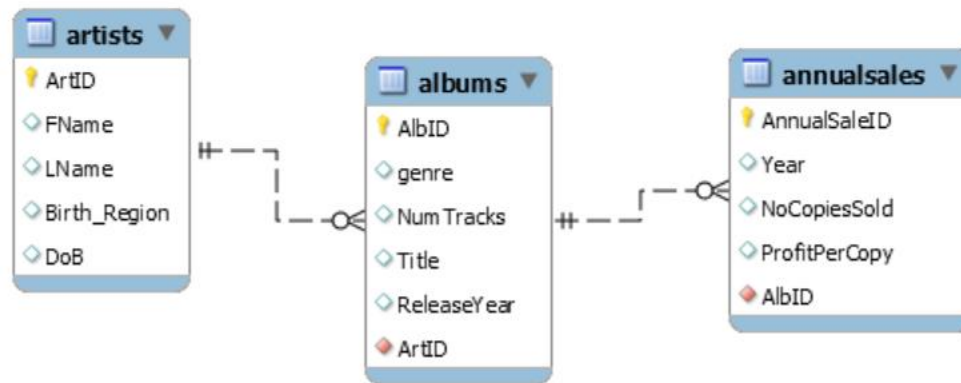


SQL Test: Musician Data Set ERD



This is a fictitious data set generated by Professor Claggett.

Artist Table – This is designed to capture all of the musicians we are tracking. This data set only contains solo artists (no bands or multiple artist collaborations). Attributes will include a unique ID, their first name, last name, region where they were born, and their date of birth. There are 350 musicians in this table.

Album Table – These are individual albums a musician may have released. Each album is associated with one and only one artist. These are “traditional” album releases – no single track releases or any other type of promotional material. In addition to a unique ID, each album will be described with a genre, a title, the number of tracks on the album, and year it was released. One quirky thing about this data set is there are a lot of the same title names across the albums. So musician X may have released a Country album named “Spooky Hearts” in 2014 and musician Y may have released a Rock album named “Spooky Hearts” in 2016. These are still two unique albums that just happen to share the same name. There are 443 albums described.

Annual Sales Table – This granularity is an album’s sales over a single year. So each row represents one year of sales for a specific album. In addition to a unique ID, expect the year of the sales it’s related to, as well as the total number of copies sold of that album that year, and the profit made for each individual sale. For example, Album #123 might have been released in 2012, and have three associated rows of data in annual sales...it’s sales in 2012, in 2013, and 2014. Each of those rows will include how many copies of album #123 were sold that year (e.g., in 2014 we sold 4,000 copies of album #123). It will also have how much profit each sale involved (e.g., in 2014, we made \$5.43 on each copy of #123). Therefore, to find the total profit made from album #123 in 2014, you would need to multiply $4,000 * \$5.43$. One additional known limitation of the dataset is not all years of sales data are known for each album. In general, this situation tracked three years of data after the release of each album. There are 1,329 rows of data.