Kevin Denny

10/04/2022

CS479

# Unit 3 Individual Project

## Introduction:

The objective of this project is to create a simple visualization in Tableau. Along the process I will demonstrate my ability to use Extract, Translate and Load (ETL) for the given dataset. In addition, I must discuss the role of databases and how the data set should be stored, then discuss the role of automation in the ETL pipeline process.

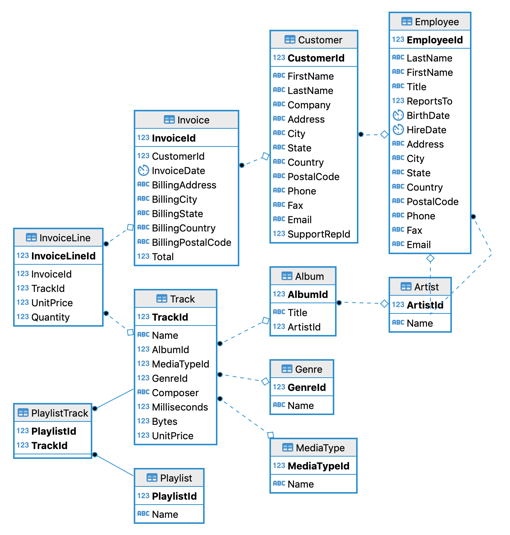
Rather than use a dataset for this project, I have opted to use a pre-defined database. This database is aptly named Chinook Database and represents a sample of transactions within an iTunes Library. Chinook is considered an Online Transaction Processing (OLTP) database.

To fulfill the requirements of this projects I will be using SQL to complete the steps I the ETL process and generate an automated pipeline from those steps. I will also be discussing how I the data is stored to create an OLAP database, then use that OLAP database to create a dashboard in Tableau.

## Chinook Database:

The Chinook Database is locally stored in a Postgres Database. This database is broken down into 11 tables. The associated link and database diagram are below.

<https://github.com/lerocha/chinook-database>



## Questions to Analyze:

The dashboard should answer questions stakeholders have about the business. Given this data is about music and sales are completed online. Stakeholders may want to know what songs to recommend improving sales. Additionally, stakeholders may want to know where most of the sales are originating. Knowing where the sales are helps prioritize areas that need more funding for marketing and where to reduce the marketing campaigns. To put it more plainly some of these questions could be:

1. Are sales improving?
2. What are the most popular tracks?
3. Where are our sales the highest?

## Target Tables:

Next, tables containing data that can answer the questions are selected. This includes the following tables.

* Customer
* Invoice
* InvoiceLine
* Track

Tableau has a way of loading data from this database directly into the application. It allows you to select which tables to pull data and will automatically detect relationship between them. There may be times where it is unable to determine the association and this association would need to be manually created. In a sense Tableau automatically handles the ETL process. It extracts the data from the database, transforms the data so that it can be easily used for analysis. Then loads the data into a temporary dataset for use with the charts.

## Conclusion:

In this assignment I used a pre-configured database that was setup in a OLTP format. This database simulated transactional data from iTunes. This data was loaded into Tableau where the software was able to conduct an ETL process that quickly formatted data to easily integrate with the visuals. The automation in the process allows charts to easily update with new information as more transactions occur. Below is the dashboard created using the database.

