

## **75.06/95.58 Organización de Datos**

### **Primer Cuatrimestre de 2019**

### **Trabajo Práctico 1: Enunciado**

Para esta primera entrega del TP, realizaremos el análisis exploratorio del set de datos provisto por la empresa Jampp.

Los datos a analizar estan disponibles en

<https://drive.google.com/drive/folders/1-7ACCWhS3sWVh0vU273YwbS74WmzEhLa?usp=sharing>

En el directorio podemos encontrar:

- installs
- clicks
- events
- auctions

El objetivo del primer TP es realizar un análisis exploratorio del set de datos del TP. Queremos ver qué cosas podemos descubrir sobre los datos que puedan resultar interesantes. Los requisitos de la primera entrega son los siguientes:

- El análisis debe estar hecho en Python Pandas o R.
- El análisis debe entregarse en formato papel en una carpeta en donde se incluya el reporte completo y todas las visualizaciones generadas. Es altamente recomendable que las visualizaciones se impriman en color.
- Informar el link a un repositorio Github en donde pueda bajarse el código completo para generar el análisis.
- Agregar en Kaggle un kernel con el análisis exploratorio realizado (deseable pero no mandatorio)

La evaluación del TP se realizará en base al siguiente criterio:

- Originalidad del análisis exploratorio.
- Calidad del reporte. ¿Está bien escrito? ¿Es claro y preciso?
- Calidad del análisis exploratorio: qué tipo de preguntas se hacen y de qué forma se responden, ¿es la respuesta clara y concisa con respecto a la pregunta formulada?
- Calidad de las visualizaciones presentadas.
  - ¿Tienen todos los ejes su rótulo?
  - ¿Tiene cada visualización un título?

- ¿Es entendible la visualización sin tener que leer la explicación?
- ¿El tipo de plot elegido es adecuado para lo que se quiere visualizar?
- ¿Es una visualización interesante?
- ¿El uso del color es adecuado?
- ¿Hay un exceso o falta de elementos visuales en la visualización elegida?
- ¿La visualización es consistente con los datos?
- Conclusiones presentadas.
  - ¿Presenta el grupo un listado de "insights" aprendidos sobre los datos en base al análisis realizado? ¿Es interesante?
  - ¿Identificaron patrones o funnels de usuarios?
  - ¿El análisis realiza un aporte a Jampp?

El grupo que realice el mejor análisis exploratorio obtendrá 10 puntos para cada uno de sus integrantes que podrán ser usados en el parcial además de ser publicado en el repositorio de la materia como ejemplo para los siguientes cuatrimestres.

Como dato a tener en cuenta, para el TP2 se trabajará con estos mismos datos intentando determinar:

- En un instante dado, estimar  $St(d)$  el tiempo hasta que un dispositivo  $d$  aparezca de vuelta en una subasta RTB
- En un instante dado, estimar  $Sc(d)$  el tiempo hasta que un dispositivo  $d$  convierta

La idea del TP1 es obtener información que facilite luego plantear una solución para el TP2.

## Glosario

- Convertir: el objetivo de mostrar publicidad es que un dispositivo instale una aplicación, a ese evento se le llama conversión.
- Dispositivo: *entidad con un id de publicidad asociado. Por ejemplo: un celular Samsung J6 con Android tiene un [id único](#), un Apple iPhone tiene un [identificador único](#).*
- Evento: cualquier tipo de acción categorizada dentro de una aplicación. *Por ejemplo, en una aplicación de e-commerce un [funnel de eventos](#) muy común puede ser del estilo "abrir\_app" → "buscar\_producto" → "revisar\_catalogo" → "agregar\_a\_carrito" → "efectuar\_compra". Cada uno de estos pasos es un **evento**.*
- Subasta: en el momento que una aplicación quiere mostrar una publicidad, ese espacio se vende en una subasta (generalmente de segundo precio) donde todos los interesados en mostrar una publicidad ofertan un precio y gana quién más ofrece.