Important : The code should be production-ready (properly packaged , unit-tested , etc.)

The code should be in an executable format (no python notebook), it should output the

results into files in the file-system.

It must also come with the required instructions on how to use it.

Please implement the following 3 exercises using the Spark python API. The implementation should be done using **Spark DataFrame** and **Spark RDD**.

Please download the required dataset for these exercises here -
http://ocelma.net/MusicRecommendationDataset/lastfm-1K.html

You will also find the necessary information about the dataset in the link above

Exercises:

1. Say we define a user's "session" of Last.fm usage to be comprised of one or more songs played by that user, where each song is started within 20 minutes of the previous song's start time. Create a list of the top 10 longest sessions, with the following information about each session: userid, timestamp of first and last songs in the session, and the list of songs played in the session (in order of play).

2. Create a list of the 100 most popular songs (artist and title) in the dataset, with the number of times each was played.

3. Create a list of user IDs, along with the number of distinct songs each user has played.