To replicate the code in the provided file, follow these instructions:

1. Install Required Libraries

* Make sure you have Python installed on your system.
* Install the required libraries by running the command given at end using CMD or install the requirements file attached in the folder.
* Requirements.txt created using pip freeze > requirements.txt

# pip install pandas numpy seaborn matplotlib plotly

2. Dataset

* Dataset is attached in the folder.
* we are creating a varible file\_path to store the path of the dataset.
* The ‘pd.read\_csv()’ function will read the dataset from the specified file path.
* Dataset has been attached in the folder, (songs\_normalize.csv)

3. Data Cleaning and Preprocessing

* The code performs various data cleaning and preprocessing steps, such as:
* Checking for missing values and their proportion
* Converting the duration from milliseconds to seconds
* Filtering the data to include only years between 2000 and 2019.

4. Analysis and Visualization

The code performs the following analyses and visualizations:

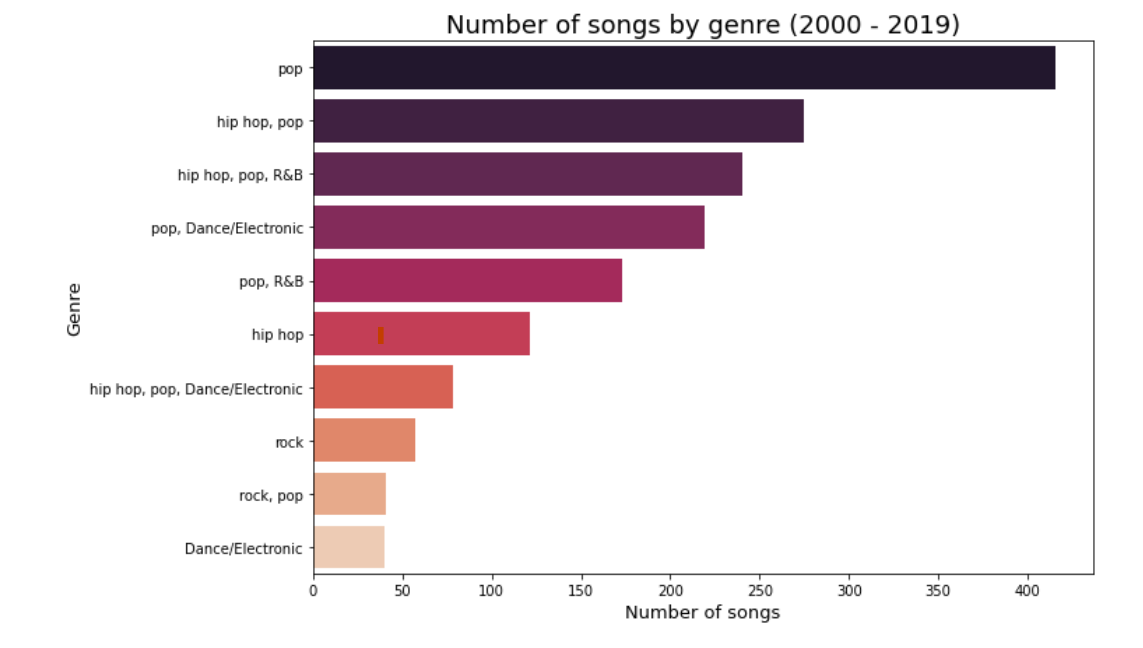
* Grouping songs by genre and counting the occurrence
* Plotting the songs produced in different genre between the period of 2000 to 2019 using horizontal bar chart.
* Visualizing the trend of the top 5 genres over time using line plots

5. Explore the Outputs

* The script will generate various outputs, including:
* Printed statements and statistical summaries
* Horizontal Bar Chart depicting the number of songs produced in the different genre and by clubbing different genre.
* Line plots showing the trend of the top 5 genres over time.

The visualisation created from the dataset are attached below: -

1. Horizontal Bar Chart:



1. Line Chart:

A group of graphs showing different colored lines

Description automatically generated