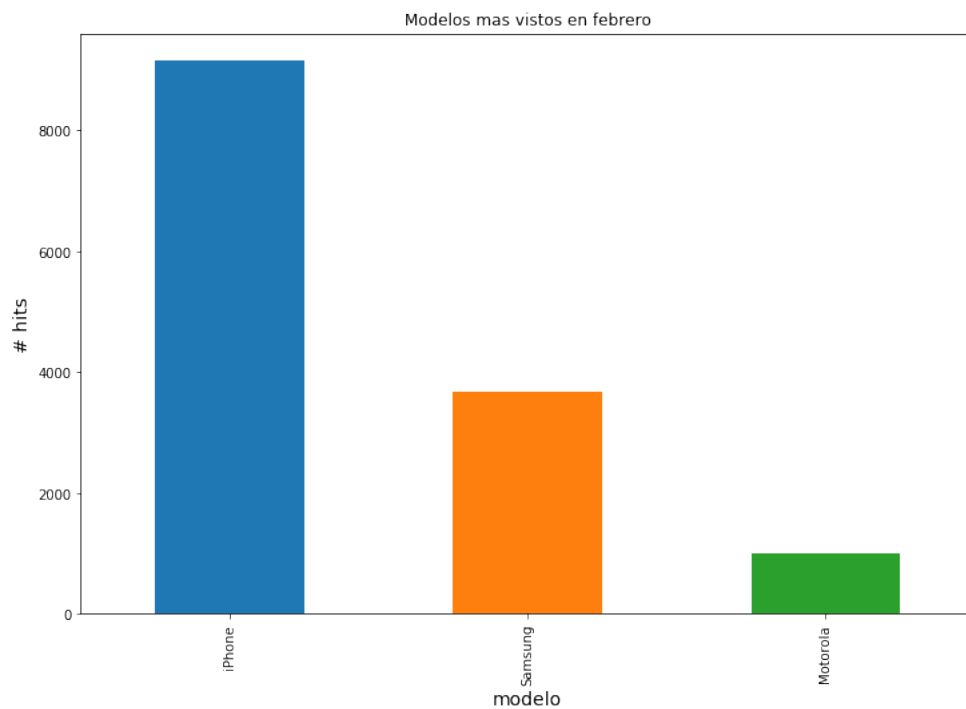


Figura 6: Modelos mas frecuentes en Marzo

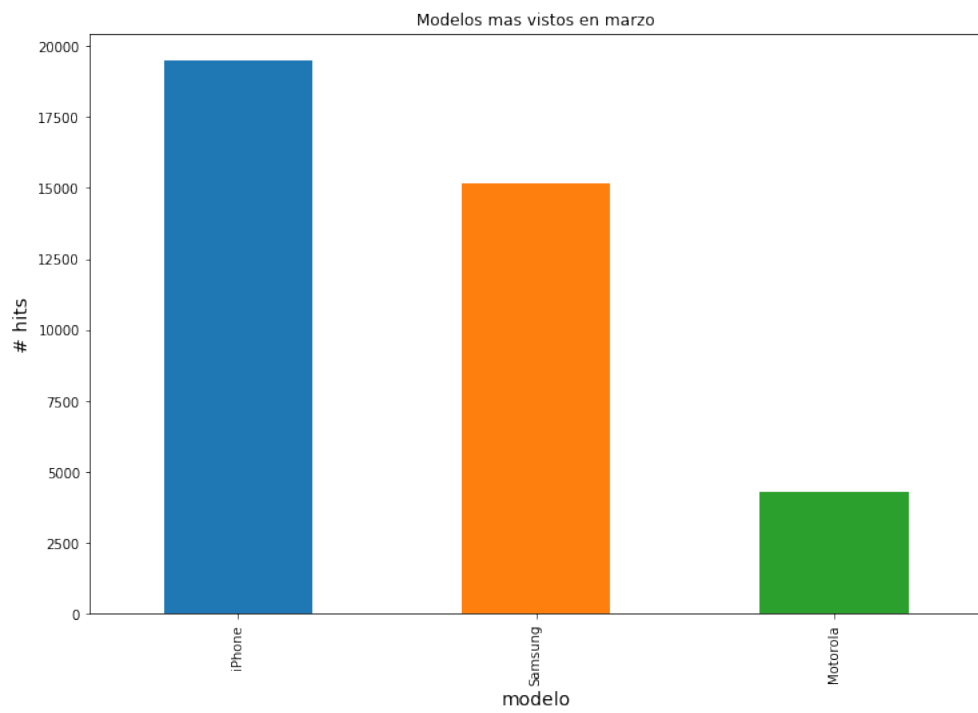


Figura 7: Modelos mas frecuentes en Marzo

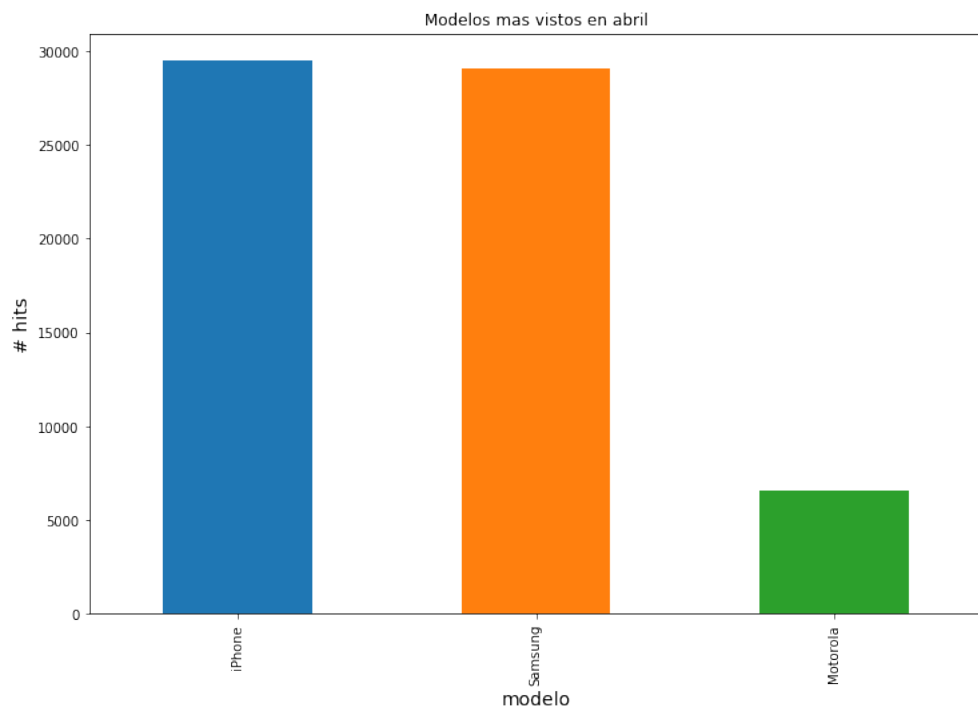


Figura 8: Modelos mas frecuentes en Abril

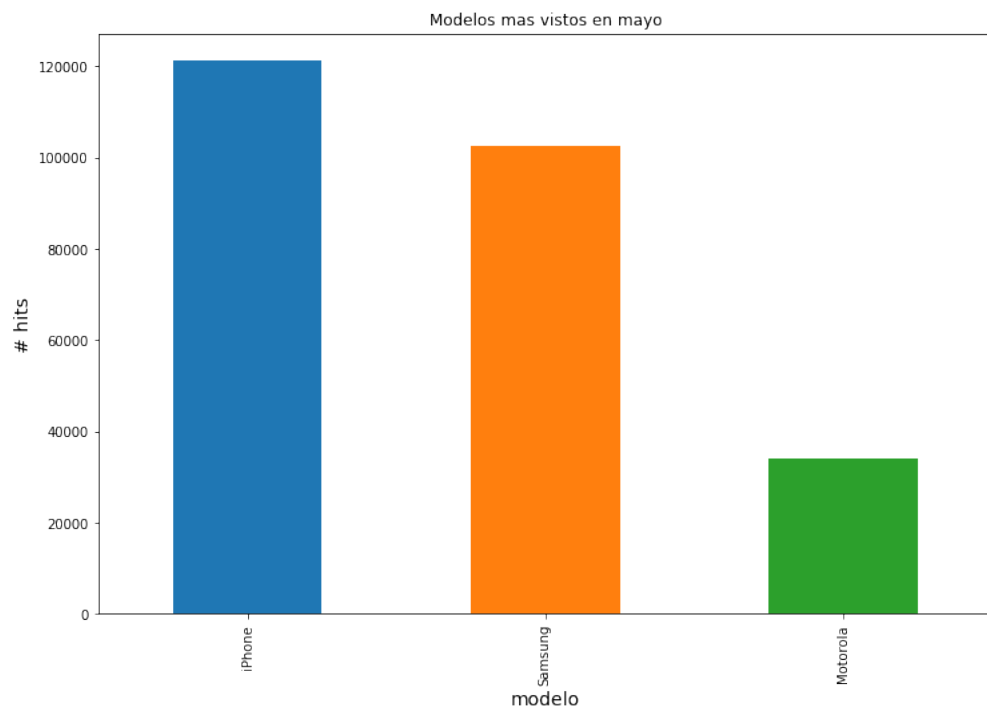


Figura 9: Modelos mas frecuentes en Mayo

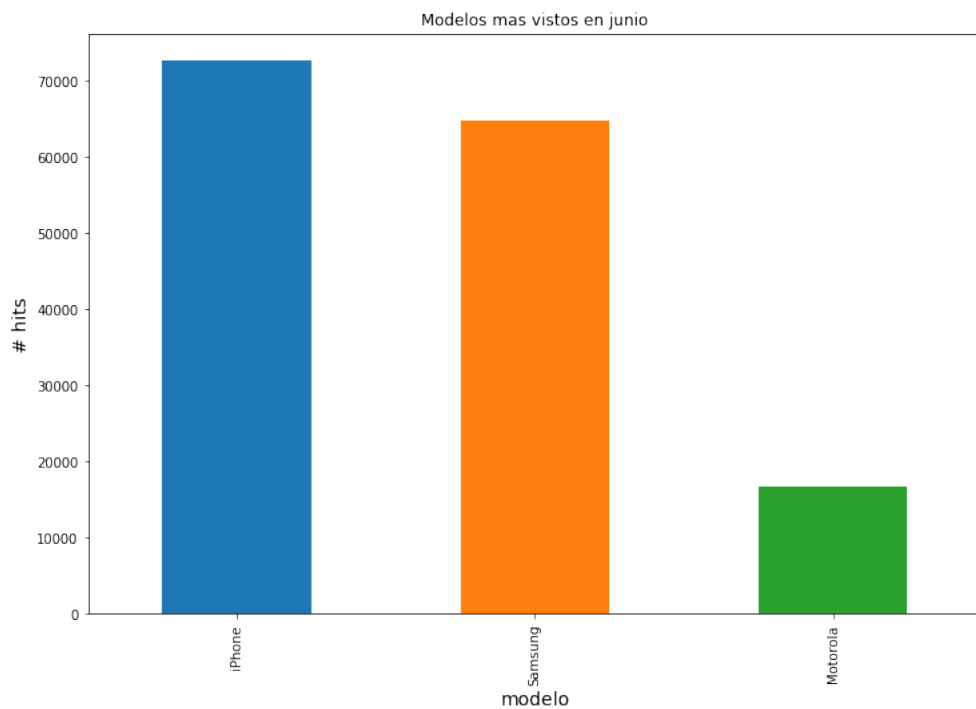


Figura 10: Modelos mas frecuentes en Junio

Al final como se puede apreciar en las imágenes los iPhones siguen siendo los mas vendidos siendo el iPhone el que predomina. Abril fue el único mes en el que Samsung tenia cifras similares a las del iPhone. Y los Motorola estuvieron muy por debajo de los iPhone y Samsung.

3.5. Frecuencia de los colores para los modelos

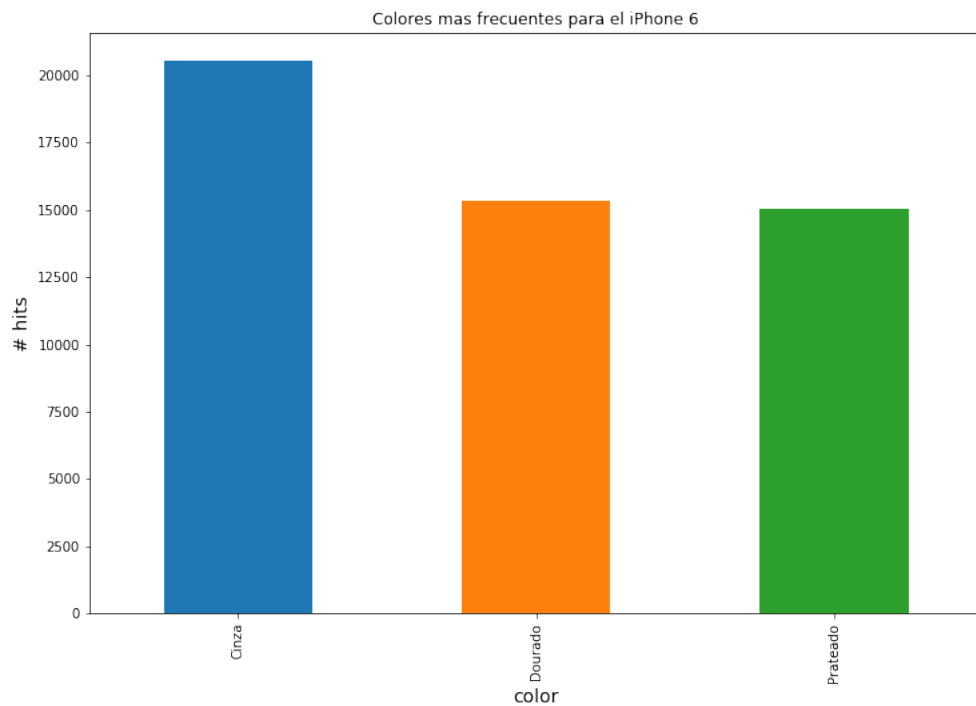


Figura 11: Colores mas frecuentes para iPhone 6

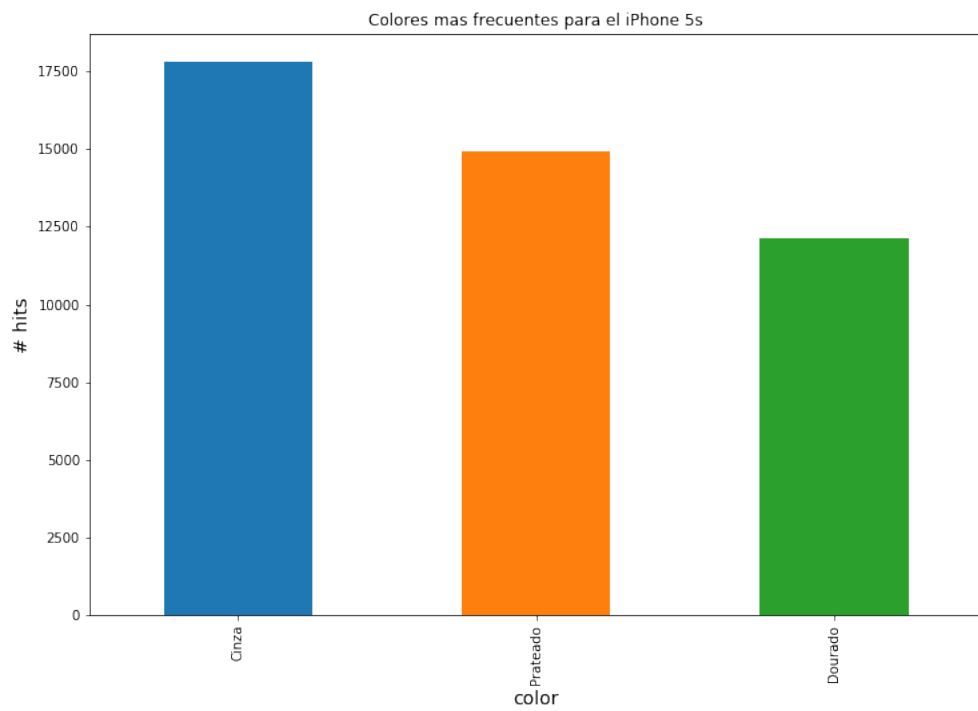


Figura 12: Colores mas frecuentes para iPhone 5S

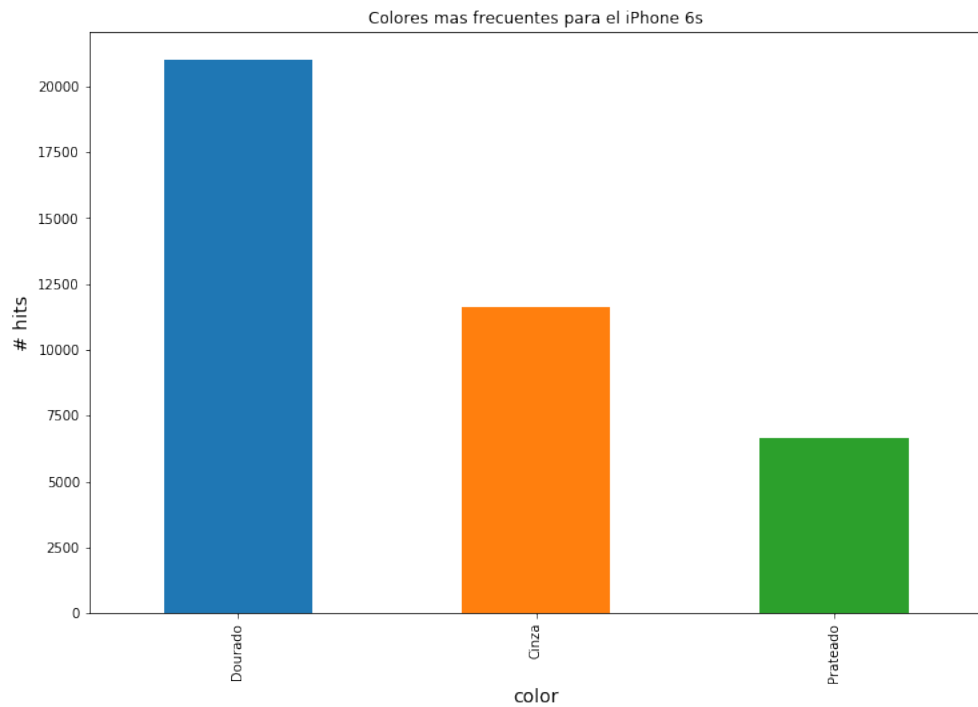


Figura 13: Colores mas frecuentes para iPhone 6S

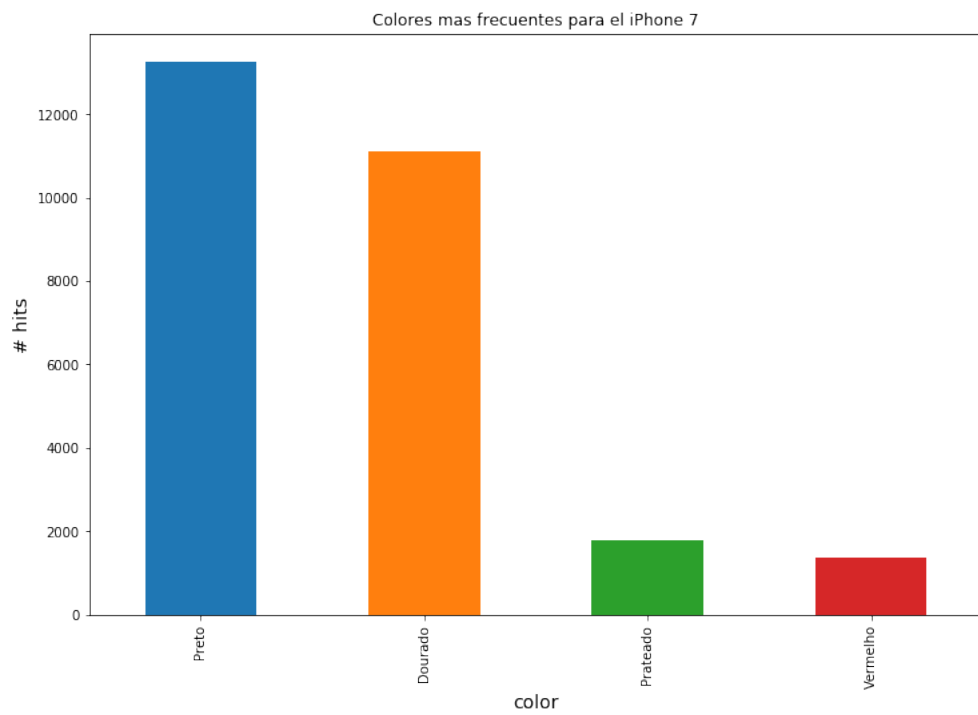


Figura 14: Colores mas frecuentes para iPhone 7

Para todos los iPhones en general solo se vieron 3 colores distintos que son el Cinza, Dorado y Plateado. Lo cual indica que es muy probable que los distintos iPhone solo vengan en esos 3 colores.

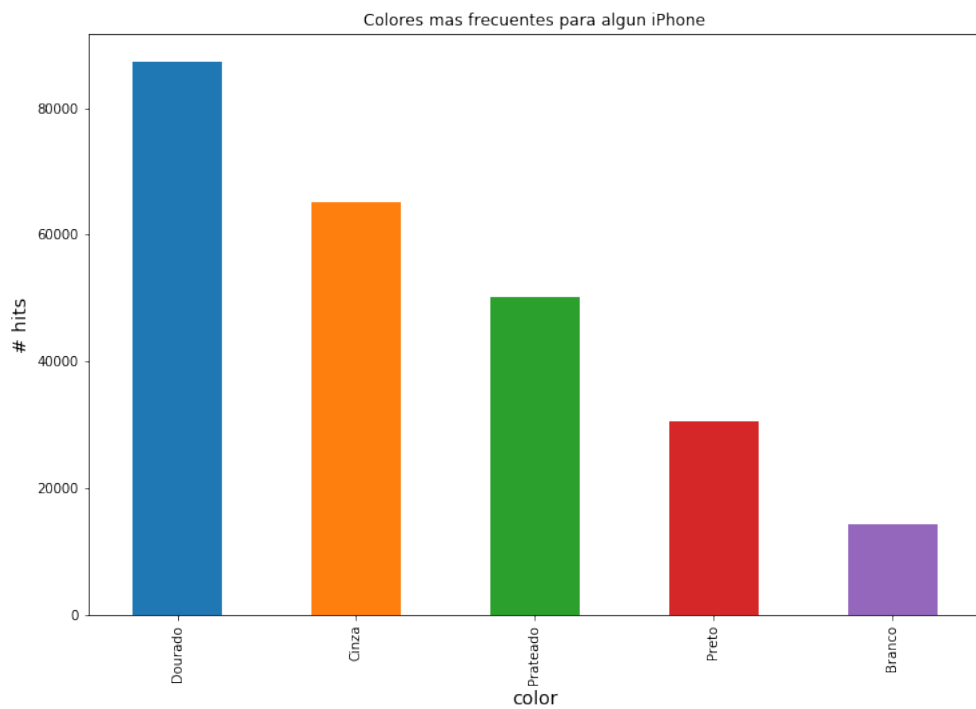


Figura 15: Colores mas frecuentes los iPhone

Pero como se puede ver también están en otros colores, pero en mucha menor medida que los primeros.

Para los Samsung lo que se obtuvo es lo siguiente:

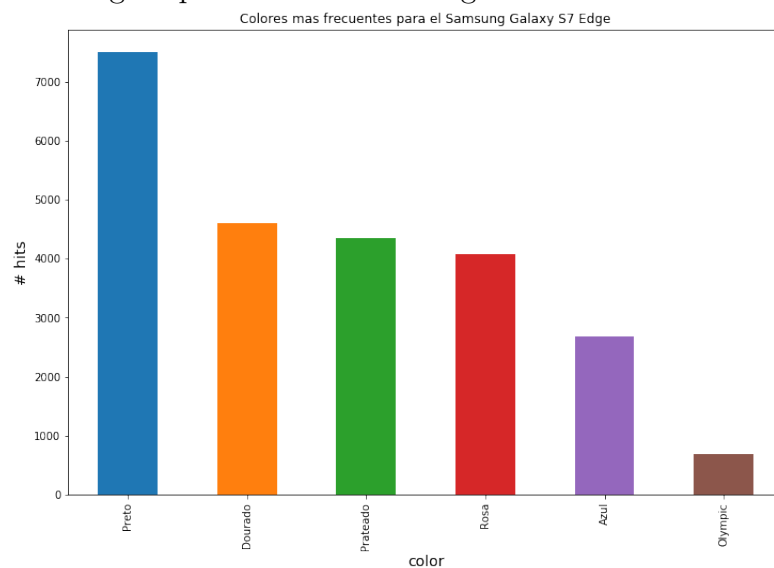


Figura 16: Colores mas frecuentes para Samsung Galaxy S7 Edge

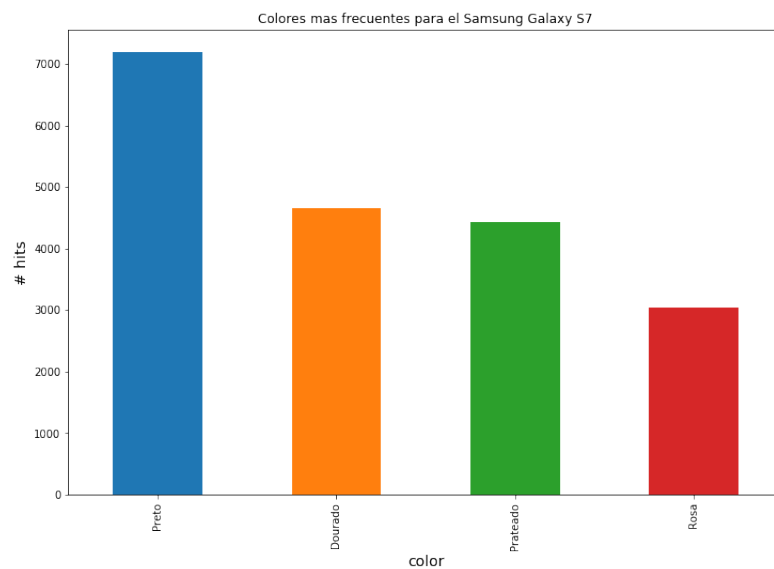


Figura 17: Colores mas frecuentes para Samsung Galaxy S7

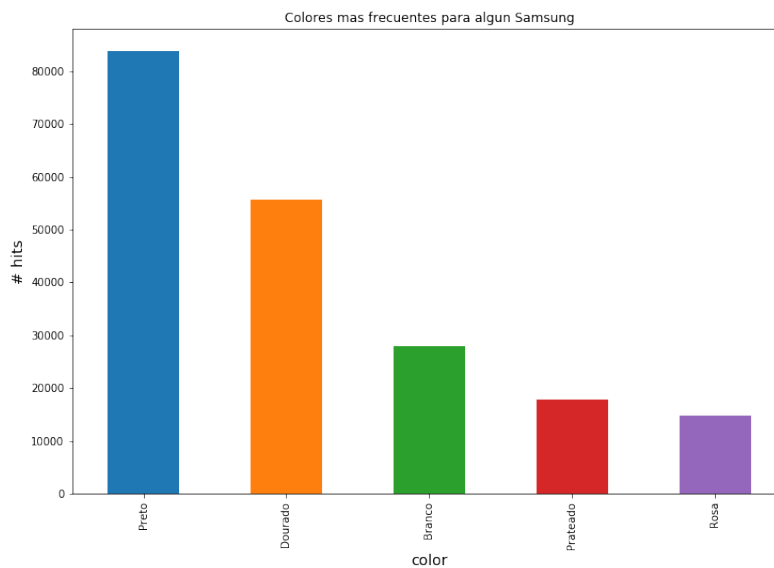


Figura 18: Colores mas frecuentes para los Samsung

Los primeros dos colores en todos los casos son el colores Preto y Dorado, que son mucho mas frecuentes que el resto de los colores.

Para los Motorola:

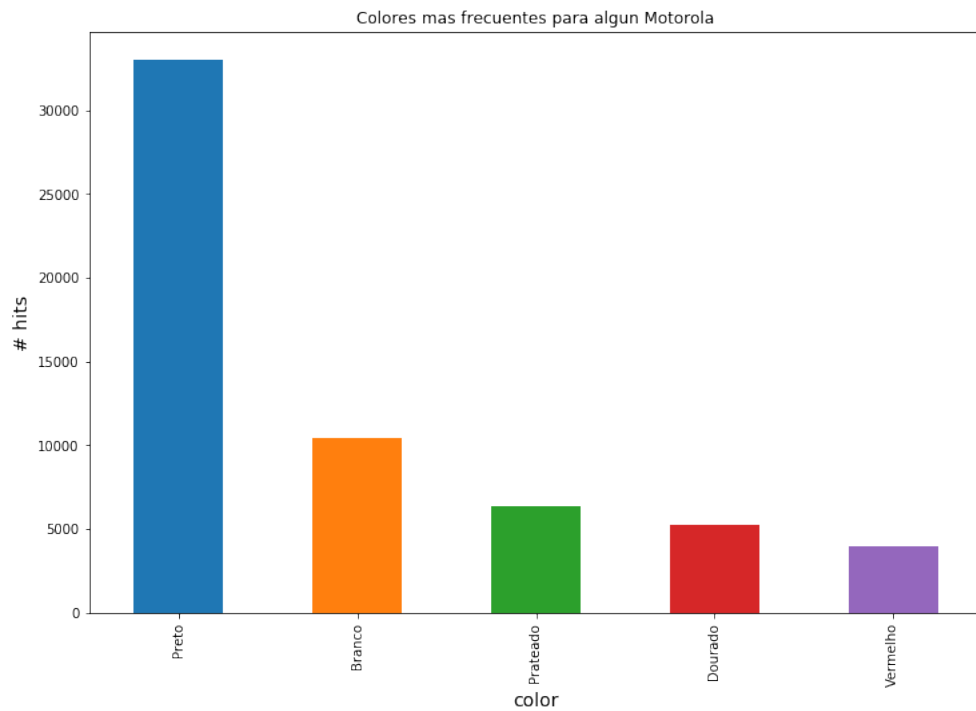


Figura 19: Colores mas frecuentes para los Samsung

Se puede ver que el Preto es muchísimo mas común que el resto. Lo que se puede pensar la próxima vez que aparezca uno que sea un Motorola es muy probable que sea de ese color.

3.6. Colores según mes

Ya habiendo analizado los colores de cada modelo podemos analizar la aparición de cada color a la largo de los meses.

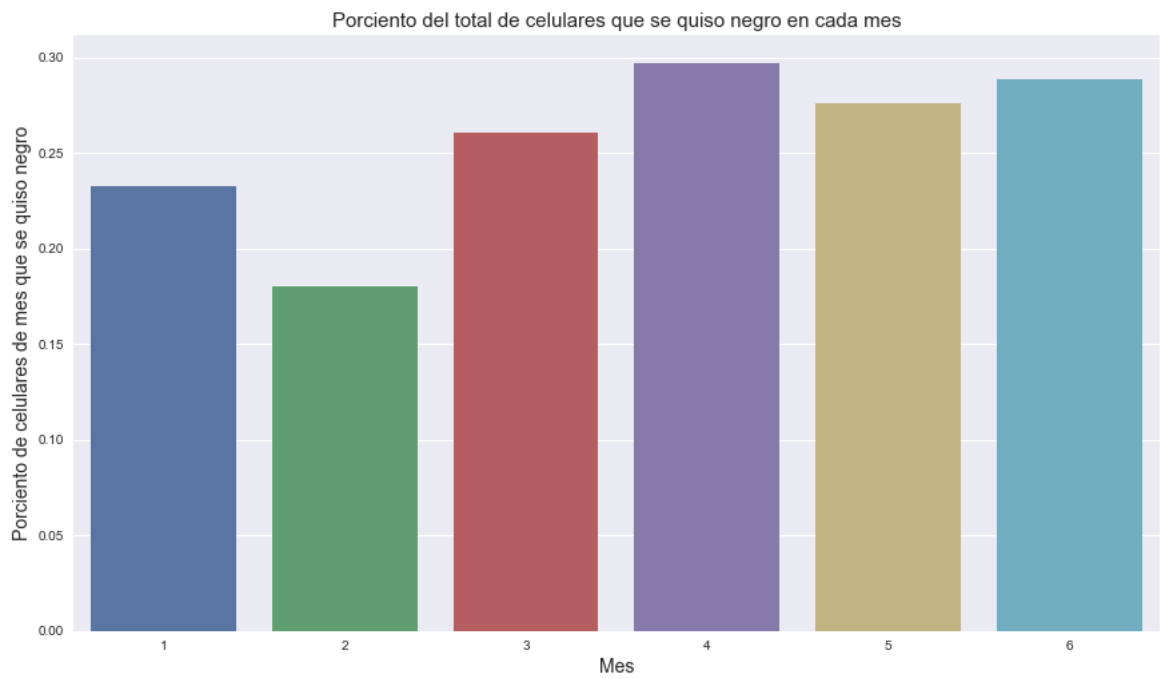


Figura 20: Porciento de vistas que recibió el color negro por mes

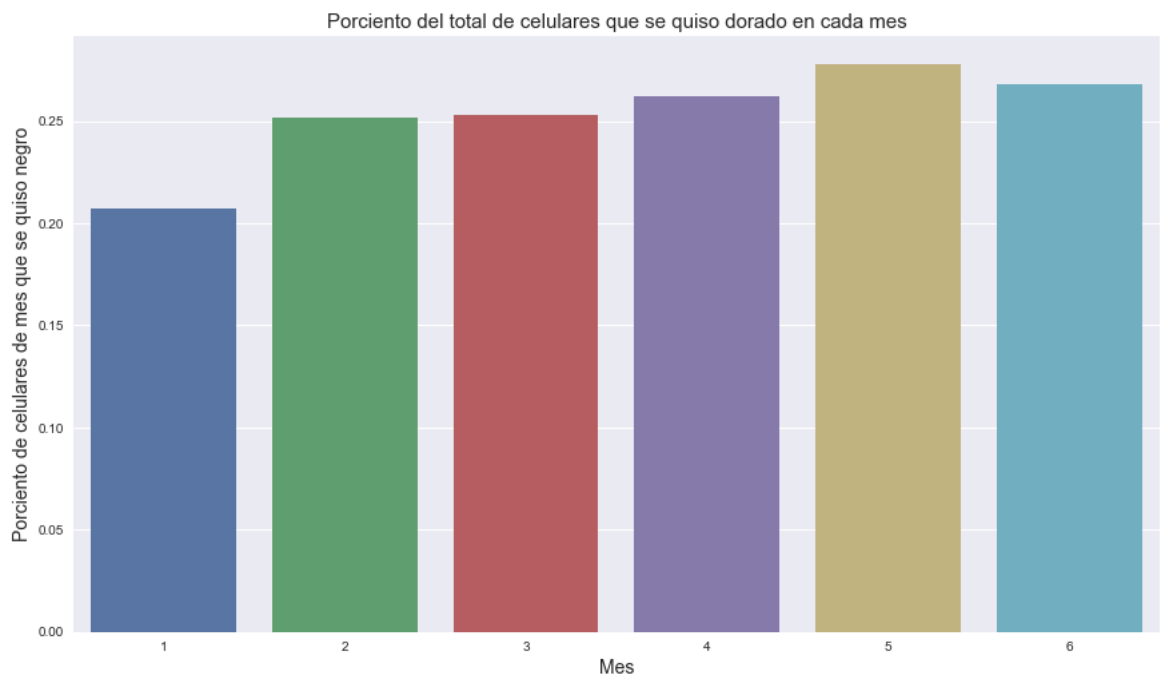


Figura 21: Porciento de vistas que recibió el color dorado por mes

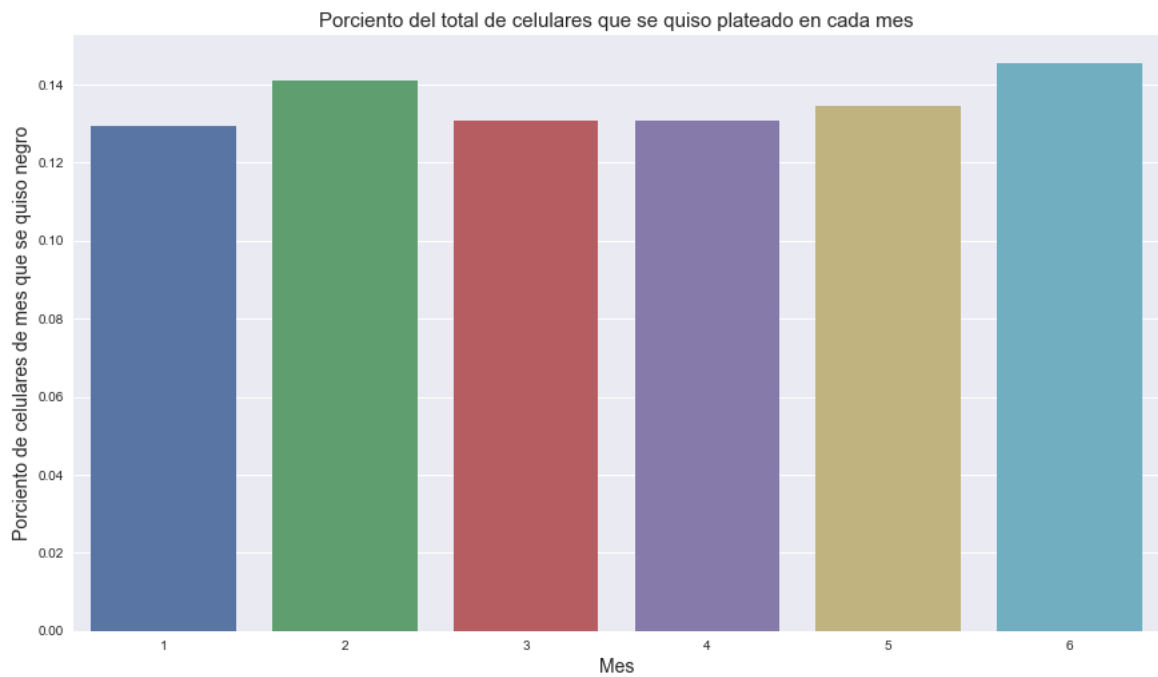


Figura 22: Porciento de vistas que recibió el color plateado por mes

Es notable que a diferencia de los otros dos gráficos, el color negro tiene un agujero en el mes 2, debido a esto investigamos si a simple vista hubo alguna anomalía en este mes, lo primero que intentamos es investigar si en este mes en particular cambiaron los modelos mas vistos.

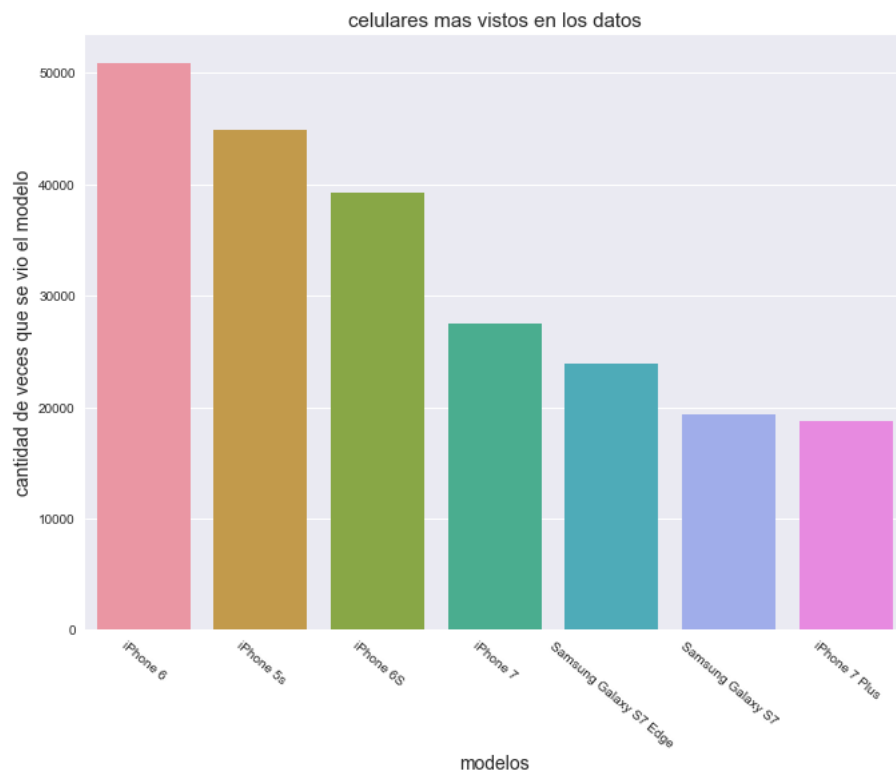


Figura 23: Modelos mas vistos en el transcurso de los datos

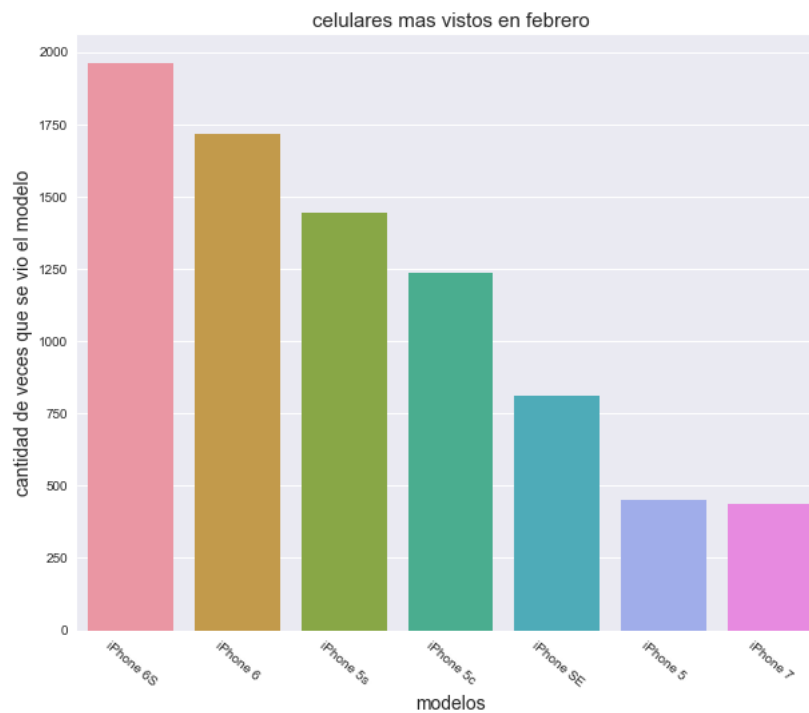


Figura 24: Modelos mas vistos en febrero

No es sorpréndete ya que los modelos mas vistos sean iPhone, sin embargo se ve que en febrero no hay ni un Samsung, que suele acompañar a los iPhone en el top del ranking. Revisando los colores podemos recordar que en iPhone no tiene la opción negro, mientras que para Samsung es la opción mas común. Esto nos deja en claro que o bien no hubo mucha demanda de Samsungs en febrero, o bien hubo mucha demanda de iPhones en el mismo mes.

3.7. Condición del Celular

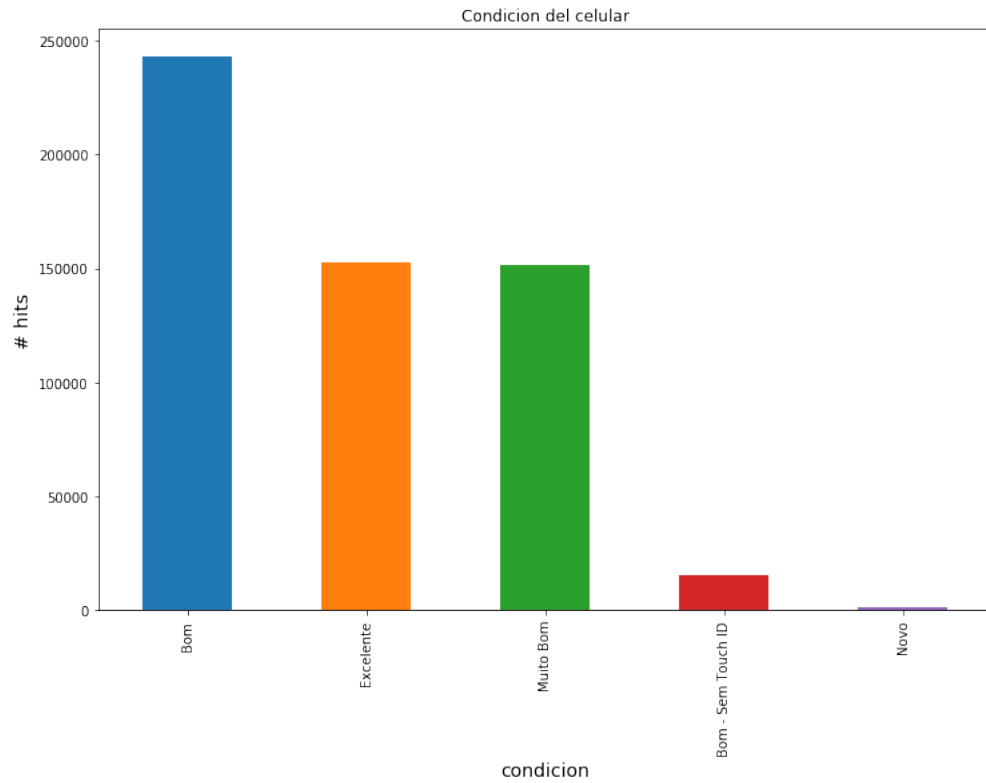


Figura 25: Condición de los celulares

Observando el previo gráfica se comprende que todos los modelos en general se encuentran en buenas condiciones por lo menos. Y muy pocos son los celulares que son nuevos.

En la siguiente sección se analizara la condición para los siguientes modelos en particular: iPhone, Samsung y Motorola

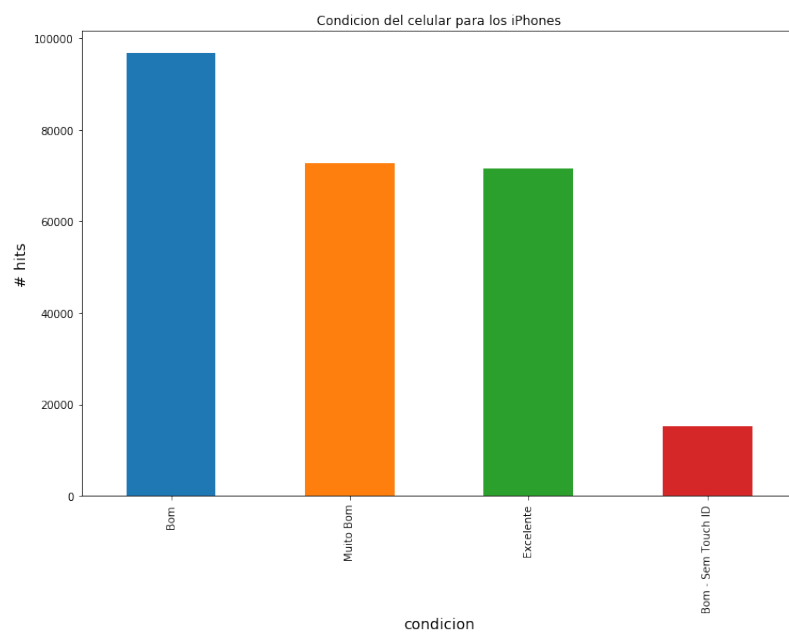


Figura 26: Condición de los iPhone

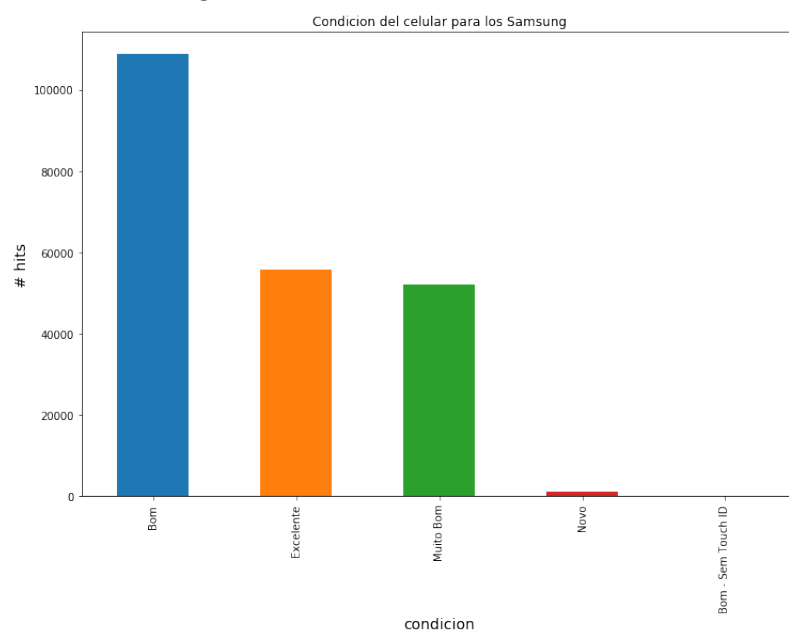


Figura 27: Condición de los Samsung

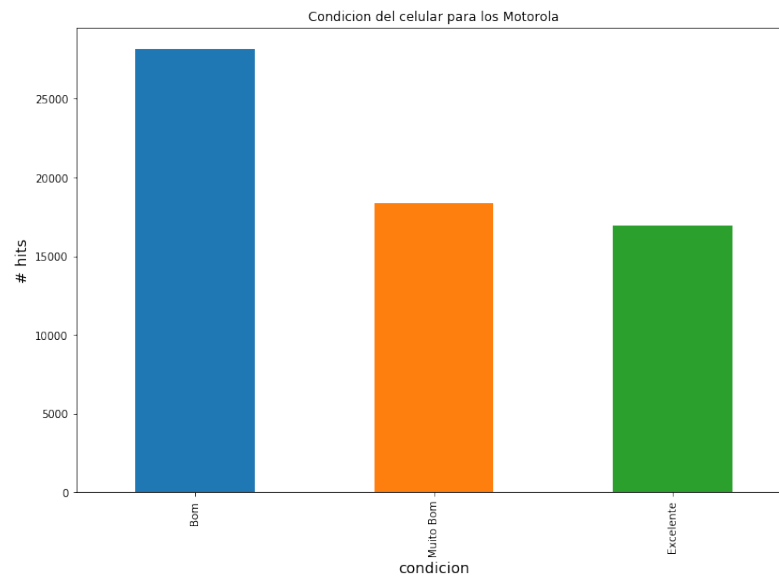


Figura 28: Condición de los Motorola

Viendo los gráficos solo los Samsung tienen algunos en condición nuevo. Y todos los modelos se encuentran en buena condición en su mayoría.

3.8. Hits de search engines y ads

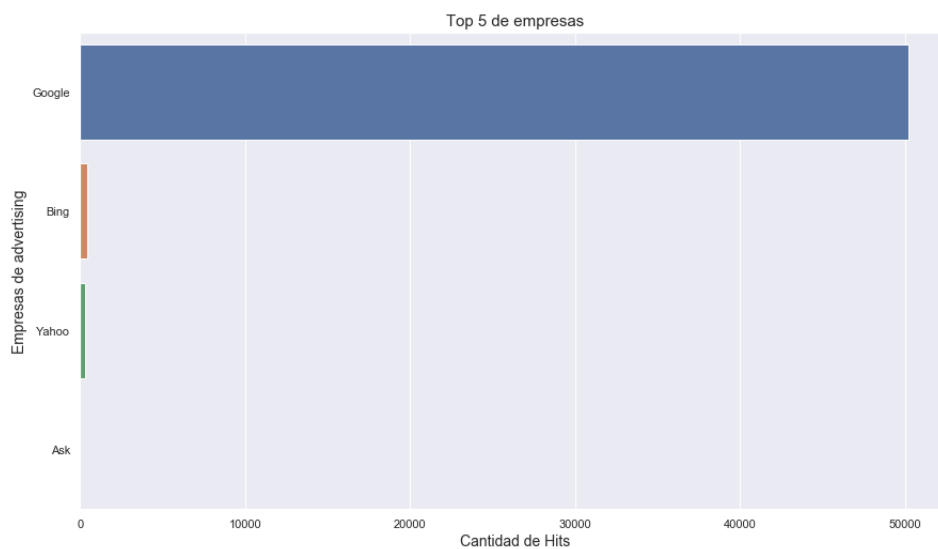


Figura 29: Cantidad de veces que un usuario ingresa desde cada buscador online

El previo gráfico deja en claro que no es muy valioso analizar los diferentes intereses de usuarios en función al buscador desde el que ingresan, sin

embargo existe un gráfico que sale de estos eventos al observar en qué día de la semana ocurren, y deja una cosa bastante en claro cuando se va de fiesta, es sabio dejar el celular en casa.

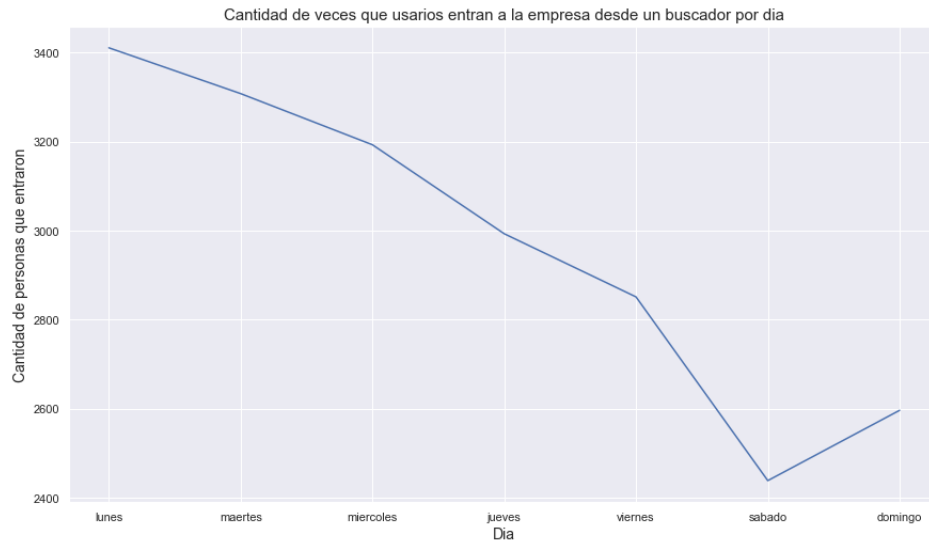


Figura 30: Visitas a la página a través de un buscador online por día
Para el caso de anuncios se observa el mismo patron.

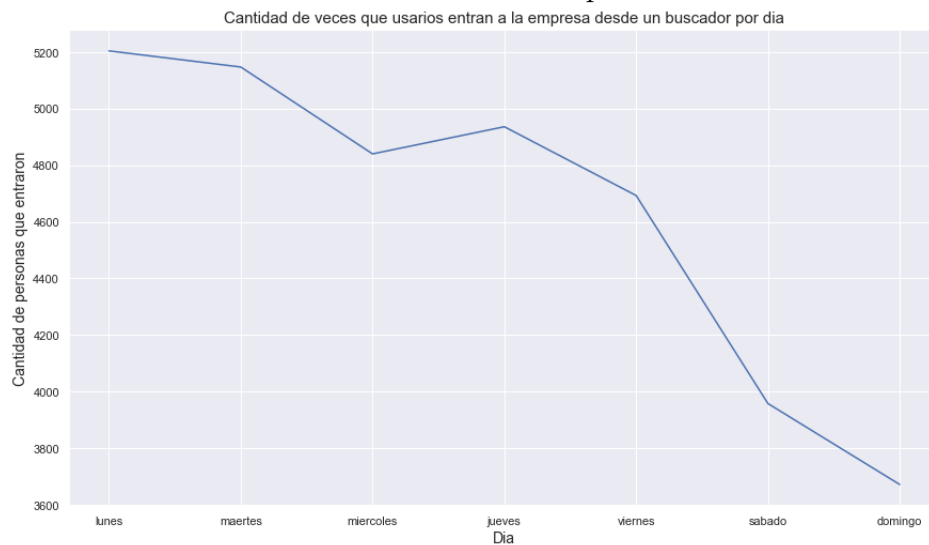


Figura 31: Visitas a la página a través de un anuncio por día

Tambien para el caso de visitas a la pagina debido anuncios se ve existe una pequeña competencia contra Google.

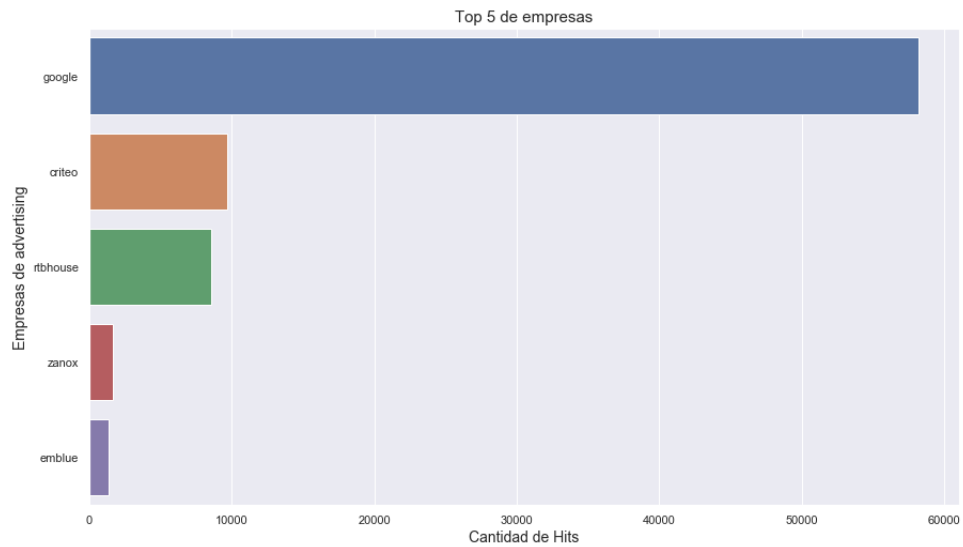


Figura 32: Visitas a la pagina por distintas empresas de anuncios
Queda claro que la fuente mas grande visitantes es Google ads. Se puede uno preguntar pensando en anuncios, cuando se los ve?

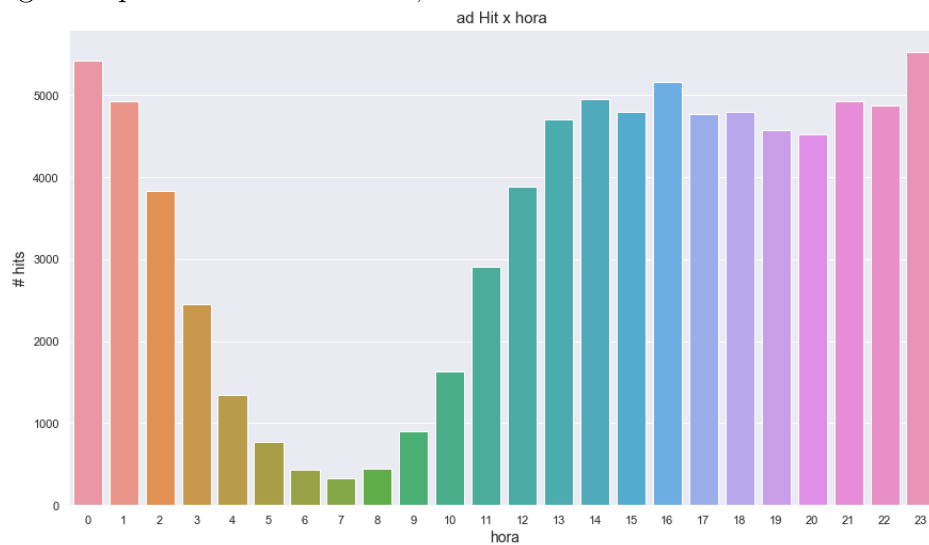


Figura 33: Visitas a la pagina por anuncios en cada hora del día
Los horas mas populares para una visitas mediante un anuncio son entre el mediodía y la noche. Luego se quiere aprender si este comportamiento se comparte entre todas las empresas de anuncios. Para lograr esto se implemento un stack-barplot.

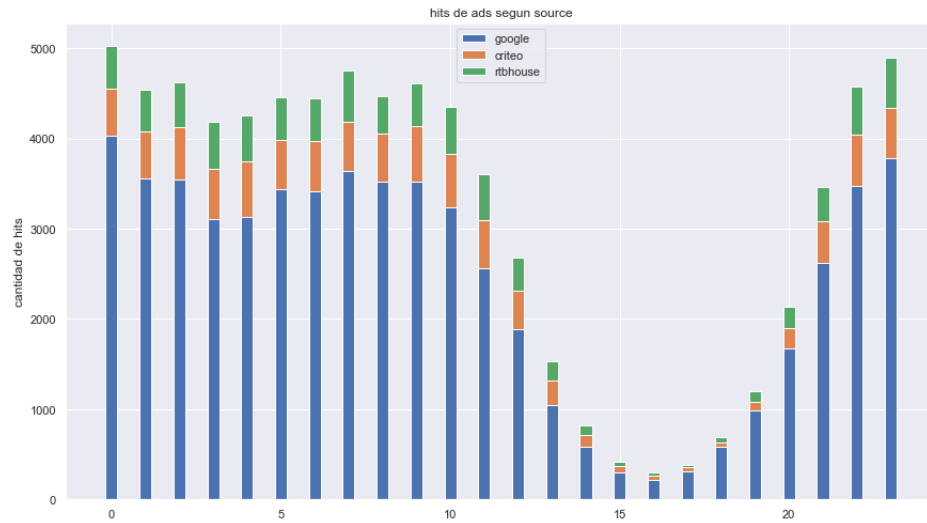


Figura 34: Visitas a la pagina por distintas empresas de anuncios en cada hora

Se observa que el comportamiento por hora de todas las empresas de anuncios es parecido.

La última pregunta que se puede hacer de este subset de datos es si los usuarios son siempre los mismos. Para ambas situaciones la respuesta es si, la cantidad de usuarios que retornan por el mismo medio es mínima como se puede observar a continuación.

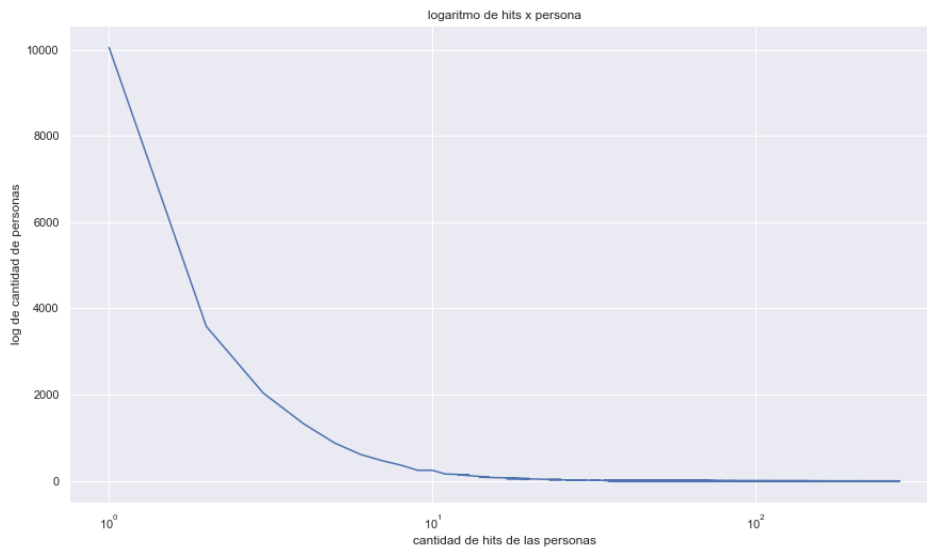


Figura 35: Logaritmo de cantidad de personas por cantidad de veces que acceden al sitio mediante un anuncio

3.9. Relación de Regiones por compras

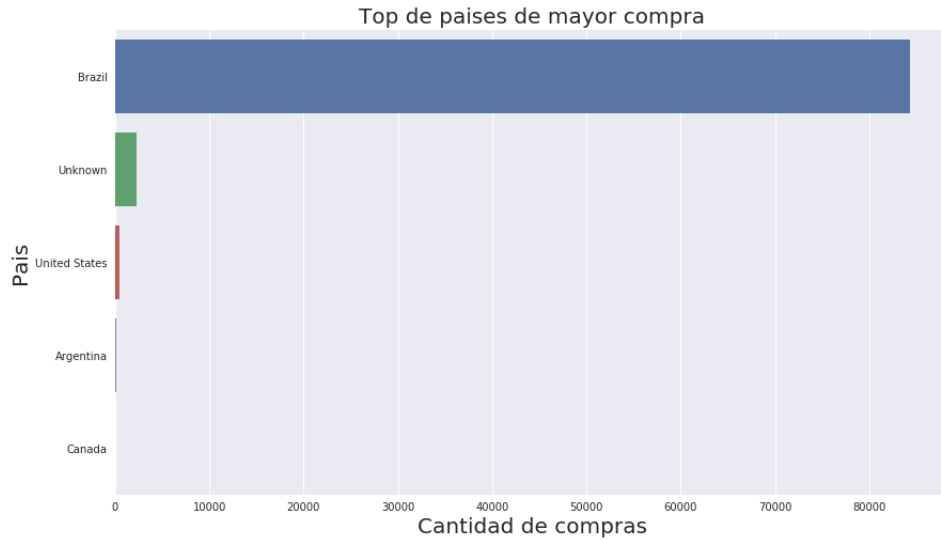


Figura 36: Top 5 de países de mayor compra de celulares (incluyendo compras sin país especificado)

Del gráfico anterior se puede apreciar que de las compras con país especificado donde se hizo la compra, la gran mayoría provienen de Brasil. De ahí la distribución de ventas baja abruptamente, siendo el segundo país con mas compras Estados Unidos aunque muy por detrás de Brasil. Los demás países no llegan ni a 100 ventas de celulares. Entonces se puede concluir que la mayor fuente de clientes de la empresa proviene de Brasil. Por otro lado se tienen casi 5000 celulares comprados en lugares desconocidos.

Veamos ahora como se distribuyen los compradores en los 3 países principales (Brasil, Estados Unidos y Argentina).

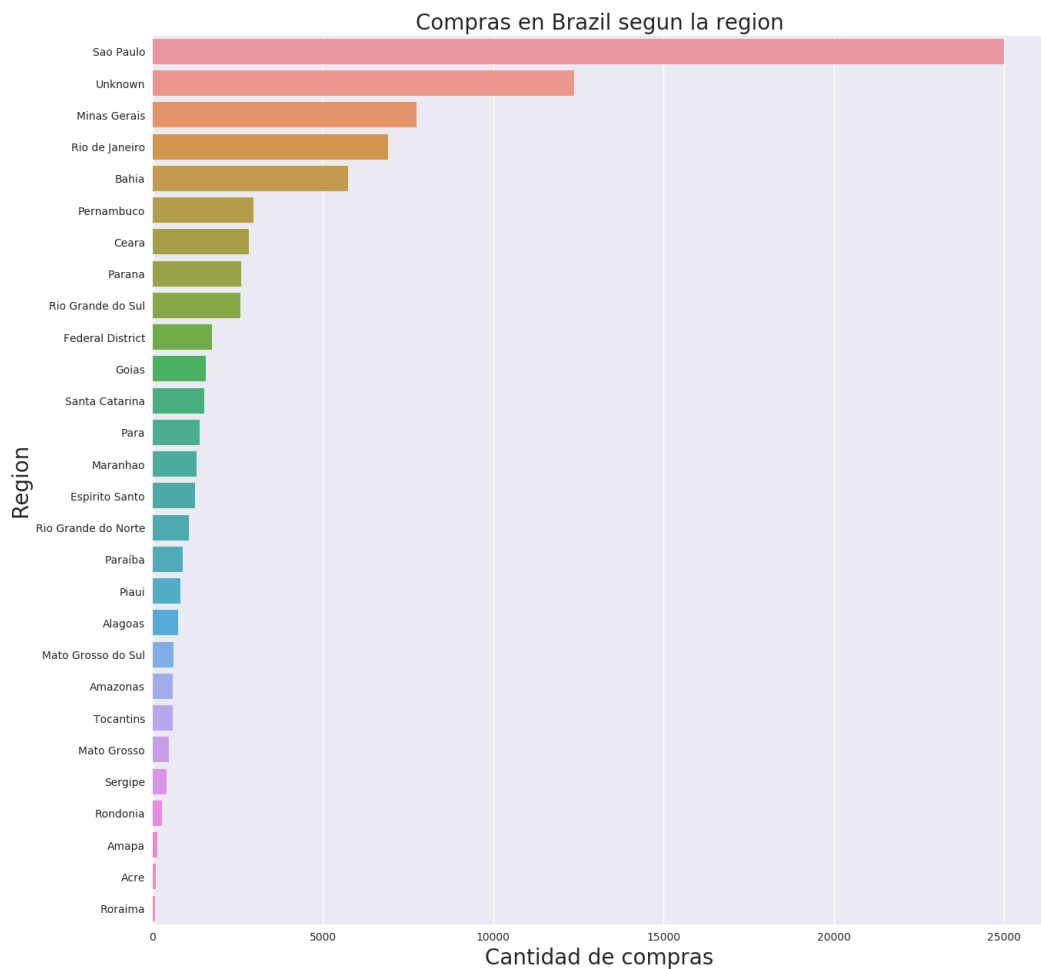


Figura 37: Cantidad de compras en Brasil por las regiones donde se realizaron las compras (incluyendo compras sin ciudad especificada)

Se puede ver que la mayoría de las compras hechas en Brasil se realizaron en grandes ciudades tales como San Pablo o Rio De Janeiro. Por lo que se podría concluir que los brasileños tienden a comprar celulares de Trocafone en las ciudades mas importantes de Brasil y de mayor población, ya que las mayores ventas por cantidad de celulares se realizaron en ciudades con poblaciones entre los 10 y 20 millones de habitantes, salvo por Rio de Janeiro que tiene 6 millones, no obstante es un gran centro turístico y una de las ciudades mas importantes de Brasil. Por lo que los brasileños en general compran celulares y en proporcionalidad a la cantidad de habitantes de cada ciudad.

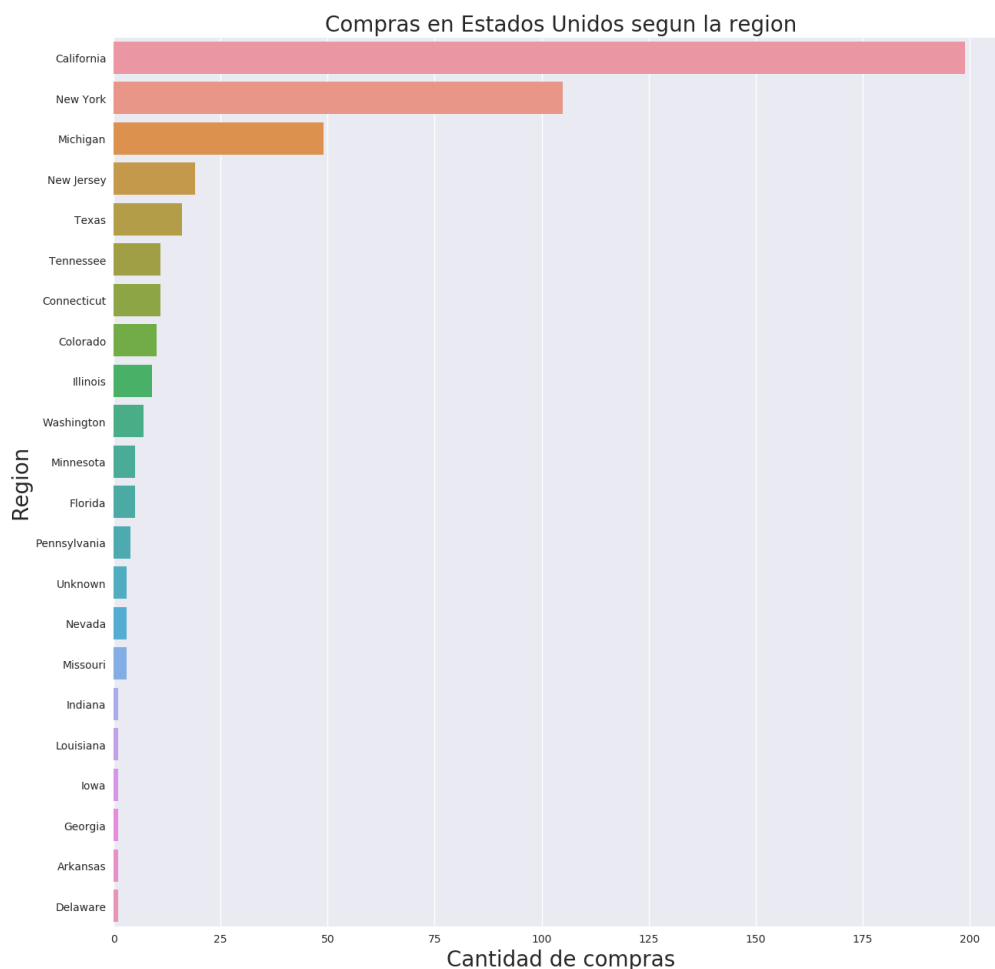


Figura 38: Cantidad de compras en Estados Unidos por los estados donde se realizaron las compras (incluyendo compras sin estado especificada)

Analizando de forma similar al gráfico de Brasil, se puede ver que la mayoría de compras se realizaron en California y la zona de New York y New Jersey. Aunque a diferencia de Brasil acá la brecha de compras es significativamente menor, habiendo apenas 200 compras en California y unas 100 en New York. Por lo que no es muy concluyente este gráfico salvo que un punto de ventas para la empresa seguro es California.

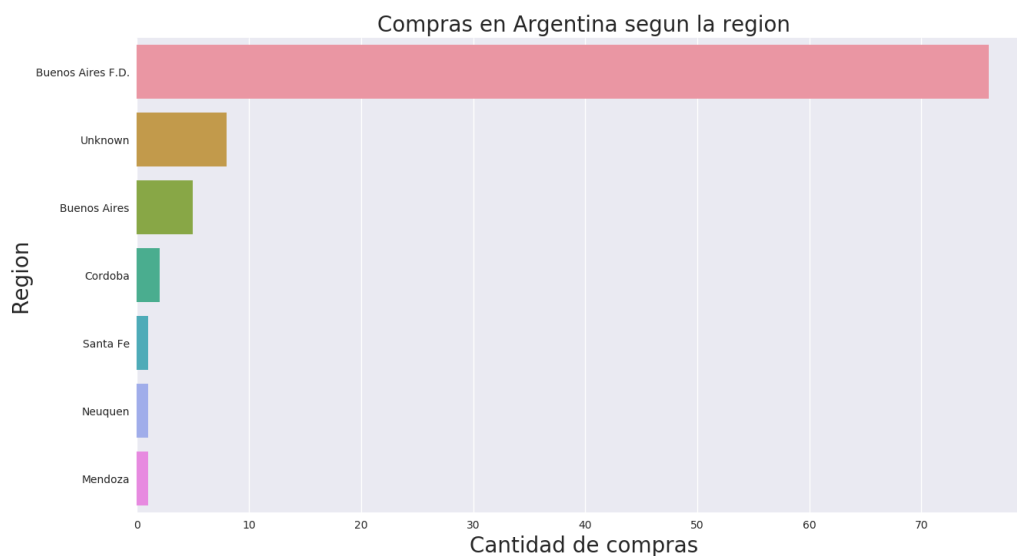


Figura 39: Cantidad de compras en Argentina por las zonas donde se realizaron las compras (incluyendo compras sin estado especificada)

Del gráfico de Argentina se puede ver que el único lugar con ventas significativas es Capital Federal, siendo el resto de ventas puntos aislados con apenas 1 venta realizada.

3.10. Canales de ingreso

En esta sección podemos ver cuales son los canales más frecuentes para ingresar al sitio web, que como podemos ver los más frecuente son los canales pagos, que suponemos que son anuncios en otras páginas web. Además, podemos ver que no hay muchos ingresos por e-mail, lo que da a entender que no hay mucha gente registrada al newsletter del sitio.

También hay un gráfico que muestra los canales con más frecuencia en los primeros seis meses del 2018, en el que podemos ver que el canal pago siempre es el más frecuente de cada mes, destacando una gran crecida en el mes de mayo.

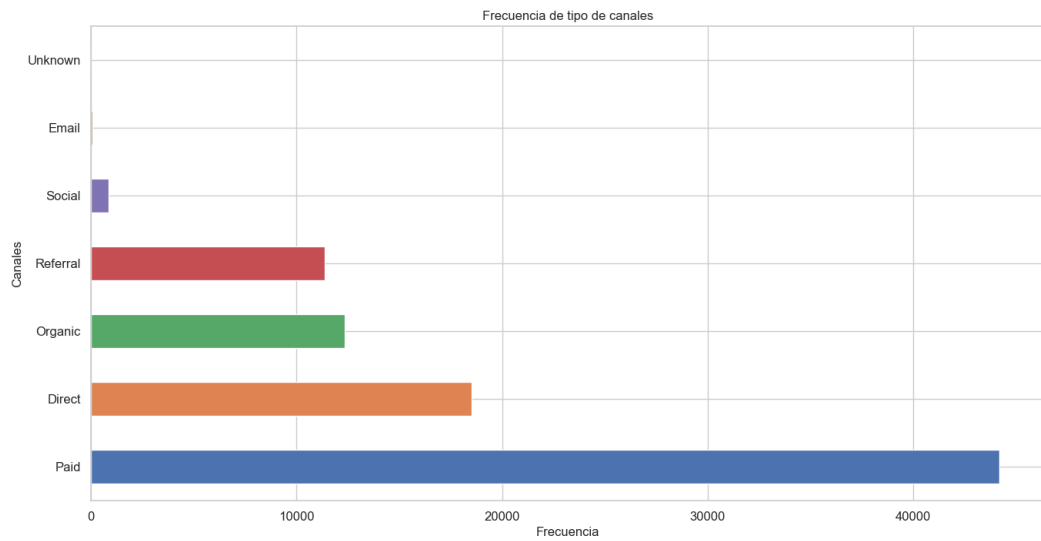


Figura 40: Canales y su frecuencias

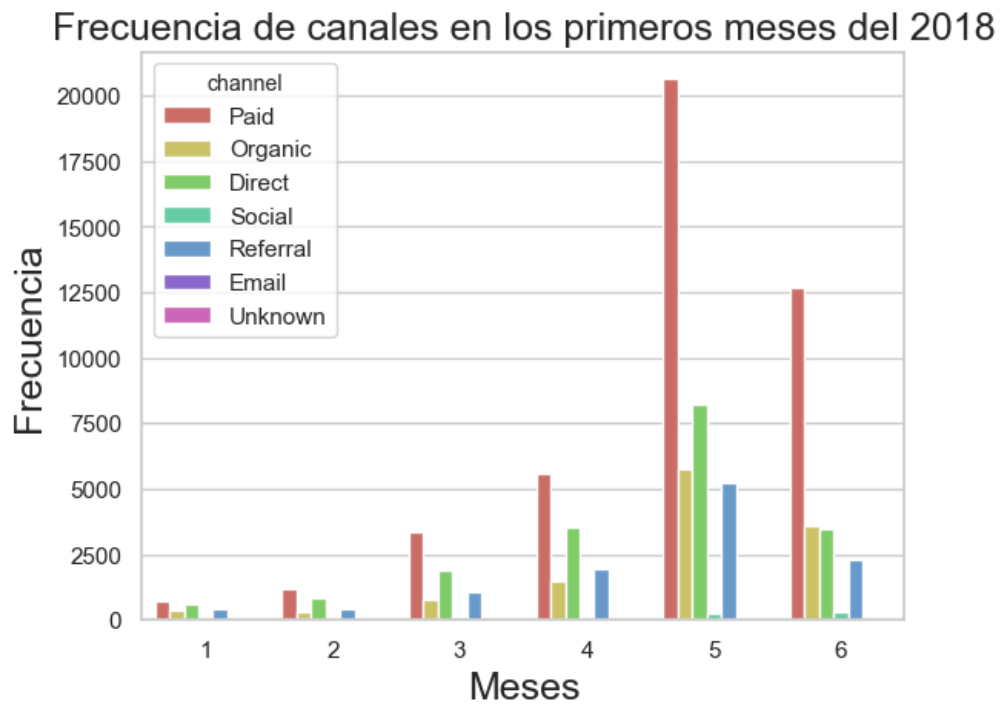


Figura 41: Canales y su frecuencias por mes

3.11. Términos de búsqueda

En esta sección se puede ver gráficos con los cinco términos más buscados, siendo los celulares más frecuentes los iPhones, siendo lo más buscados lo que

sería su línea de gama media como lo son el iPhone 5 y iPhone 6. Sin embargo, hay términos ajenos al iPhone que se cuelan en el top 5, como por ejemplo ‘j5’, aludiendo al Samsung Galaxy Prime J5, otro celular de gama media.

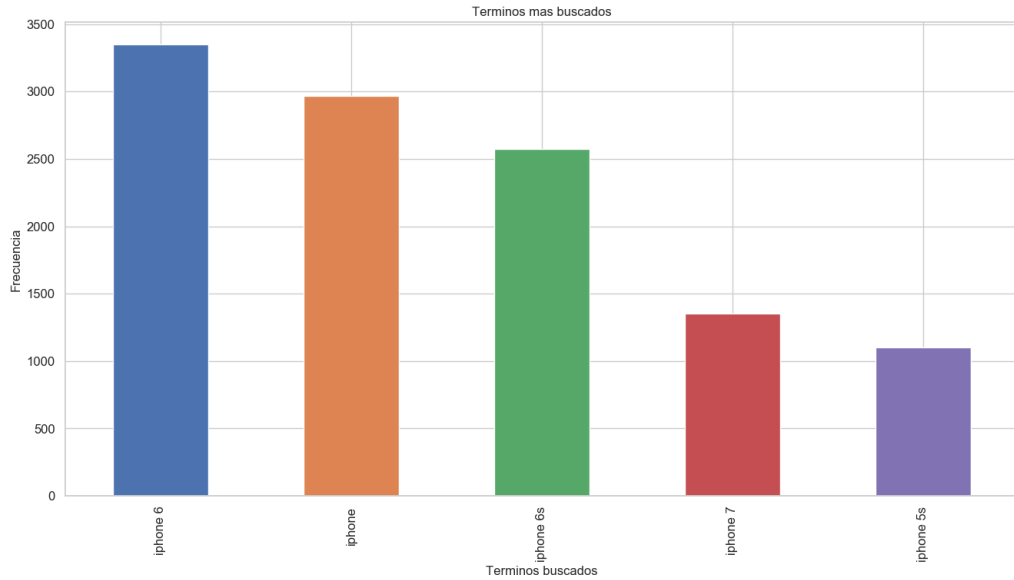


Figura 42: Top 5 términos frecuentes

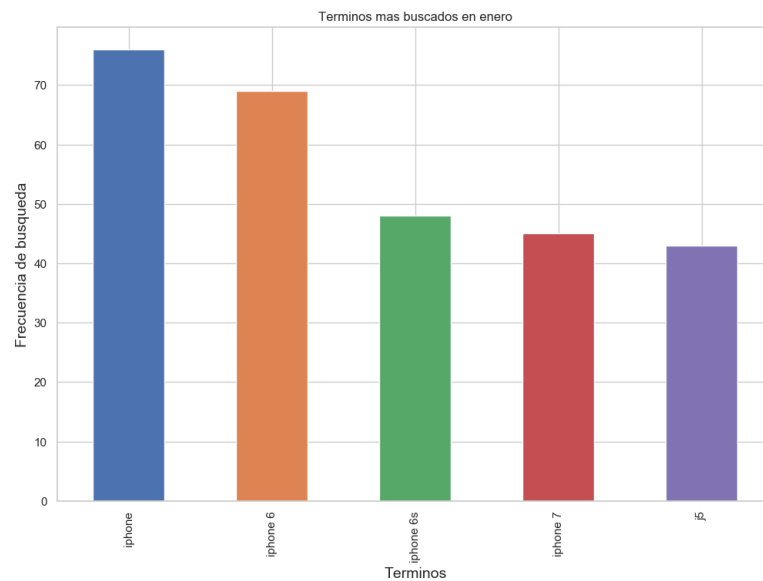


Figura 43: Top 5 términos frecuentes en Enero

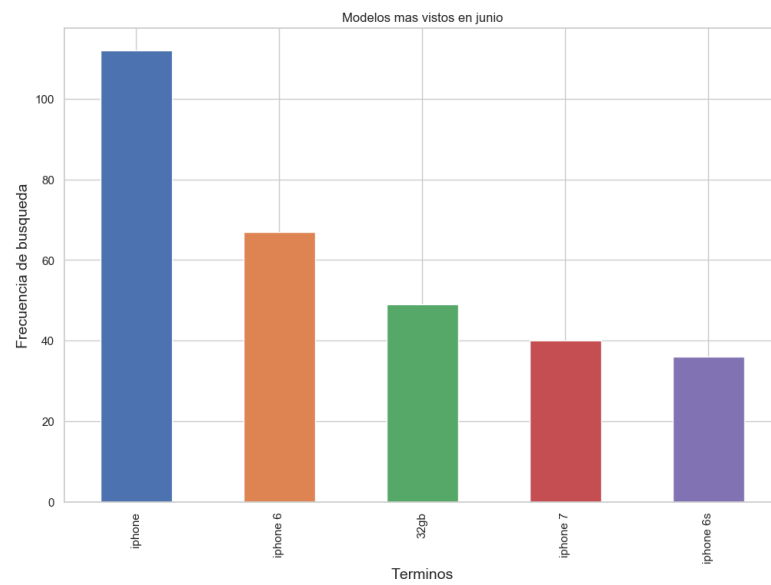


Figura 44: Top 5 términos frecuentes en Febrero

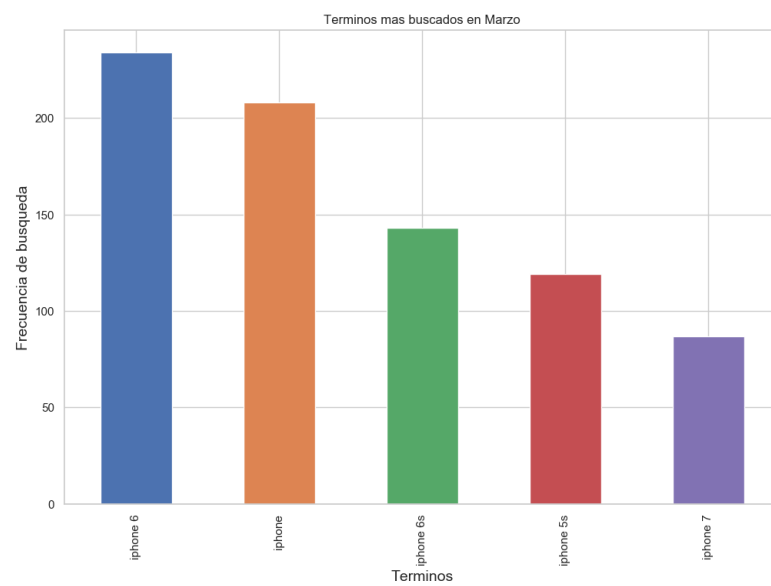


Figura 45: Top 5 términos frecuentes en Marzo

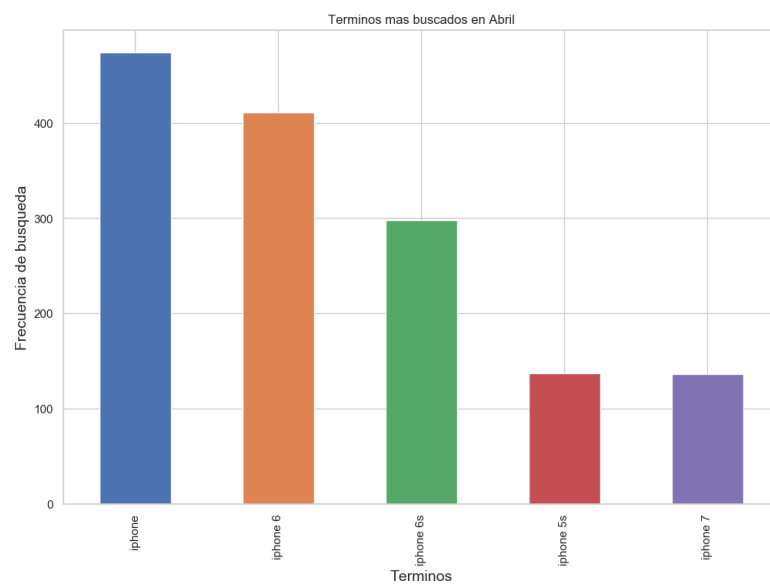


Figura 46: Top 5 términos frecuentes en Abril

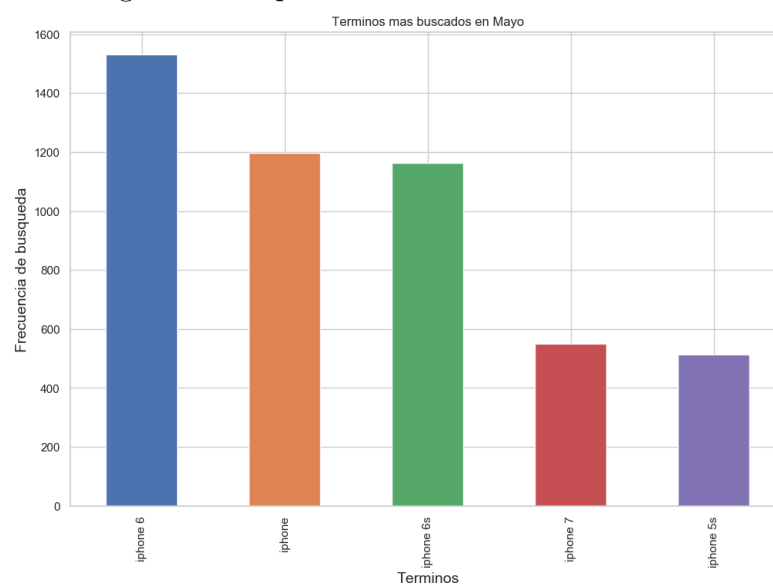


Figura 47: Top 5 términos frecuentes en Mayo

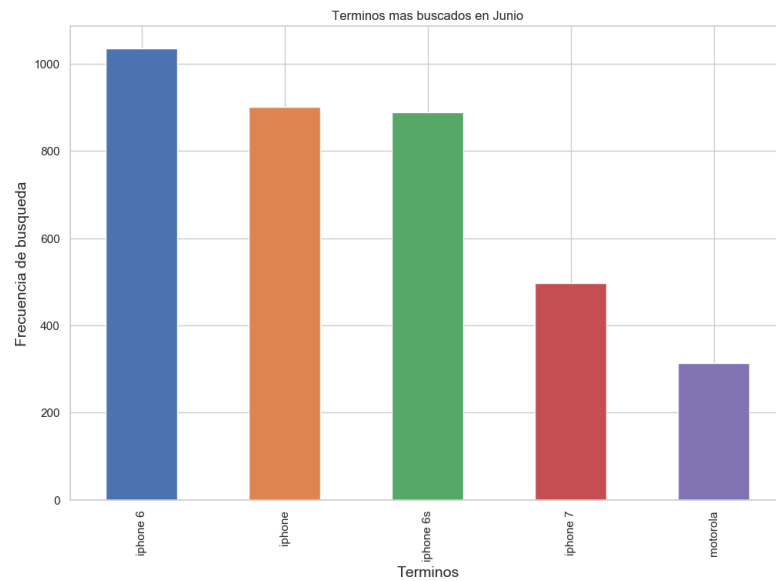


Figura 48: Top 5 términos frecuentes en Junio

3.12. Usuarios Nuevos vs Usuarios que regresan

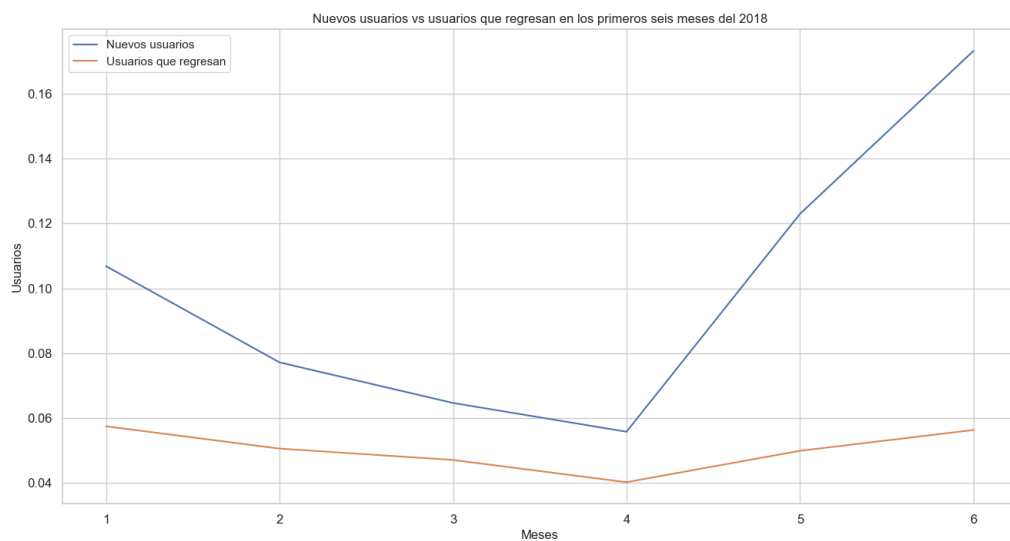


Figura 49: Usuarios Nuevos contra Usuarios que regresan

En el gráfico previo se puede observar que son mas los usuarios nuevos que los que regresan, por lo tanto dando entender que no hay un muy buena retención de los usuarios. Este gráfico esta normalizado, siendo dividido por la cantidad de visitas mes a mes.

4. Relación entre eventos

Mientras se analizaba el set de datos para la realización del Trabajo Práctico se noto que le faltaba relación entre los diferentes eventos. Imposibilitando responder preguntas como, los usuarios de safari buscan más por un iPhone que el promedio general?. Las conexiones sin embargo se pueden realizar separando los datos por eventos, y luego uniéndose por persona. Así se logró conseguir dibujar gráficos que se pueden ver a continuación como, porcentaje de vistas que tiene un modelo, dependiendo de navegador. Debido a la cantidad de datos utilizado nos pareció correcto no distinguir entre eventos como compra o visto, la hipótesis por la que nos parece correcto esto, es que la proporción entre vistas, compras u otro tipo de operaciones para un modelo se mantienen igual entre todos los modelos. Por lo tanto, cuando se hable de visto, se considera tanto visita a la pagina del modelo, tanto como la compra del mismo. Como argumento a favor de esta hipótesis se presentan los siguientes gráficos mostrando la relación cuantitativa entre diferentes acciones para diferentes modelos.

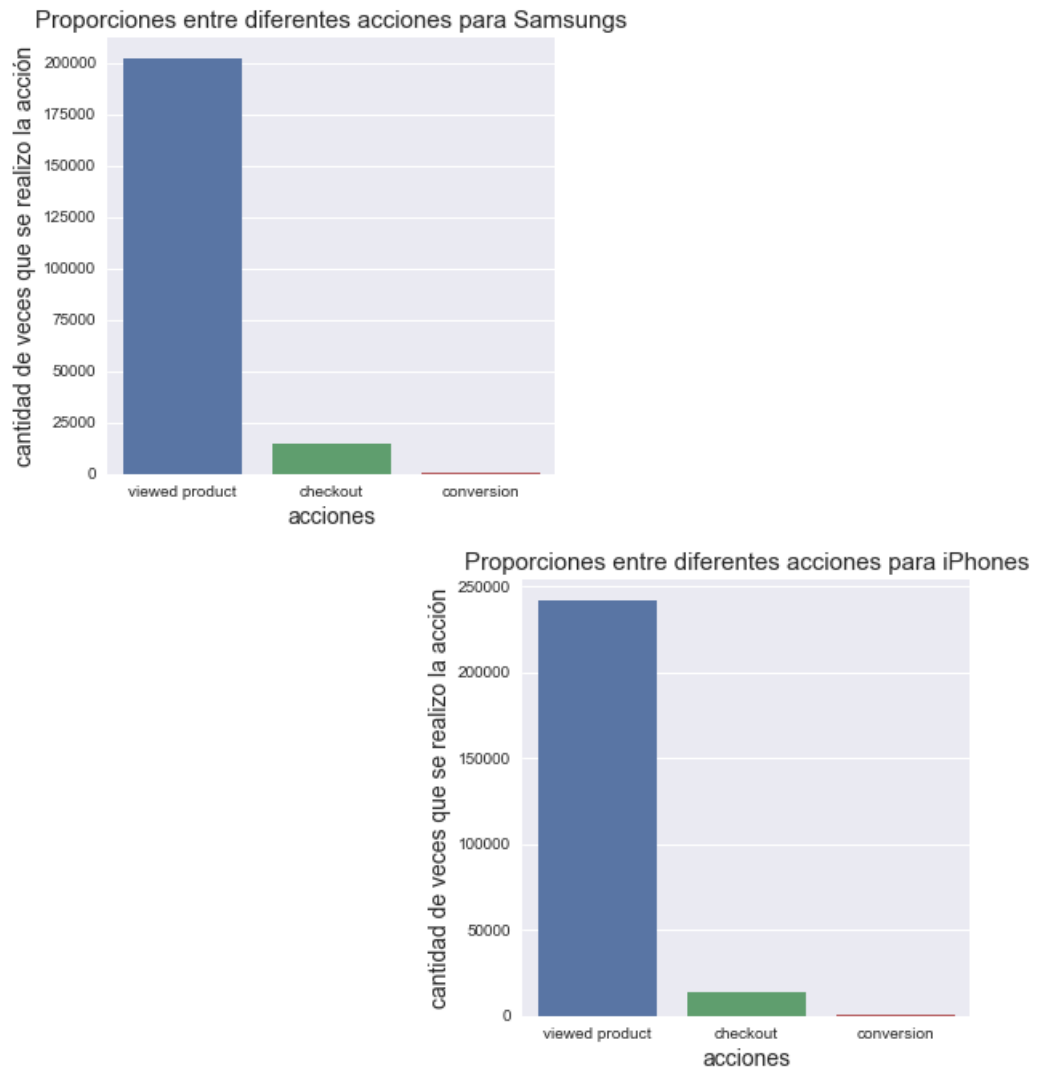


Figura 50: Muestra cuantitativa de las proporciones de diferentes acciones para Samsungs y iPhones

4.1. Navegadores y Modelos

Como ya se indico anteriormente es posible al asignarle a una persona un navegador y agregar estos datos a otro evento, crear relaciones como que marca tienen de preferencia los usuarios de cada navegador. Antes que nada, observemos cuanto se utiliza cada navegador, y también que tantas visitas en promedio tiene los usuarios de cada navegador.

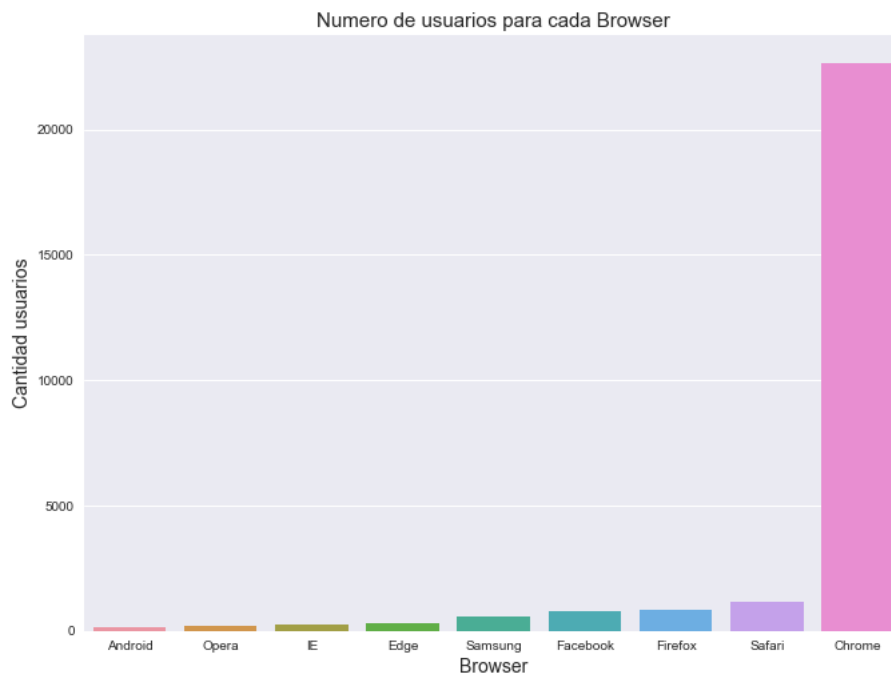


Figura 51: numero de usuarios de cada navegador

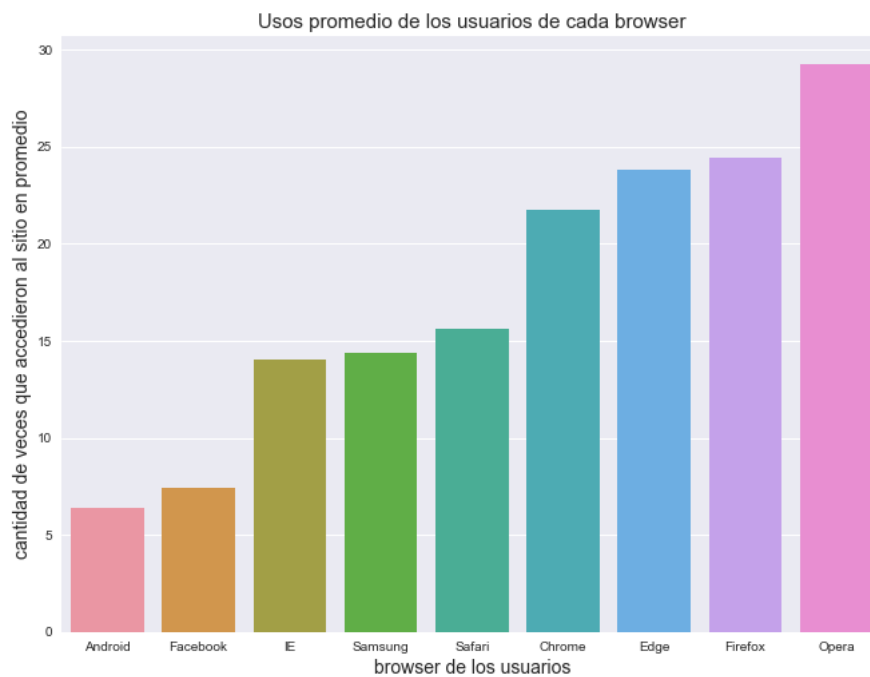


Figura 52: Promedio de visitas de usuarios de diferentes navegadores
 Antes que nada lo primero que se aprende de los gráficos observados es que se debe trabajar con porcentajes en este preciso análisis de otra manera

todos los resultados van a tener la misma forma que el número de usuarios de cada navegador, debido a la desproporcionada cantidad de usuarios de Google Chrome.

La siguiente observación que se debe hacer es que lo mismo sucede para la cantidad de visitas por navegador, a pesar de que no se ve una desproporción tan grande, sigue siendo cierto que los usuarios en promedio para un navegador Android realizan 5 visitas a la página, mientras que los usuarios de Opera realizan 30.

A continuación observamos que porcentaje de visitas de los usuarios de un navegador en específico reciben diferentes marcas.

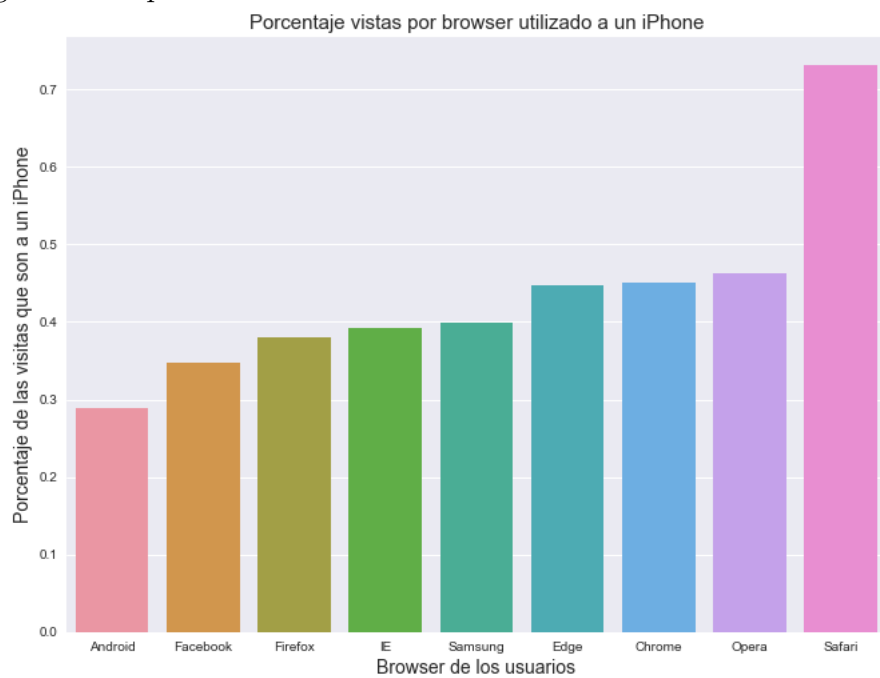


Figura 53: Porcentaje de las visitas de cada navegador a modelos iPhone

Figura 54: Porcentaje de las visitas de cada navegador a modelos Samsung

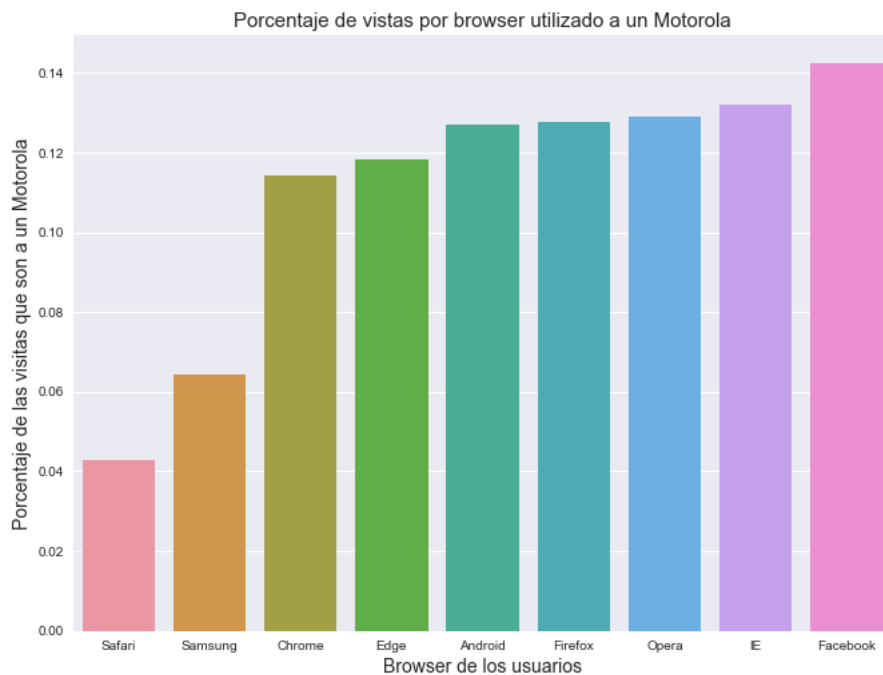


Figura 55: Porcentaje de las visitas de cada navegador a modelos Motorola

Lo que uno supone lo que van a dar los resultados de estos gráficos imagina que los usuarios de Safari serán los que más iPhone buscan, y que los usuarios del navegador Samsung son los que más Samsung buscan. Y esto fue cierto. Por el otro lado algo mucho más llamativo fue revelado, tanto los usuarios de navegadores Samsung como Safari tienen un interés muchísimo más bajo en modelos de otras Marcas, como en el ejemplo expuesto, Motorola.

4.2. Navegadores y Estado

Podemos seguir el análisis relacionando los usuarios de cada navegador y el estado de celular que visitan. En particular como ya se había visto antes, los 3 estados relevantes son Excelente, muy bien y bien. Los resultados para estos estados en porcentaje a continuación.

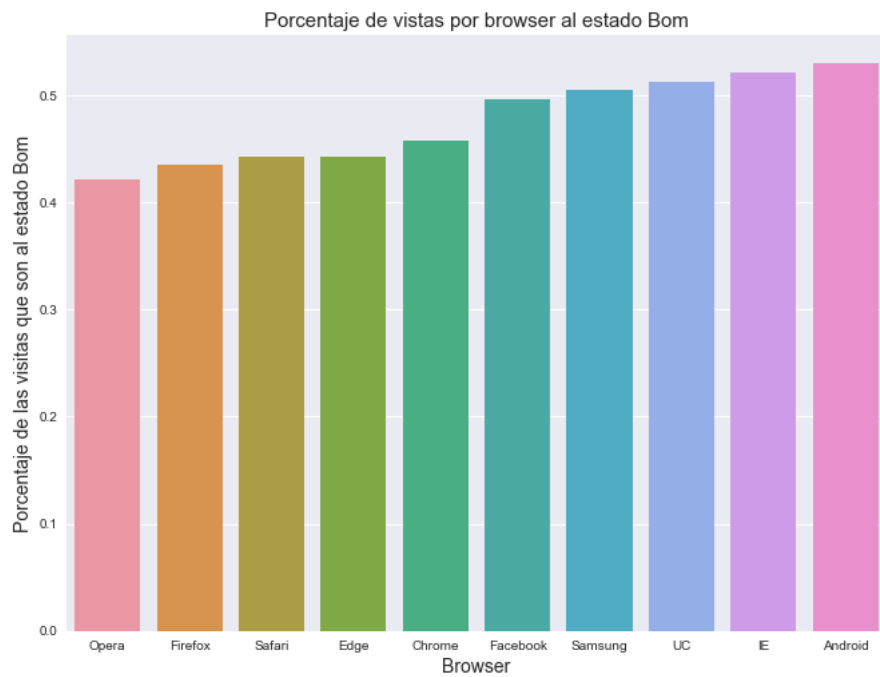


Figura 56: Porciento de vistas para usuarios de cada navegador que pertenecen a modelos en estado bueno

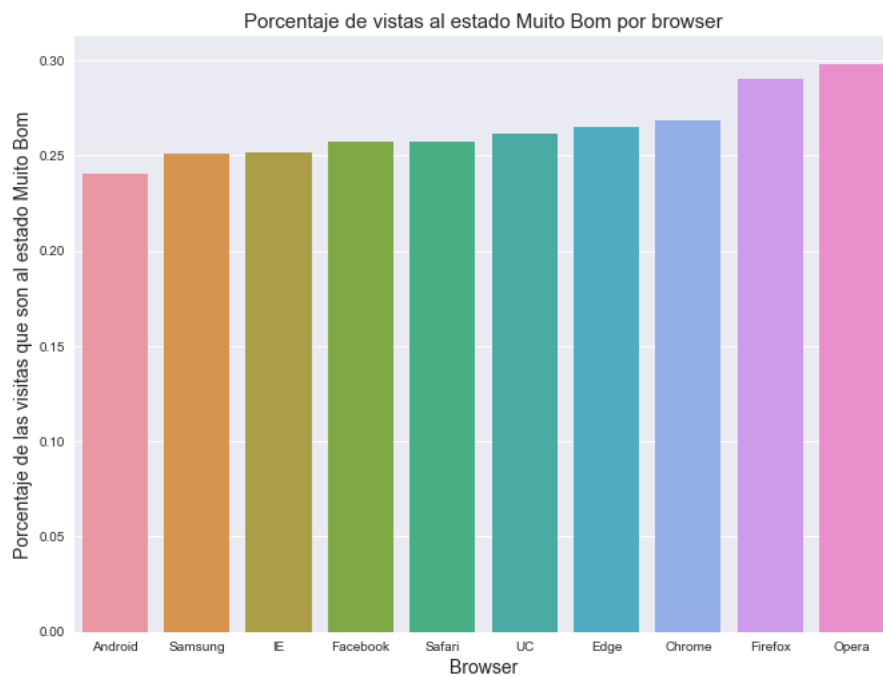


Figura 57: Porciento de vistas para usuarios de cada navegador que pertenecen a modelos en estado muy bueno

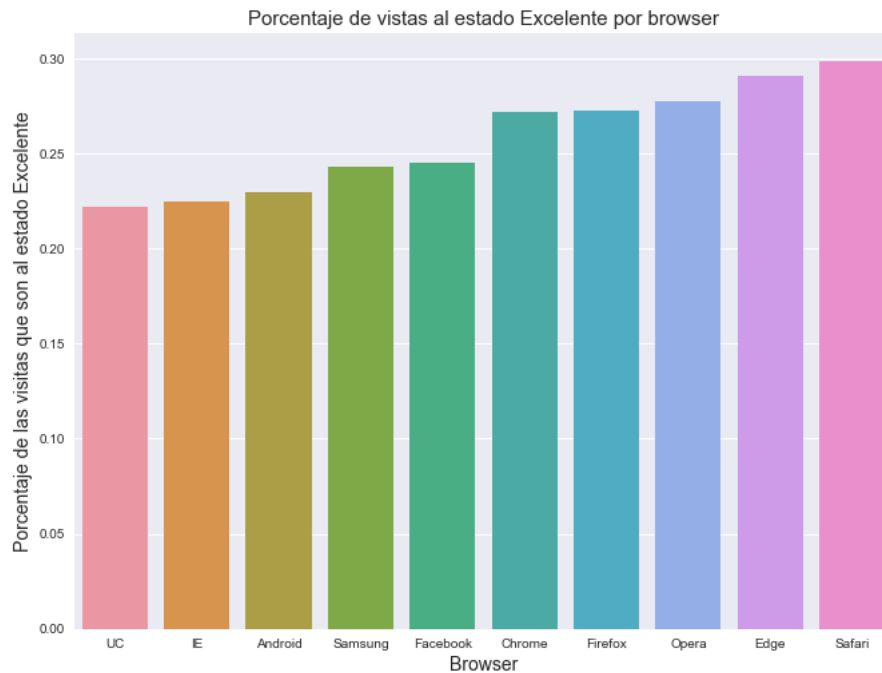


Figura 58: Porcentaje de vistas para usuarios de cada navegador que pertenecen a modelos en estado excelente

No se presenta ninguna gran irregularidad en lo visto, lo único que se puede constatar es que usuarios de Safari, Edge y Opera parecen optar mas por dispositivos en mejor estado, mientras que usuarios de Android ,Internet Explorer y el navegador de software chino UC les interesan mas dispositivos en condición buena, a su vez también es notable que los usuarios del navegador de Facebook siempre se mantienen en el medio.

4.3. Checkout vs Compras

En este apartado se pueden ver gráficos comparativos entre los eventos ‘checkout’ y ‘conversion’, es decir, las compras realizadas. Se puede ver que hay una gran diferencia, además de una caída de compras de Abril a Mayo, que puede deberse a un mayor caudal de visitas al estar normalizado, siendo dividido por los ingresos de cada mes.

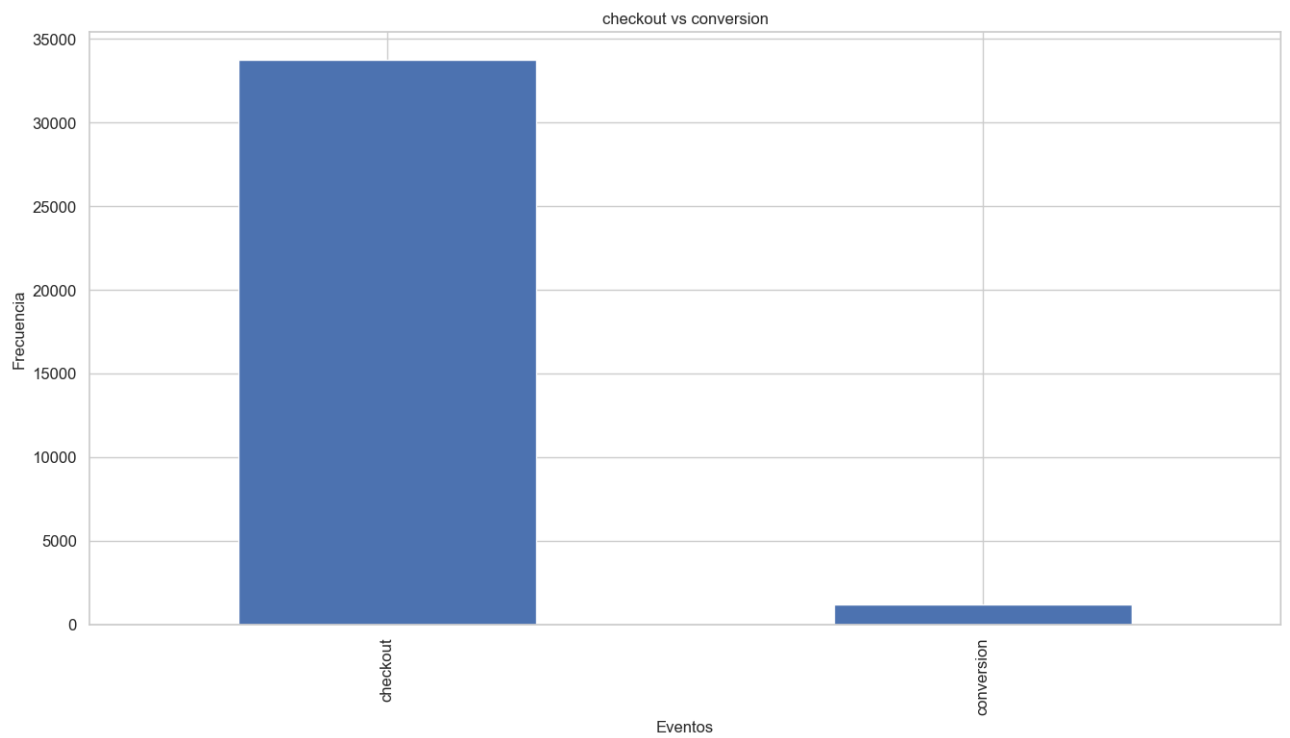


Figura 59: Checkout vs Compras

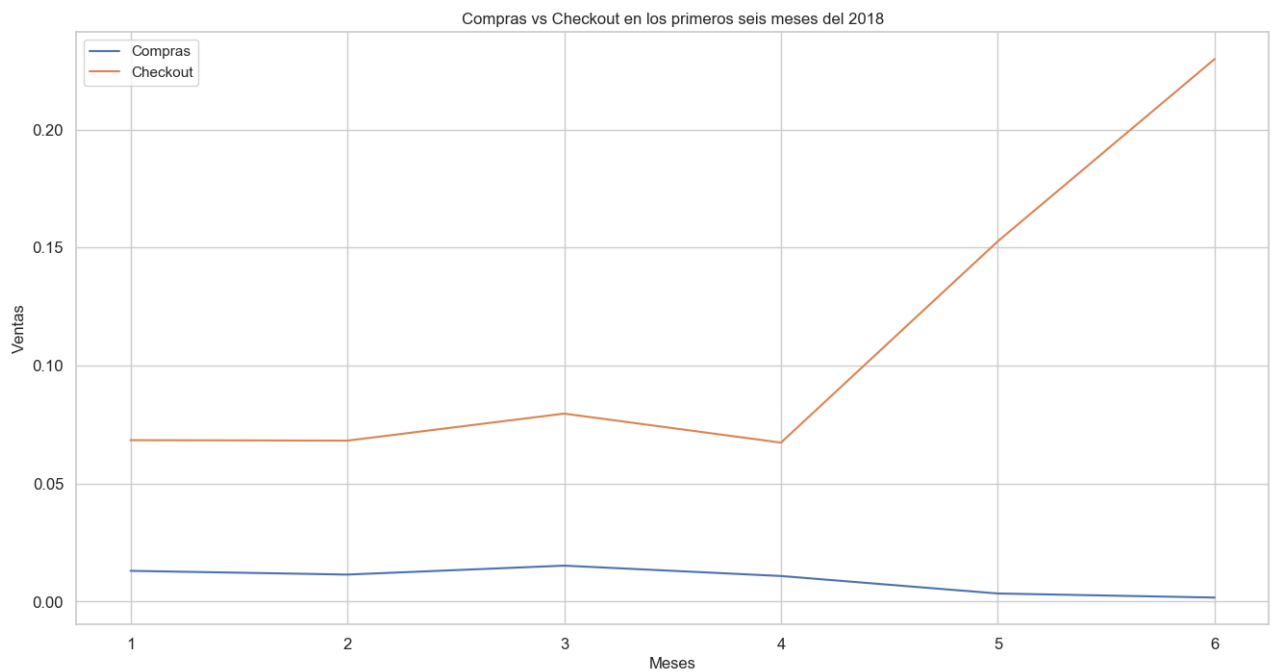


Figura 60: Checkout vs Compras mes a mes

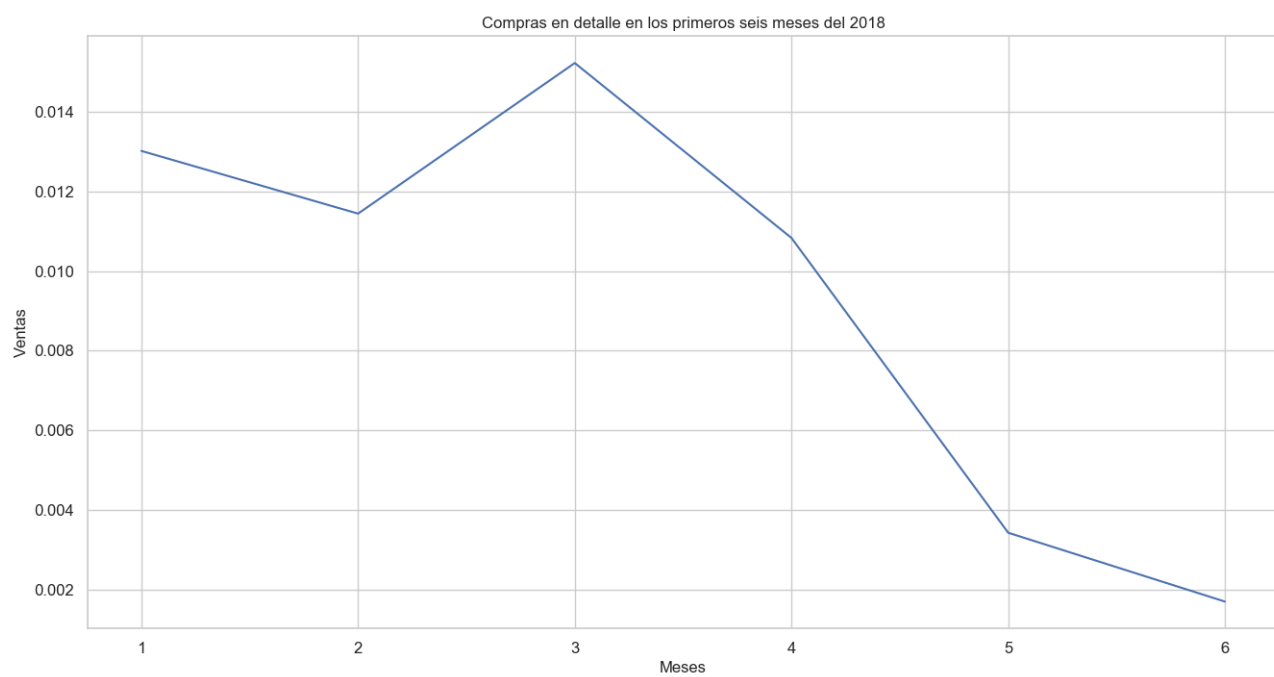


Figura 61: Compras en detalle