

UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE INGENIERÍA

**75.06 Organización de Datos**

**Trabajo Práctico 1**

**Primer Cuatrimestre de 2019**

Grupo 18

|  |  |
| --- | --- |
| Camussi, Alan |  |
| Catolino, Lucas |  |
| Zugna, Federico | 95758 |

**Link de GitHub:** <https://github.com/LucasCatolino/Orgnizacion-de-Datos> **(Privado)**

Índice

[1. Introduccion 3](#_Toc6781156)

[1. ¿Qué es Jampp? 3](#_Toc6781157)

[2. Análisis Set de Datos Clicks 4](#_Toc6781158)

[3. Análisis Set de Datos Auctions 5](#_Toc6781159)

[4. Análisis Set de Datos Events 6](#_Toc6781160)

[5. Análisis Set de Datos Installs 7](#_Toc6781161)

[1. Implicit Install 7](#_Toc6781162)

[2. Atributed Install 10](#_Toc6781163)

[3. Install with WiFi 11](#_Toc6781164)

[4. Instalación por Sistemas Operativos 11](#_Toc6781165)

[5. Instalaciones por día 13](#_Toc6781166)

[a. Lunes 13](#_Toc6781167)

[b. Martes 14](#_Toc6781168)

[c. Miércoles 15](#_Toc6781169)

[d. Jueves 15](#_Toc6781170)

[e. Viernes 16](#_Toc6781171)

[f. Sábado 16](#_Toc6781172)

[g. Domingo 17](#_Toc6781173)

[6. Conclusión 17](#_Toc6781174)

[6. Conclusión Final 18](#_Toc6781175)

# Introduccion

Este informe se encarga de analizar los datos sobre un conjunto de set de datos provisto por la empresa Jampp.

Dentro del directorio del set de datos podemos encontrar:

* installs
* clicks
* events
* auctions

El objetivo del TP es realizar un análisis exploratorio del set de datos.

## ¿Qué es Jampp?

JAMPP es una plataforma de marketing de rendimiento para adquirir y atraer clientes móviles. La compañía combina datos de comportamiento con tecnología predictiva y programática para generar ingresos para los anunciantes al mostrar anuncios personales y relevantes que instan a los consumidores a comprar por primera vez o más a menudo.

Fundado en 2003, nuestro equipo atiende a una base global de clientes desde oficinas en San Francisco, Londres, Berlin, San Pablo, Singapur, Ciudad del cabo y Buenos Aires.

# Análisis Set de Datos Clicks

# Análisis Set de Datos Auctions

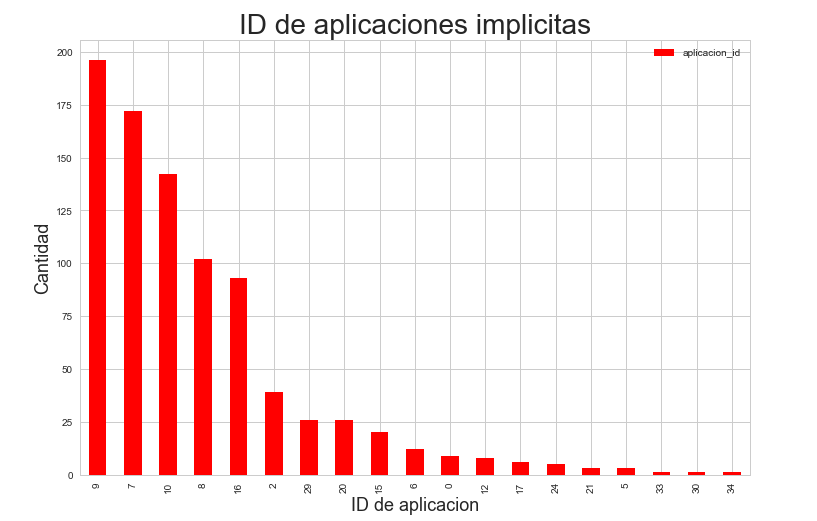
# Análisis Set de Datos Events

# Análisis Set de Datos Installs

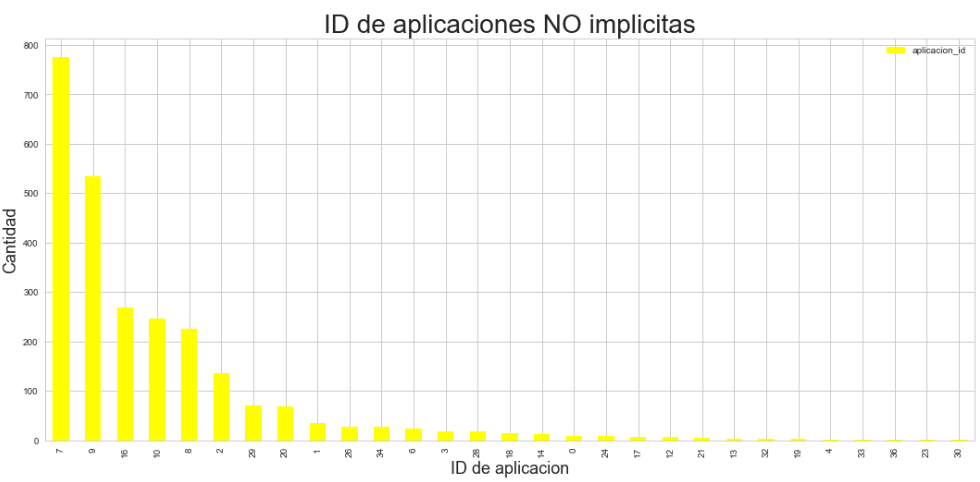
Para comenzar, se carga el archivo con los datos que contiene los datos de las instalaciones de aplicaciones hechas por los usuarios y se estudia que campos tenemos y verificamos las descripciones de cada campo con el archivo de descripciones que fue provisto especialmente para el trabajo práctico.

## Implicit Install

Lo primero que se realiza es un filtrado de todas las aplicaciones que se instalaron de forma implícita, esto significa que la instalación fue realizada por un dispositivo que no se ha instalado de acuerdo con la plataforma de seguimiento.

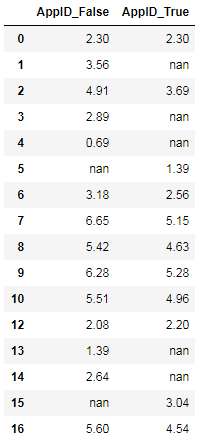
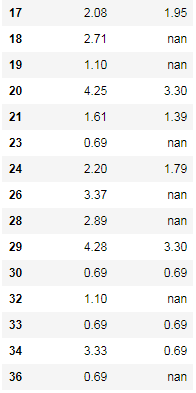


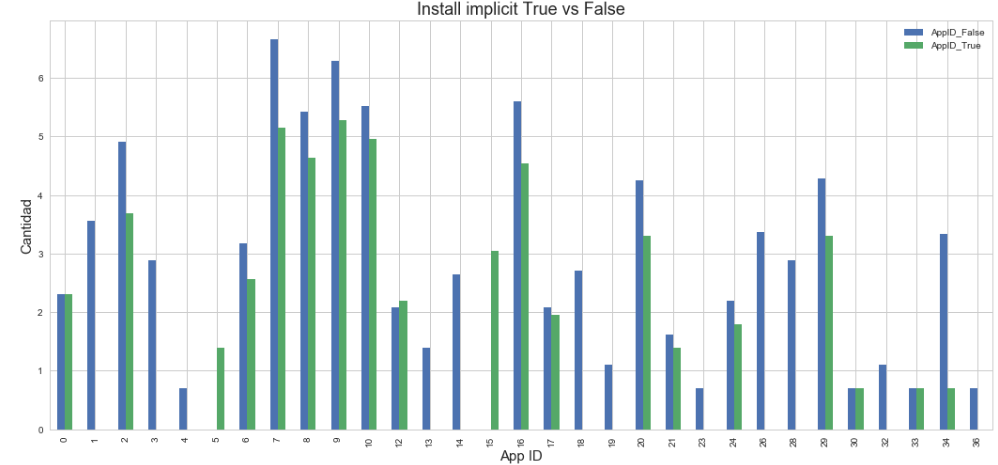
Luego, se realiza el mismo análisis para las instalaciones que no fueron de forma implícita.



En este segundo gráfico, vemos que hay mayor cantidad de aplicaciones instaladas de forma no implícita, hay 29 aplicaciones instaladas sobre un total de 30 estudiadas en el set de datos.

Haciendo una comparación del total de las aplicaciones que tanto las aplicaciones n°7 y n°9 son las aplicaciones más instaladas en ambas formas.



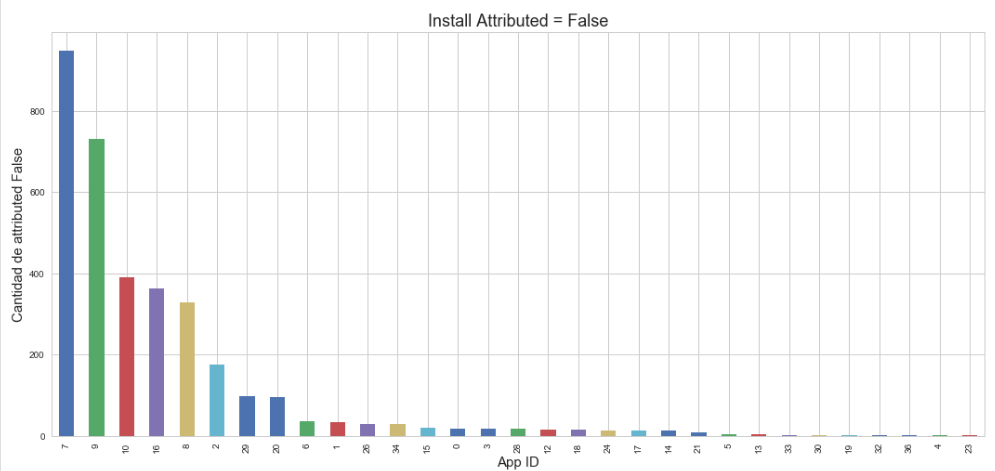


Hubo más instalaciones de más aplicaciones de forma no implícita, por no ser dispositivos que hicieron las instalaciones pero no por la plataforma de seguimiento, o sea por las publicidades que aparecen y el usuario le hace click, sino que lo hace por otra vía.

## Atributed Install

Las instalaciones hechas pueden ser atribuidas a Jampp o no.

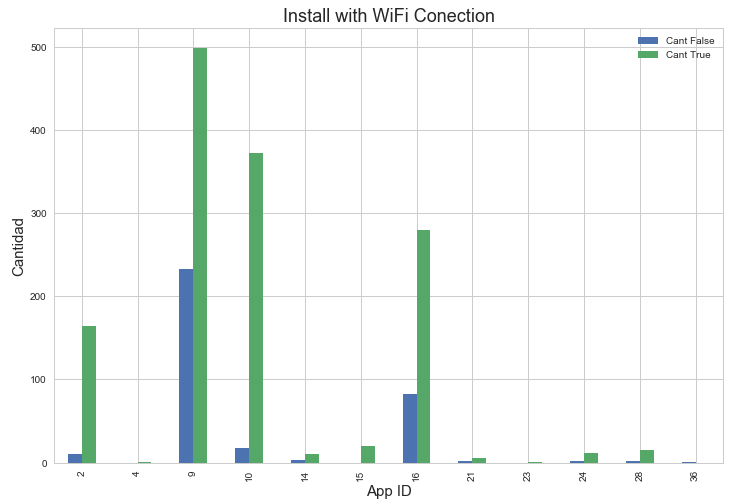
Al realizar el estudio de los datos, se verifica que ninguna instalación fue atribuida a Jampp, por los que todas las instalaciones hechas de todas las aplicaciones fueron hechas por fuera de las plataformas de Jampp.



El gráfico está ordenado de mayor cantidad a menor cantidad de instalaciones sin atribuir a Jampp y seguimos viendo como la aplicación 7 y 9 son las que lideran las instalaciones.

## Install with WiFi

Para este análisis, lo que tenemos es varios datos de si la instalación se hizo en dispositivos con conexión de WiFi o no.



Como se puede apreciar en el gráfico, hay muy pocos datos explícitos de las instalaciones con dispositivos que permiten acceder a la información dela conexión de WiFi, tanto si está conectado o no a la red.

De igual manera, en todos los casos la cantidad de instalaciones hechas con WiFi activado son mayores que las que no, por lo que el usuario aprovecha para descargar e instalar aplicaciones cuando está conectado a una red inalámbrica para no hacer uso de sus propios datos móviles, y poder hacer uso de los mismos en su rutina diaria y también poder reducir sus gastos.

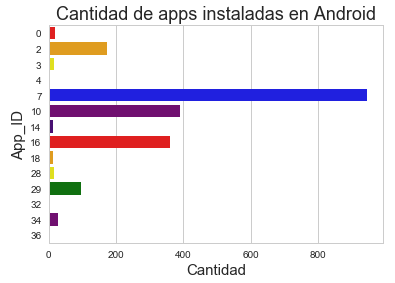
## Instalación por Sistemas Operativos

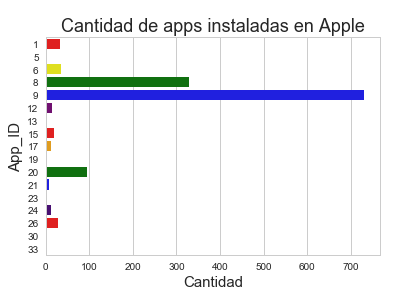
En la columna del set de datos “ref\_Type” donde hay un número que puede ser tanto Apple\_ifa o google\_advertising\_id según el archivo de descripciones.

Al juntar y contar los valores, se observa que solo hay dos tipos de “ref\_type” por lo que se supone que el de mayor valor es el de google\_advertising\_id por lo que son dispositivos con sistema operativo Android y el otro tipo corresponde a Apple\_ifa entonces deducimos que son dispositivos con iOS como sistema operativo.



Del total de instalaciones en el período estudiado (9 días) se notan que las instalaciones en dispositivos Android supera ampliamente a las de iOS.



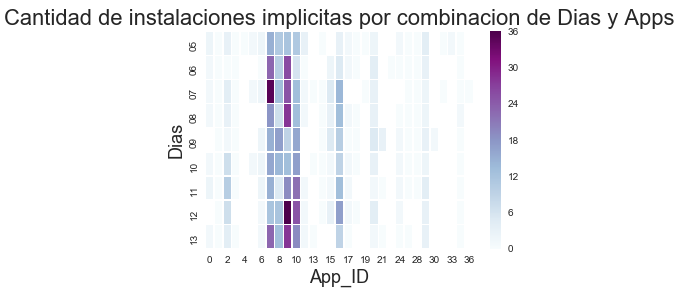


Lo que se observa de estos gráficos, es que las dos aplicaciones más instaladas, la n°7 y la n°9, son instaladas solamente en dispositivos con Android y con Apple respectivamente. Lo mismo sucede con el resto de las aplicaciones, solo son instaladas en cada dispositivo y nada más.

## Instalaciones por día

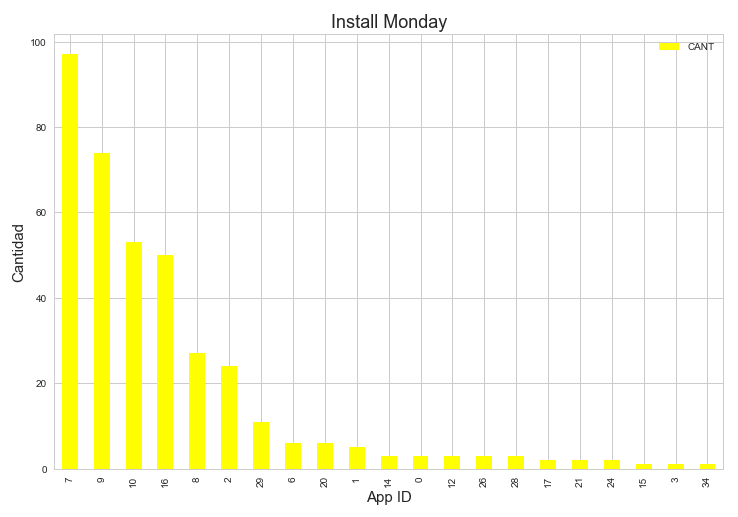
Todas las instalaciones vienen en el set de datos con un campo de creación. Ese campo viene en el fomato de aaaa-mm-dd. Por lo que se separa la fecha y se divide en nuevas columnas donde se guardan solamente el año, el mes y el día por separado.

Al tener estos datos por separado y sabiendo que el año y el mes no cambian en todos los datos, solo tomo los días para analizar.



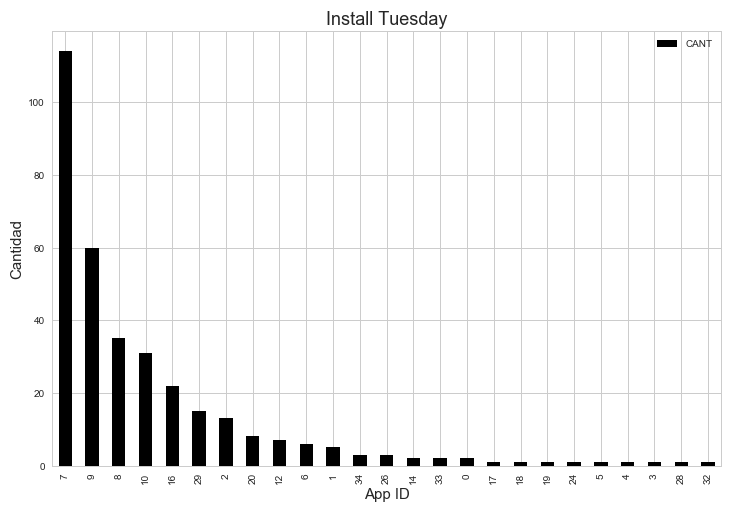
Viendo este HeatMap se observa que de las dos aplicaciones más instaladas, la n° 7 es instalada con mayor fuerza en el día 7 del mes y la n° 9 el día 12 del mes. Luego, el resto de las aplicaciones se van instalando durante todo el ciclo estudiado.

### Lunes



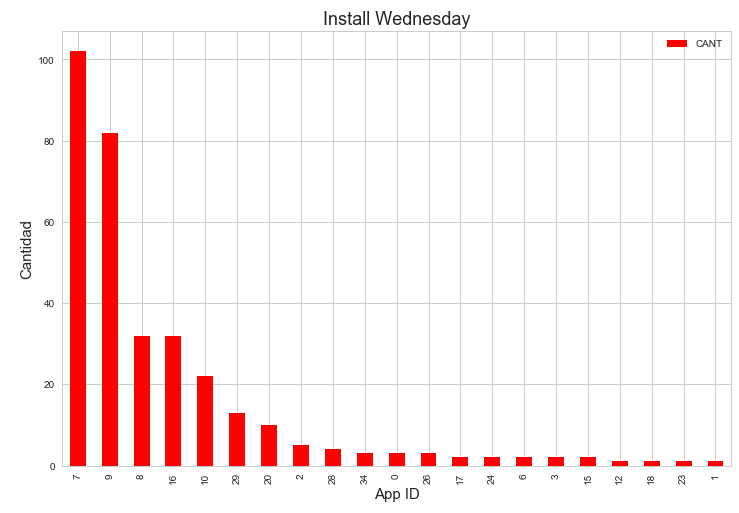
### Martes

Al ser el ciclo de 9 días, hay dos martes dentro de del período, el 5 y 12 de Marzo del 2019. Por lo que al filtrar los días por separado, luego los junto para analizar en conjunto los martes.

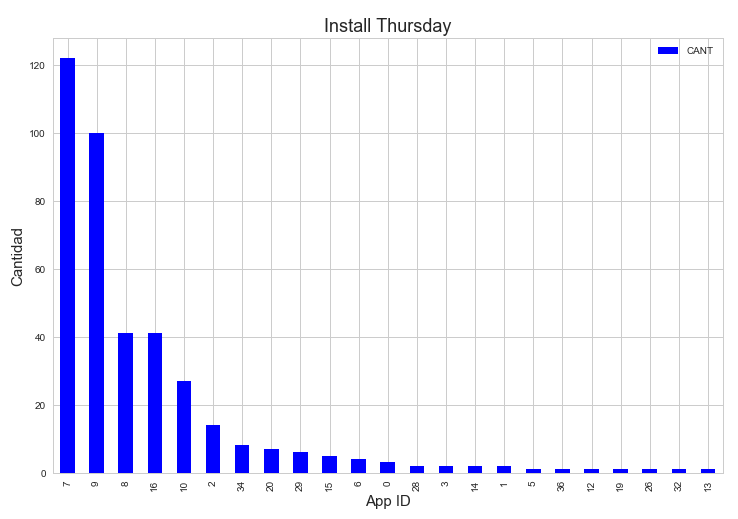


### Miércoles

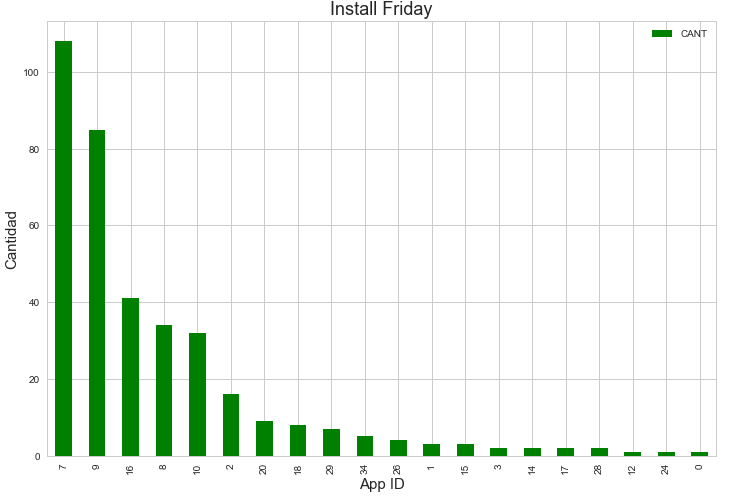
Al igual que el martes, es el otro día que se repite dentro del período a analizar. Se actuó de la misma manera que con el día martes.



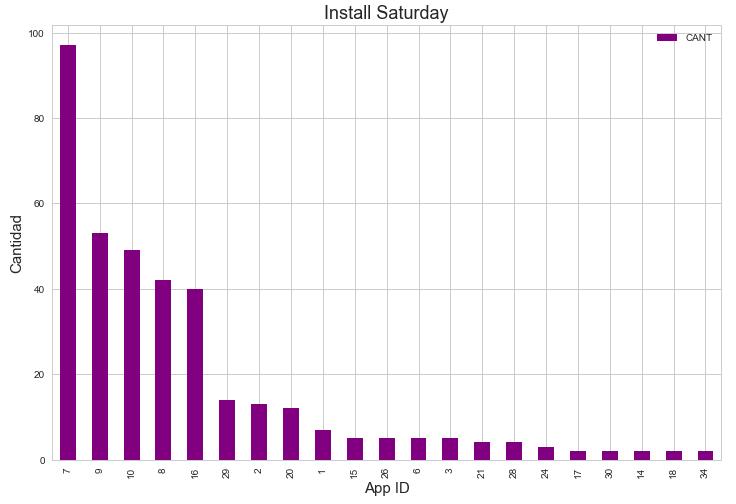
### Jueves



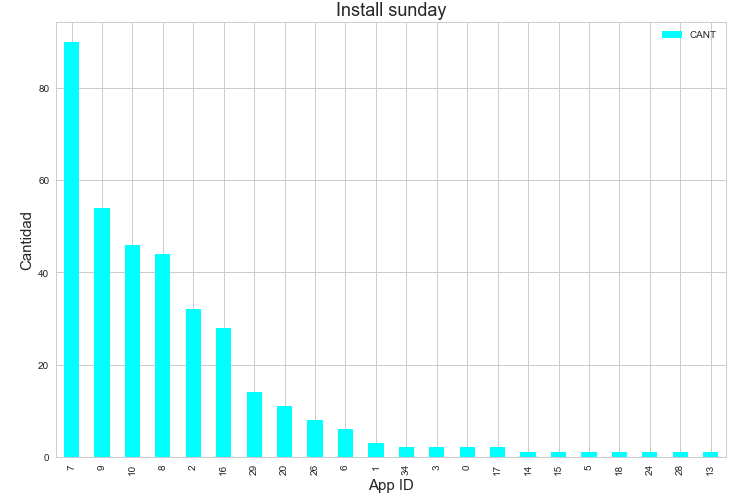
### Viernes



### Sábado



### Domingo



Según los gráficos de cada día, los jueves es cuando hay mayor cantidad de instalaciones y en el fin de semana, sábado y domingo, es cuando hay menos cantidad. Esto puede ser porque baja la actividad en el uso del celular entonces hay menos instalaciones.

## Conclusión

A la hora de las instalaciones hechas por los dispositivos, las no implícitas superan a las que sí lo son, esto puede ser porque el usuario llega a la aplicación por otro medio que los eventos que aparecen en un juego u otro lado. Esta puede ser por publicidad gráfica, la televisión o la recomendación de otra persona. Lo importante para el cliente es que la aplicación sea instalada. A su vez, ninguna es atribuida a Jampp, por lo que habría que mejorar las formas para que sea rentable el trabajo.

Es común que la mayoría sean en dispositivos con Android, ya que son muchas las marcas que eligen el sistema operativo de Google, en cambio solo los productos Apple tienen a iOS como sistema operativo, y son muchos menos comparados con la gran variedad del resto.

Los días jueves son los de mayor descarga, puede ser porque es un día próximo al fin de semana donde el usuario común disminuye el uso de los dispositivos.

# 6. Conclusión Final

* Mayor instalaciones NO implícitamente hechas
* Mayor cantidad de instalaciones en dispositivo Android, porque tiene más variedad de dispositivos disponibles con ese sistema operativo
* Los fin de semana disminuye la cantidad de instalaciones
* NINGUNA instalación se le atribuye a Jampp
* De los datos que tenemos, predominan las instalaciones hechas con el WiFi conectado, esto puede ser porque el usuario prefiere no gastar sus propios datos móviles para hacerla descarga y la instalación de las aplicaciones.