

# ***Spotify & Youtube Song Track Analysis***

***About the project:***

In this project, you will be working on a real-world dataset of Spotify and Youtube combined. This project aims on cleaning the dataset, analyze the given dataset and mining informational quality insights.

##### ***Project Description:***

This project will help you understand how a real-world database is analyzed using SQL, how to get maximum available insights from the dataset, pre-process the data using Python for a better upcoming performance, how a structured query language helps us retrieve useful information from the database

The Project will consist of 2 modules:

***Module 1***: Pre-processing data using Python

***Module 2***: Analyzing data using SQL

# ***Prerequisites*** ***for the Project (mandatory):***

1. SQL (MYSQL)
2. Excel
3. Python

## Module 1 Pre-processing data using Python

### Pre-processing the data

Data Pre-processing is one of the important steps in data analytics because data that is not processed can lead to different unwanted results when the data will be used for further applications. This task includes sub-tasks such as handling null values, deletion or transformation of irrelevant values, datatype transformation, removing duplicates, etc. The tasks to be performed for cleaning the data set are given below:

#Task 1: Remove columns that are not needed in our analysis. Remove Url\_spotify, Uri, Key, Url\_youtube, Description

#Task 2: Handle the null values

#Task 3: CHECK FOR DUPLICATES AND REMOVE THEM KEEPING THE FIRST VALUE

#Task 4: CONVERT millisecond duration to minute for a better understanding

#Task 5: Rename the modified column to Duration\_min

#Task 6: Remove irrelevant track name that starts with ?

#Task 7: Calculate the Energy to Liveness ratio for each track

#Task 8: change the datatype of views to float for further use

#task 9: compare the views and stream columns to infer that the song track was more played on which platform, youtube or Spotify. Create a column named most\_playedon which will have two values. Spotify and Youtube, If a song track is most played on youtube then the most\_played on column will have youtube as the value for that particular song

## *Module 2 Analyzing Cleaned Data Using Sql*

**MySQL problem statements:**

**1. Which is the most viewed song track on youtube?**

**2. Which Song track is streamed most on Spotify?**

**3. EnergyLiveness ratio is one of the popular ways to measure the quality of the song, which are the top 5 songs that have the highest energyliveness ratio**

**4. let us assume a situation where an artist named Black Eyed Peas wants to analyze his songs. The artist wants to know which platform is capable of keeping his song track more engaged. To check this he assigns you this task and wants you to report to him where his song tracks are more played on. compare the platforms.**

**5. Gorillaz wants to know their most liked song on youtube. Report to them with their most liked song along with the Energy and Tempo of the song.**

**6. Which Album types are more prominent on Spotify?**

**7. Spotify's most loved song tracks are to be declared soon. Help Spotify choose the top 5 most streamed+youtube viewed song track.**

#### 

#### 

## 

## 

.