



## Análisis Exploratorio de Datos

### Trabajo Práctico 1 - Organización de Datos

Nombre de grupo

Poner fecha

Nombre	Padrón	Mail
Álvarez, Federico	99266	fede.alvarez1997@gmail.com
La Torre, Gabriel	87796	latorregab@gmail.com
Medrano, Lucas Nicolás	99247	lucamedrano97@gmail.com
Piro Martino, Ariel	99469	ariel.piro@hotmail.com

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Análisis individual de archivos</b>	<b>3</b>
2.1	Subastas . . . . .	3
2.1.1	Análisis general . . . . .	3
2.1.2	Subastas por día de Marzo . . . . .	3
2.1.3	Subastas por día de Marzo por sistema operativo . . . . .	5
2.1.4	Subastas por hora del día . . . . .	5
2.1.5	Subastas por día por sistema operativo . . . . .	5
2.1.6	Subastas por sistema operativo . . . . .	5
2.1.7	Subastas por source . . . . .	5
2.2	Clicks . . . . .	5
2.3	Eventos . . . . .	5
2.4	Instalaciones . . . . .	5
<b>3</b>	<b>Análisis de archivos en conjunto</b>	<b>5</b>
<b>4</b>	<b>Conclusion</b>	<b>5</b>

# 1 Introduction

En el trabajo se hace un análisis exploratorio de un set de datos provistos por la empresa Jampp. En el mismo se encuentra información de subastas, instalaciones, clicks, entre otros.

Primero se hará una visión general de los archivos (installs.csv, clicks.csv, auctions.csv, events.csv) para entender la distribución y la cantidad de datos, el significado de las columnas, reconocer las columnas que no aportan información (por ejemplo, las que tienen todos sus valores nulos), reconocer el tipo de datos en cada columna, seguido de un análisis mas profundo para obtener mas información de los datos. Luego se hará un análisis global, buscando información relativa a los archivos en conjunto, permitiendo obtener otro tipo de información.

## 2 Análisis individual de archivos

### 2.1 Subastas

#### 2.1.1 Análisis general

El archivo 'auctions.csv' contiene información acerca de subastas. Hay dos columnas que no nos aportan información significativa. 'auction\_type\_id' tiene todos sus valores nulos, por lo que no fue tomada en cuenta para el análisis. 'country' informa un solo valor posible, que se supone que debe ser Argentina.

La columna platforms tiene dos valores posibles (1 y 2) que se supone son Android e iOS. Va a ser importante para el análisis que hagamos más adelante.

Por último, 'source' nos da información acerca del exchange de donde surge la subasta.

Además vemos que ningún valor de este archivo, excluyendo la columna 'auction\_type\_id', es nulo, por lo que no es necesario tomar ninguna decisión respecto a eso.

#### 2.1.2 Subastas por día de Marzo

En el gráfico siguiente se pueden observar varias cosas:

1. La cantidad de subastas parece, en general, aumentar con el paso de los días.
2. El valor del último día es mas del doble del valor del primer día.
3. El mayor aumento se da del sexto al septimo día.
4. La cantidad de subastas se mueve entre unos valores extremos de un millón y 3 millones.

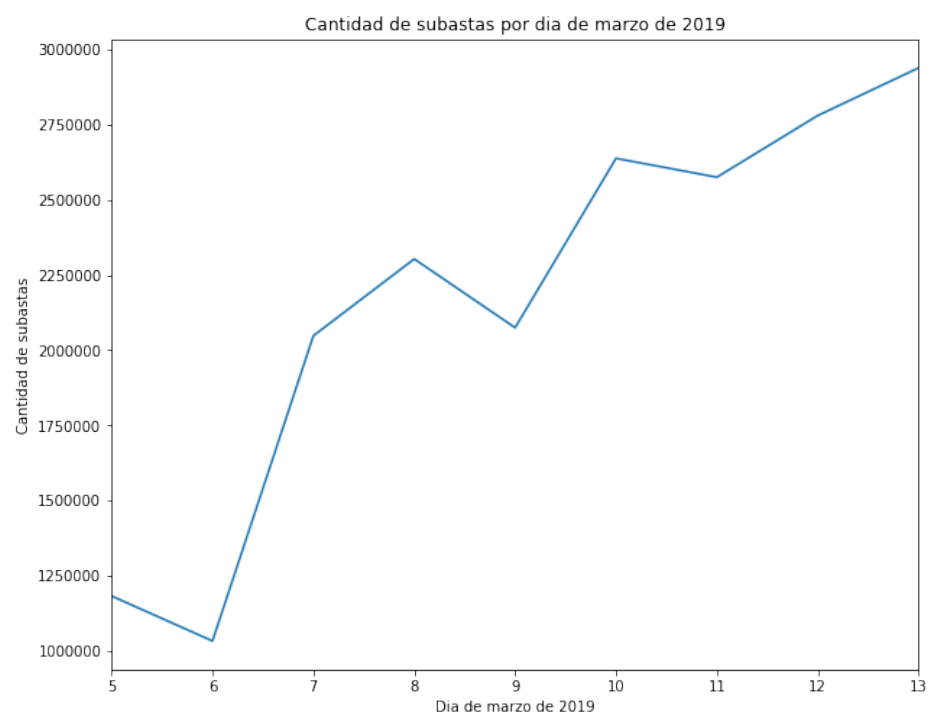


Figure 1: Subastas por día de Marzo

- 2.1.3 Subastas por día de Marzo por sistema operativo
- 2.1.4 Subastas por hora del día
- 2.1.5 Subastas por día por sistema operativo
- 2.1.6 Subastas por sistema operativo
- 2.1.7 Subastas por source
- 2.2 Clicks
- 2.3 Eventos
- 2.4 Instalaciones
- 3 Análisis de archivos en conjunto
- 4 Conclusion