

INST 327: Database Design & Modeling

Section 0103

Final Project

Team 6

Jay Mathur, James Baik, Krish Guru, May Kyaw

## **Final Report**

### **Introduction:**

This is the final report for our team's database project. Our team chose to construct a database that holds information about songs in Spotify. Songs have a lot of different categories and they have a lot of attributes that allow us to build a database that we can then use to categorize and analyze information about songs.

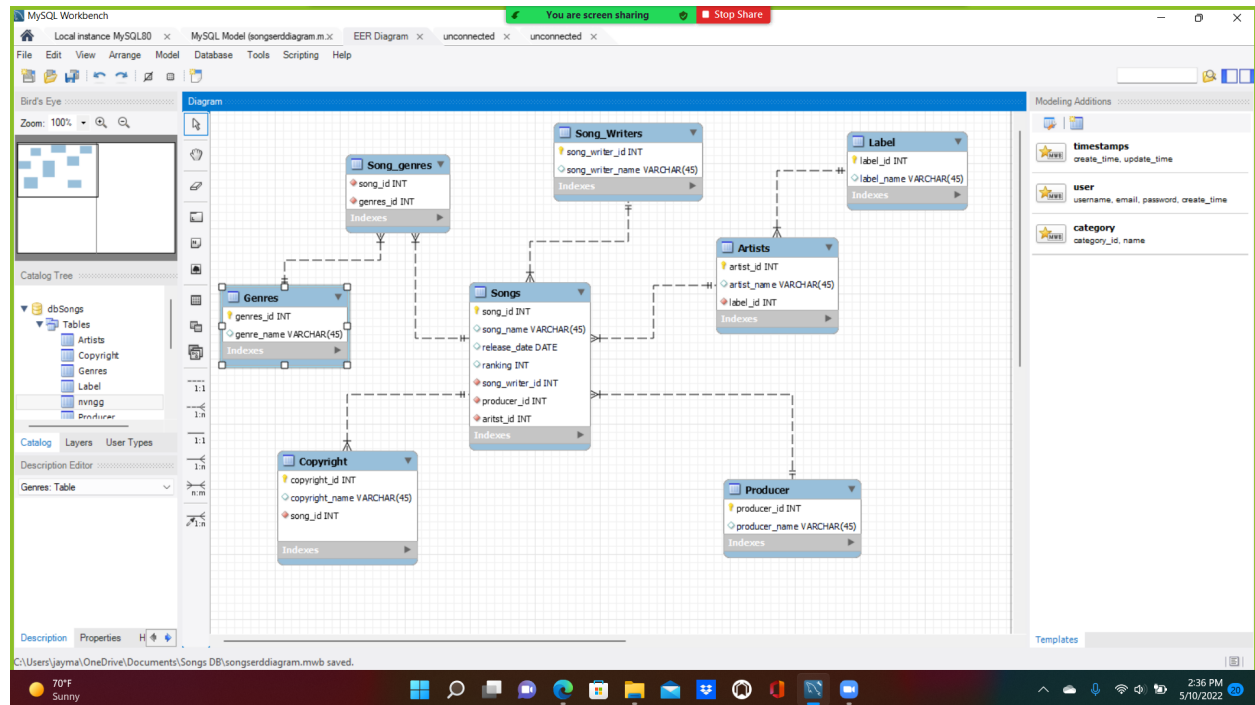
Our idea was to build a database that would capture this information for a wide variety of songs so that we could quickly learn information about songs and to run queries that let us answer questions about how songs are related and how certain people are involved in multiple songs.

We chose this topic because it interested us and it allowed us to use a database to answer questions of interest that we had about how different songs are related.

### **Database Description:**

### **Logical Design:**

Our final database design wound up having 8 tables: Artist, Genres, Label, Producer, Songs, Copyright, Artists\_Genres, and Song\_Writers. Each table contains various foreign keys in order to facilitate proper table relationships. The foreign key allows the database to be easily utilized and allows for easier and more informed query writing. Each table is connected in a logical manner where data is separated but still allows for easy understanding for a variety of users. We followed the methodology from the course to reach our final database design and our design is highlighted in the ERD diagram below:



## Physical Design:

Our database description and scope have not changed for the majority, and it is still based on one that holds information about songs but specific information on the top 20 artists on the Spotify playlist. The target audience for our database will still be people who are in the music business because of the interrelationships that exist in the industry. Using our database will allow someone to seek which artists have worked with what production companies and vice versa. In addition to people in the music industry, our database will target anyone who is interested in music and Spotify in general. For people who want to know which artists have been featured on another artist's album and similar types of information, this database can be useful. The scope of our database still remains the same because of our interest in the extent of the common relationships that occur in the music industry. Our current plan is to finalize our ERD diagram and make sure that all the tables included are relevant and correct. Additionally, we are looking to start bringing over the data of the top 20 artists on Spotify into our database by gathering all relevant information.

## Sample Data:

The data that was selected for our database came from the Spotify Top Charts list on Spotify's website. We chose the top 20 songs that users listened to the most which would become

our starting point. This data would serve a purpose for some users to see specific trends from certain types of artists and songs. From there we manually imported what we thought were the most significant factors of information for each column that contributed to the making of each song. Some of these columns included but are not limited to the song name, the name of the artist, producer name, the genre of the song, release date, and where that song ranked on the list.

### Sample Data Photo:

	Ranking	Song_name	Artist_name	Producer_name	Release_date	Genres_id	
▶	1	As it Was	Harry Styles	Kid Harpoon	2022-03-31	1	
	2	First Class	Jack Harlow	Jack Harlow	2022-04-08	2	
	3	WAIT FOR U (feat. Drake & Tems)	Future	ATL Jacob	2022-04-29	2	
	4	PUFFIN ON ZOOTIES	Future	TM88	2022-04-29	2	
	5	About Damn Time	Lizzo	Blake Slatkin	2022-04-14	3	
	6	Heat Waves	Glass Animals	Dave Bayley	2020-08-06	4	
	7	IM ON ONE ( feat. Drake)	Future	Torey Montana	2022-04-29	2	
	8	712PM	Future	TM88	2022-04-29	2	
	9	INDUSTRY BABY (feat. Jack Harlow)	Lil Nas X	Kanye West	2021-09-17	2	
	10	Stay	Justin Tranter	Blake Slatkin	2021-07-23	5	
	11	No Role Modelz	J. Cole	Dreamville	2014-12-09	2	
	12	good 4 u	Olivia Rodrigo	Daniel Nigro	2021-05-21	6	

### Views and Queries:

The number of queries we ended up creating for our final database project was 5. Each query asks and gives different result sets for questions we thought may interest users utilizing our database. Based on our original project proposal questions and the information we put into our database some of those questions can not be answered with our database such as the total record sales of a certain artist and which label has worked with certain video directors. Overall, all queries have been saved as views and the 5 queries we did come up with give a result set that is more practical information given the data we imported into our database.

### View Description:

1. The first view (hip\_hop) shows all the songs in the top 20 list that fall under the hip-hop genre.
2. The second view (multiple\_producers) shows how many different songs the producer “TM88” has been involved with.
3. The third view (recent\_release) shows how many songs in the top 20 were released in the current year “2022”.

4. The fourth view (same\_artist) shows how many songs the artist “Future” has in the top 20 list because out of the 20 songs he appears the most.
5. The fifth view (singer\_writer1) shows which artists on the top 20 list have written their own songs.

View Name	Req. A	Req. B	Req. C	Req. D	Req. E
hip_hop	x	x	x	x	
multiple_producers	x	x		x	
recent_release	x	x			
same_artist	x	x			
singer_writer1		x			x
TM88_top_songs		x	x		
jbsongs		x	x		

### **Changes from Initial Proposal:**

One major change from our initial proposal is how we were going to obtain our sample data. Originally, hypothetical data was going to be used but we decided that using our favorite artists and albums may not have been the best way to get data. Rather, we as a group decided to use the Spotify top 20 lists as our sample data as it shows no bias because it is determined by the number of times a song has been listened to. Another change from the initial proposal is the target audience. The original target audience will remain the same but we decided to add any group who is interested in music. We decided to broaden the target audience because the top 20 list on Spotify may not be as attractive to just one group of people.

### **Diversity, Equity, and Inclusion Considerations:**

After solidifying the plan for our database, and revisiting the database it seems to be an

inclusive information resource in its domain. In terms of bias of the data as we chose to examine the top 20 songs on Spotify in the U.S. there would be a bias amongst songs in English as most of the population speaks the language. There would also be a bias amongst songs and artists they only stream on Spotify as the specific content is only available on the streaming platform.

### **Data, Privacy, Fair Use, Other Ethical Considerations:**

As we stated in our proposal we do not see any problems or legal issues with our database or anything that could come up. We are not using any audio or copyrighted information from Spotify and we should not have any problems in doing so. Our database aims to combine existing public data that Spotify has provided and make it easier to analyze in one complete database instead of two websites. There is nothing proprietary about the information we are using.

### **Lessons Learned**

Working on this database during our course was a very good and practical experience that helped us to get hands-on experience with working with databases. We feel that the DataBase Normalization process was the most difficult aspect of the project and we wound up reducing the number and scope of our tables based on this. Once we had worked with our TA to finalize our design for the database, using SQL to build it and then test it was fairly straightforward. We did run into syntax errors at many points in our project and we were able to resolve them ourselves or get help from our TA to resolve them. We encountered numerous problems that we have come to solve as we completed the database. There are technical errors that we had to fix along with some problems with finding the data. We initially had a finished ERD with tables of different data items that we could place in them. Once we tried to get that data, not all of it was available on the internet. We had to change our tables up, what they had in them, and how they would be connected. We realized that each song had numerous writers who each had their own label. We could connect the label to the songwriters, artists, and producers but not the song. We had to first configure this. We also could not find some information for tables like copyright and genre on the Spotify app. Using the unique Spotify ID of each song, we used the Spotify developer website to gain more information. Once we had all our data we had a problem forward engineering it as we did not look over some repetitions in columns and some columns did not

have the right connecting keys. Overall we successfully finished the database after solving our problems.

### **Potential Future Work**

In the future this project has a lot of potentials to be used. This database only expands on the 20 songs and the information about them. Even with the 20 songs available, we have been able to find some connections between the songs and their metadata. It is easy to look for another song with the same genre as another song and more. This will allow users to find more songs that they may be interested in. If Spotify or other music industry giants implement this with every song, it would be easier for all users to find new music based on genre, producer, label, and more. This could also be expanded to include more tables of information about each song.