**ENUNCIADO ACTIVIDAD 1**

# Nombre del caso o enunciado de la actividad de entrega 1:

***Análisis de datos de navegación y conversión de usuarios.***

Los usuarios cuando quieren contratar un producto lo buscan en internet, y llegan en la mayoría de los casos a unas páginas web especializadas en conversión llamadas landings. Estas landings suelen contener un teléfono al cual el usuario puede llamar al call center de la compañía y/o también un formulario en el cual dejan su nombre y su teléfono para que dicho call center se ponga en contacto con ellos.

Existen diferentes tipos de conversiones siendo los más comunes:

* simplemente el contacto con el call center ya sea a través de una llamada o del envío del formulario.
* Contratación del producto Nuestro ejercicio

El cliente es una compañía de coches y nos pide que le realicemos informes sobre las visitas a las landings de sus productos y las conversiones.

Datos

Tenemos dos tipos de ficheros:

1) De navegación, en el que se recogen los datos de navegación de los usuarios. Este fichero contiene las siguientes columnas:

uuid(string): Es utilizado para crear identificadores únicos universales que permitan reconocer e distinguir un objeto dentro de un sistema, o el mismo objeto en diferentes contextos. En este caso el objeto es el usuario. gclid(string): El ID de clic de Google (GCLID) es un parámetro que se transfiere a la URL a través de los clics en anuncios.

**id\_user**(string): identificador único de usuario de la aplicación que se comunica con el call center

user\_recurrent(boolean): nos indica si el usuario es recurrente (true), es decir, si ha entrado más de una vez al conjunto de landings de la compañía. Si

no es recurrente (false), significa que el usuario está entrando por primera vez al conjunto de landings.

url\_landing (string): cuando el usuario hace click en un anuncio o sitelink se lo envía a una landing con una url que contiene varios datos interesantes para su análisis. A continuación, explicamos cada parte de esta url.

Ejemplo:

https://[www.metropolis.com/es/home/gclid=Cj0KCQjw1dGJBhD4ARIsANb6OdmV6XYIc](http://www.metropolis.com/es/home/gclid%3DCj0KCQjw1dGJBhD4ARIsANb6OdmV6XYIc) MpvAhDvnmHLRGVelRzIeciTG3j1ItnthWfYtV\_XFWKDhLsaAr4sEALw\_wcB&iduser=ec 1eef0d-6141-4a85-86cc-979a653362eb&uuid=1a05bd93-b939-4661-924b- a643cfcec3de&camp=732187328&adg=46724581628&device=m&sl=&adv=533655604 [703&rec=false](https://www.metropolis.com/es/)

https://[www.metropolis.com/es/](http://www.metropolis.com/es/)  es la url base

Después de la url base le sigue la landing específica a la cual ha ido el usuario: home (que sería la web principal), o una landing de algún modelo de coche.

Luego siguen: gclid, iduser, uuid que ya los hemos definido anteriormente Finalmente tenemos los datos de las campañas:

* camp: id de la campaña
* adg: id del adgroup
* device: dispositivo desde el cual el usuario está accediendo a la landing. c: computer, t: tablet, m: móvil
* adv: id del anuncio
* sl: id del sitelink. Este puede ser un número si el usuario ha hecho click en el sitelink, o 0 o vacío si el usuario ha hecho click en el anuncio.

1. En el caso del fichero de conversiones tenemos las siguientes columnas:

date: fecha de la conversión hour: hora de la conversión

**id\_lead**: identificador único de la conversión

lead\_type: tipo de conversión, CALL el usuario llama directamente al call center, FORM el usuario rellena un formulario y espera que el call center lo

llame

result: resultado después de haber hablado con el call center

* + ilocalizable: no logran hablar con el usuario porque justo el usuario cuelga la llamada o no atiende
  + Positivo: que le interesa y va a contratar
  + No le interesa



En este esquema vemos como se organizan los datos en Adwords.

Una campaña puede contener uno o varios grupos de anuncios. Y un grupo de anuncios puede contener uno o varios anuncios.

Los sitelinks suelen depender solo de la campaña. Es decir, una campaña puede contener uno o varios sitelinks.

# Preguntas a responder

* Cuántas visitas recibe en el día el cliente

**o** Cuántas de ellas convierten y cuántas no (en %)

* Por tipo de conversión (CALL o FORM), ¿cuántas hay de cada una?
* Porcentaje de usuarios recurrentes sobre el total de usuarios
* Coche más visitado. ¿Es el que más convierte?

# Descripción de la actividad

Pasos para realizar la práctica

1. Leer los datos con python. Ficheros de navegación y de conversión
2. Separar los datos en columnas, y obtener para cada línea de navegación: campaña, adgoup, advertisement y site link que se obtiene de la columna URL.
3. Identificar si hay usuarios repetidos: id\_user, gclid, cookie

Para los que no tienen id\_user, hay que mirar el gclid, y si tampoco está hay que mirar la cookie

Y ordenar los datos según ts

1. Unir los datos de navegación ya tratados con los datos de conversiones,creando una columna de 0 y 1 indicando si el usuario no ha convertido o si ha convertido. La unión se hace a partir de la columna id\_suite, si esta está vacía sería por la de gclid, y si esta está vacía también por cookie.

Nota: tenemos 2 opciones

* 1. Si hay usuarios repetidos nos quedamos con un solo dato
  2. Si nos quedamos con todos los datos repetidos buscar el más cercano a la conversión (este punto es más complicado, es solo para quienes se atrevan)
     1. Con estos datos ya unidos y tratados debemos realizar diferentes informes que se proponen en el apartado siguiente (Entrega individual)