# Open-ended Capstone – Step One – Project ideas

Set up ETL pipeline and database schema for the million-song dataset(<http://millionsongdataset.com/>). It contains metadata for a million songs. There are additional datasets available which can be used to further extend the primary one.

The entire dataset is 280 GB. A subset of 1.8GB is available for initial test set up.