



**Misk Academy**  
**أكاديمية مسك**  
Empower The Next Generation

# Data Foundations Nanodegree Program

Project#3 [Query a Digital Music Store Database](#)

Name: Marwan Saeed Alsharabbi

Date: 09-11-2018

2018

## INTRODUCTION

**Chinook Database Introduction** In this project, you will query the Chinook Database. The Chinook Database holds information about a music store. For this project, you will be assisting the Chinook team with understanding the media in their store, their customers and employees, and their invoice information. To assist you in the queries ahead, the schema for the Chinook Database is provided below. You can see the columns that link tables together via the arrows.

All of the below instructions are discussed in detail as we work through this lesson on your way to completing this project. The below serves as a quick reference of what you will be doing as you progress through the completion of this project.

**Why the name Chinook?** The name of this sample database was based on the Northwind database. Chinooks are winds in the interior West of North America, where the Canadian Prairies and Great Plains meet various mountain ranges. Chinooks are most prevalent over southern Alberta in Canada. Chinook is a good name choice for a database that intends to be an alternative to Northwind

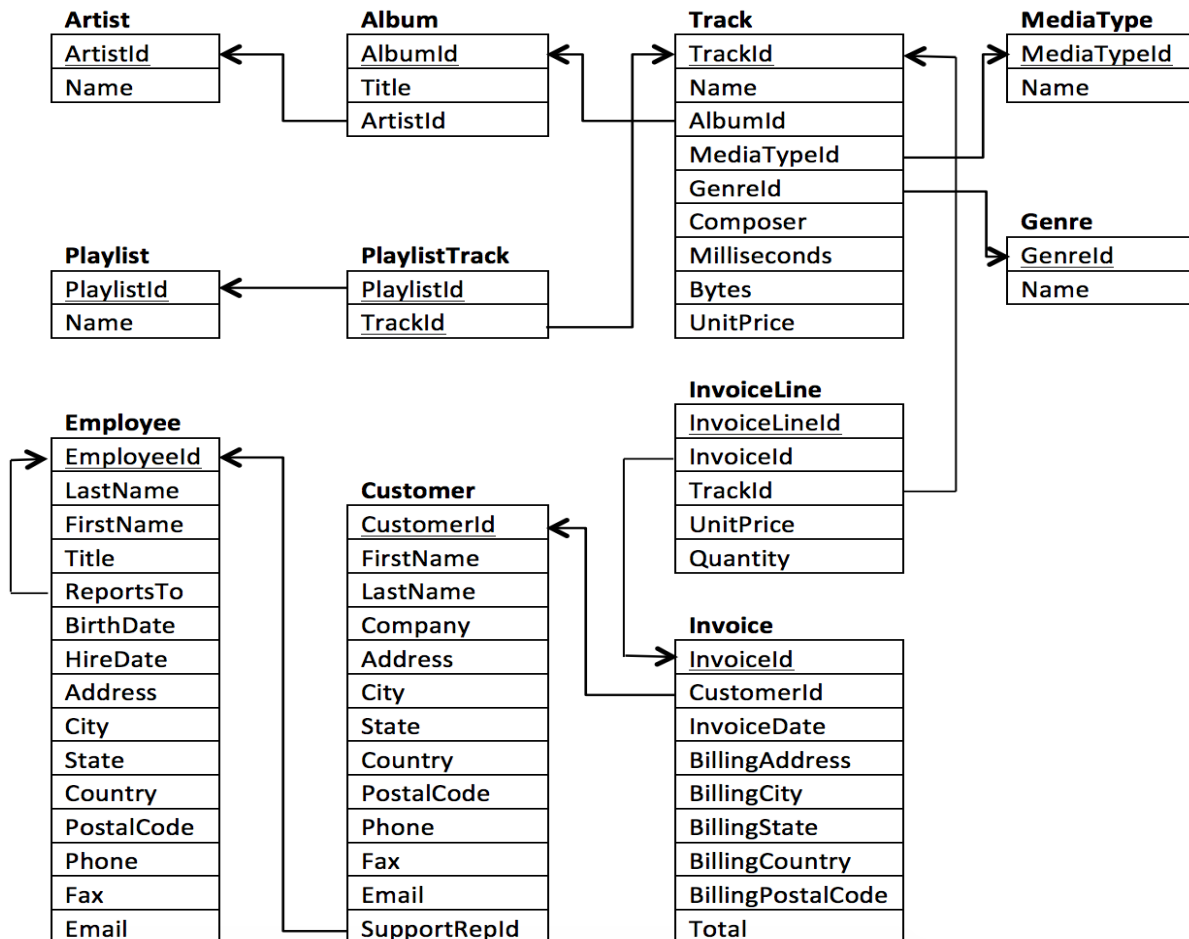
Chinook sample database tables

There are 11 tables in the chinook sample database.

- 1- Employees table stores employee's data such as employee id, last name, first name, etc. It also has a field named ReportsTo to specify who reports to whom.
- 2- Customers table stores customer's data.
- 3- Invoices & invoice items tables: these two tables store invoice data.
- 4- The invoices table stores invoice header data and the invoice items table stores the invoice line items data.
- 5- Artists table stores artist's data. It is a simple table that contains only artist id and name.
- 6- Albums table stores data about a list of tracks. Each album belongs to one artist. However, one artist may have multiple albums.
- 7- MediaTypes table stores media types such as MPEG audio file, ACC audio file, etc.
- 8- Genres table stores music types such as rock, jazz, metal, etc.
- 9- Track table store the data of songs. Each track belongs to one album.
- 10-Playlist & PlaylistTrack tables: playlists table store data about playlists. Each playlist contains a list of tracks. Each track may belong to multiple playlists.

The relationship between the playlists table and tracks table is many-to-many. The playlist track table is used to reflect this relationship.

You can see the columns that tables together via the arrows.



All you'll need to do here is...

Download your new database

Download DB Browser for SQLite

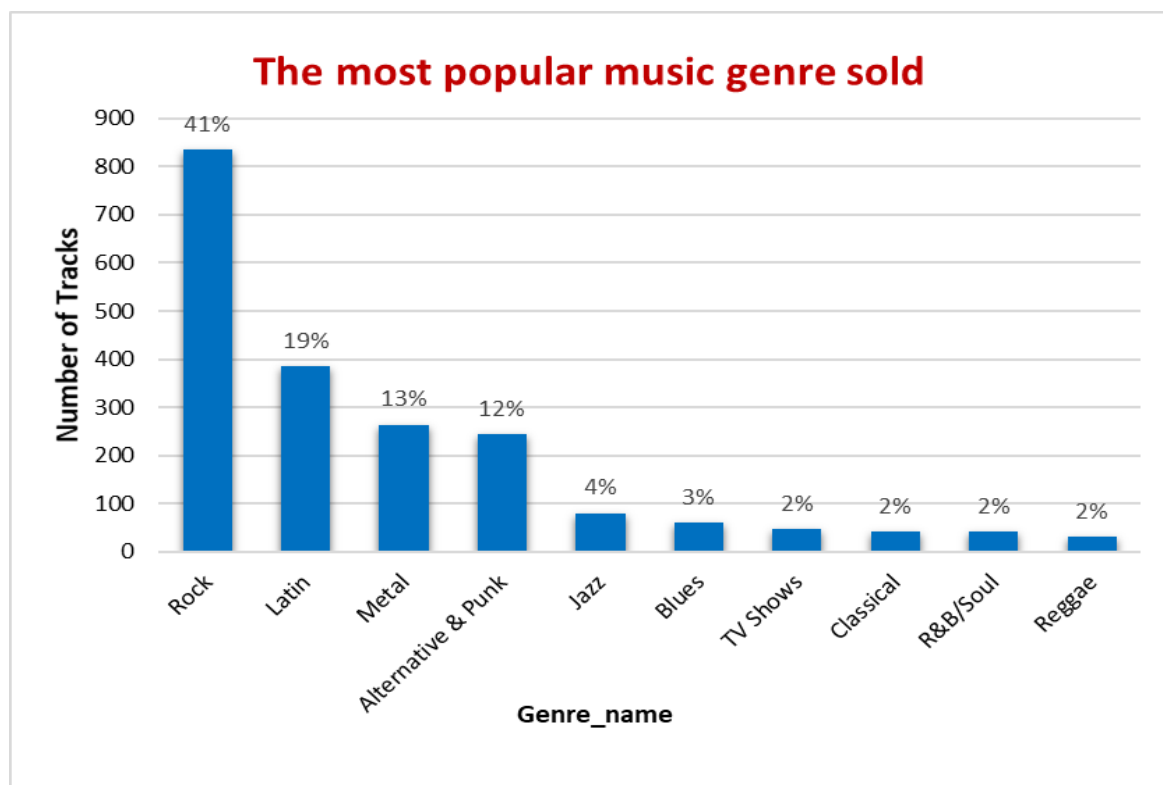
DB Browser for SQLite can be downloaded here: <http://sqlitebrowser.org/>

### Query1: for the first question

Find out what's the most popular music genre sold I joined Track onto InvoiceLine  
Using the primary key TrackId and did number of tracks sold based on genre.

```
SELECT Genre.Name Genre_name, COUNT(Genre.Name) NumTracks
FROM InvoiceLine
JOIN Track
ON InvoiceLine.TrackId = Track.TrackId
JOIN Genre
ON Track.GenreId = Genre.GenreId
GROUP BY Genre_name
ORDER BY NumTracks DESC
LIMIT 10;
```

Result 24 rows returned in the query, I'm take 10 rows only. Graphic visualization of data column chart



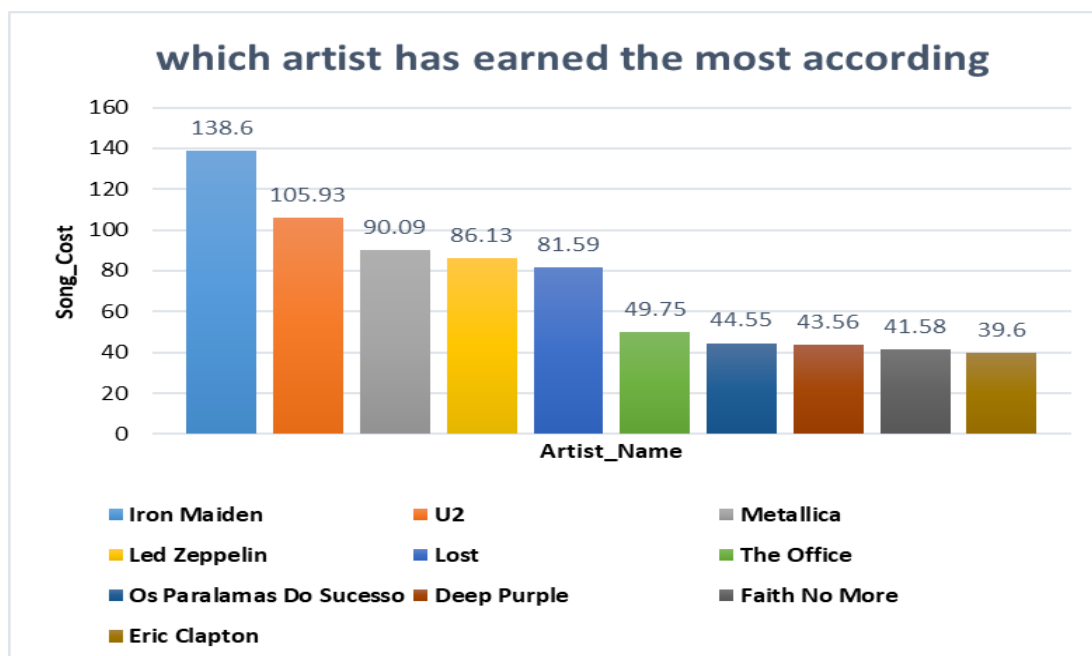
The most popular music genre is rock which has double of songs sold as the next most popular genre, Latin we can further group rock metal and alternative & punk together, which account for approximately 85% of total tracks sold.

### Query2: for the second question

Find which artist has earned the most according to the InvoiceLines?  
I'm will need to use the Invoice, InvoiceLine, Track  
Customer, Album, and Artist tables.

```
SELECT Artist.Name AS Artist_Name , SUM(InvoiceLine.UnitPrice*InvoiceLine.Quantity)
AS Song_Cost
FROM Invoice
JOIN InvoiceLine
ON Invoice.invoiceld = InvoiceLine.Invoiceld
JOIN Customer
On Invoice.CustomerId = Customer.CustomerId
JOIN Track
ON Track.trackId = InvoiceLine.TrackId
JOIN Album
ON Album.AlbumId= Track.AlbumId
JOIN Artist
ON Artist.ArtistId= Album.ArtistId
GROUP BY Artist.Name
ORDER BY Song_cost desc
LIMIT 10
```

Result 165 rows returned in the query, I'm take 10 rows only. Graphic visualization of data column chart



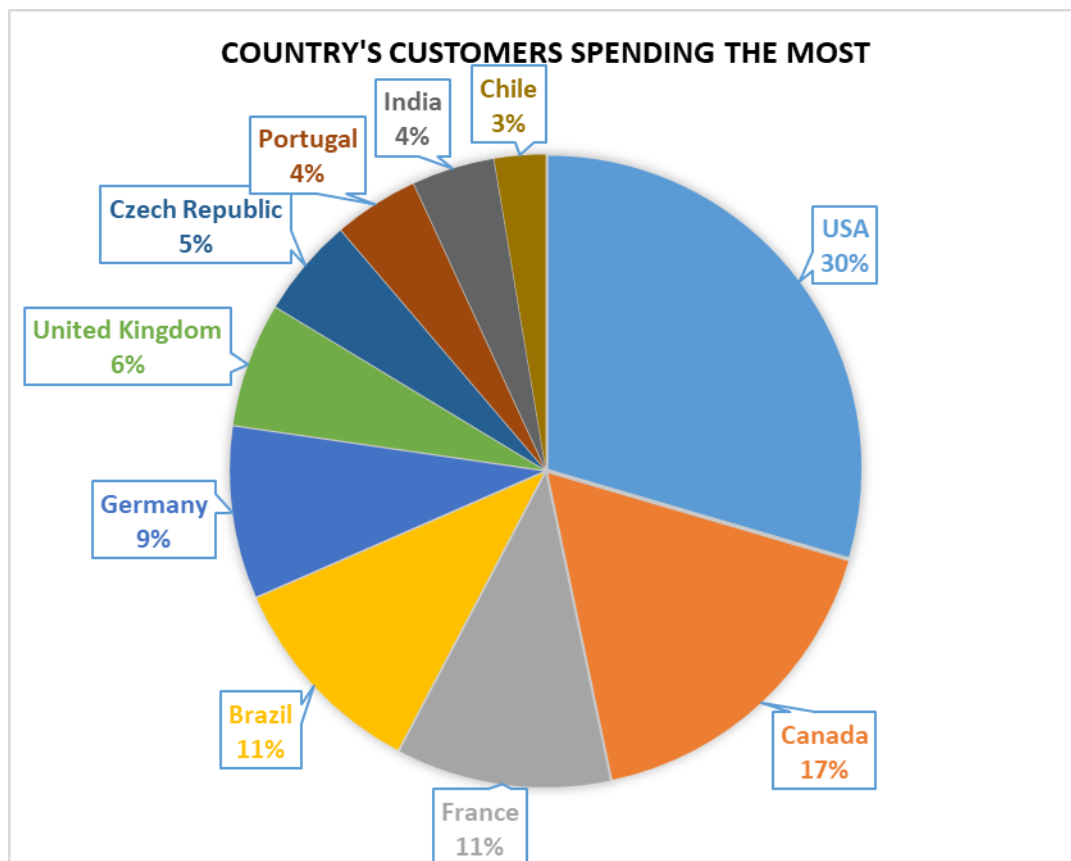
Iron Maiden artist has earned the most according to the InvoiceLines In second place, the artist U2 is more profitable and in the last position is Eric Clapton.

### Query3: for the third question

Find which country's customer's spending the most?  
I'm will need to use the Invoice Employee, Customer

```
SELECT * FROM  
(SELECT Customer.Country, SUM(Invoice.Total) as salesTotal  
FROM Employee, Invoice , Customer  
WHERE Employee.EmployeeId= Customer.SupportRepId  
AND Customer.CustomerId= Invoice.CustomerId  
GROUP BY Customer.Country)  
ORDER BY salesTotal DESC  
LIMIT 10;
```

Result 24 rows returned in the query, I'm take 10 rows only. Graphic visualization of data pie chart



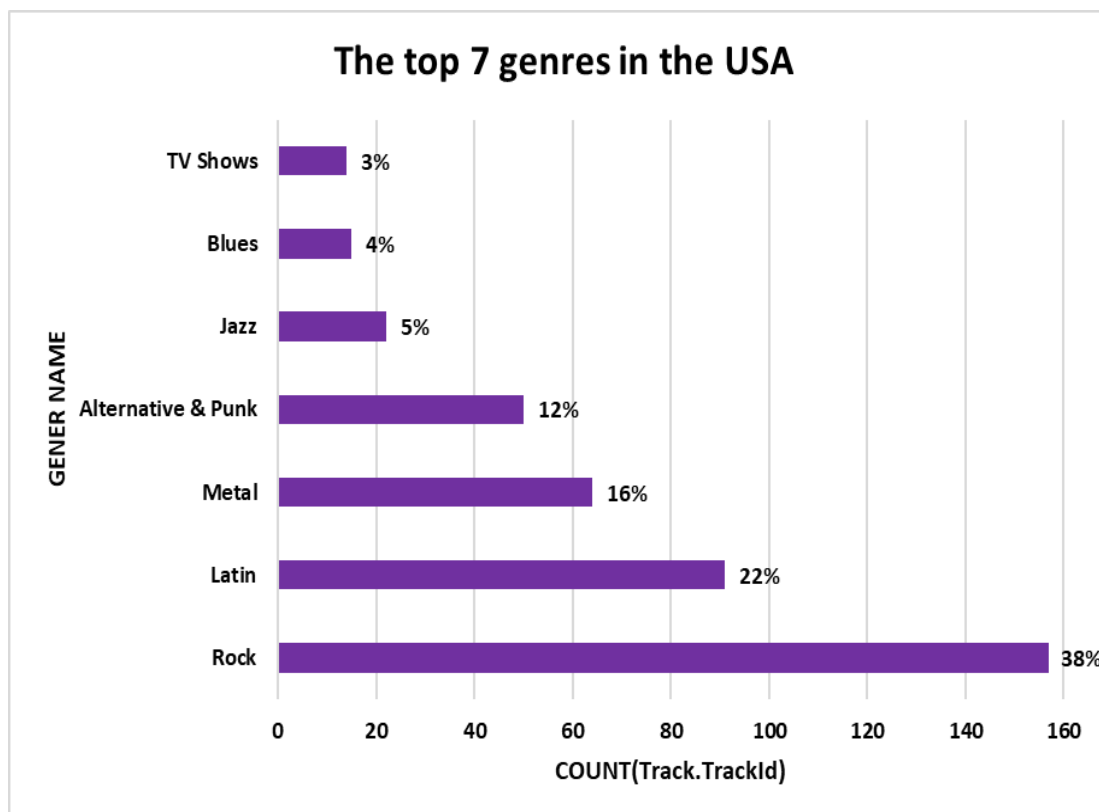
The USA country customers spending the most  
According to the database by up to 31% first in customers spending and  
second customers spending the Canadian state by 17% and in the last  
position Chile by 3%

#### Query4: for the fourth question

What are the top 7 genres in the USA?

```
SELECT Genre.Name, COUNT(Track.TrackId) FROM Track
JOIN Genre
ON Track.GenreId = Genre.GenreId
JOIN InvoiceLine
ON Track.TrackId = InvoiceLine.TrackId
JOIN Invoice
ON InvoiceLine.InvoiceId = Invoice.InvoiceId
WHERE Invoice.BillingCountry = "USA"
GROUP BY Genre.GenreId
ORDER BY COUNT(Track.TrackId) DESC
LIMIT 7;
```

Result 22 rows returned in the query, I'm take 7 rows only. Graphic visualization of data Bar chart



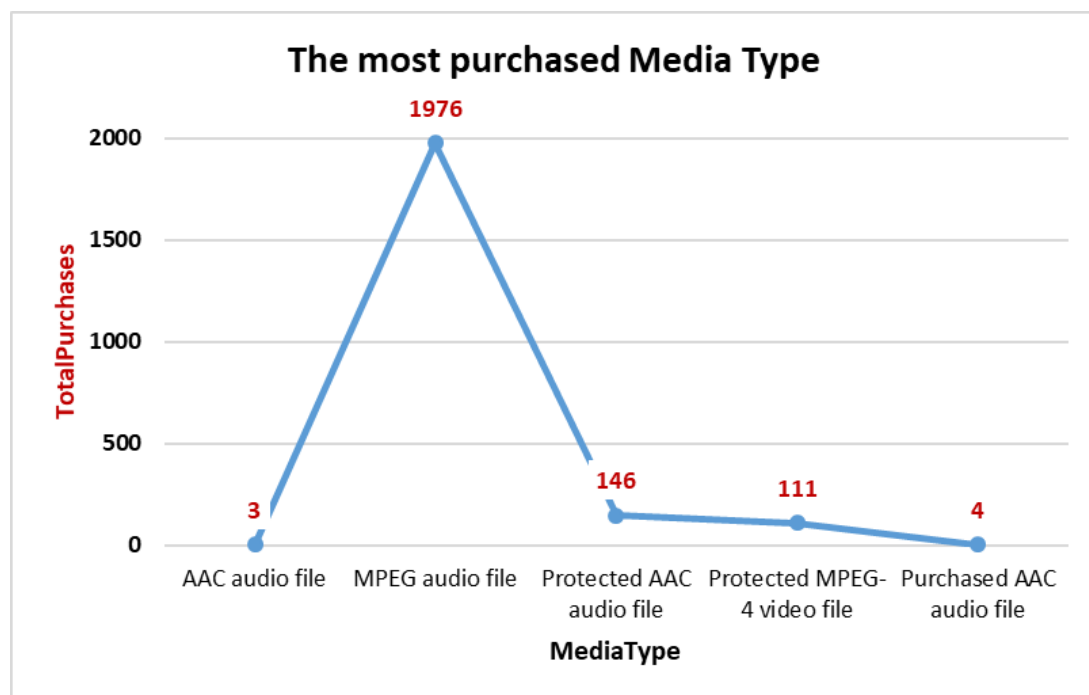
The most music genre is rock first in USA and Latin in second place of songs the most popular genre, we can further group rock metal and alternative & punk together, which account for approximately 88% of genres in the USA.

### Query5: for the fifth question

Provide a query that shows the most purchased Media Type?

```
SELECT
MediaType.Name AS MediaType,
COUNT (InvoiceLine.InvoiceLineId) AS TotalPurchases
FROM Track
JOIN InvoiceLine
ON Track.TrackId = InvoiceLine.TrackId
JOIN Invoice
ON InvoiceLine.InvoiceId = Invoice.InvoiceId
JOIN MediaType
ON Track.MediaTypeId = MediaType.MediaTypeId
GROUP BY MediaType.Name
```

Result 5 rows returned in the query, I'm take all rows. Graphic visualization of data line chart



MPEG audio file most purchase Media Type and he Larger than the other media type according to the accompanying visualization diagram and the query output in data base.

I wish success to all.

Marwan Saeed Alsharabbi