ECHO TEAM PROJECT REPORT

A COMPILATION OF THE REPORT MADE BY TEAM MEMBERS:

- **TOYIBAT ADELE**
- **CHIDERA JENNIFER UDOH**
- **NNENNA EDEH**
- **OKORIE UGONMA REGINA**
- **OSARUONAMEN OLU-OSAWE**

REPORT DETAILS

This document is a compilation of all the reports each team member made, and provides more details on the project. It is sectioned into the following:

1. Project Overview
2. Pre-querying
3. Report on Task 1
4. Report on Task 2
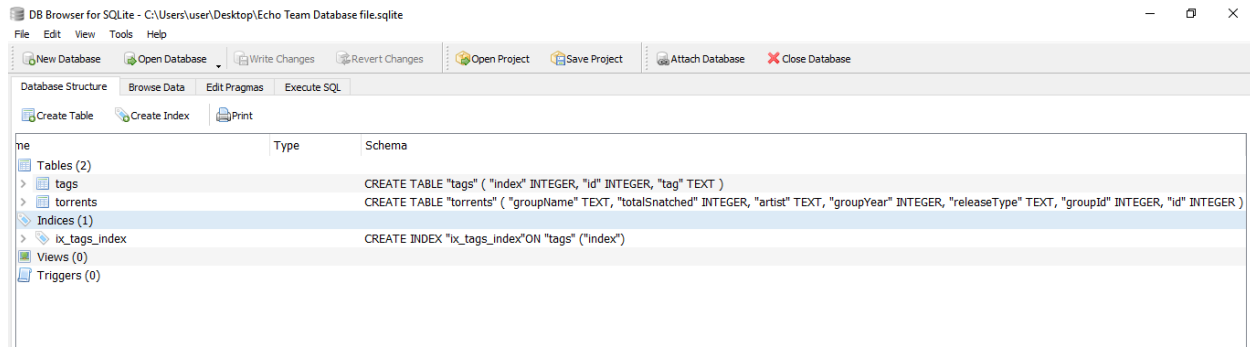5. Report on Task 3
6. Report on Task 4
7. Conclusion

PROJECT OVERVIEW

The Echo Team members were tasked with using this SQLite database file to run queries and provide answers to these questions:

● What is the most popular release type?
● For the last three years, what release type has the most torrents?
● Who are the top five most common artists?
● What is the average number of downloads?

Each question was assigned to the team members as an individual task. They had to write queries to get an answer to the question assigned to them, and provide a well-detailed report showing their steps and thought processes.

PRE-QUERYING

Before carrying out the tasks, the team looked at the dataset. We deduced that the dataset was complete as no missing value existed in any field. The schema of the dataset showed that each feature(column) had the correct datatype. View the screenshot below.

The number of rows in the dataset was also in tally with that of the source file. After those checks, we were convinced that the dataset was fit for use.

Report on Task 1

This task was assigned to Okorie Ugonma Regina. Her task was to write a query to answer the question, 'What is the most popular release type?'. The following are the steps she took:

- I selected the *releasetype* column using the DISTINCT keyword, along with the count of each of them. Then, I grouped by released type and ordered by count. The query I used is below:

> *SELECT DISTINCT releasetype, COUNT(*)*
> *FROM torrents*
> *GROUP BY releasetype*
> *ORDER BY COUNT(*) DESC;*

- After running my queries, I found out that the release type that is the most popular is '**album**'. **Album** has a **total count** of **35,233**.

The screenshot below shows how the query was written and the result gotten.

```
1  SELECT DISTINCT releasetype, COUNT(*)
2  FROM torrents
3  GROUP BY releasetype
4  ORDER BY COUNT(*) DESC;
```

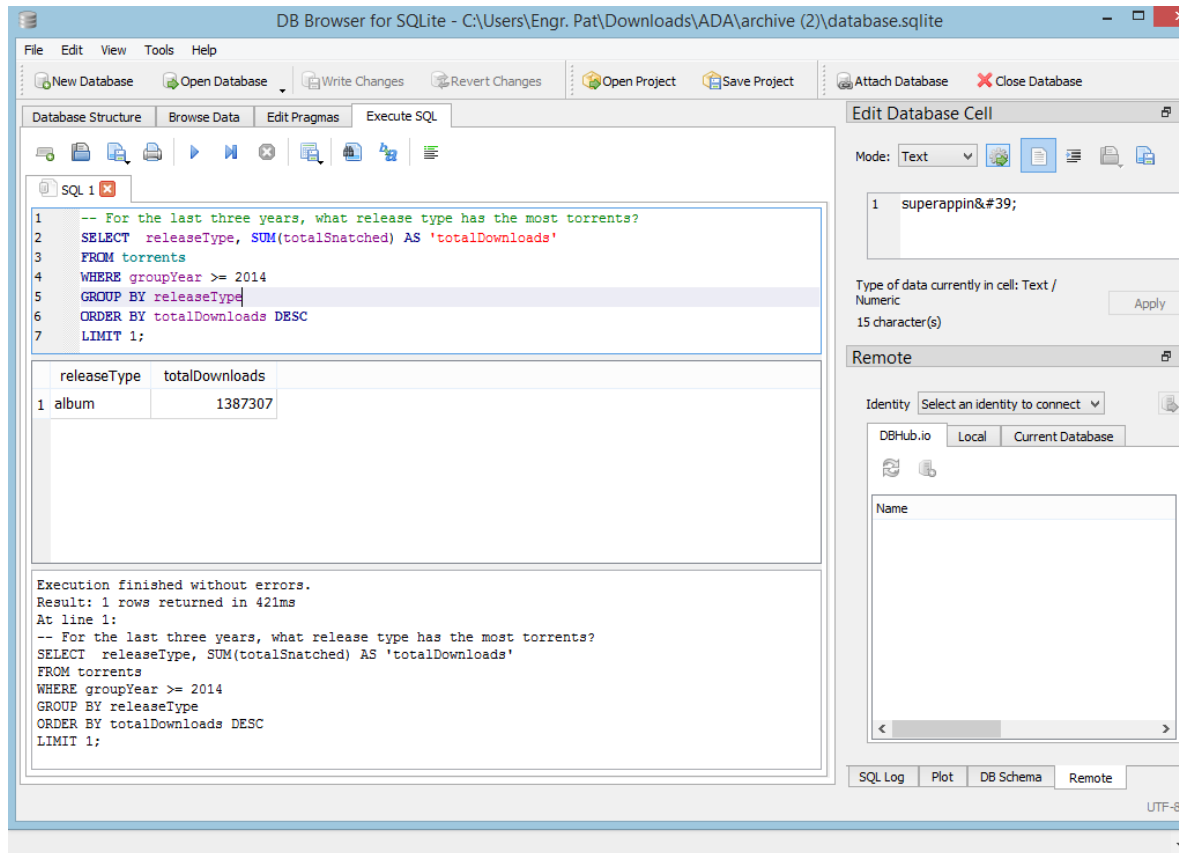| releaseType	COUNT(*)
album	35233
single	12228
ep	10260
compilation	5691

To access the original file of the report on task 1, click here.

Report on Task 2

This task was assigned to Osaruonamen Olu-Osawe. Her task was to write a query to answer the question, 'For the last three years, what release type has the most torrents?'. The following are the steps she took:

- I selected the *releasetype* column along with the sum of the values in the *totalsnatched column* under an alias called '*total downloads*', using the SELECT statement.
- I selected the records for only the last 3 years using the WHERE clause, grouped the records of the *releasetype* column with the GROUP BY keyword and ordered the total downloads in descending order with the ORDER BY and DESC keywords.
- Then I used a limit of 1 to get the highest total download with the LIMIT keyword.

After running my queries, I found that the release type with the most torrent(*totalsnatched*) from the last three years was **Album** with a **total download** of **1,387,307**.

To access the original file of the report on task 2, click [here.](here)

Report on Task 3

This task was assigned to Chidera Jennifer Udoh. Her task was to write a query to answer the question, 'Who are the top five most common artists?'. The following are the steps she took:

- Firstly, I downloaded the DB Browser SQLlite Database software and used it to open the document.
- Then I clicked on the execute query tab. Here, I typed in the query that will help me get the top 5 most common artists. The query is:

SELECT artist, COUNT(artist)
FROM torrents
GROUP BY artist
ORDER BY COUNT(artist) DESC
LIMIT 5;

To simplify it, I'll break the query down into sections:

1. "SELECT artist, COUNT(artist)

FROM torrents"

This is to select the column named *artist* from the *torrents* table, and from that column selected, count how many times a particular artist was mentioned.

2. "GROUP BY artist
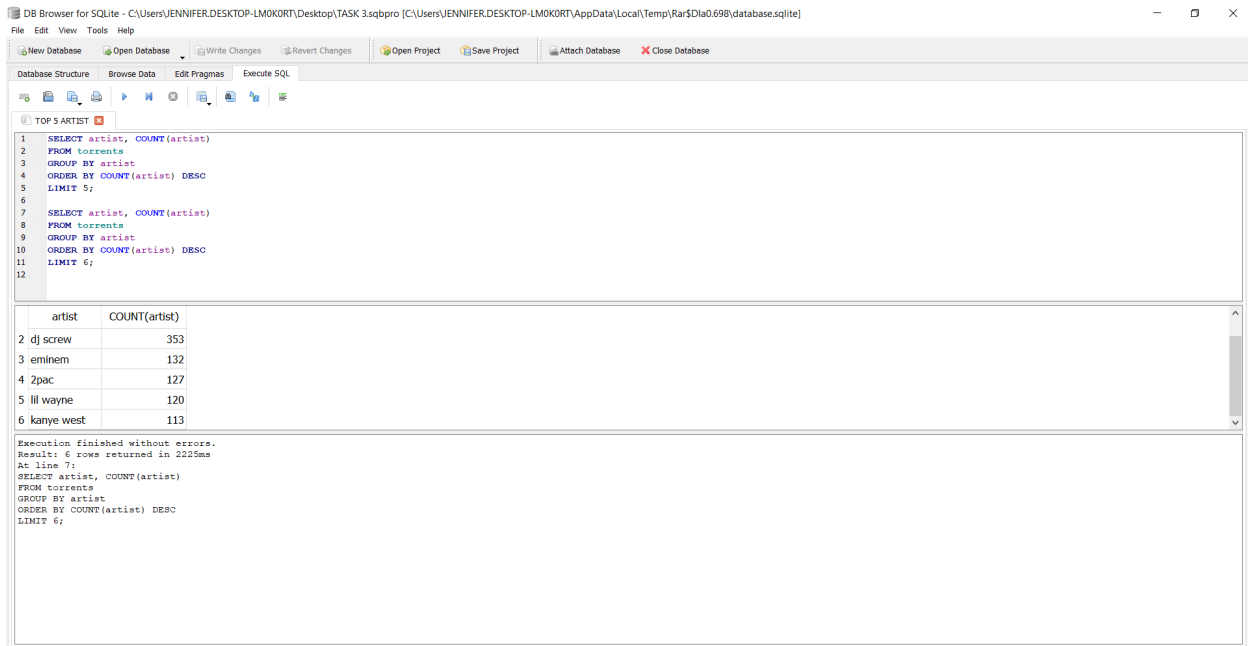ORDER BY COUNT(artist) DESC"

This means that these artists selected in the previous section of the query were counted and then grouped. This is done according to their names and the number of times they appeared in Descending order (from the highest to the lowest).

3. "LIMIT 5;"

Then finally I added the 'Limit 5' statement to the other part of the query above and concluded with ';'. This will only produce the first 5 in the table, giving us our top 5 most common artists.

According to the result, the first one which is '*various artist*' is not a name of an artist but a general term for unnamed artists, so I will exclude it and look at the top 5 after 'various artist'. This means the 'LIMIT 5' command will now be replaced with 'LIMIT 6' as shown below:

SELECT artist, COUNT(artist)
FROM torrents
GROUP BY artist
ORDER BY COUNT(artist) DESC
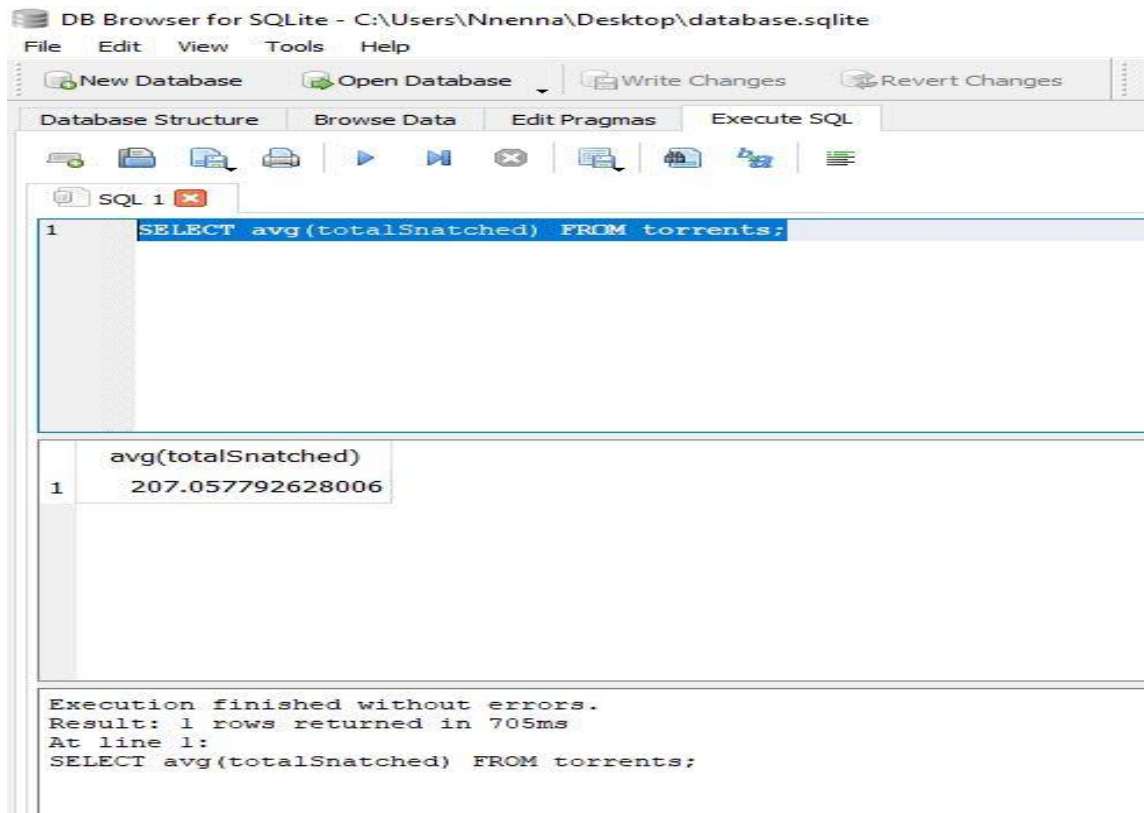LIMIT 6;

Therefore, the Top 5 most common artists will now be:

1. **DJ screw**
2. **Eminem**
3. **2pac**
4. **Lil Wayne**
5. **Kanye west**

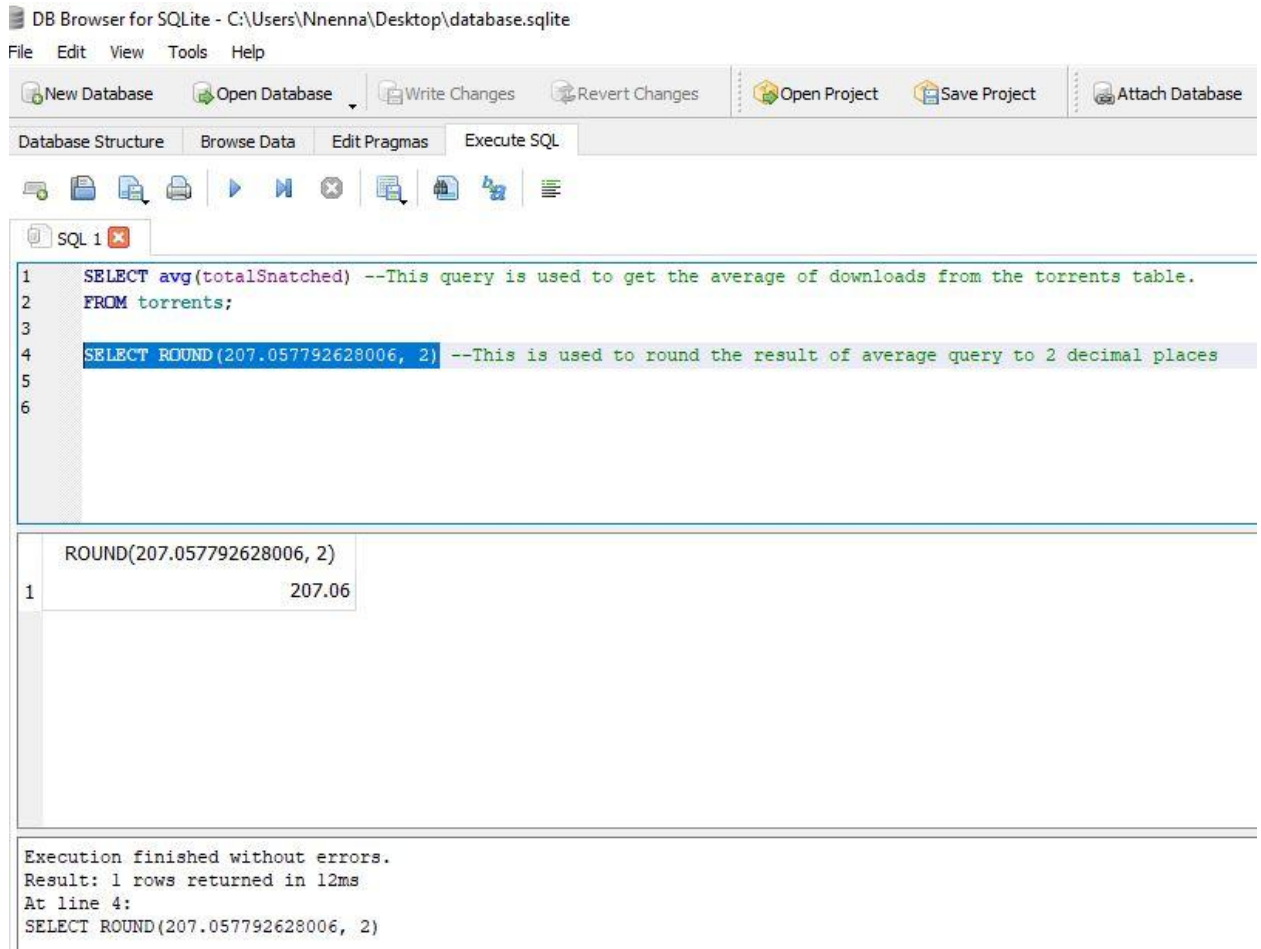To access the original file of the report on task 3, click here.

Report on Task 4

This task was assigned to Nnenna Lucy Edeh. Her task was to write a query to answer the question, 'What is the average number of downloads?'. The following are the steps she took:

● I used the SQL function, *avg*, as seen in the screenshot below, to get the average downloads from the torrents table. Also, *totalSnatched* is an integer column that shows the number of times the release has been downloaded.

- I used the *Round* SQL function to round the result of the average, to two decimal places. See this in the screenshot below.

- After executing my queries, I found out that **207.06** average downloads were made, from the torrents table.

To access the original file of the report on task 4, click <u>here.</u>

<u>Conclusion</u>
After successfully writing and running queries, and receiving the report after each analysis, the following are the answers provided to the questions asked at the start of this document:

1. What is the most popular release type? -> The most popular release type is '**Album**' and it has a **total count** of **35,233**.

2. For the last three years, what release type has the most torrents? -> The release type with the most torrent from the last three years was **Album,** with a **total download** of **1,387,307**.

3. Who are the top five most common artists? -> Top 5 most common artists are **DJ screw, Eminem, 2pac, Lil Wayne, and Kanye west.**

4. What is the average number of downloads? -> The average number of downloads is **207.06.**