



**UNIVERSIDAD DE BUENOS AIRES**  
**FACULTAD DE INGENIERÍA**

**75.06 Organización de Datos**

**Trabajo Práctico N°1**

**Análisis exploratorio de datos de Jampp**

**Primer Cuatrimestre de 2019**

**Grupo 43**

**Repositorio:** [www.github.com/cruzpa95/Datos2019](https://www.github.com/cruzpa95/Datos2019)

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

**Contenido**

INTRODUCCIÓN..... 3

1.1 ACERCA DE JAMPP ..... 3

1.2 OBJETIVO ..... 3

2. ANÁLISIS DEL SET DE DATOS ..... 4

2.1 EVENTOS..... 4

2.2 INSTALACIONES ..... 12

2.3 CLICKS..... 17

2.4 SUBASTAS ..... 23

3. CONCLUSIONES ..... 26

4. FUENTES DE CONSULTA..... 27

## INTRODUCCIÓN

### 1.1 ACERCA DE JAMPP



Jampp es una plataforma de marketing para adquirir y captar clientes móviles. La compañía combina comportamiento de datos con tecnología predictiva para generar beneficios para los anunciantes al mostrar anuncios relevantes y personales que inspiran a los consumidores a comprar por primera vez o más a menudo.

Fundada en 2013, Jampp atiende a una base de clientes global desde las oficinas de San Francisco, Londres, Berlín, São Paulo, Singapur, Ciudad del Cabo y Buenos Aires.

### 1.2 OBJETIVO

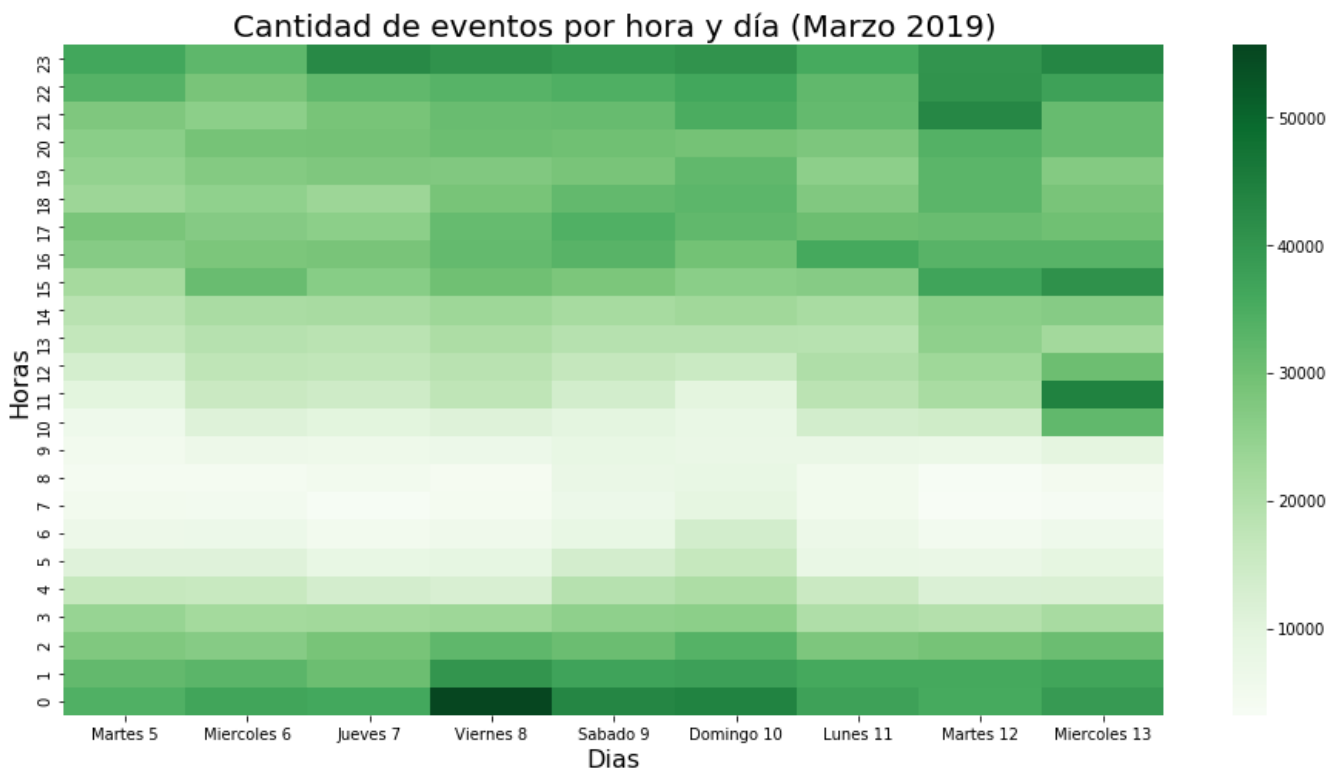
Este trabajo práctico tiene como objetivo extraer la información que se considere más significativa para la empresa Jampp a partir de un conjunto de datos facilitados por un integrante de la compañía y realizar un análisis en profundidad que pueda resultarle útil a la misma.

## 2. ANÁLISIS DEL SET DE DATOS

La mayoría de los datos brindados fueron hasheados por una decisión de seguridad de la empresa, sólo las fechas se pueden ver reales y de ahí se pudo observar que los datos fueron recolectados en un periodo de nueve días con inicio el martes 5 y finalización el miércoles 13 de marzo de 2019.

### 2.1 EVENTOS

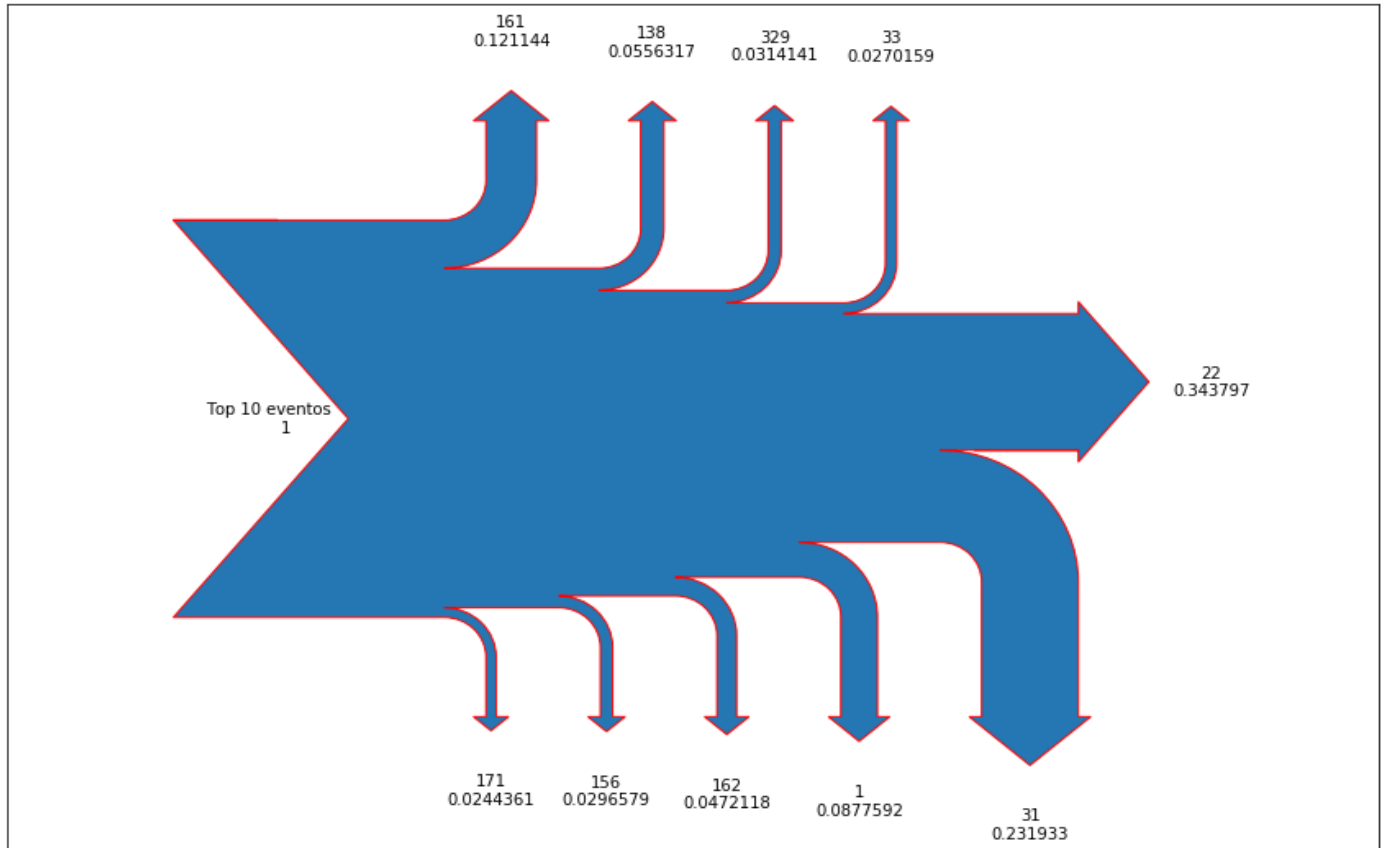
Comenzando por el set de datos de eventos, la distribución de los mismos temporalmente se ve detalladamente en el siguiente gráfico:



Se puede observar que, como es de esperar, la cantidad de eventos es casi nula en la franja horaria de las 4 a las 9 de la mañana, ya que es el momento donde la mayoría de la gente duerme o recién comienza a despertarse. También se puede notar que el día viernes 8 a las 00:00 horas se da el mayor número de eventos. Esto puede deberse a que ese día se celebró el Black Friday. Lo curioso es que sólo hay una enorme cantidad de eventos en esa única hora y luego durante todo el viernes la actividad es muy parecida a la de los demás días.

Para el siguiente gráfico se decidió tomar los 10 eventos más frecuentes, ya que a partir del 11, la relación con los primeros era enorme y la figura se dibujaba de manera desproporcionada.

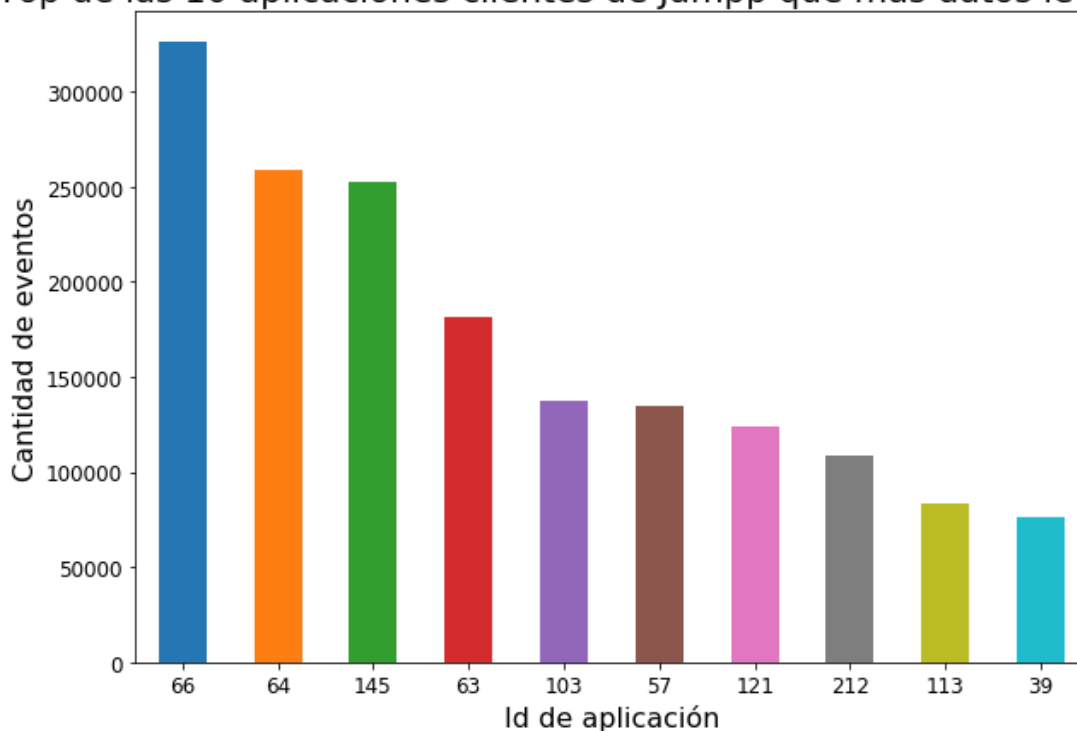
### Los 10 eventos más frecuentes



El evento con el id 22 es el que más ocurrencia tiene, seguido por el 33 y el 31. Es muy probable que estos eventos correspondan a acciones de los usuarios tales como “abrir aplicación”, “cerrar aplicación”, es decir cosas que siempre debe hacer un usuario al usar una aplicación.

Se consideró conveniente saber cuál era la aplicación cliente de Jampp que más datos le provee:

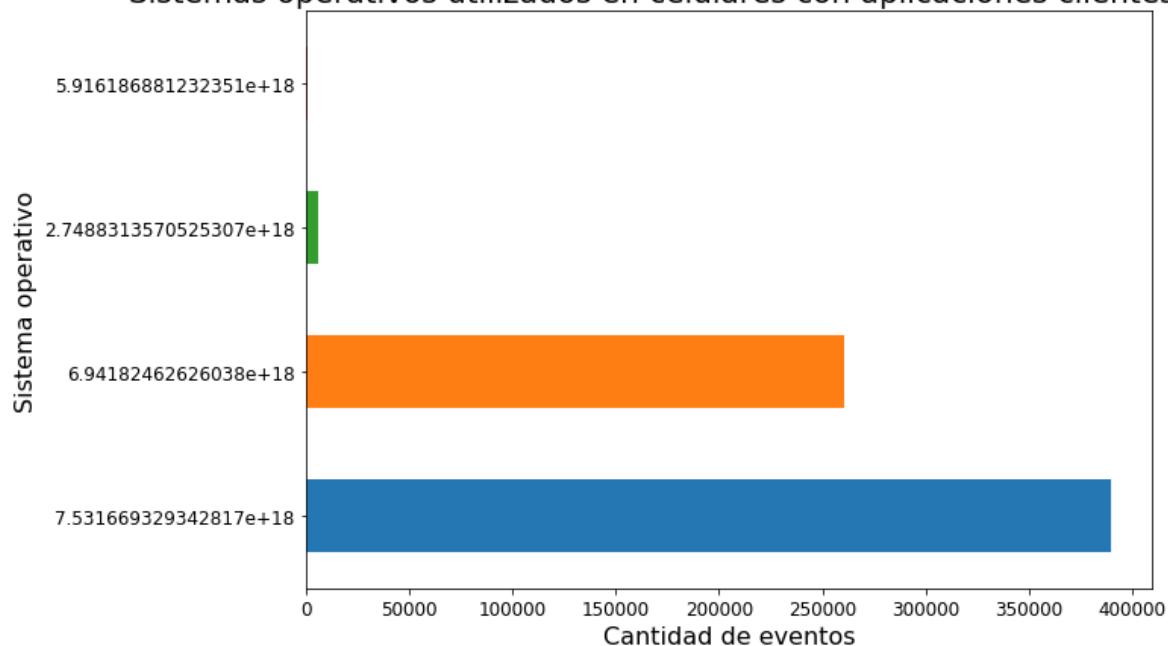
### Top de las 10 aplicaciones clientes de Jampp que más datos le proveen



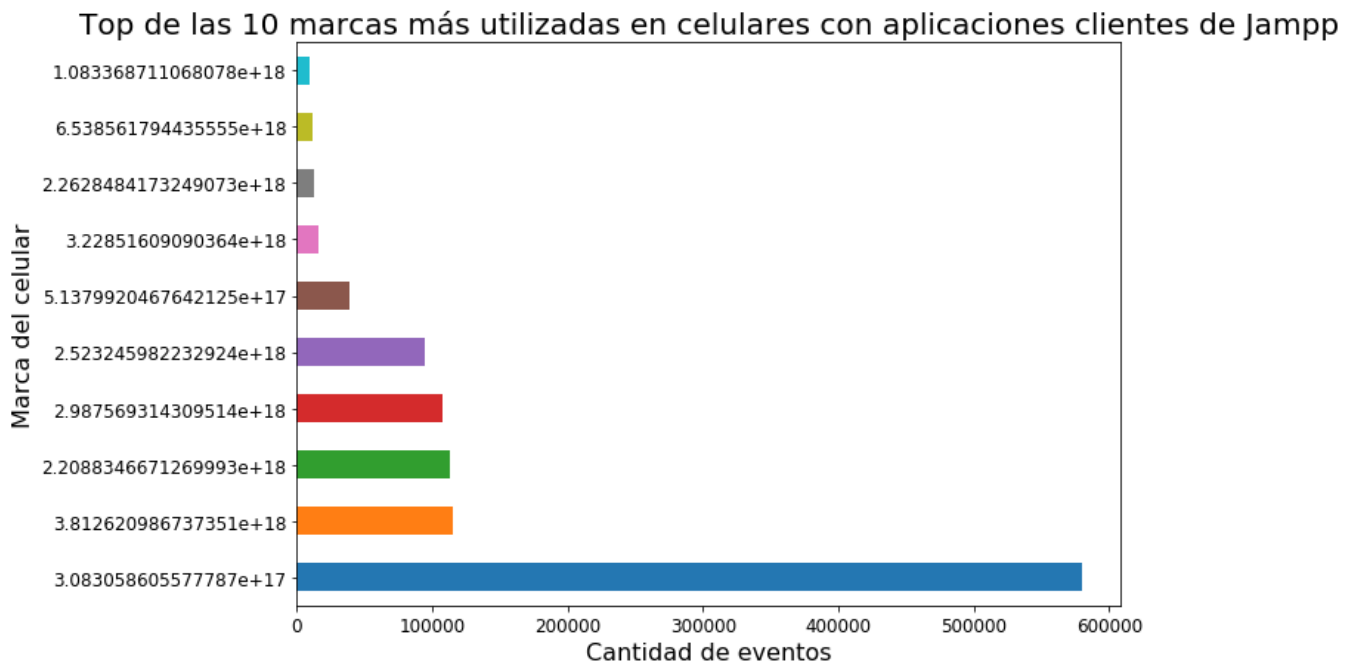
La aplicación con el id 66 es la que más datos le provee a la empresa, se puede pensar entonces, que es la que más uso tiene por parte de las personas que tienen dicha aplicación en su celular.

Analizando los celulares de los usuarios de las aplicaciones clientes de Jampp, los datos sobre los sistemas operativos utilizados también fueron hasheados. De todas maneras, en el siguiente gráfico se puede observar claramente que hay dos que lideran los eventos. Sería muy lógico suponer que son iOS y Android.

### Sistemas operativos utilizados en celulares con aplicaciones clientes de Jampp

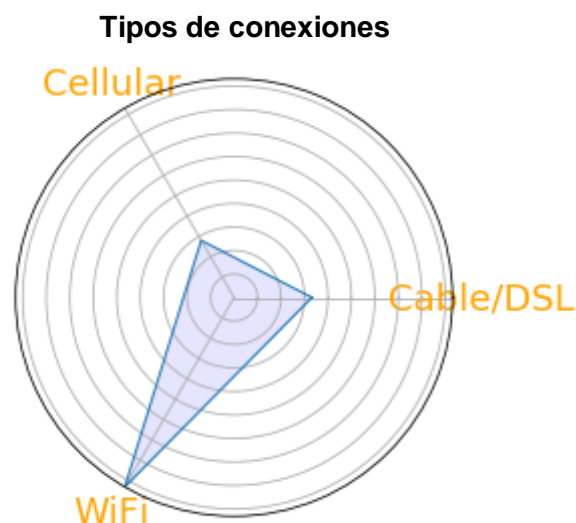


En cuanto a las marcas de los celulares, el siguiente gráfico muestra el top 10:



Se puede observar que hay una marca que supera a las demás en un gran porcentaje. Conociendo un poco sobre celulares, asumimos que es Samsung debido a que es líder mundial en venta de celulares.

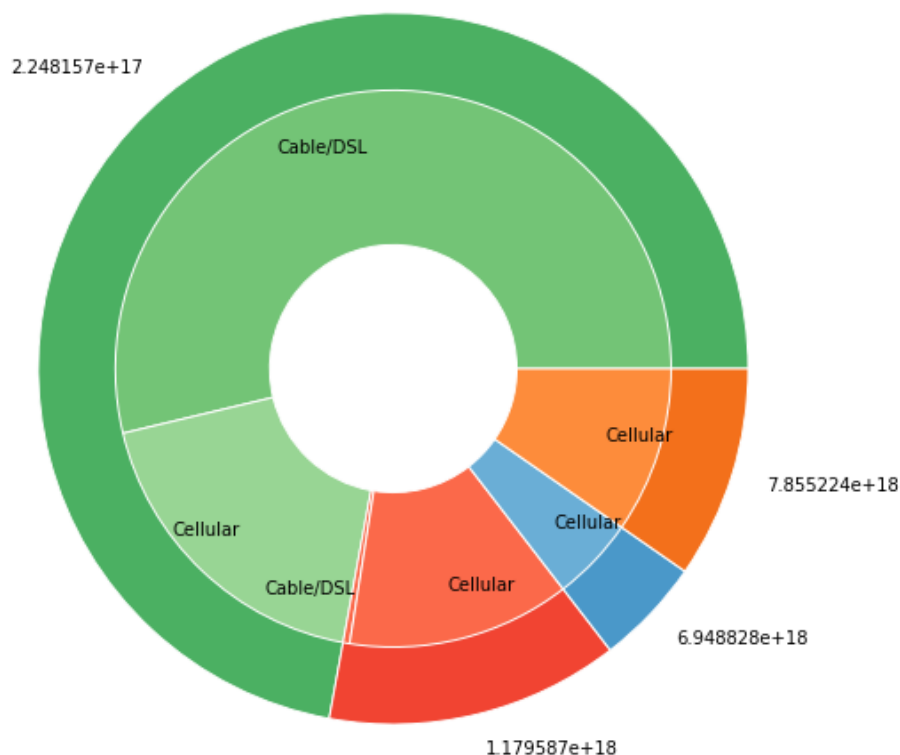
Para el análisis de los tipos de conexión en los dispositivos, se consideró interesante presentar un gráfico de radar. En este tipo de esquemas se define según escala un valor a cada radio del radar y se divide la circunferencia en tantas partes como variables se quieran mostrar. Luego se le asigna el valor que alcanza cada variable y se genera el gráfico, en el cual se podrá ver a simple cuáles son las conexiones más frecuentes y su relación proporcional entre ellas.



Siguiendo el gráfico, como era de esperar, en la mayoría de las aplicaciones clientes de Jampp se navega con una conexión de WiFi. En segundo lugar, no muy distante del tercero, se ubica la conexión mediante Cable/DSL, lo que se puede suponer es que, es muy probable que éstos sean juegos ya que en ellos se necesita una conexión más directa para que rinda al máximo. También se podría suponer que son computadoras que no tienen tecnología WiFi, pero esto es muy poco frecuente a esta altura de la tecnología. En tercer lugar, quedó la conexión mediante red Celular, que debe ser 3G, 4G, etc. Lo cual es lógico ya que generalmente se trata de ahorrar los datos móviles que uno tiene en el celular debido a que son costosos.

Siguiendo con el análisis de conexiones, se analizarán ahora, las compañías proveedoras. En los datos, hay cuatro compañías que se diferencian mucho con las restantes en cuanto a usuarios, por lo que se graficarán sólo éstas.

### Compañías proveedoras de internet y telefonía móvil

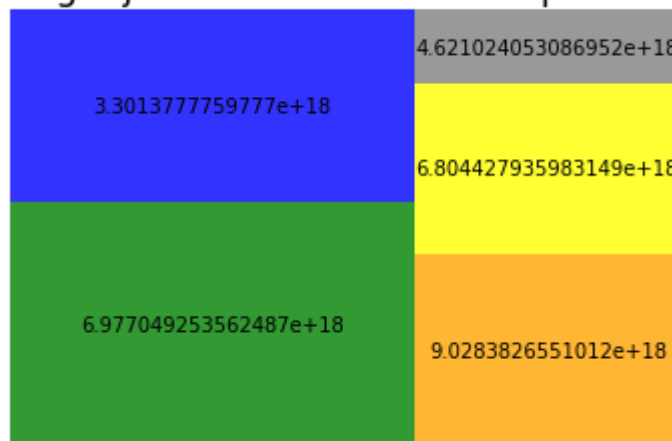


De las cuatro compañías analizadas, se puede observar que hay una que contiene a la gran mayoría de los usuarios, a los cuales les provee tanto el servicio de telefonía móvil así como el de internet. La segunda empresa con más usuarios (representada por el color rojo) también brinda los dos servicios, pero mayormente el de celular. Las dos restantes sólo tienen servicio para dispositivos móviles.



En cuanto a los lenguajes utilizados en los dispositivos se observa una curiosidad, hay demasiada diversidad de lenguajes. Sólo se consideraron 5, ya que los demás eran utilizados en relativamente muy pocos dispositivos.

### Lenguajes utilizados en los dispositivos



El lenguaje que está representado por el cuadrado con más área (el verde) se podría afirmar que es el "español", seguido del "inglés" (representado por el cuadrado de color azul), si se sigue la lógica de los idiomas más hablados en Occidente. Vale aclarar que todos los eventos fueron originados en un solo país. Lo curioso es que los lenguajes que se utilizan en los dispositivos son muchos, es por ello que en un primer momento se había pensado que el país de los datos era Estados Unidos, ya que allí hay mucho turismo y también mucha gente viviendo que habla diferentes idiomas, pero luego se supo que el país era Uruguay. Resulta muy raro entonces, que en un país tan pequeño haya celulares con tantos idiomas distintos.

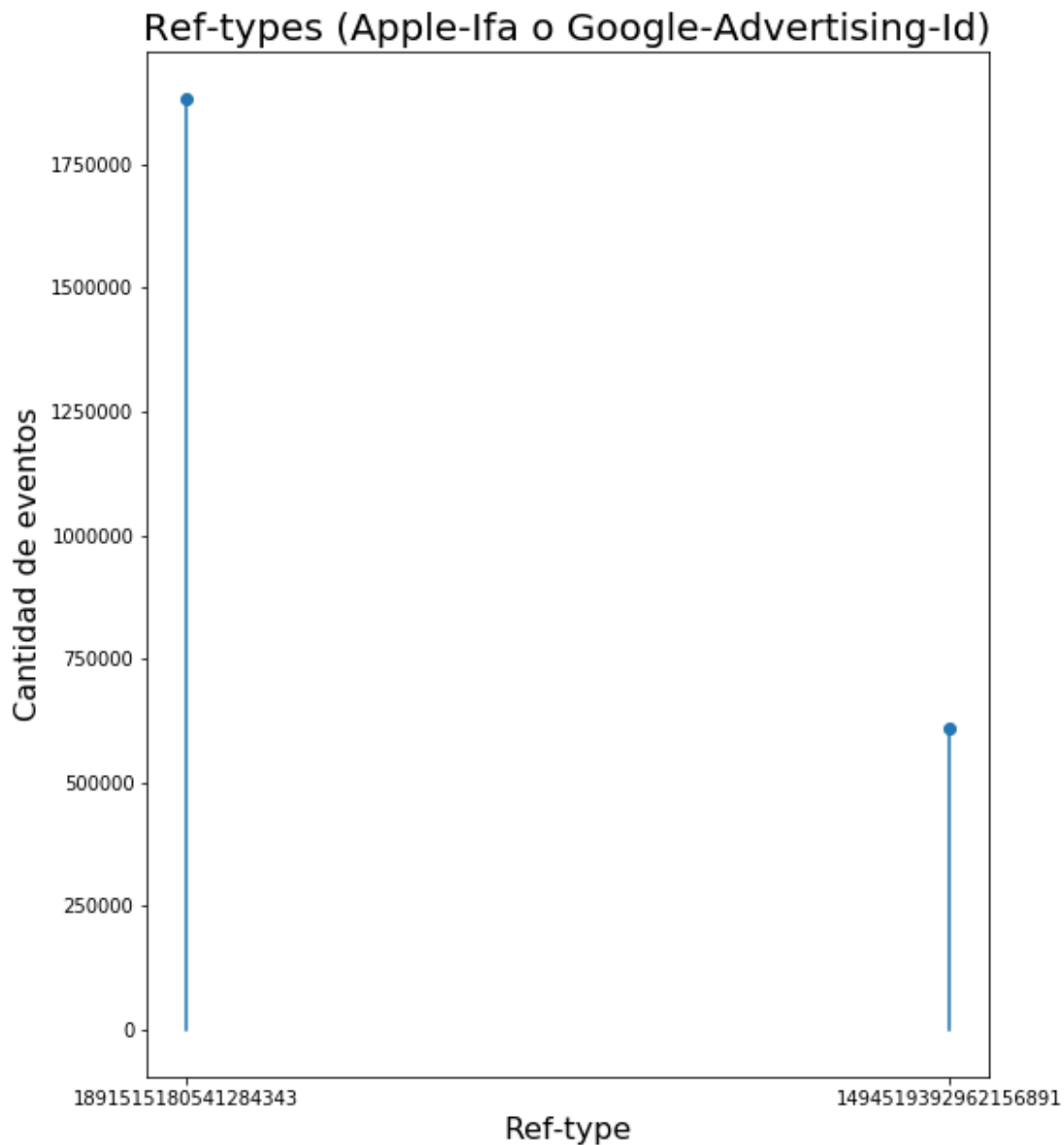
Otra conclusión que se planteó, fue que los idiomas representan distintas distribuciones del español, ejemplo: "Español (España), Español (Latinoamérica), Español (Argentina), etc."

Otro tipo de análisis que se puede realizar es sobre las ciudades con más eventos, resultando el siguiente gráfico:



Se puede observar que hay una ciudad en la que se originan casi todos los eventos. Es muy probable que esta sea la capital del país de los datos. Sabiendo que dicho país es Uruguay, se podría deducir que la ciudad es Montevideo y sería muy lógico ya que la población charrúa se concentra principalmente allí.

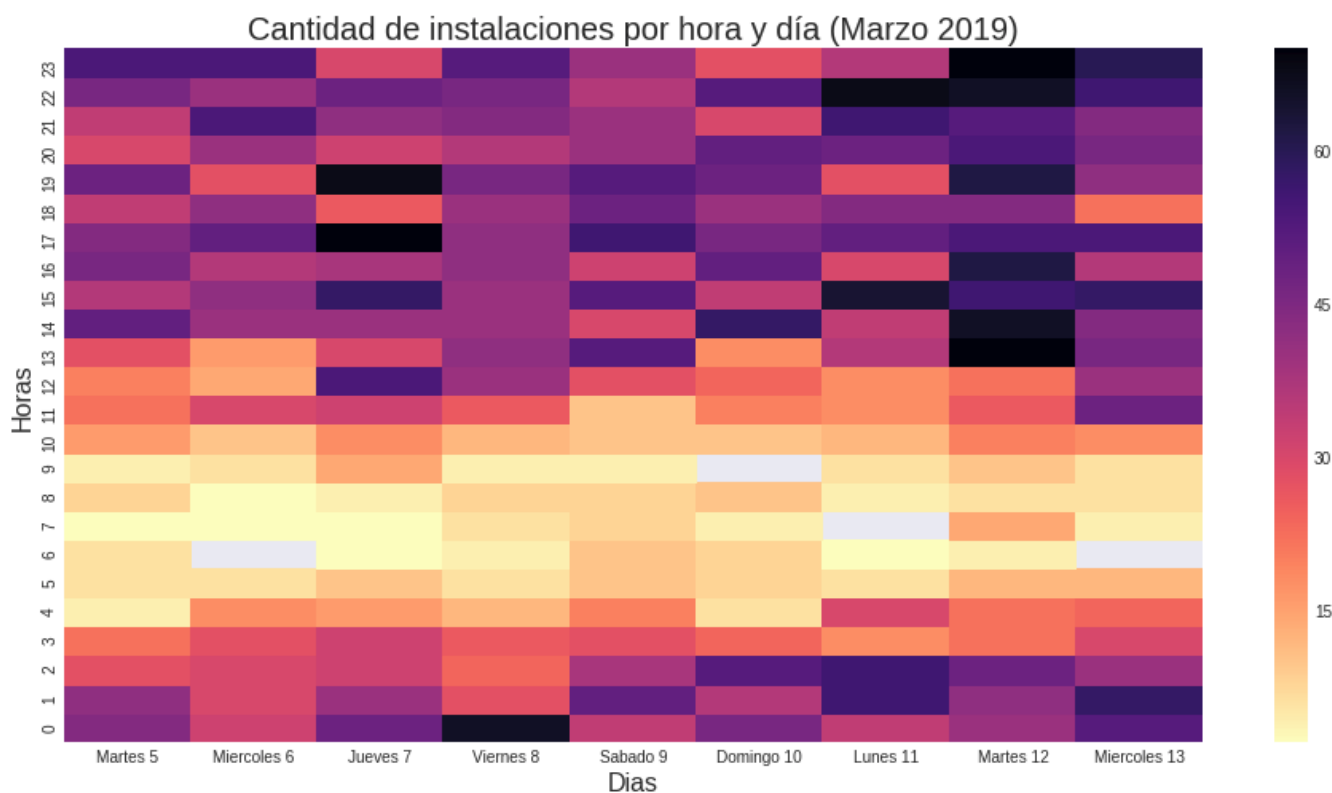
Finalizando con el set de datos de eventos, se analizaron los ref-types, en donde se puede visualizar que hay sólo dos y uno de ellos genera muchísimos más eventos que el otro. Sabiendo que son Apple-Ifa (celulares iPhone) y Google-Advertising-Id (celulares con Android), se supone que el mayor corresponde a los dispositivos con Android ya que en el mundo hay más celulares con ese sistema operativo.



## 2.2 INSTALACIONES

Como comentario introductorio, se quiere decir que resulta sospechoso, pero ninguna de las instalaciones del set de datos fue atribuidas a Jampp, todas figuran como 'False'.

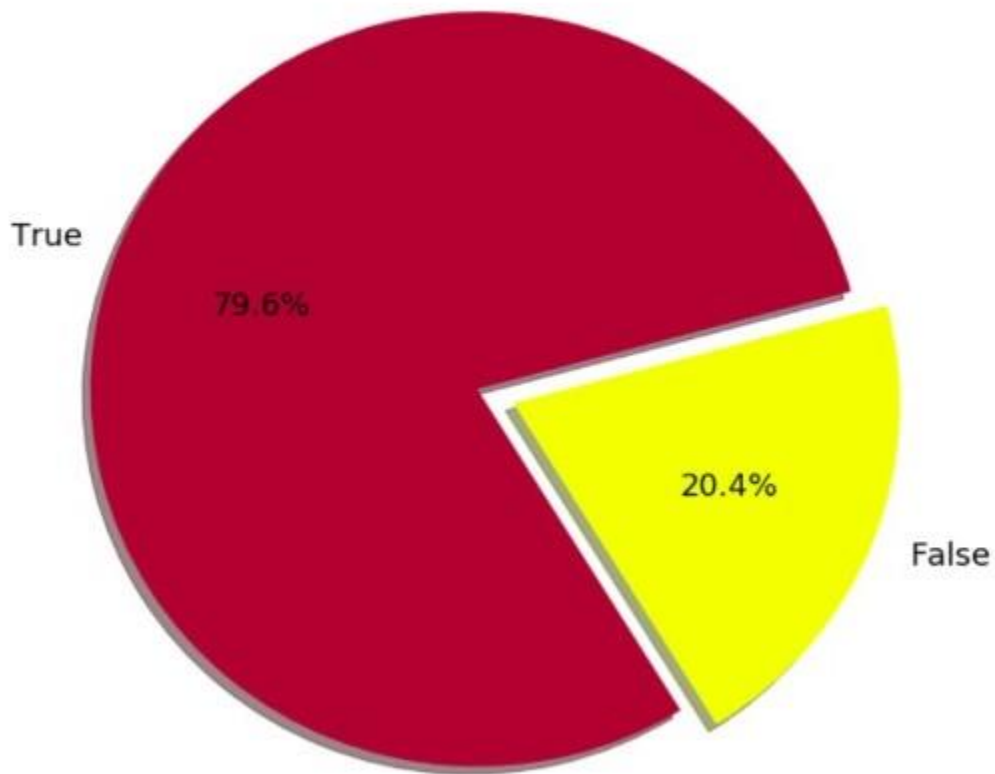
Temporalmente, las instalaciones se llevaron a cabo de la siguiente manera:



Es muy curioso la distribución de las instalaciones durante los 9 días recolectados, como se puede observar en el gráfico hay 5 horarios en los que se dio la mayor cantidad, pero no siguen ningún patrón, ya que son horarios totalmente distintos. Lo que se puede ver claramente es la disminución de las mismas en la franja horaria de las 4 a las 10 de la mañana.

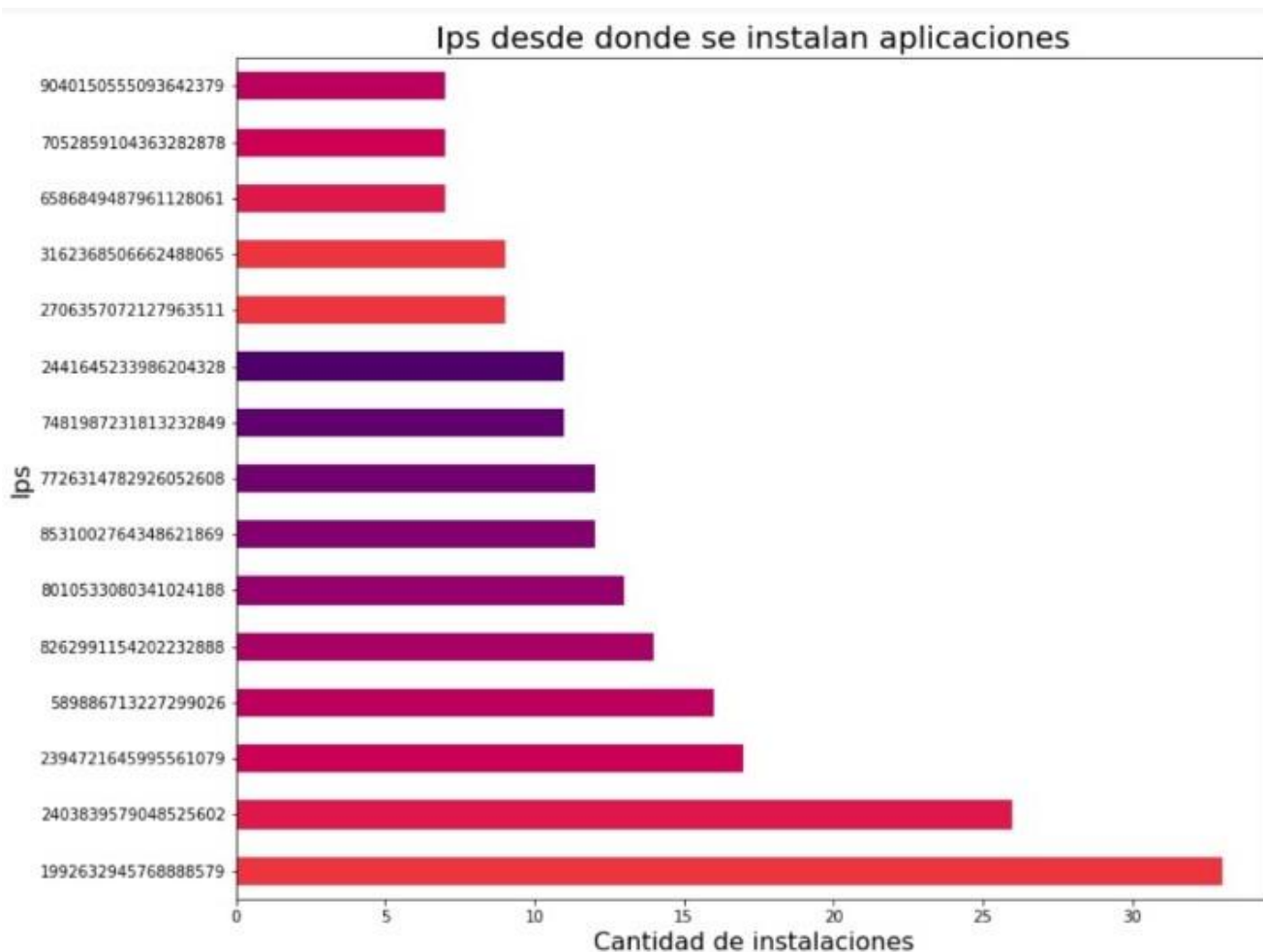
Como se puede observar en el siguiente gráfico, la mayoría de las instalaciones fueron hechas con una conexión wifi en el dispositivo, como era de esperarse.

### Instalaciones realizadas desde un dispositivo con o sin Wifi



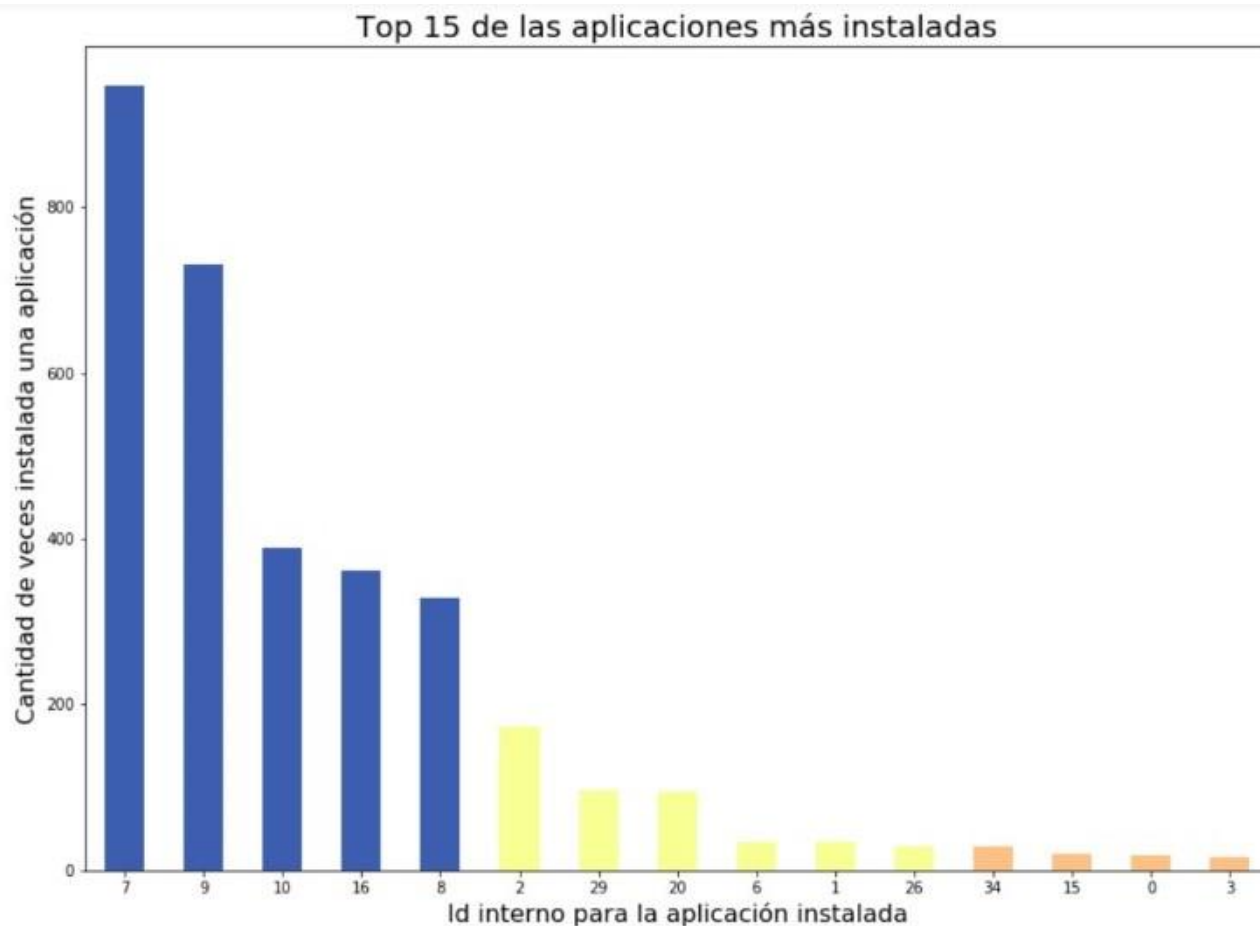
Instalar una aplicación con una conexión Wifi es muy lógico ya que si no se gastan datos móviles, los cuales son limitados y necesarios para otros motivos.

Siguiendo con lo relacionado a la conexión, se analizaron las ips desde donde se realizaron instalaciones, resultando:



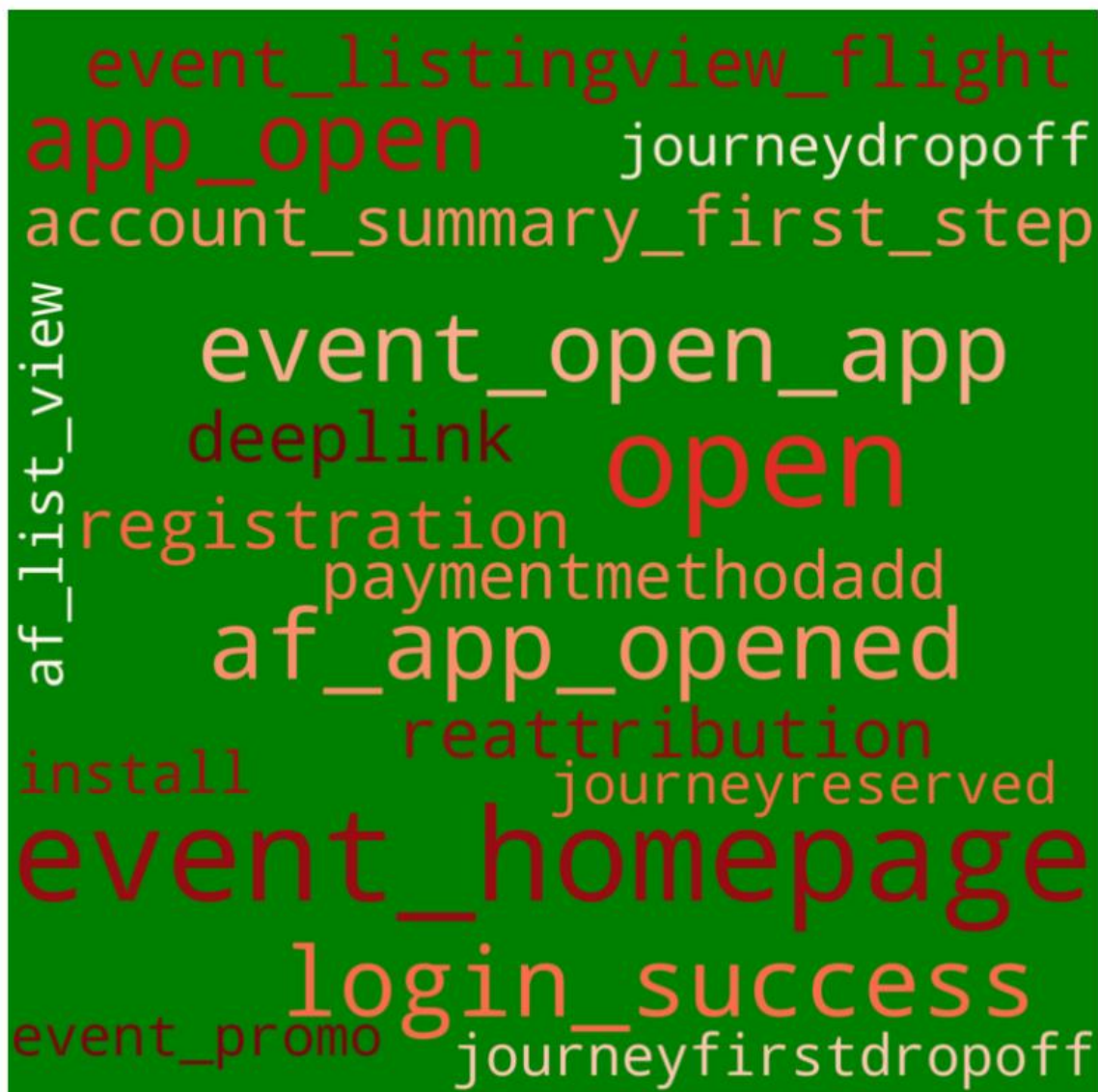
Se puede notar que hay algunas ips que realizaron muchas instalaciones de aplicaciones en los pocos días registrados, se supone que éstas son de dominio público o de algún lugar donde se comparte la red Wifi, por ejemplo, bares, restaurantes, centros comerciales, etc.

En cuanto a las aplicaciones más instaladas, se tomaron las primeras 15, ya que como se ve en el siguiente gráfico a partir de ésta, la cantidad de instalaciones de una aplicación es muy poca.



Se puede observar en el gráfico que hay 2 aplicaciones que son las favoritas de los usuarios, las cuales tienen id '7' y '9' respectivamente. Luego las siguientes 3 están muy cerca entre ellas, pero tienen la mitad de las instalaciones que las 2 primeras. Las aplicaciones rankeadas en el sexto, séptimo y octavo lugar también compiten entre ellas de cerca. Pero a partir de la novena se puede ver que son muy pocas las instalaciones.

Finalmente, se analizaron los tipos de instalaciones (columna 'kind' del dataframe), y se realizó el siguiente gráfico:

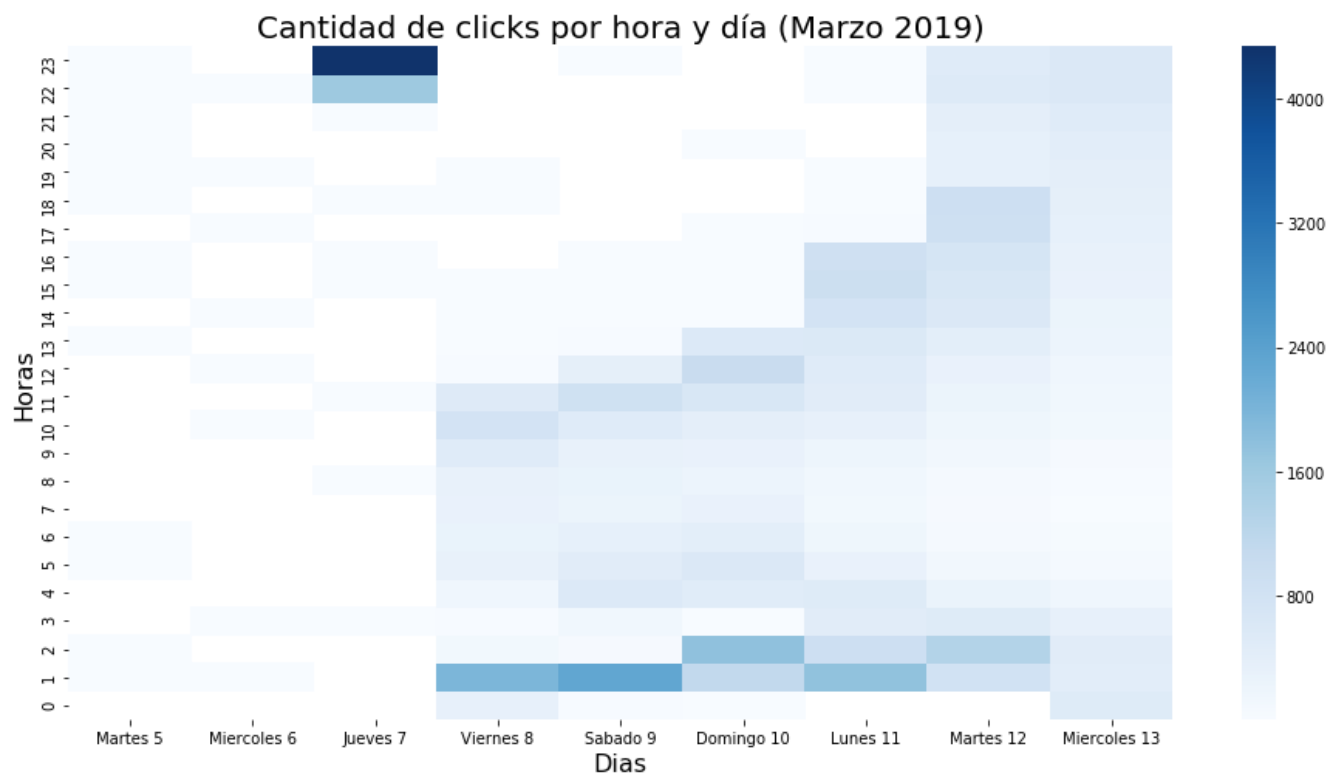


Como se puede observar, las instalaciones realizadas en Jampp se centran en muchas ocurrencias que están vinculadas con OPEN (ya sea desde eventos para abrir la app) y LOGUEOS en la página.



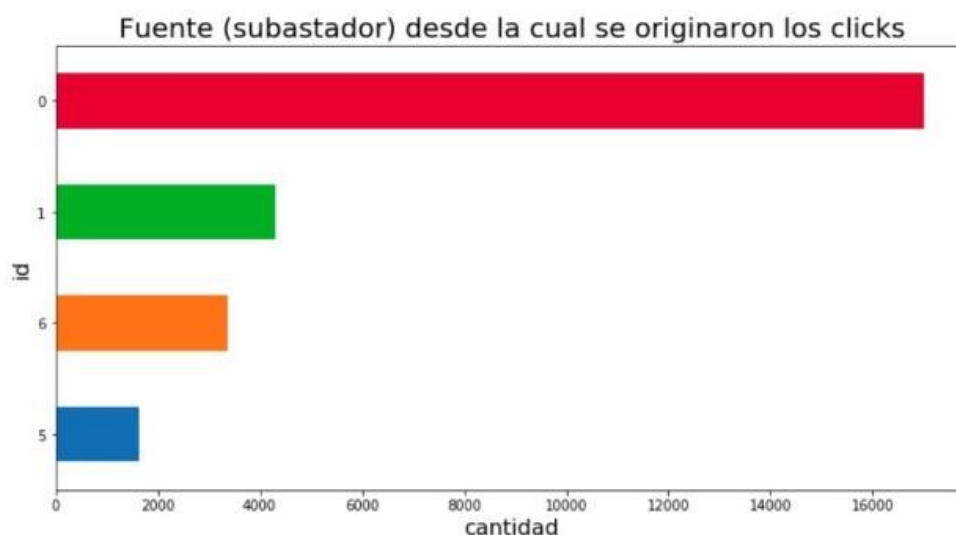
## 2.3 CLICKS

El siguiente gráfico muestra cómo se distribuyeron los clicks a lo largo de los 9 días:



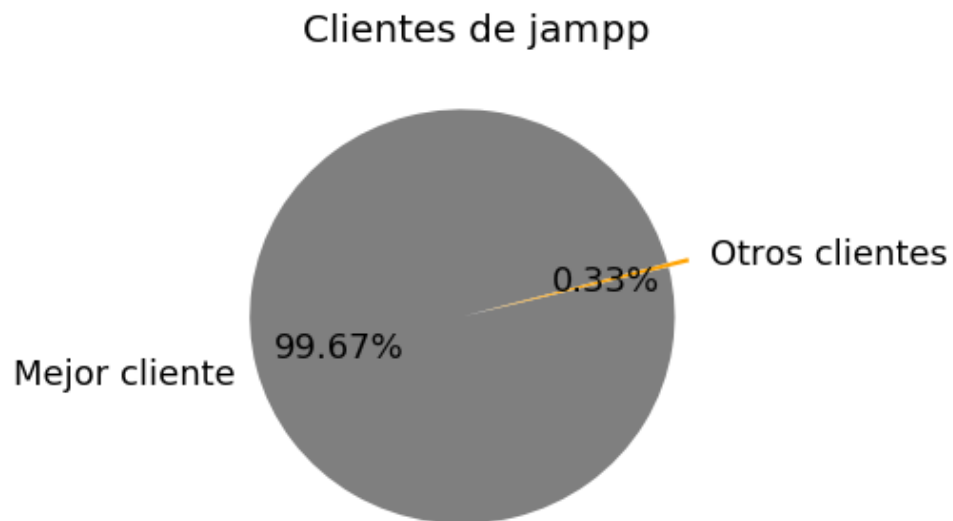
Como es de esperarse, casi ninguna persona hace click en una publicidad que le aparece en el celular. Pero el día jueves 7 a las 23 hubo muchísimos clicks, esto puede deberse a que a la hora siguiente (a las 00:00 del viernes 8) comenzaba el Black Friday.

Una fuente obtiene un espacio para una publicidad y lo subasta entre competidores de Jampp, hay varias fuentes de subastas, las cuales fueron representadas el siguiente esquema:

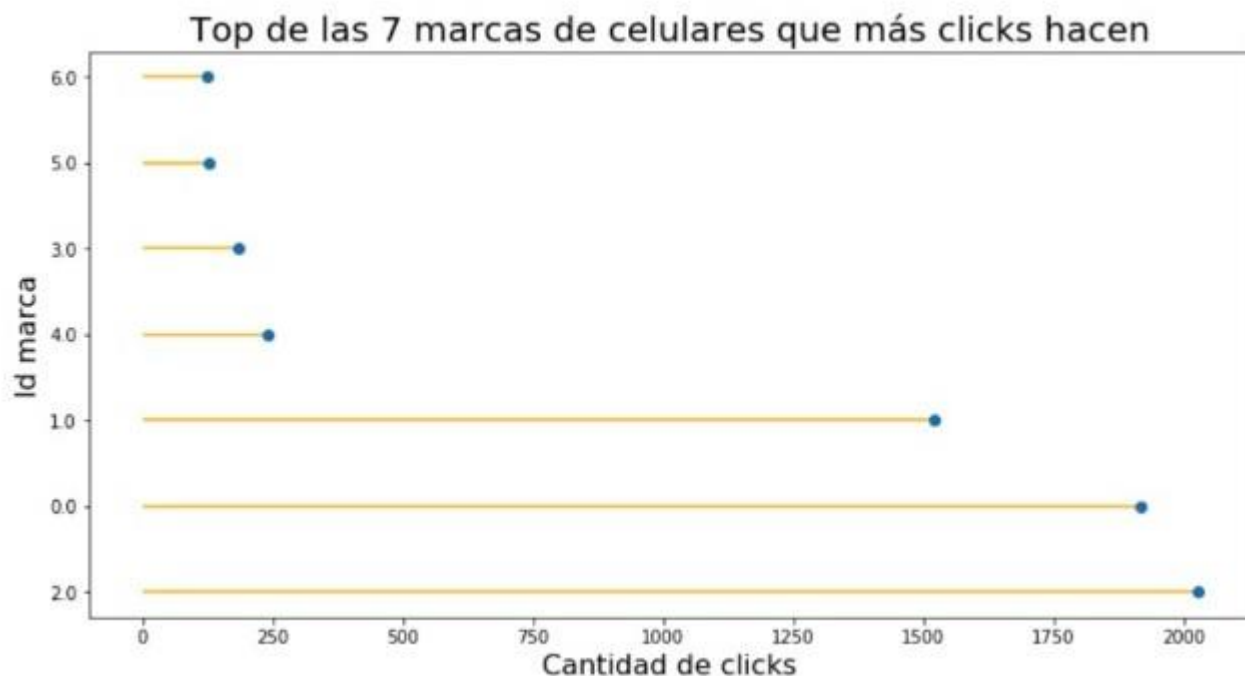


Se puede observar que hay un subastador que obtiene la mayor cantidad de clicks. Se podría decir que es el más efectivo y quizá sea una tendencia la que los usuarios en los que el subastador con id '0' imprime una publicidad son más factibles a hacer click.

Los clientes de Jampp son los que le pagan para que la empresa les consiga el mejor lugar donde colocar una publicidad. En el siguiente gráfico, se muestra que hay un cliente que obtiene casi la totalidad de los clicks.

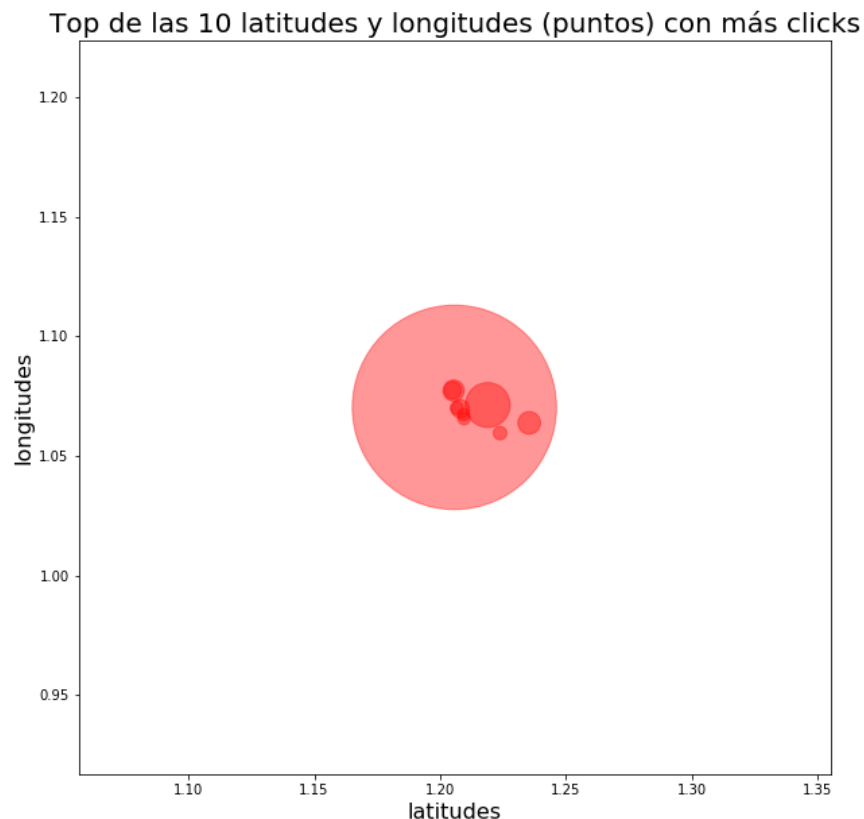


Analizando los celulares que más clicks hacen, resultó el siguiente gráfico sobre marcas:

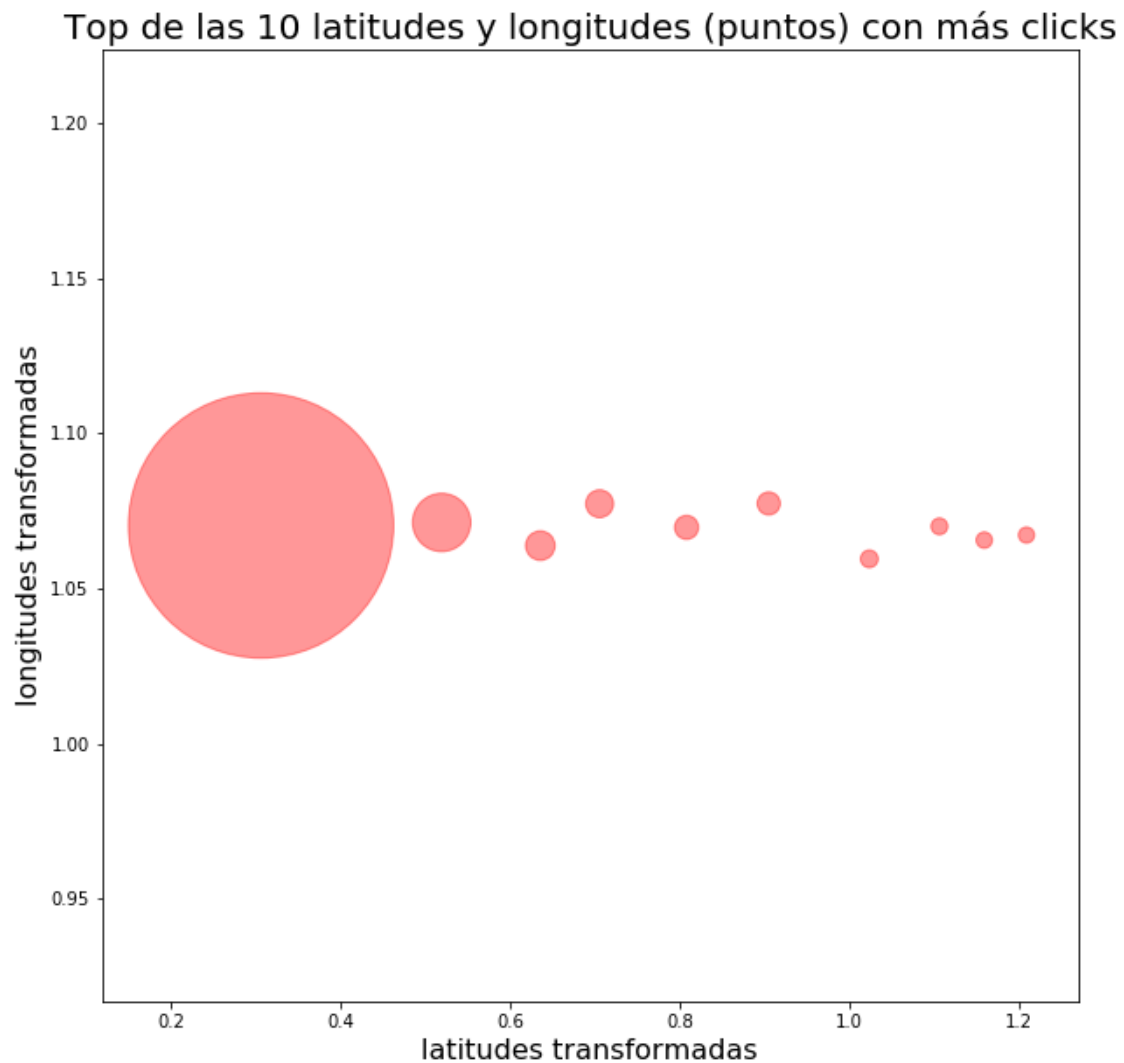


En el mismo, se puede observar que hay 3 marcas de celulares en las que más clicks se hacen. Se podría intuir que las éstas son Samsung, iPhone y Motorola, ya que son las líderes del mercado.

También se analizaron las coordenadas desde se originaron clicks. Se pudo notar que no hay ninguna latitud que tuviera el mismo hash que una longitud. Entonces se tomó una tercera coordenada que se llamó 'punto' y resultó de la suma de 'latitud' + 'longitud', para luego hacer un contador de apariciones y determinar cuántos clicks se realizaron desde una misma 'posición' las cuáles se decidieron representar con un gráfico de burbujas, en el que cada burbuja tiene un área, la cual representa la cantidad de veces que se hizo un click desde una misma coordenada.

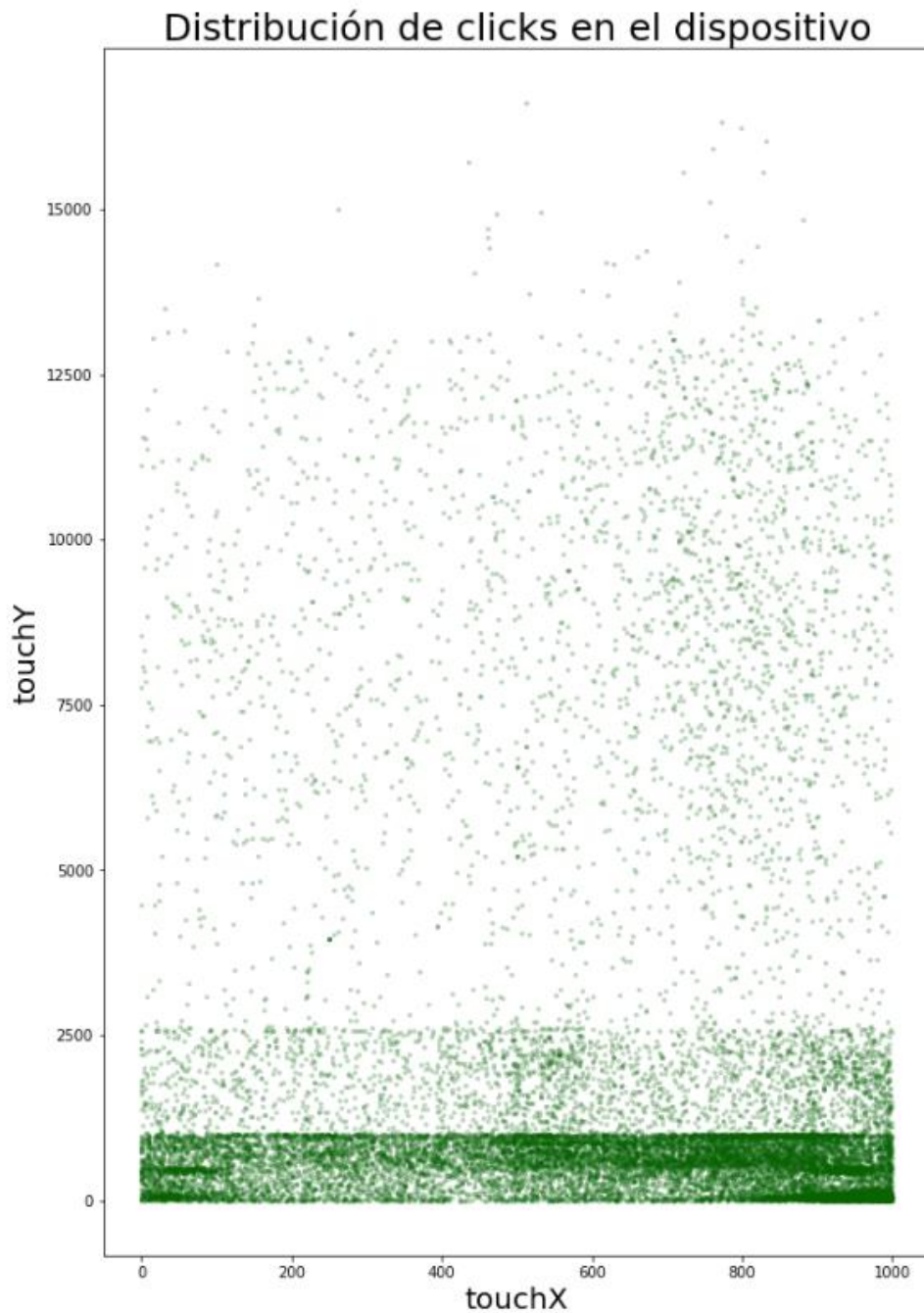


Como los puntos están muy cerca y hay un punto que concentra la gran mayoría de los clicks, la burbuja que representa a ese punto engloba a todos los demás. Es por ello, que para que se note la diferencia de área (de cantidad de clicks) se decidió separar las burbujas, moviéndolas sobre el eje x.



Luego de la transformación aplicada, resultó el gráfico de arriba en el que se puede ver claramente que hay una burbuja que supera ampliamente a las demás en área. Se supone al igual que en el gráfico "Top de las 10 ciudades con más eventos", que ésta sería la capital (Montevideo).

Finalmente, se analizó la distribución de los clicks en las pantallas de los dispositivos, resultando el siguiente gráfico:

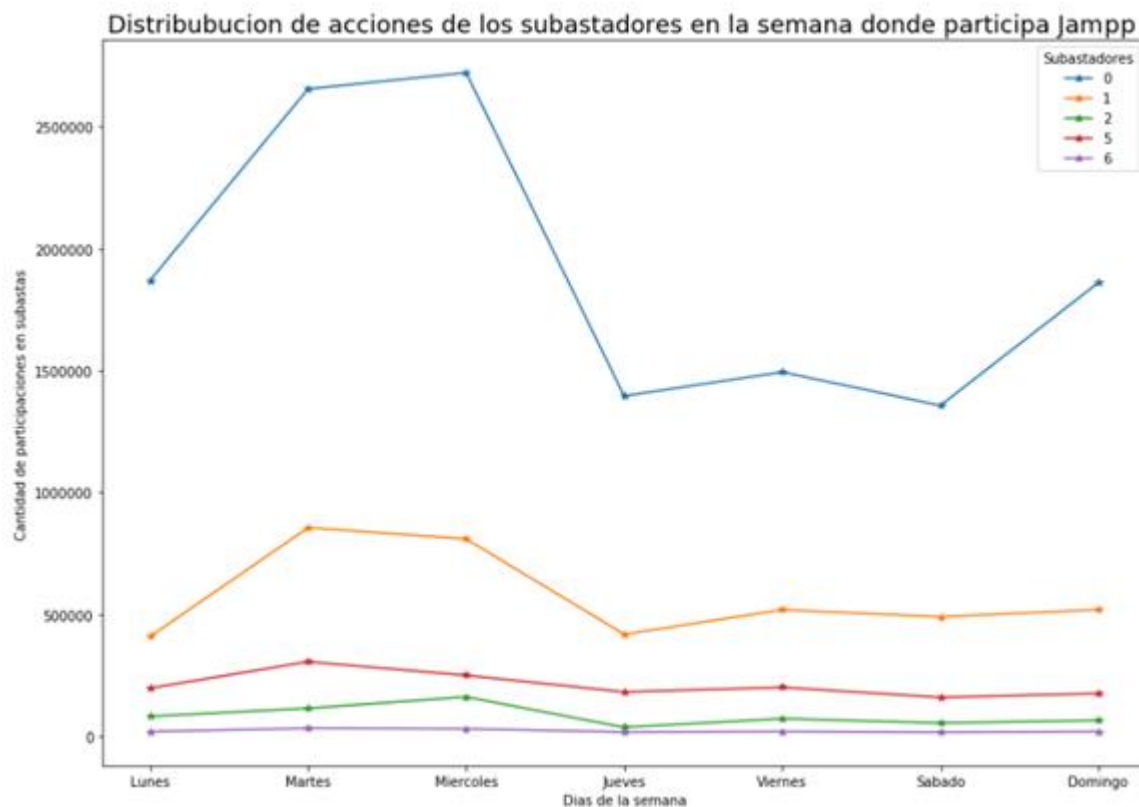


Se puede ver claramente que la mayoría de los clicks se realizaron en la parte baja de las pantallas. Teniendo en cuenta que casi nadie mira las publicidades que le aparecen en el celular, se supone que, en esas posiciones, seguramente, se encontraban los botones para “saltar publicidad”, la “x” para cerrar la publicidad, etc.

Otra conclusión sería que los clicks nos dicen el tamaño promedio de las publicidades que se tomaron en cuenta para el set de datos, es decir el espacio que brindan los proveedores para un anuncio subastado en los que participa Jampp.

## 2.4 SUBASTAS

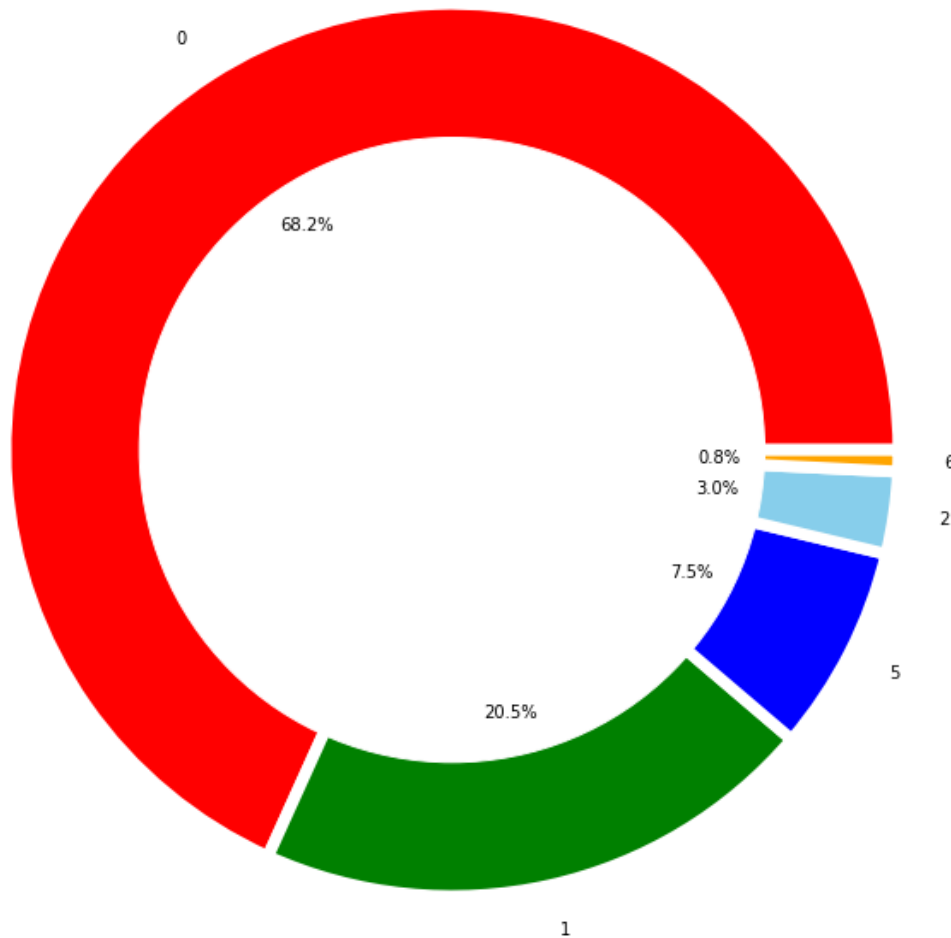
Realizando un análisis temporal de las subastas, se obtuvo el siguiente gráfico:



Se puede notar que los subastadores registran un alza en sus acciones durante los primeros días de la semana, por lo cual, Jampp debe participar de manera muy activa para ganarse ese espacio en disputa, no así durante los siguientes días dónde el flujo de acciones de estos subastadores merma con respecto a los demás días.

Analizando las diferentes fuentes de subastas (subastadores), se pudo notar que con los datos provistos, hay una cierta preferencia (ocupando un 68.2% de los intercambios) por un motor de subastas, donde seguramente Jampp hace su subasta para poder ganarse un espacio. El siguiente gráfico muestra lo dicho anteriormente:

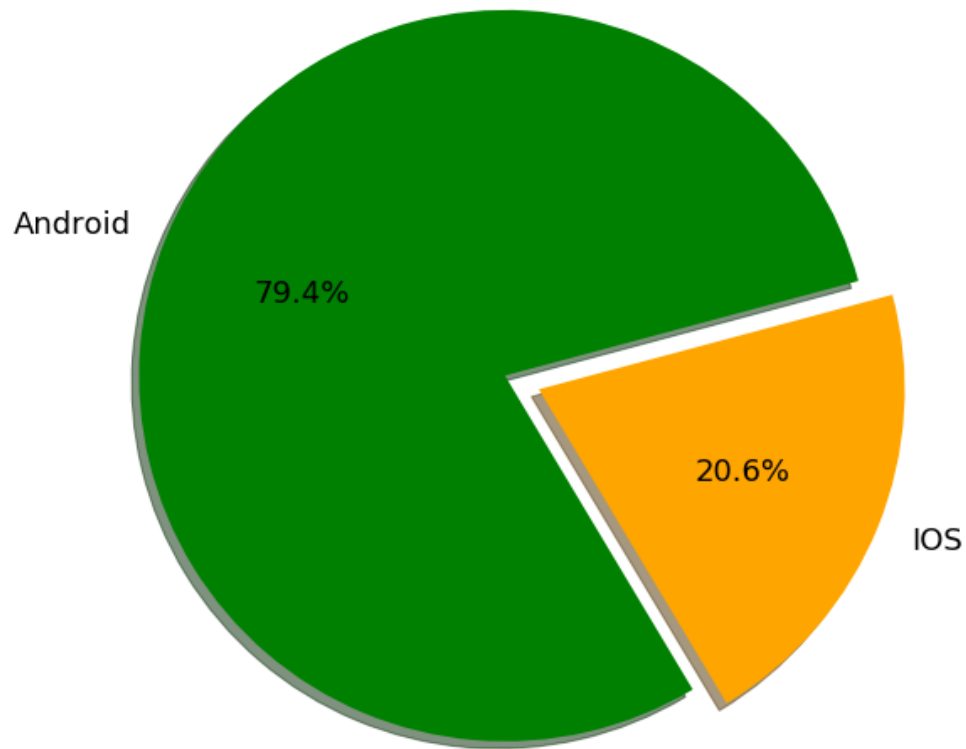
**Distribución de fuentes de subastas**





Finalmente, se estudiaron las diferentes plataformas desde donde se realizan subastas:

### Distribución de las plataformas desde donde se realizó la subasta



Se puede observar, como era de esperarse, que los uruguayos tienen en forma mayoritaria dispositivos Android, desde donde salen las subastas donde participa Jampp.

### 3. CONCLUSIONES

Las conclusiones son en base a los datos correspondientes al país Uruguay entre las fechas 5 al 13 de marzo del 2019.

- ❖ No es conveniente participar en subastas publicitarias en la franja horaria de 4 a 9 am ya que es donde menos eventos se registran.
- ❖ El momento que más clicks se registró fue una hora antes del Black Friday, y además, sólo la primera hora del Black Friday es donde se registra una enorme cantidad de eventos, luego tienden a normalizarse. Jampp debería enfatizar en esas 2 horas.
- ❖ Los 2 sistemas operativos más usados son IOS y ANDROID por lo cual, se recomendaría compatibilidad con estos sistemas operativos.
- ❖ Hemos registrado que hay una marcada preferencia por la instalación de 2 aplicaciones. Como un usuario instaló a través de una publicidad, suponemos que es probable que si en esas aplicaciones haya otra publicidad, ese usuario también llegue a instalarla.
- ❖ Se puede notar que los subastadores registran un alza en sus acciones durante los primeros días de la semana, por lo cual, Jampp debe participar de manera muy activa para ganarse ese espacio en disputa, no así durante los siguientes días donde el flujo de acciones de estos subastadores merma con respecto a los demás días.

## 4. FUENTES DE CONSULTA

- Jampp - <https://jampp.com/> - Utilizada para obtener información acerca de la empresa.
- StackOverflow - <http://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.