**SQL Data Challenge**

**Goal**

The goal of this task is to assess your familiarity with data manipulation using SQL, and some basic skills in data analysis.

**Rules**

Submission date: your results must be in our mailbox within the shared due-date.

Submission format: depending on your favorite tools, you can submit a Jupyter Notebook or a pdf file with your analysis. In the second case, please attach the source code (sql scripts and other) that you used to get your results. Nice to have: business-oriented presentation including the analytical solution proposed and the followed methodological approach (ppt based).

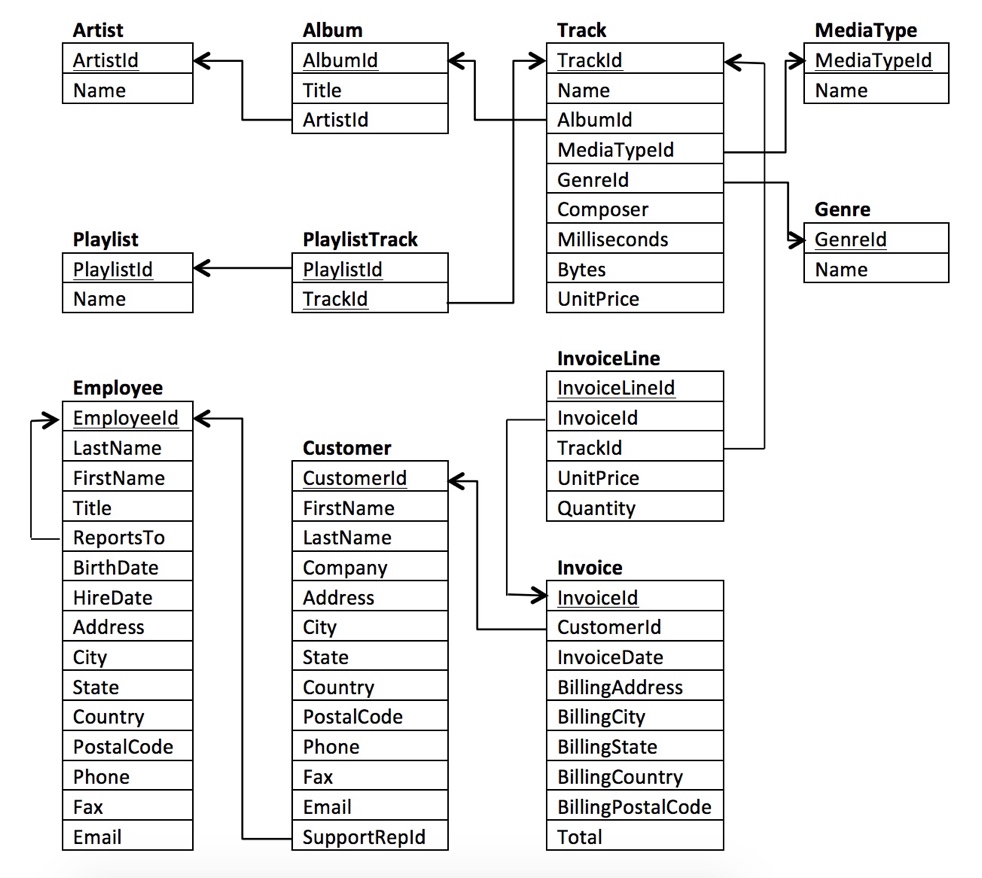
Time boxing: do not overstretch your effort. Try to limit your dedicated time to 8 hours maximum, including the creation of the report.

**Data**

The database that your asked to use is the open source Chinook database. The Chinook data model represents a digital media store, including tables for artists, albums, media tracks, invoices and customers.

Below you find a schematic of the data model, that is composed of several relations.

In order to access the data, go to the official git repository of the data model and follow the instructions: <https://github.com/lerocha/chinook-database>. We recommend SQLite as engine but feel free to develop your analysis using other solutions like Postgres and such.



**Required tasks**

1. SQL for single-question data analysis
2. Who is the employee that served more customers?
3. Who is the customer that bought more tracks?
4. Who is the employee that manages more employees?
5. Find the top-3 playlists in terms of profit for the media store
6. SQL for data pre-processing
   1. Build a *data matrix* (a single table) having in mind the following analysis goal: we want to analyze the customers behavior in order to optimize the strategies of the media store and to improve the satisfaction of the customers. For example, we would like to understand if there are specific patterns in the consumption of specific media type for given demographic groups.
7. Ad hoc data analysis
   1. Using the data matrix that you built in the previous step, implement with a tool of your choice (Python, R, Excel, you name it) the proposed analysis.

**What we expect from your results**

For the tasks group A, we expect a *correct answer*, but we also evaluate your coding style, the presence and clarity of comment in your code and, ideally, some consideration about the computational load that your query have (try to think about what would happen with your query if your media store scaled up overnight to billions of users).

For the tasks group B, we expect a clear and reproducible data preprocessing pipeline, clear and well commented queries, and a documented thought process (why did you make some decision?).

For the tasks group C, we would like to see curiosity, an investigative approach as well as clear and commented code.