# Evaluación inicial de los datasets

A continuación, se describen las características principales del conjunto de datasets de los que disponemos.

Enlace: <https://drive.google.com/file/d/1d6Ax3h1ZpC0mDEIJKKJ-Bfw13G-qClzh/view?usp=sharing>

## Demográficos:

Este dataset consta de 171 registros y cuatro variables con las características demográficas de los usuarios. La primera variable es el identificador del usuario, la segunda es la edad, la tercera el género y la cuarta y, última, el país de origen del usuario. Todas las variables mencionadas anteriormente son categóricas, incluida la edad que se recoge en rangos.

A simple vista parece que el dataset tiene un registro por usuario. Sin embargo, se comprobará mas adelante en la etapa de limpieza de datos.

Por otro lado, se observa que el último Id de usuario corresponde al número 306 pero en el dataset solo hay 171 registros, por lo tanto, existen muchos usuarios de los cuales no se tiene la información demográfica.

## Obras:

Este dataset consta de 30 registros y 13 variables. Las variables se pueden dividir en tres grupos bien definidos. Por un lado, se tiene las variables con las características principales de la obra: Identificador, categoría artística, artista, título, localización (museo, ciudad y país) y año de creación. Esta última variable se recoge, de forma general, cómo un único año, pero existen registros donde se recoge con un rango de años. Estos casos se supone que corresponden a las obras de las cuales se desconoce el año exacto de creación.

El segundo grupo de variables corresponde a los enlaces (url) con la información siguiente: Imagen de la obra, información de la obra, información del artista.

Por último, las variables restantes corresponden a los distintos prefijos (wd:) necesarios para realizar consultas en SPARQL: identificador de la obra, dirección URL de la obra, identificador del artista y dirección URL del artista.

## Sentimientos:

Este dataset consta de 1759 registros y 4 variables. La primera y la segunda variables corresponden al identificador del usuario y artista respectivamente. La tercera recoge una emoción provocada al usuario por la obra. La última, corresponde a una variable de clasificación de la emoción.

A simple vista se observa que existen varios registros por usuario (un mismo usuario opina sobre diferentes obras) y varios registros por usuario y obra (un mismo usuario registra sobre la misma obra diferentes emociones).