



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

1	Introduction	3
2	Análisis individual de archivos	3
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	4
2.1.4	Subastas por hora del día	5
2.1.5	Subastas por día por sistema operativo	6
2.1.6	Subastas por sistema operativo	6
2.1.7	Subastas por source	6
2.2	Clicks	6
2.3	Eventos	6
2.4	Instalaciones	6
3	Análisis de archivos en conjunto	6
4	Conclusion	6

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. De ahora en más, platform y sistema operativo serán sinónimos en este informe.

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

Como primer acercamiento a este set de datos, es interesante ver cómo se distribuye la cantidad de subastas en los días que incluye el archivo (05/03/19 al 13/03/19). El gráfico 1 representa dicha distribución. En él se pueden observar varias cosas:

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

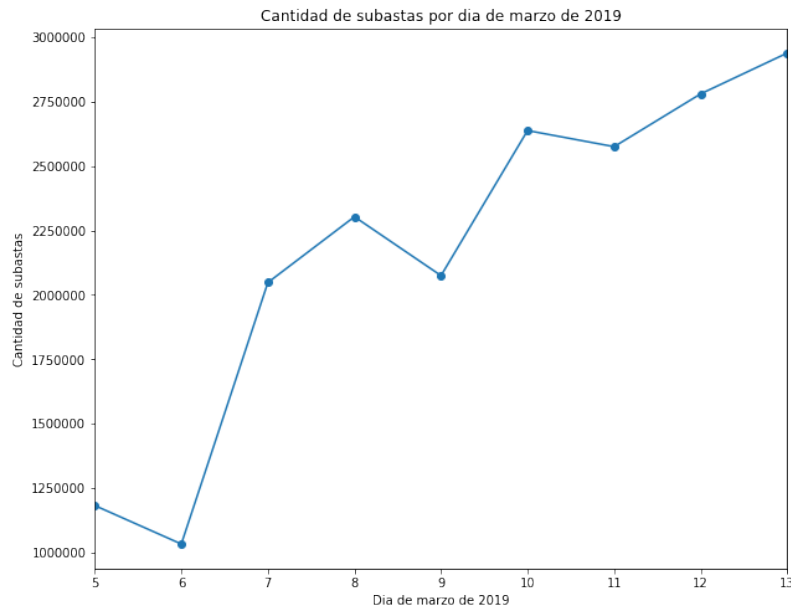


Figure 1: Cantidad de subastas por día de Marzo

2.1.3 Subastas por día de Marzo por sistema operativo

Otro punto interesante es dividir el problema. Obtener la distribución de subastas en los días con datos disponibles para cada plataforma (Android e iOS). En la imagen 2 se pueden observar las cantidades. Obsérvese que llamamos '1' y '2' a las plataformas, ya que no sabemos cuál es Android y cuál es iOS.

Puntos interesantes a reconocer:

- La cantidad de subastas para la plataforma '1' es, salvo en el cuarto y el quinto día, considerablemente mayor a la cantidad para la plataforma '2'.
- La figura de la plataforma '2' es mucho mas "chata" que la de la plataforma '1', la cual representa mas picos y saltos.
- La figura de la plataforma '1' es muy parecida a la del gráfico 1, mientras que la de la plataforma '2' no lo es. Esto es resultado, principalmente, de lo indicado en el primer ítem. Este análisis puede llegar a ser muy útil para reconocer partes de los datos que son representativas del total.

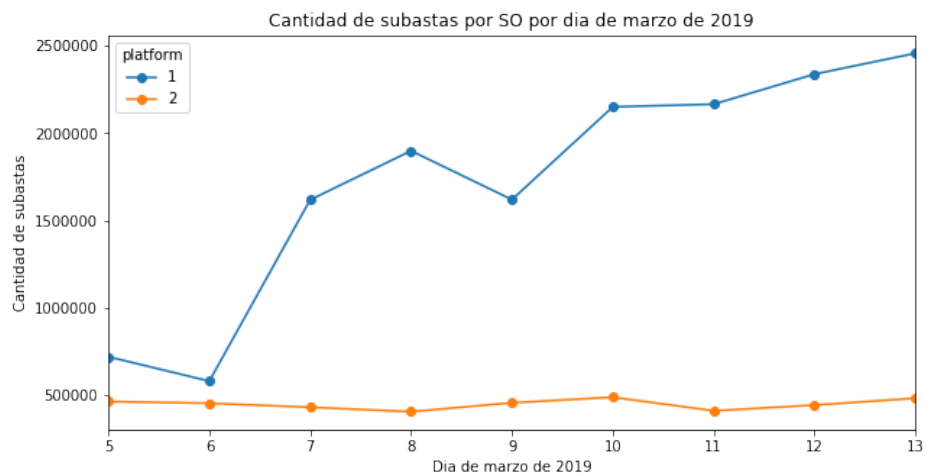


Figure 2: Cantidad de subastas por día de Marzo y por sistema operativo

2.1.4 Subastas por hora del día

Ahora vamos a analizar cómo se distribuyen las subastas a lo largo del día. Para esto hacemos un gráfico de hora del día contra cantidad de subastas (Gráfico 3).

Desde un análisis cualitativo se pueden observar algunos puntos:

- Parece ser que la mayor cantidad de subastas se distribuyen por la noche y la madrugada.
- La cantidad de subastas es poca en horas de la mañana y el mediodía. En el gráfico se puede ver un gran valle en esa parte del día.
-

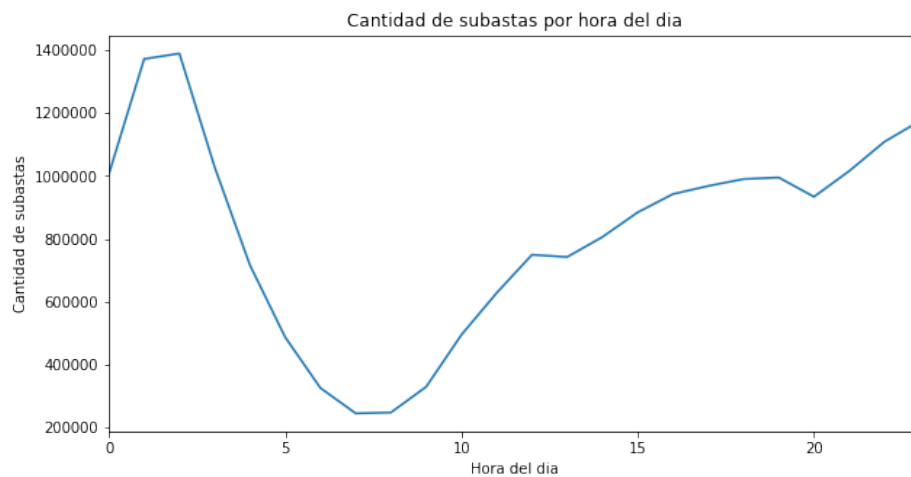


Figure 3: Subastas por día de Marzo

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