# 1) SQL task

In the following test, you are expected to write SQL statements that would answer the questions below. You can submit your answers in a Word, Gdoc or txt file.

The file **works.csv** contains two columns: 1) uuid\_work is a unique identifier for every work 2) primary\_title is the work title.

The file **royalty\_statement.csv** represents royalty income for a client. It relates to the works.csv data through the foreign key uuid\_work.

**Questions:**

1. Which work titles are included in the royalty\_statement.csv?
2. Which work has generated the most total income (received\_royalty)?
3. Get all works that generated income in synch (income\_type= synch)

# 2) Python task

In the following test you are expected to write a python script or a jupyter notebook that reads the “songs.csv” file, does the necessary data manipulation and answers the following questions.

**Tasks/questions:**

1. Clean the **duration** column to only have integer values as total duration in seconds
2. Delete the songs with an empty **instrumental** column
3. Create a new field called “tag” and assign
   1. “energetic” if the song has **energy** higher than 0.7
   2. “positive” if the song has **energy** higher than 0.6 and **danceability** higher than 0.7
4. Answer the following questions
   1. What is **the average danceability** of a Backstreet Boys song?
   2. What is the **artist with the highest average popularity**?
   3. What is the average duration in seconds of mainstream (popularity > 50) and indie (popularity ) songs?