

ASSIGNMENT 4

This assignment will cover SQL queries. Please submit the Python notebook. As usual, we begin by loading the necessary libraries. You should not import any other libraries.

QUESTION 1

- a.) Create a sqlite database called **nanoim.db** with the tables given in the Python notebook
 - All numbers should be integers.
 - IDs should be the primary key.
 - All other entries should be of type TEXT.
 - Be sure to close the table once you have created it.
- b.) Reconnect with **nanoim.db** in a connection named `conn`.

For the rest of this question you will construct SQL queries and run them.

 - Make sure to store the result in the variable name given.
 - Make queries so that they would work if the tables are expanded or updated.
 - Use only sqlite3 functions.
- c.) Construct a query to return names of the actors who are ranked 3rd or higher (**The highest rank is 1**). Store the result in a variable called *actors*.
- d.) Construct a query to return the name of movies starring Keanu Reeves. Store the result in a variable called *keanu*.
- e.) Construct a query to return the directors of movies starring Tommy Wiseau. Store the result in a variable called *tommy*.
- f.) Close the connection.

QUESTION 2

We will connect to the database given in the file `music.db`. This database has quite a large number of tables. The relevant tables and their respective columns are listed below:

- **album:** album id, title, artist id
 - **artist:** artist id, name
 - **genre:** genre id, name
 - **track:** track id, name, album id, media type id, genre id, composer, milliseconds, bytes, unit price
- a.) Construct a query which lists the albums Red Hot Chili Peppers have in the database and put it into variable called *rhcp*.
 - b.) Construct a query which lists all the tracks on the album Achtung Baby and store the result in a variable called *u2*.
 - c.) Construct a query which returns how much it would cost to buy all Metallica songs to the nearest pound and store it in a variable called *rockon*.
 - You should assume the track costs are in pounds.
 - d.) Construct a query which returns the name and the length of the longest Rock song and transforms it in seconds (to the nearest second) and stores the name in a variable called *longsong* and the length in a variable called *time*.

QUESTION 3

We will connect to the database given in the file **`eateries.sqlite`**. This database has health inspection violations for San Francisco stored in 3 tables

- 'businesses'
- 'inspections'
- 'violations'

A function has been included that will print out the columns of the tables as pandas dataframes so you can examine them and then construct the queries. Unless otherwise stated you must only use SQL to get full marks.

- a.) How many restaurants do not have a precise geographic location given. Store the number (and only the number) in a variable called *locations*.
- b.) How many spellings of San Francisco are there? (You can assume that all variations count including modifiers such as north, south, east, west) also refer to San Francisco. **You can obtain the result through visual inspection.** Store the number in a variable called *SF*.
- c.) Create a histogram plot of the Inspection Scores for January 2014.
 - Here you should use the function `sql_to_frame()` which takes a query and returns a Pandas dataframe.
 - Remove any null values.